Investigating the Role of Paralogous Gene Expansion and Functional Homology of Membrane-Trafficking Machinery in Organellogenesis: Apicomplexan Parasites as a Model Case

by

Christen Markus Klinger

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

Doctor of Philosophy

Department of Medicine
University of Alberta

Abstract

The ease with which large multicellular organisms, such as plants and animals, may be observed belies an unappreciated wealth of diversity; the majority of eukaryotes are unicellular, and display a dazzling array of morphologies and lifestyles. Some eukaryotes are free-living or symbiotic, including heterotrophs that feed on other cells and phototrophs that harness solar energy through biochemical reactions; still others are parasitic, relying on host organisms for nutrients. This biodiversity is underpinned by a concomitant wealth of cellular diversity, and begets questions: 1) how did the diversity of extant eukaryotes arise? 2) What are the underlying mechanisms that give rise to this diversity? 3) Given this diversity, to what extent can eukaryotic features and cellular machinery (genes) be considered conserved?

One cellular system that is instructive in answering these questions is the membrane-trafficking system (MTS), which encompasses the set of membrane-bound organelles and machinery that mediates movement of material between them. This system is a eukaryotic innovation, absent from prokaryotic and archaeal cells, and is critical in defining cellular ultrastructure, homeostasis, and interaction with the extracellular environment. In an attempt to answer the third question posed above, numerous studies have elucidated the core set of organelles and machinery across eukaryotes, which, by parsimony, are also presumed to have been present in the Last Eukaryotic Common Ancestor (LECA). MTS machinery is well-conserved, and the LECA is inferred to have possessed a complex trafficking system, similar to what is observed in extant eukaryotes. However, all such studies have relied on an implicit assumption of functional homology, that orthologous genes identified through *in silico* analyses of genomic data from diverse eukaryotes perform the same function; this assumption has never been formally tested. In addition, it is unclear how differences in the MTS across eukaryotes, both in terms of organelles and machinery, could arise based solely on the presence of conserved MTS machinery.

Hence, the first two questions posed above become key considerations. As the MTS is a eukaryotic innovation and a complex MTS is inferred in LECA, it must have arisen sometime

between the advent of eukaryotes (eukaryogenesis) and the radiation of extant eukaryotes from LECA. One hypothesis to explain how a complex system could be generated from primordial components is the Organelle Paralogy Hypothesis (OPH), which posits that gene duplication and divergence would have resulted in a simultaneous increase both in number of distinct endomembrane compartments and trafficking machinery. Although the OPH was originally described to explain the ancient origins of eukaryotic complexity, it represents a viable hypothesis for the emergence of cellular complexity since the LECA, including in parasitic eukaryotes.

The focus of this thesis is on understanding the relationship between gene duplication, gene function, and organelle complement in extant eukaryotes. Although the OPH presents an attractive hypothesis to explain the continued emergence of novel organelles across eukaryotes, reliable inference of such events relies on the functional constraint of machinery inherited from the LECA (i.e. functional homology). Without this constraint, individual pieces of machinery could perform diverse functions, and the predictive significance of machinery gained in a lineage since the LECA for explaining novel organelles is essentially lost.

Hence, this thesis investigates in detail an enigmatic group of parasites, the Apicomplexa, which possess unique secretory organelles in addition to a "core set" of eukaryotic organelles and are therefore attractive candidates for studying the OPH. Chapter 2 demonstrates the utility of including high-quality genomes of closely related free-living taxa for mapping evolutionary events during apicomplexan evolution. Chapter 3 presents a pan-eukaryotic literature analysis focussing on the question of functional homology within the MTS. Chapter 4 introduces novel data that systematically demonstrate, for the first time, the presence of additional paralogues within some MTS families in the Apicomplexa; Chapter 5 then investigates three novel paralogues in the model apicomplexan *Toxoplasma gondii*, providing data to support that one such paralogue is involved in trafficking to additional organelles within the parasite.

The data presented in this thesis provide an initial basis to explore the questions posed in the first paragraph and suggest that the OPH mechanism is at least partly responsible for generating organelle diversity across eukaryotes. It is expected that similar future studies will allow this model to be expanded and refined.

Preface

(Mandatory due to collaborative work)

Some of the experiments conducted in this work would not have been possible without the contributions of others. Chapters 4 and 5 describe work performed for a collaborative effort, jointly led by Dr. Joel B. Dacks at the University of Alberta and Dr. Markus Meissner, formerly of the University of Glasgow and currently at Ludwig-Maximilians-Universität (LMU). In particular, initial design of vector constructs and overall assistance with experimental and technical details was assisted by Dr. Elena Jimenez-Ruiz, and all electron microscopy work was performed by Dr. Leandro Lemgruber. In addition, Lucas Paoli assisted with the choice of statistical tests for analysis. All presented data (with the exception of EM work) are the results of experiments I performed and analyzed.

Section 1.6 of this thesis, as well as figures 1.8 and 5.11, are modified from a book chapter in preparation: S. Besteiro, C.M. Klinger, M. Meissner, and V.B. Carruthers, "Endomembrane trafficking pathways in *Toxoplasma"*, in *Toxoplasma gondii* 3rd edition by L.M. Weis and K. Kim, Copyright Elsevier (2020). I am the sole author of the materials included in this thesis.

Section 2.2 of this thesis contains work published as part of Y.H. Woo, H. Ansari, T.D. Otto, C.M. Klinger, M. Kolísko, et al., "Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites" (2015), *eLife*, e06974. I was involved in performing an analysis of membrane-trafficking machinery, as well as writing and editing part of the manuscript. Due to space constraints, all other author contributions cannot be listed here, but are available at the end of the aforementioned publication.

Section 3.2 of this thesis was published as C.M. Klinger, I. Ramirez-Macias, E.K. Herman, A.P. Turkewitz, M.C. Field, and J.B. Dacks, "Resolving the homology-function relationship through comparative genomics of membrane-trafficking machinery and parasite cell biology" (2016), *Molecular and Biochemical Parasitology*, vol. 209, issue 1-2, 88-103. I was responsible for writing the manuscript, together with input from all other authors, and for generating tables and figures. All authors read and edited the manuscript.

Section 4.4.5 of this thesis presents work on TBC-N and TBS proteins; I previously performed a similar analysis that has been published as part of S.V. Pipaliya, A. Schlacht, C.M. Klinger, R.A. Kahn, and J.B. Dacks, "Ancient complement and lineage-specific evolution of the Sec7 ARF GEF proteins in eukaryotes" (2019), *Molecular Biology of the Cell,* vol. 30, issue 15,

1846-1863. The analysis presented in this section confirms and extends this previous analysis, including more relevant alveolate taxa.

To Fernanda,

for your love and support in all things

Acknowledgments

In the course of the last eight years, both as an undergraduate and graduate student, working as a member of the Dacks lab, it has been my great pleasure to work and interact with a number of very exceptional individuals.

First and foremost, I would like to thank my supervisor Joel Dacks. Joel is one of the most dedicated and understanding scientists I have ever met. From the first days of sitting at the back of his CELL 310 class to the current day, he has vastly expanded my idea of what constitutes "Biology". As Theodosius Dobzhansky wrote, "nothing in biology makes sense except in the light of evolution"; I could not agree more. I thank Joel for his mentorship, especially for allowing me to take charge over my research projects and encouraging me to explore new areas and methods. There is little doubt in my mind that I would not be the scientist I am today without his guidance.

I want to thank all of the members of the Dacks lab, both past and present, for their assistance and insight throughout these last eight years. It has been a pleasure working alongside you. I want to extend a huge thank you to Alex Schlacht and Emily Herman, for providing guidance and support during the tumultuous beginnings of a graduate degree. I would also like to especially thank Beth Richardson, for including me in your ongoing obsession with dinoflagellate plastid genomics, and for being a wonderful friend both inside and outside the lab.

Over the course of my graduate studies I had the privilege of spending over two years in the lab of our collaborator Markus Meissner at the University of Glasgow learning how to culture and genetically manipulate *Toxoplasma*. I am incredibly grateful to Markus and the whole Meissner group for welcoming "the Canadian" into their group. Thank you to Markus for acting as a secondary supervisor while I was in the lab, to Matthew Gow for all the cells you split and for all the gigs we attended, to Elena Jimenez-Ruiz and Simon Gras, for helping me through the initial confusion and teaching me a great deal about molecular biology, to Mario del Rosario, for your willingness to discuss literally anything at all times, and to Leandro Lemgruber, for answering all my microscopy questions and for your EM expertise. Finally, I would like to thank all the scientists who shared materials with the Meissner lab, both those that I used and those that I did not; science is at its best when collaboration trumps competition.

Over the course of my graduate degree I ended up switching departments; I would like to thank all the members of the Cell Biology Department, and in particular the office staff, for their assistance and insight. Thank you to the Department of Medicine for accepting me during a trying time, and for making the process as painless as possible.

I would especially like to thank my supervisory committee members, both past and present: Gary Eitzen, Stephanie Yanow, and Paige Lacy. Your continued input has always been a source of both knowledge and motivation. I would like to thank those individuals who sat on my candidacy committee, Richard Rachubinski and Kinga Kowalewska-Grochowska, for helping to make a stressful examination a surprisingly enjoyable event. Special thanks to John Parkinson for agreeing to be the external examiner for my defence and to my entire examining committee for critically reading this thesis.

Over the last eight years I have had the pleasure of meeting, and in some cases directly working with, many amazing people. There are too many names to mention them all here, but I would like to especially thank a few of them. Thank you to Lucas Paoli; although you were not in the Dacks lab long, I always look forward to our conversations regarding both life and science. Thank you to Richard Dorrell, for inspiring me to always perform the most thorough analysis possible in my work. A special thank you to Kannan Venugopal, both for your support and for many insightful discussions both in the lab and in the pub. I would also like to thank those individuals who openly welcomed me to visit their labs while in the UK: Anastasios Tsaousis, Chris Howe, and Ross Waller. Thank you to those members of the Wellcome Centre for Integrative Parasitology who were always open to discussion.

This work would not have been possible without funding from numerous agencies: thank you to the Women and Children's Health Research Institute (WCHRI), Alberta Innovates Health Solutions (AIHS), and the Canadian government through the Canadian Institutes for Health Research (CIHR) for helping to fund this work.

Finally, I have to thank those that have made this journey possible through their support outside the lab. To my parents, thank you for raising me to know the value of hard work, for your support in all things, and for your interest in my research. To my brother, thank you for always being a friend first and a brother second, and for putting up with me during the hard times. To my friends, both inside and outside the lab, thank you for reminding me that there are more important things in life than what goes on at the bench.

Finally, I give my most sincere thanks to my partner Fernanda. From our first meeting at a party in a basement flat in Glasgow's west end until today, you remain a constant source of support and inspiration. You are an amazing scientist and an even more amazing person; your smile brightens every room, your personality infects even the dourest of individuals, and your presence makes even the most insurmountable problems seem manageable. This thesis would not have been possible without you.

Table of Contents

List of Tables ..... *xv*

List of Figures ..... *xvi*

List of Abbreviations ..... *xviii*

1. Introduction *1*

1.1 A brief introduction to the introduction 2

1.2 Eukaryotic diversity and phylogeny 2

1.3 The Membrane-trafficking system (MTS) 7

1.3.1 MTS gene families 8

1.3.1.1 Arf, Sar, and Arl G proteins and their regulators 8

1.3.1.2 Adaptor proteins (APs) 11

1.3.1.3 COPI, COPII, Clathrin, and Retromer 12

1.3.1.4 Rab GTPases and their regulators 13

1.3.1.5 Tethering factors 13

1.3.1.6 SNAREs and SM proteins 15

1.3.1.7 ESCRTs 15

1.3.2 The MTS and eukaryotes 16

1.4 Apicomplexan parasites 16

1.4.1 Apicomplexan evolution and chromerid algae 16

1.4.2 Brief overview of molecular research into apicomplexans 19

1.5 *Toxoplasma gondii*: the model apicomplexan 20

1.5.1 *T. gondii* ultrastructure 20

1.5.1.1 The apical complex and Inner Membrane Complex 21

1.5.1.2 The nucleus, endoplasmic reticulum, and Golgi complex 24

1.5.1.3 The endosomal system 24

1.5.1.4 Micronemes 24

1.5.1.5 Rhoptries 25

1.5.1.6 Dense granules 25

1.5.1.7 Endosymbiotic organelles: apicoplast and mitochondrion 25

1.5.2 Lifecycle of *T. gondii* 26

1.5.3 *T. gondii* prevalence and pathogenesis 32

1.5.4 Genetic manipulation of *T. gondii* 33

1.6 Membrane trafficking in Apicomplexa 38

1.6.1 Coding Complement of Apicomplexa 38

1.6.1.1 Rab and Arf GTPases 38

1.6.1.2 Adaptor Proteins and Cargo Adaptors 40

1.6.1.3 Coats 40

1.6.1.4 Tethers 40

1.6.1.5 SNAREs and ESCRTs 41

1.6.2 Organization, Markers, and Function of the Apicomplexan MTS 41

1.6.2.1 Overview 41

1.6.2.2 Endosymbiotic organelles – apicoplast and mitochondrion 42

1.6.2.3 ER, Golgi, and IMC 42

1.6.2.4 The Dense Granules 43

1.6.2.5 The Endosomal System 43

1.6.2.6 A Model for the Apicomplexan MTS 44

1.7 Mechanisms of MTS evolution 50

1.8 Rationale and hypothesis 51

2. Chromerid genomes and the evolution of Apicomplexa 55

2.1 Chromerids as free-living apicomplexan relatives 56

2.2 Endomembrane trafficking system 57

2.2.1 Materials and Methods 57

2.2.2 Results and discussion 58

2.2.2.1 ESCRT machinery 61

2.2.2.2 APs 61

2.2.2.3 MTCs 62

2.3 Additional discussion 64

3. Exploring the relationship between gene homology and function 70

3.1 Preface 71

3.2 Resolving the homology-function relationship through comparative genomics of membrane-trafficking machinery and parasite cell biology 73

3.2.1 Introduction 73

3.2.1.1 The membrane-trafficking system: a modern molecular view 74

3.2.1.2 Evolution of membrane-trafficking: LECA complement and modern innovations 77
3.2.2 Emerging model organisms 78
3.2.3 Examining the case for functional homology 80
3.2.4 Functional homology in trafficking machinery between divergent organisms 80
3.2.4.1 Adaptor proteins 80
3.2.4.1.1 AP-1 80
3.2.4.1.2 AP-2 84
3.2.4.1.3 Functional homology in adaptor proteins 85
3.2.4.2 The ESCRT complexes 85
3.2.4.2.1 Functional homology in ESCRT complexes 87
3.2.4.3 Retromer 87
3.2.4.3.1 Functional homology in retromer 89
3.2.4.4 Rab GTPases 89
3.2.4.4.1 Rab5 90
3.2.4.4.2 Rab7 91
3.2.4.4.3 Rab11 92
3.2.4.4.4 Functional homology in Rab GTPases 94
3.2.5 Discussion 95
3.2.5.1 Overview 95
3.2.5.2 Functional homology of trafficking machinery in diverse eukaryotes 98
3.2.5.3 Evolutionary precedent of conserved and novel features 99
3.2.5.4 Conclusions and future perspectives 100
3.3 Additional discussion 100
4. An Informatics Screen to Identify Novel Paralogues in Apicomplexa 104
4.1 An argument for novel paralogues in Apicomplexa 105
4.1.1 Loss: the common pattern observed previously 105
4.1.2 Apicomplexa possess organelles not found in other eukaryotes 105
4.1.3 Rationale and hypothesis 107
4.2 Materials and Methods 108
4.3 An informatics screen to identify novel paralogues 109
4.4 Results 110
4.4.1 Arf regulators: ArfGEFs and ArfGAPs 113

4.4.2 SM proteins 116
4.4.3 SNAREs 119
4.4.3.1 Qa-SNAREs 119
4.4.3.2 Qb- and Qc-SNAREs 126
4.4.3.3 R-SNAREs 133
4.4.4 Rab GTPases 140
4.4.4.1 Rab1-related GTPases 143
4.4.4.2 Rab5-related GTPases 148
4.4.4.3 Rab11-related GTPases 153
4.4.4.4 RabX1, an uncharacterized myzozoan-specific Rab 153
4.4.5 RabGAPs (TBCs) 156
4.4.5.1 TBC-N and TBS 161
4.4.5.2 TBC-PI and TBC-X2 166
4.4.5.3 TBC-Q and other novel TBC LSPs 166
4.4.6 Arf and Arl G proteins 166
4.5 Discussion 176
5. Characterization of three novel Arl proteins in T. gondii 187
5.1 Overview of novel paralogue characterization in *T. gondii* 188
5.2 Materials and Methods 189
5.2.1 Parasite and host cell culture, and transfection 189
5.2.2 Genomic DNA Isolation, cloning, and PCR 190
5.2.3 Induction of the ArlX3 inducible knockdown 191
5.2.4 Immunofluorescence analysis (IFA) 191
5.2.5 Western blotting and protein detection 192
5.2.6 Structured illumination microscopy (SIM) 192
5.2.7 Quantitative fluorescence microscopy 192
5.2.8 Electron Microscopy 193
5.2.9 Five-day plaque assay 193
5.2.10 Gliding assay 194
5.2.11 Invasion assay 194
5.2.12 Egress assay 194
5.2.13 Marker mis-localization analysis 194
5.2.14 Statistical analysis 195
5.3 Results 195

5.3.1 Endogenous tagging and localization of Arl LSPs in *T. gondii* 195
5.3.2 Generation of an inducible TgArlX3 knockdown line 204
5.3.3 TgArlX3 is crucial for asexual growth 214
5.3.4 TgArlX3 knockdown impairs each step within the lytic cycle 214
5.3.5 TgArlX3 knockdown causes mis-localization of resident microneme and rhoptry proteins 220
5.4 Discussion 224
6. Final Discussion 232
6.1. Overview 233
6.2. The organelle paralogy hypothesis 233
6.3. Functional homology 234
6.4. The OPH in the FECA-LECA transition and extant eukaryotes 235
6.5. Apicomplexa as a model to study the OPH in extant eukaryotes 235
6.6 Identification of novel MTS paralogues in Apicomplexa 236
6.7 Novel Arl paralogues: an explicit test of the OPH? 239
6.8 An overall picture of apicomplexan MTS evolution 242
6.9 A philosophical plea for continued study of organellogenesis 242
6.10 Proposals for future work 243
6.11 Conclusions 244
References 246
Appendix 1: Chapter 4 supplementary material 300
Appendix 2: Chapter 5 supplementary material 304
Online Appendix 319

List of Tables

Table 3.1 Functional homology across model systems 96

Table 4.1 Novel Arl Paralogues in T. gondii. 180

List of Figures

Figure 1.1 Current view of eukaryotic phylogeny 4

Figure 1.2 Overview of trafficking steps 9

Figure 1.3 Alveolate phylogeny 17

Figure 1.4 Schematic of a T. gondii tachyzoite 22

Figure 1.5 The asexual lytic cycle of T. gondii 27

Figure 1.6 Replication by endodyogeny 30

Figure 1.7 Overview of genetic manipulation techniques in T. gondii 35

Figure 1.8 Proposed model of T. gondii endosomal trafficking 48

Figure 1.9 Model for the role of gene duplication in organellogenesis 52

Fig. 2.1. An overview of endomembrane trafficking components 59

Figure 3.1 Model Organisms Across Eukaryotes 75

Figure 3.2 Function of select membrane-trafficking machinery in a model endomembrane system 81

Figure 4.1. Phylogenetic workflow to identify lineage-specific paralogues 111

Figure 4.2 ArfGEF and ArfGAP conservation 114

Figure 4.3 SM protein conservation 117

Figure 4.4 Qa-SNARE conservation 120

Figure 4.5 Qa-SNARE phylogeny 122

Figure 4.6 Detailed phylogeny of Stx2 and Stx12 124

Figure 4.7 Qb- and Qc-SNARE conservation 127

Figure 4.8 Qb-SNARE phylogeny 129

Figure 4.9 Qc-SNARE phylogeny 131

Figure 4.10 R-SNARE conservation 134

Figure 4.11 Detailed phylogeny of VAMP sequences 136

Figure 4.12 Detailed phylogeny of Ykt6 sequences 138

Figure 4.13 Rab conservation 141

Figure 4.14 Rab phylogeny 144

Figure 4.15 Detailed phylogeny of Rab1-related sequences 146

Figure 4.16 Secondary and tertiary structure of Rab1K 149

Figure 4.17 Detailed phylogeny of Rab5-related sequences 151

Figure 4.18 Detailed phylogeny of Rab11-related sequences 154

Figure 4.19 Detailed phylogeny of RabX1-related sequences 157

Figure 4.20 TBC conservation 159

Figure 4.21 TBC phylogeny 162

Figure 4.22 Detailed phylogeny of TBC-N-related sequences 164

Figure 4.23 Detailed phylogeny of TBC-PI/X2 sequences 167

Figure 4.24 Detailed phylogeny of TBC-Q-related sequences 169

Figure 4.25 Arl conservation 172

Figure 4.26 Arl phylogeny 174

Figure 4.27 Detailed phylogeny of ArlX1/X2/X3 177

Figure 5.1 Endogenous gene tagging of Arl LSPs in T. gondii 197

Figure 5.2 Detailed localization of ArlX1 200

Figure 5.3 Cryo-immunogold electron microscopy of ArlX1 202

Figure 5.4 Detailed localization of ArlX3 205

Figure 5.5 Cryo-immunogold electron microscopy of ArlX3 207

Figure 5.6 Generation of an inducible ArlX3 knockdown line 210

Figure 5.7 ArlX3 knockdown kinetics 212

Figure 5.8 ArlX3 is crucial for T. gondii asexual growth 215

Figure 5.9 Effect of ArlX3 knockdown on the T. gondii lytic cycle 218

Figure 5.10 Effect of ArlX3 knockdown on localization of select organelle markers 221

Figure 5.11 Proposed model of T. gondii endosomal trafficking including ArlX3 228

List of Abbreviations

AMA1 – apical membrane antigen 1
ANOVA – analysis of variance
AP – adaptor protein
ARF – ADP-ribosylation factor
ARFRP – ADP-ribosylation factor-related protein
Arl – ARF-like
ATc – anhydrotetracycline
BB/E – basal body/extracellular
BFA – brefeldin A
BLAST – basic local alignment search tool
BSA – bovine serum albumin
CAS9 – CRISPR-associated 9
CAT – chloramphenicol acetyl transferase
CATCHR – complexes associated with tethering containing helical rods
COP – coatomer protein
COG – conserved oligomeric Golgi
CRISPR – clustered regularly interspaced short palindromic repeats
DAPI – 4’,6-diamidino-2-phenylindole
DHFR-TS – dihydrofolate reductase-thymidylate synthase
DiCre – dimerizable cre recombinase
DNA – deoxyribonucleic acid
DrpB – dynamin-related protein B
DSB – double-strand DNA break
ELC – endosome-like compartment
ER – endoplasmic reticulum
ERD – ER-retention deficient
ESCRT – endosomal sorting complex required for transport
F-actin – filamentous actin
FOV – field of view
GalNAc - UDP-GalNAc:polypeptide N-acetylgalactosaminyl-transferase
GAP – GTPase activating protein
GAP(#) – gliding-associated protein
GDP – guanosine diphosphate

GEF – guanine nucleotide exchange factor
GFP/YFP/RFP – green/yellow/red fluorescent protein
GOI – gene of interest
GRASP – Golgi reassembly-stacking protein
GTP – guanosine triphosphate
HFF – human foreskin fibroblast
HMM – hidden Markov model
HSD – honest significance difference
HXGPRT – hypoxanthine-xanthine-guanine phosphoribosyl transferase
iEM – cryo-immunogold electron microscopy
IFA – immunofluorescence analysis
IMC – inner membrane complex
KD – knockdown
KO – knockout
LECA – last eukaryotic common ancestor
LIC – ligation-independent cloning
LRO – lysosome-related organelle
LSP – lineage-specific paralogue
M2AP – MIC2-associated protein
(m)AID – (mini) auxin-inducible degron
MIC – microneme protein
MIP – maximum intensity projection
MTC – multi-subunit tethering complex
MTS – membrane-trafficking system
MVB – multi-vesicular body
NHEJ – non-homologous end joining
OPH – organelle paralogy hypothesis
PBS – phosphate-buffered saline
PCR – polymerase chain reaction
PLV – plant-like vacuole
Rab – Ras-like proteins from rat brain
ROI – region of interest
RON – rhoptry neck protein
ROP – rhoptry bulb protein
SAG – surface antigen

SAR – Stramenopiles, Alveolates, and Rhizaria

SM – Sec1/Munc18-like

SNARE – soluble N-ethylmaleimide-sensitive factor attachment protein receptor

Stx – syntaxin

TATi – trans-activator trap identified 1

TEM – transmission electron microscopy

TetR – tetracycline repressor

TGN – *trans*-Golgi network

UPRT – uracil phosphoribosyl transferase

VAC – *Toxoplasma* vacuole

Vps – vacuolar protein sorting

1. Introduction

1.1 A brief introduction to the introduction

This thesis is intended to cover a wide array of topics in order to fully address the overarching questions and hypotheses laid out at the end of this chapter. In order to provide sufficient background, this chapter begins with an overview of eukaryotic diversity, describing major groups currently recognized, based primarily on data from molecular phylogenetic studies. Following this, an overview of the eukaryotic membrane-trafficking system, including the basic mechanisms by which material moves between cellular compartments and descriptions of some of the machinery involved, is provided. Following this, the focus shifts to a group of enigmatic parasites, the Apicomplexa, wherein the model organism *Toxoplasma gondii*, used for characterization of several genes in Chapter 5, resides. The biology and utility of *T. gondii* as a model organism are discussed. A detailed view of the current state of knowledge regarding membrane trafficking in the Apicomplexa, with emphasis on *T. gondii*, is presented. Finally, a mechanistic model for the emergence of organellar complexity is described and the main aims and hypotheses of the work are described.

As this thesis focusses on evolutionary mechanisms involved in generating the complexity of endomembrane compartments within eukaryotes, this chapter begins with a discussion of eukaryotic diversity.

1.2 Eukaryotic diversity and phylogeny

Increases in the capacity of environmental sampling, culturing, and sequencing have vastly increased knowledge of eukaryotic diversity, including that of unicellular eukaryotes (protists), which comprise a large swathe of currently appreciated diversity. Traditional systematics divided eukaryotes into "basal" and "crown" groups, suggesting that eukaryotes radiated in a ladder-like fashion with more "primitive" organisms, such as parasites, branching prior to more complex organisms such as humans[1–3].

This classification was supported by other observations that appeared to reinforce the notion that some eukaryotes were primitive. This included the observed lack of stacked Golgi bodies and peroxisomes, as well as a paucity of introns and odd ribosome/ribosomal RNA features, in some taxa[4,5]. However, the primary piece of evidence used was the apparent lack of mitochondria in some basal taxa: archamoeboids, metamonads, microsporidians, and parabasalids; together, these observations gave rise to the "archezoa" hypothesis that posited a late acquisition of the mitochondrion into eukaryotes, after the branching off of these amitochondriate lineages[6,7].

It was subsequently shown that the observed phylogenetic relationships were artefacts due to low taxon sampling and the use of simplistic models of evolution for phylogenetic

reconstruction (so-called "long-branch attraction")[8]. Modern phylogenetic methods have evolved to overcome these early limitations, including the widespread adoption of robust statistical frameworks for inference (maximum likelihood[9,10] and Bayesian[11]) and the ability of phylogenetic models to account for evolutionary rate variation across sites[12]. Additionally, relationships between lineages are now commonly inferred from the concatenation of multiple genes with consistent phylogenetic signal (commonly referred to as "phylogenomics") rather than a single gene, greatly increasing the amount of data available for exploring internal nodes within phylogenies[13].

Improved phylogenies placed the archezoan taxa within clades of organisms possessing canonical mitochondria, confirming that their "amitochondriate" state was secondarily derived[14]. Furthermore, enigmatic organelles, generally referred to as either hydrogenosomes or mitosomes, were discovered in almost all amitochondriate lineages and shown to be divergent mitochondria[15]. Although at least one example of a eukaryote that truly lacks a mitochondrion has been described[16], it is clear that this represents a secondary loss. Together, these considerations have led to modern eukaryotic systematics, which combine large-scale phylogenomic inference, smaller focussed analyses, and morphological characters to provide a framework for classification.

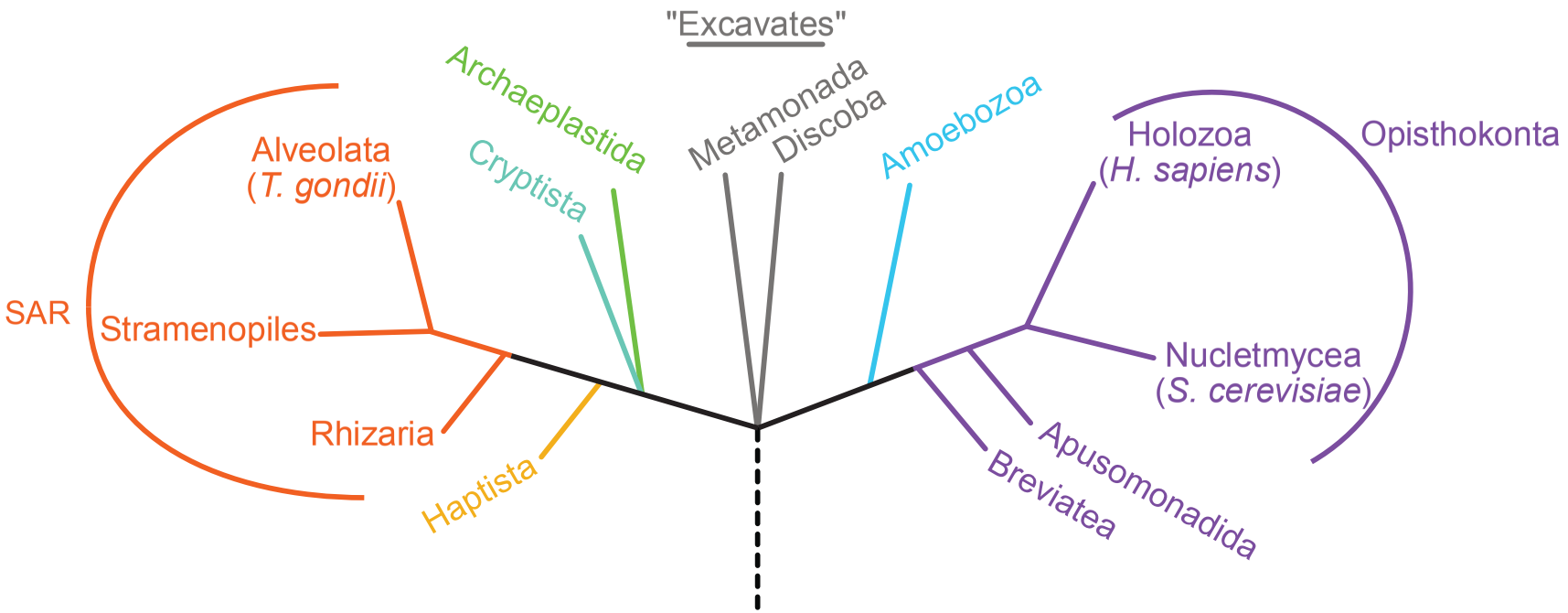
1.2.1 Modern view of eukaryotic diversity

Community-wide efforts have recently culminated in an updated view of eukaryotic diversity (Figure 1.1). Modern classification efforts are hierarchical, but without relying on strict orders or ranks, and places emphasis on monophyly as assessed by molecular phylogenetics[17].

A previous classification recognized five supergroups, together with several groups whose phylogenetic affinity was uncertain and multiple *incertae sedis*, taxa whose affiliation was largely unknown[18]. These groups appear stable, as they remain in the most recent taxonomy, with the exception of the Excavata, whose monophyly has never been strongly supported (discussed below). Still, the majority of known eukaryotes fall within one of seven large groups: Obazoa, Amoebozoa, Discoba, Metamonada, Archaeplastida, Haptista, and Stramenopiles, Alveolates, and Rhizaria (SAR, Figure 1.1). In order to understand the significance of character evolution both within smaller groups and across eukaryotes, a basic description of each group is provided below (information based on Adl et al. (2012), (2019), *inter alia*[17,18]).

Figure 1.1 Current view of eukaryotic phylogeny

This figure provides an overview of the relationships among taxa referenced in this thesis. Large assemblages ("supergroups") are indicated and colour-coded: Obazoa (purple; note the reference to Opisthokonta as encompassing Holozoa and Nucletmycea), Amoebozoa (cyan), "Excavates" (grey; informal designation), Archaeplastida (green), and SAR (orange, an abbreviation for stramenopiles, alveolates, and rhizarians). In addition, well-supported groups of Cryptista (teal) and Haptista (yellow) are shown. For orientation, the location of common organisms, including *Homo sapiens* and *Saccharomyces cerevisiae* (both opisthokonts), and of *Toxoplasma gondii* (a parasitic alveolate important to this thesis) are shown. This phylogeny is based on that shown in Figure 1 of Adl et al. (2019), which itself is based on a synthesis of a large volume of phylogenomic studies performed by the community.


"Excavates"
Archaeplastida
Cryptista
Metamonada
Discoba
Amoebozoa
Holozoa
(H. sapiens)
Opisthokonta
Alveolata
(T. gondii)
SAR
Stramenopiles
Rhizaria
Haptista
Nucletmycea
(S. cerevisiae)
Apusomonadida
Breviatea

In general, eukaryotes are grouped into two main domains, Amorphea and Diaphoretickes. Amorphea, previously referred to as Unikonta, comprises the Obazoa and Amoebozoa. The Obazoa includes the Opisthokonta, which itself comprises animals and their unicellular relatives (Holozoa), along with fungi (Nucletmycea), as well as several recently described basal taxa such as apusomonads and breviates[19].

Obazoa by and large contain a single posteriorly inserting cilium and have mitochondria with flat cristae. Metazoan body forms differ drastically, though some cell types bear a collar of cilia similar to choanoflagellates. Basal Holozoa like ichthyosporids and Filasterea are amoeboid or contain long tapering pseudopodia and some, like *Capsaspora owczarzaki,* parasitize animals. Fungi are generally mycelial and contain chitinous cell walls, though there is a broad range of morphologies. Parasites are found in numerous lineages. Microsporidia are basal fungi which parasitize animals and contain mitosomes. Apusomonads are biflagellated with an organic sheath covering the anterior end of the cell. Breviates are amoeboid flagellates with unusual mitochondrial structures.

Amoebozoa contain a large number of amoeboid organisms, along with the enigmatic slime moulds. Cells are often naked, though some are testate, mitochondrial cristae are usually tubular but some irregular branching structures have been observed. Tubulinea and Discosea contain pseudopodia with either unidirectional or polydirectional cytoplasmic flow. Slime moulds like myoxogastrids often accumulate to form unique stalk and body structures. Archamoebae are flagellated and contain hydrogenosomes/mitosomes; *Entamoeba histolytica* is an important human parasite.

SAR, Archaeplastida, Haptista, and Cryptista together comprise the Diaphoretickes. In general, this group is well-resolved, but the placement of Cryptista is uncertain, branching basally to the archaeplastids in some analyses[20]; for simplicity, this branch is collapsed and hence no absolute claims regarding this relationship are inferred (Figure 1.1).

The Archaeplastida contain land plants, mosses, lichens, and green algae, all united by the acquisition of a plastid through a single primary endosymbiotic event with a cyanobacterium. The group can be roughly divided into the green algae, red algae, and the glaucophytes, or blue-green algae, which branch basally and contain numerous unique features, such as a peptidoglycan layer in the plastid membrane. Rhodophyceae include the Cyanidiales, extremophile heterotrophs or photoautotrophs found in acidic or high temperature environments, as well as the Bangiales, pluricellular red algae which are among the earliest eukaryotic fossils known. Green algae of the Chlorophyta include colonial organisms like *Volvox carteri*, as well as the human parasite *Prototheca*. Charophyta includes a large diversity of organisms, including multicellular plants.

SAR contains three main lineages, the stramenopiles, alveolates, and rhizarians. The rhizarians diverged first and are arguably the least understood of the major groups of eukaryotic diversity. Rhizarian morphology varies drastically, including amoeboid and flagellated forms; foraminiferans possess filose pseudopodial networks, radiozoans have axopodia that erupt from organic capsules surrounding the cell, and chlorarachniophytes have corkscrew cilia. Stramenopiles are equally diverse, spanning from unicellular heterotrophs and photoautotrophs to multicellular brown algae, pathogens of plants and humans, and the calcaceous or silicaceous tests of diatoms. Notable members of the stramenopiles include *Blastocystis*, an opportunistic pathogen of humans, along with *Phytophthora spp.*, species of which cause potato blight and sudden oak death. Alveolates are unified by the presence of membranous sacs (alveoli) subtending the plasma membrane, but are otherwise diverse, both morphologically and trophically; alveolates will be discussed in more detail in section 1.4.1.

Finally, there are groups of organisms which do not fall into either domain, Metamonada, Discoba, and Malawimonada, that were formerly known as "Excavata" (and still informally known as excavates). These represent a diverse group of predominantly heterotrophic flagellates, many of which live in oxygen poor environments and/or are important parasites. Metamonads are generally endosymbionts/parasites and lack classical mitochondria, while discobids are generally free-living and have classical mitochondria. Additionally, the euglenozoans contains the only photosynthetic excavate, *Euglena gracilis*.

One of the key features common to all eukaryotic cells is the presence of an elaborate set of membrane-bound compartments, and the cellular machinery that mediates movement between them, which is described in the next section.

1.3 The Membrane-trafficking system (MTS)

The MTS comprises the distinct intracellular organelles within eukaryotic cells together with the protein machineries mediating trafficking between them. In most cells, this includes the endoplasmic reticulum (ER), one or more Golgi complexes, a population of endosomes, a terminal compartment for digestion (vacuole/lysosome), the plasma membrane, and various other organelles such as peroxisomes and acidocalcisomes. As such, the MTS is responsible for many essential cellular functions linked to metabolism and homeostasis, as well as uptake from, and secretion to, the extracellular environment.

Movement between organelles is achieved through trafficking steps; although the participants in such steps vary in terms of morphology, the following description uses the most common form, the vesicle. Trafficking can be broken down into the following steps: 1) recruitment of soluble and trans-membrane cargo into a nascent forming vesicle at the donor

compartment, 2) recruitment of coat proteins that induce membrane curvature and budding, 3) scission of the vesicle from the donor compartment, 4) movement of the vesicle from the donor to acceptor compartment, 5) initial long-range tethering of the vesicle to the acceptor compartment, and 6) closer apposition of the vesicle/acceptor followed by 7) SNARE-mediated membrane fusion and cargo release; at some point between scission (3) and fusion (7), the vesicle coat is shed[21] (Figure 1.2). These steps are carried out by members of large gene families, most of which are paralogous.

1.3.1 MTS gene families

Despite the complexity of trafficking pathways required to maintain organelle identity, the individual steps are mediated by and large by members of large paralogous gene families, wherein each paralogue performs a similar function at a distinct intracellular location (discussed in section 1.7). Other sets of MTS machinery do not represent such paralogous families, although some still contain paralogous proteins, and instead contain numerous individual factors that are not detectably related to one another. Select MTS families are discussed below, in relative order of the trafficking step in which they take part; the ESCRT machinery, which does not easily fall under the scheme outlined in Figure 1.2, is described last. For the sake of brevity, none of the descriptions are exhaustive; in the case of some machinery, additional detail, including the function of specific paralogues/subunits, is discussed throughout this thesis.

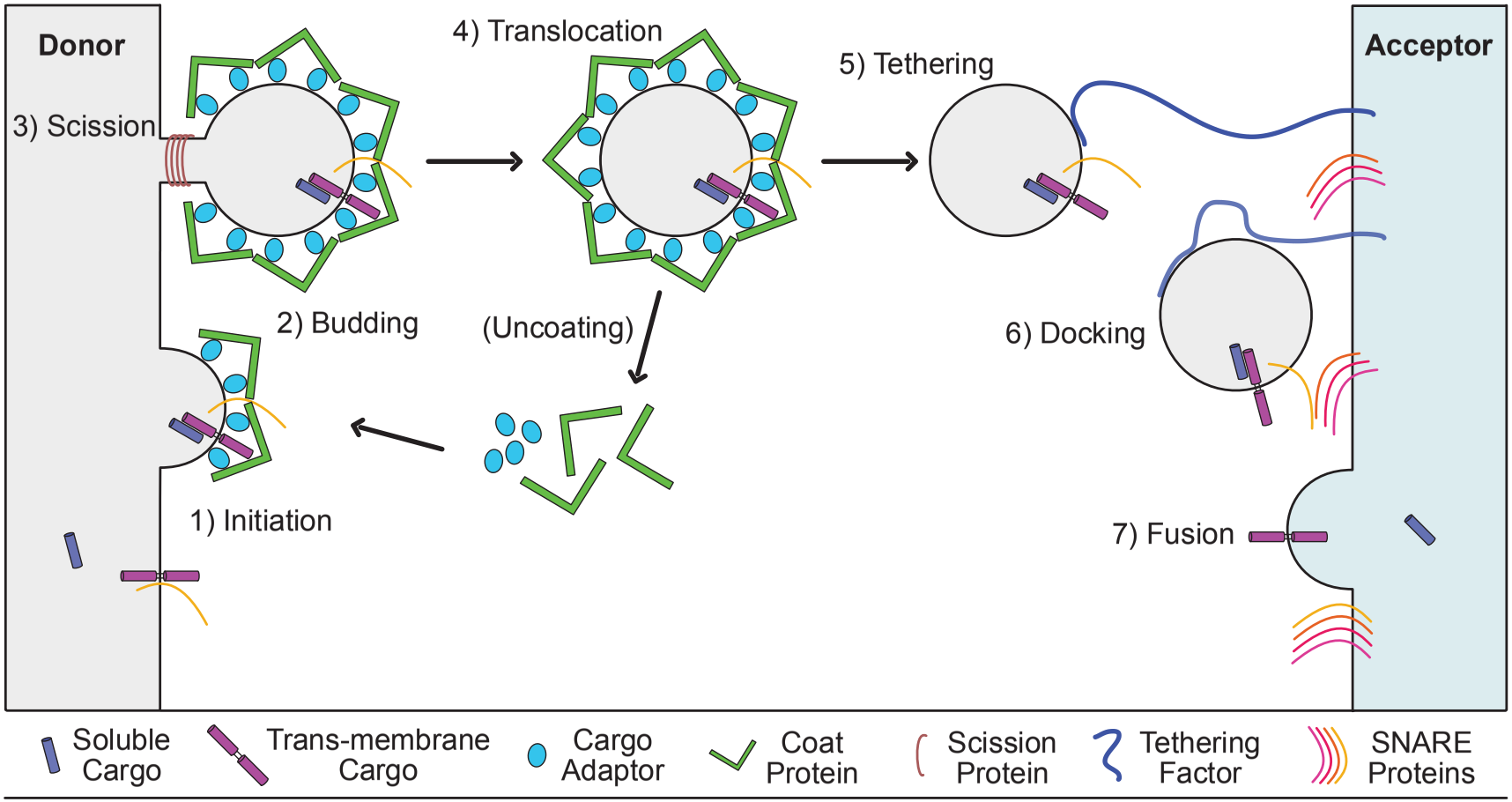
1.3.1.1 Arf, Sar, and Arl G proteins and their regulators

The Arf family of small G proteins includes the ADP-ribosylation factor (ARF) proteins, ARF-like (Arl) proteins, and Sar1, whose members are characterized by the presence of an N-terminal amphipathic helix that is exposed to allow membrane insertion when the G protein is bound to GTP[22,23]. Together with Rab family proteins (discussed below), as well as the Ras, Rho, and Ran families, these form the Ras superfamily[24,25]. Like other G proteins, Arf family members cycle between GTP- and GDP-bound forms, assisted by guanine nucleotide exchange factors (GEFs, which facilitate GDP release) and GTPase-activating proteins (GAPs, which facilitate GTP hydrolysis); these regulatory proteins are especially important for Arf family members, which lack appreciable GTPase activity on their own[26].

Arf family members possess an N-terminal amphipathic helix to mediate membrane association; residues in the N-terminus are also frequently subject to myristoylation or, in some members, acetylation, to aid in membrane association[27–30]. Therefore, unlike Rab proteins, which are C-terminally prenylated and function ~7-8 nm away from membranes,

Figure 1.2 Overview of trafficking steps

This figure provides a cartoon depiction of the basic steps involved in membrane trafficking between a donor (grey) and acceptor (light blue) compartment. Vesicle formation begins at the donor membrane with concentration of trans-membrane and soluble cargo via cargo adaptors, which subsequently recruit coat proteins (1, initiation). Coat proteins, in concert with other membrane deformation proteins, result in increasing curvature and budding of a nascent vesicle from the donor membrane face (2, budding). Eventually, the vesicle is separated from the donor membrane, for example by dynamins (3, scission), and travels from the donor to acceptor compartment (4, translocation). At some point between scission and eventual fusion at the acceptor compartment, the vesicle undergoes uncoating, with the cytosolic factors free to initiate new rounds of vesicle formation (uncoating). At the acceptor compartment, the vesicle first undergoes long range tethering (5, tethering), which brings it in close apposition to the acceptor membrane (6, docking). Eventually, SNARE molecules on both the vesicle and acceptor membrane form a complex and induce membrane fusion, releasing cargo (7, fusion). The identity of all depicted components is provided in the legend at the bottom of the figure. Adapted from Bonifacino and Glick (2004)[21]


Donor
Acceptor
3) Scission
4) Translocation
5) Tethering
2) Budding
(Uncoating)
6) Docking
1) Initiation
7) Fusion
Soluble Cargo
Trans-membrane Cargo
Cargo Adaptor
Coat Protein
Scission Protein
Tethering Factor
SNARE Proteins

Arf family proteins function at the membrane face; consequently, Arf family effectors are often involved in changing the lipid composition of membranes, budding of vesicles and tubules from the membrane, and maintaining the structural integrity of organelles[26]. Hence, Arf family proteins, like Rabs, are involved in diverse processes including COPI and clathrin vesicle formation, Golgi structure and trafficking, endocytosis and plasma membrane recycling, and ciliogenesis[27].

Little comparative genomic analysis has been performed for Arf family members across eukaryotes. ARFs can be classified into Class I (ARF1-3), Class II (ARF4 and ARF5), and Class III (ARF6), yet organisms outside of the opisthokonts often encode only a single ARF homologue[31,32]. Similarly, around 20 Arls have been described, but only a few (Arl1, Arl2, Arl3, Arl5, Arl8, and ARFRP1) have been described outside of opisthokonts[31] and others (Arl4, Arl9, Arl10, Arl11, Arl14, and Arl15/ARFRP2) are reported to be restricted to this group[26]. Sar1 appears universally conserved across eukaryotes[31,33].

Arf GEF activity is associated with the ~200 amino acid long Sec7 domain[34,35], which defines a family of Arf GEF proteins with at least three ancient members (BIG, GBF, and Cytohesin)[36]. Similarly, most known Arf GAPs share a common GAP domain with a zinc finger to mediate GTP hydrolysis[37], and at least six of these (ArfGAP1, ArfGAP2/3, SMAP, ACAP, AGFG, and ArfGAPC2) are presumed ancient[32]. However, not all GEF/GAP proteins acting on Arf family members fall into these families; for example, components of the COPII coat act as both GEF and GAP for Sar1[38,39]. Similarly, the GEFs and GAPs for Arls remain poorly described. In yeast, the canonical Arf GEF Syt1p and GAP Gcs1p are reported to regulate nucleotide exchange on Arl1[40,41], whereas non-homologous proteins ELMOD1/ELMOD2 and RP2 are GAPs for ARFs/Arl2 and Arl3, repectively[42–44].

1.3.1.2 Adaptor proteins (APs)

Adaptor protein (originally termed "assembly protein"[45]) complexes are heterotetrameric complexes linking cargo and coat recruitment during the process of vesicle formation[46,47]. There are five complexes conserved across eukaryotes[48], and they are related to an additional coat-like complex TSET as well as the COPI coat[49]. Each comprises two large subunits (called, respectively, γ, α, δ, ε, and ζ, and β1-5, in the AP-1 through AP-5 complexes), one medium subunit (μ1-5), and one small subunit (σ1-5). Their role in diverse post-Golgi trafficking events is supported by the ability of AP-1, AP-2, and possibly AP-3 to bind clathrin[45,50,51]; the coats for the other two AP complexes are not well-defined, although

AP-5 has been proposed to form a coat with the hereditary spastic paraplegia proteins SPG11 and SPG15[52].

1.3.1.3 COPI, COPII, Clathrin, and Retromer

There are numerous coat and coat-like complexes within eukaryotic cells, which primarily function in the initial steps of vesicle formation by binding to cargo and/or cargo-selective proteins and inducing membrane curvature/tubulation.

The COPII coat is a complex functioning in ER to Golgi transport which comprises seven proteins: the GTPase Sar1 together with its cognate GEF Sec12, Sec13, Sec16, Sec23, Sec24, and Sec31, all of which are universally conserved across eukaryotes and presumed ancient[33,53]. Sec12 activates Sar1 by catalyzing GDP to GTP exchange[38,54,55], which subsequently recruits its cognate GAP and cargo-selective complex, Sec23/24[56–58], followed by Sec13/31[59]. Sec16 plays a role in the regulation of Sar1 GAP activity[60]. Sar1 GTP hydrolysis is thought to be a requirement for eventual vesicle uncoating[61], which occurs in part through changes in phosphorylation state and interaction with the Golgi-resident TRAPPI tethering complex[62–64].

The COPI coat complex instead functions primarily in the reverse pathway, Golgi to ER transport, but also in retrograde traffic through the Golgi itself, and comprises seven coatomer subunits (α-, β-, β'-, γ-, δ-, ε-, and ζ-COP), as well as the GTPase ARF1[65–69]. COPI subunits are detectably similar to APs, as well as the AP-like TSET complex, and are conserved across eukaryotes[49]. Similar to the case with Sec12 and Sar1, the Golgi-resident ArfGEF GBF1 activates ARF1[70], which subsequently recruits COPI to the membrane[69,71]. Cargo binds directly to the coat itself, mediated, at least in part, by N-terminal β-propellers of α- and β'-COP[72,73]. Vesicle uncoating occurs via ARF1 GTP hydrolysis[74], mediated by the Arf GAP proteins ArfGAP1 and ArfGAP2/3[37,75,76].

The clathrin coat comprises only two main subunits, the clathrin heavy and light chain, and regulates a diverse array of post-Golgi trafficking, including a classical role in clathrin-mediated endocytosis[77–83]. Unlike the COPI and COPII coats described above, clathrin does not associate with cargo proteins directly, instead relying on cargo adaptors. The set of cargo adaptors includes not only the AP complexes described in section 1.3.1.2, but also a collection of monomeric adaptors including Dab2, ARH, GGAs, stonins, epsin/epsinR, and CALM/AP180; whereas the clathrin heavy and light chains themselves are conserved across eukaryotes, many of the known monomeric adaptors are restricted within the opisthokonts[84–90]. Clathrin uncoating involves several proteins, including the HSP70 ATPase, which is recruited by auxilin/GAK, and possibly others such as ACK1 and OCRL1[91,92].

Retromer is a coat-like complex comprising both a cargo-selective module consisting of Vps26, Vps29, and Vps35, as well as a BAR domain-containing membrane deformation dimer of Vps5 (SNX1 and SNX2 in mammals) and Vps17 (SNX5, SNX6, and SNX32 in mammals)[93–96]. This pentamer is stable in yeast, but not in mammalian cells[93–95]. First identified for its role in recycling of receptors such as Vps10/sortilin and CI-MPR, retromer mediates endosome to plasma membrane and endosome to TGN recycling of numerous cargoes[97–102]. Cargoes either associate directly with one of the core retromer subunits, or with additional cargo adaptor-like sorting nexins, including SNX3 and SNX27[103–105]. In comparison to the COPI and COPII coats described above, retromer is recruited to membranes via active Rab7[106,107]; subsequent recruitment of the Rab7 GAP TBC1D5 is thought to mediate Rab7 GTP hydrolysis and uncoating[108].

1.3.1.4 Rab GTPases and their regulators

As mentioned in section 1.3.1.1, Ras-like proteins from rat brain (Rab) GTPases possess an extended, hypervariable, C-terminal extension that is C-terminally prenylated by a geranylgeranyltransferase to mediate membrane association[109]. Comparative studies indicate ~23 ancestral Rabs in eukaryotes[110,111], although frequent losses and expansions have been described. Similar to Arf family proteins, Rabs undergo GEF- and GAP-mediated cycles, which control binding to a diverse array of effectors. Through these interactions, Rabs are capable of regulating all steps of vesicular trafficking including cargo loading/coat formation, vesicle translocation, and vesicle tethering/fusion[112]; additional roles have been identified as well, including the ability to influence membrane phosphoinositide composition and modulate actin networks, for example, during cell division[113].

Rab GEFs comprise multiple complexes and small families, none of which are homologous to each other, including the Vps9[114–116] and DENN[117] families, the multi-subunit TRAPP complex[118], the Mon1/Ccz1 complex[119], and the Ric1-Rgp1 complex[120]. Comparatively, with the exception of a GAP for Rab3[121], all known Rab GAPs fall into a family defined by the presence of a Tre-2/Bub2/Cdc16 (TBC) domain, which contributes both an arginine and a glutamine finger to mediate GTP hydrolysis[122]. A comparative study demonstrated the presence of at least 10, and possibly 13, TBC ancient TBC members[123], suggesting that Rab function and regulation across eukaryotes is likely to be complex.

1.3.1.5 Tethering factors

Tethers are required for the initial long-range interaction of vesicles with an acceptor compartment, which brings the vesicle in close apposition to facilitate eventual SNARE-

mediated fusion. Within eukaryotic cells, tethers can be broadly classified into long coiled-coil proteins, for example the well-characterized early endosomal tether EEA1[124,125], and the multi-subunit tethering complexes (MTCs). This thesis only covers MTCs in detail, which are known to be broadly conserved across eukaryotes[126,127].

MTCs are a collection of heteromultimeric protein complexes involved in vesicle tethering at distinct intracellular locations. Based on sequence and structural similarity, MTCs have been divided into three groups: CATCHR complexes, comprising Dsl1, GARP, COG, and exocyst, VpsC complexes, comprising CORVET and HOPS, and the various TRAPP complexes, TRAPPI/II/III[128]; recently, it was proposed that the VpsC complexes are structurally similar to those within the CATCHR family[129].

Of the CATCHR families, Dsl1 is composed of Dsl1, Tip20, and Sec39, and functions to tether retrograde transport vesicles at the ER[130–133]. GARP comprises Vps51/Ang2, Vps52, Vps53, and Vps54, and functions at the TGN[134–137]. COG comprises eight subunits, Cog1-8, and functions in retrograde Golgi trafficking, and hence, maintenance of Golgi morphology as well[138–140]. Similarly, exocyst is also composed of eight subunits, Sec3, Sec5, Sec6, Sec8, Sec10, Sec15, Exo70, and Exo84, referred to as EXOC1-8 in mammals. The complex consists of two modules, or sub-complexes, one containing Sec3, Sec5, Sec6, and Sec8, and the other Sec10, Sec15, Exo70, and Exo84[141–143]. Originally discovered for its role in tethering secretory vesicles[143,144], it has since been implicated in diverse cellular functions including cytokinetic abscission[145], autophagosome formation[146], and ciliogenesis[147].

The VpsC complexes share a core complement of Vps11, Vps16, Vps18, and the SM protein Vps33; CORVET is defined by the additional subunits Vps3 and Vps8, while HOPS contains Vps39 and Vps41. Through interactions with either Rab5 (CORVET) or Rab7 (HOPS), these complexes mediate similar tethering function at early, and late, endosomal compartments, respectively[148–152].

The TRAPP complexes comprise numerous subunits: Bet5/TRAPPC1, Trs20/TRAPPC2, Tca17/TRAPPC2L, Bet3/TRAPPC3, Trs23/TRAPPC4, Trs31/TRAPPC5, Trs33/TRAPPC6, Trs85/TRAPPC8, Trs120/TRAPPC9, and Trs130/TRAPPC10. Current understanding of TRAPP complex organization in yeast suggests a core TRAPPI complex, comprising Bet3, Bet5, Trs20, Trs23, Trs31, and Trs33, to which other subunits join to form the TRAPPII (additionally comprising Tca17, Trs65, Trs120, and Trs130) and TRAPPIII (with the addition of Trs85) complexes; the organization and existence of equivalent distinct TRAPP complexes in mammalian cells remains unresolved[153–160]. In yeast, the distinct TRAPP complexes mediate trafficking at the early Golgi (TRAPPI), late Golgi (TRAPPII), and also function in autophagy (TRAPPIII)[155,161,162].

1.3.1.6 SNAREs and SM proteins

Soluble N-ethylmaleimide-sensitive factor (NSF) attachment protein receptor (SNARE) proteins are short (~200-300aa) coiled-coil proteins, usually attached to membranes through a C-terminal transmembrane domain, involved in membrane fusion throughout eukaryotic cells. SNARE-mediated fusion is thought to occur via the interaction of SNAREs on both the vesicle (within this process referred to as v-SNAREs), and target (t-SNAREs) membranes. The SNARE domains of v- and t-SNAREs interact, "zippering" to form a helical bundle (often referred to as a *trans*-SNARE complex) that leads to the formation of a hemifusion intermediate stage, wherein the two opposing membranes are fused by only one of the two leaflets of the membrane bilayer[163,164]. Molecular simulations suggest that entropic forces cause these *trans*-SNARE complexes to move away from each other, forcing closer apposition, and eventual fusion, of the remaining membrane leaflet[165]. Regardless of the exact mechanism, membrane fusion leads to the presence of assembled SNARE complexes (now referred to as *cis*-SNARE complexes) in the fused membrane, which are subsequently disassembled by NSF concurrently with ATP hydrolysis[166]. Although originally grouped as v- and t-SNAREs, as mentioned above, currently accepted classification of SNAREs separates them into four subfamilies, based on the residue present at the "0-layer" of the SNARE domain: Q-SNAREs (usually encoding glutamine and which themselves fall into Qa, Qb, and Qc subfamilies) and R-SNAREs (usually encoding arginine); all of these subfamilies are generally conserved across eukaryotes[167,168].

The Sec1/Munc18-like (SM) proteins are a family of four (Sec1, Sly1, Vps33, and Vps45) proteins that interact with SNAREs to regulate and assist SNARE-mediated membrane fusion[169–171]. All four SM proteins are conserved across eukaryotes[126].

1.3.1.7 ESCRTs

The endosomal sorting complex required for transport (ESCRT) machinery is a set of protein complexes and associated machinery that mediates membrane invagination that is the topological inverse of the coat complexes discussed in section 1.3.1.3 (i.e. away from the cytosol). These complexes are: ESCRT-0, comprising Vps27/Hrs and Hse1/STAM1/2[172], ESCRT-I, comprising Vps23/TSG101, Vps28, Vps37, and Mvb12[173–175], ESCRT-II, comprising Vps22/EAP30, Vps25/EAP20, and Vps36/EAP45[176], and ESCRT-III, comprising Snf7/CHMP4, Vps2/CHMP2, Vps20/CHMP6, and Vps24/CHMP3[177]. A complex including the AAA ATPase Vps4/SKD1, as well as other associated machinery including Vta1, Vps31/Bro1/ALIX, Vps46/Did2, and Vps60/CHMP5, is referred to throughout this thesis as ESCRT-IIIa

("associated")[178–181]. ESCRTs are conserved across eukaryotes, with the exception of the opisthokont-specific ESCRT-0 complex, which may be functionally replaced in other eukaryotes by Tom1[182,183]. Although primarily known for its role in cargo degradation through the formation of intraluminal vesicles at multivesicular bodies (MVBs), ESCRTs have been implicated in a range of functions including cytokinesis, viral budding, and autophagy[184–187].

1.3.2 The MTS and eukaryotes

Although, as eluded to in the various parts of section 1.3.1, the core MTS machinery in eukaryotes is well-conserved, there are frequent cases of both gene loss and gain. Additionally, as will be discussed more fully in Chapter 3, the organization and organelle complement of the MTS varies substantially between diverse eukaryotes, including in the Apicomplexa, which will be discussed in the following sections.

1.4 Apicomplexan parasites

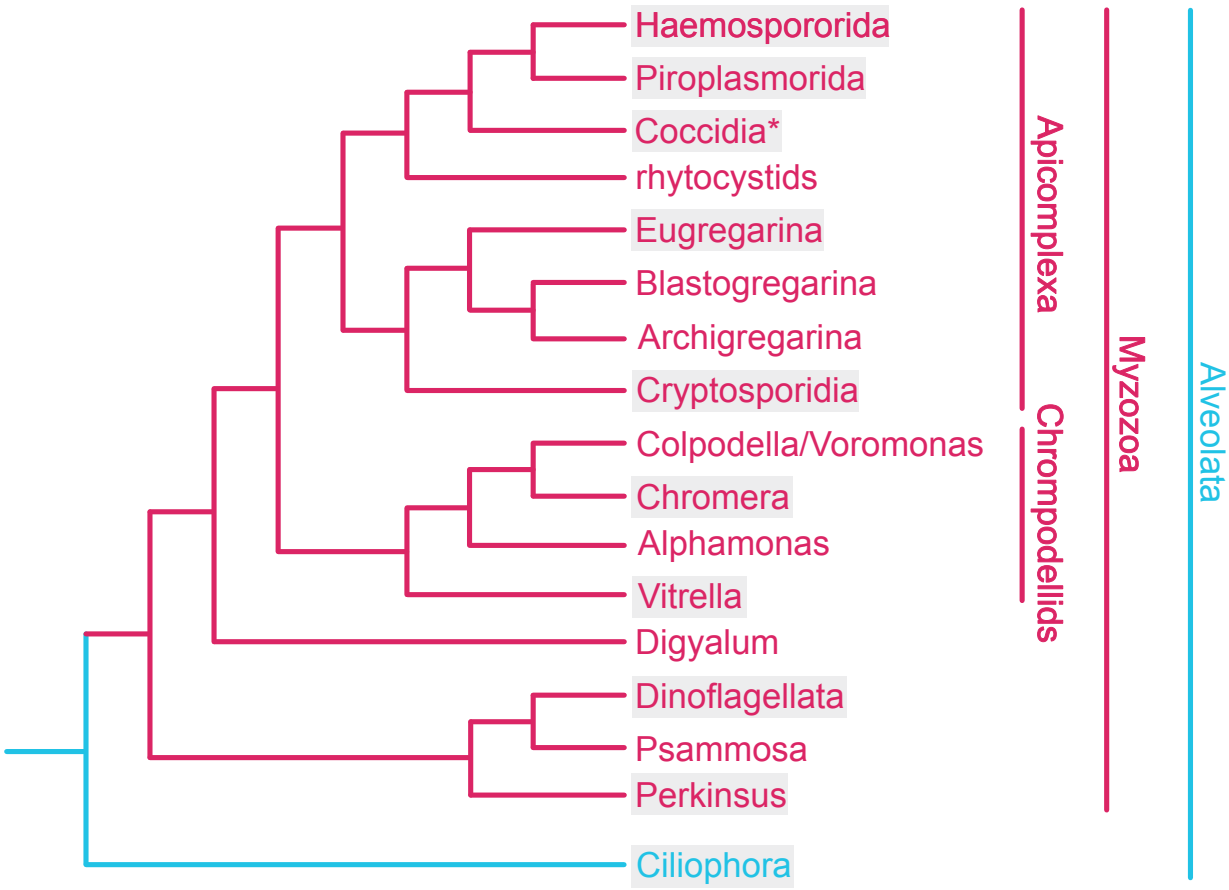
The Apicomplexa are a phylum of unicellular eukaryotes including important pathogens of humans and economically valuable animals. Included in this group are the various species of *Plasmodium*, causative agents of malaria in humans and other animals. Despite increased preventative measures and widespread research into vaccination and treatment, the most recent World Health Organization (WHO) numbers suggest that malaria is still responsible for ~435, 000 deaths annually[188]. *Cryptosporidium* infects humans and is responsible for enteric disease; infected individuals usually present with diarrhea and other associated symptoms such as nausea, vomiting, and weight loss[189]. Along with *Toxoplasma gondii*, discussed below, the highest disease burden is in conjunction with immunocompromised individuals, where the disease can be fatal[190,191]. Other prominent apicomplexans include members of the genera *Eimeria, Neospora, Theileria,* and others, which primarily infect livestock and other domesticated animals and hence are responsible for socioeconomic damage[192–194].

1.4.1 Apicomplexan evolution and chromerid algae

Apicomplexa are alveolates, one of three large groups that comprise the SAR clade[17] (Figure 1.3). Within the alveolates, the majority of known diversity is found within one of three groups: ciliates, apicomplexans, and dinoflagellates. Ciliates branch basally within the alveolates (Figure 1.3) and are generally large cells with rows of cilia arrayed in regular arrangements around the cell[17]. However, ciliates also display a huge diversity of morphological characteristics and are comparatively under-studied. The most common examples are *Tetrahymena thermophila* and *Paramecium tetraurelia*, oblong cells notably with

Figure 1.3 Alveolate phylogeny

This figure provides a detailed overview of alveolate phylogeny, with an emphasis on the Myzozoa. The basal ciliate branch is shown in cyan, and the myzozoan group in red. Note the main split at the base of Myzozoa that divides Apicomplexa and associated basal taxa, including chromerids and colpodellids on one side of the divide from dinoflagellates and basal taxa such as perkinsids on the other. Grey boxes around each group denote those for which genomic data was available for the analyses presented in Chapter 4 (a subset of which are also present in Chapter 2). The asterisk following the group Coccidia denotes the location of *T. gondii*, the model organism used for analysis in Chapter 5. Topology based on that in Janouškovec et al. (2019)[195].


Haemosporidida
Piroplasmorida
Coccidia*
rhytocystids
Eugregarina
Blastogregarina
Archigregarina
Cryptosporidia
Colpodella/Voromonas
Chromera
Alphamonas
Vitrella
Digyalum
Dinoflagellata
Psammosa
Perkinsus
Ciliophora
Apicomplexa
Chrompodellids
Myzozoa
Alveolata

a feeding groove for phagocytosis, a contractile vacuole for osmoregulation, and multiple regulated secretory organelles subtending the plasma membrane[196]. Additionally, ciliates are genetically complex, possessing two kinds of nuclei: micronuclei for germline propagation of genetic material, and polyploid macronuclei that are assembled after each division and used for somatic gene expression[197].

Dinoflagellates represent a sister group to Apicomplexa, and, together with numerous poorly studied basal taxa, comprise the Myzozoa (Figure 1.3)[198]. Dinoflagellates, like ciliates, display a huge diversity of morphology, including, in some members, the presence of solid thecal plates surrounding the cell. Additionally, dinoflagellates employ an array of trophic strategies, including phagocytosis/endocytosis, pallium feeding, in which a membrane protrudes from the cell and covers the prey to dissolve it, photoautotrophy, and mixotrophy; some are additionally symbionts of ecologically important organisms such as coral[199,200]. Approximately half of all dinoflagellates possess a functional plastid, although dinoflagellates are unique among eukaryotes in having frequently replaced their ancestral (peridinin) plastid through serial endosymbiosis[201]. The ancestral red algal-derived plastid, containing the accessory light harvesting pigment peridinin, is thought to be homologous to the relict plastid in some apicomplexans (the apicoplast; discussed further below), although its exact provenance remains under debate[202–204].

The putative homologous relationship between the functional peridinin-containing plastids of dinoflagellates and the apicoplast suggests that the myzozoan ancestor was photosynthetic. Additionally, environmental sampling revealed a huge diversity of uncultured organisms branching basally to the Apicomplexa associated with coral, presumably involved in photosynthesis[205,206]. Two examples of these photosynthetic relatives have been described as a paraphyletic group of "chromerid" algae (Figure 1.3), *Chromera velia* and *Vitrella brassicaformis*[207,208]. To date, *C. velia* and *V. brassicaformis* represent the closest known free-living photosynthetic relatives of Apicomplexa, making them invaluable as outgroups for comparative analysis (see Chapter 2).

1.4.2 Brief overview of molecular research into apicomplexans

Despite that multiple sequenced genomes are available across the major apicomplexan lineages through EuPathDB (https://eupathdb.org/eupathdb/), the bulk of research is performed in two genera (*Plasmodium* and *Toxoplasma*). Of the *Plasmodium* species, focussed studies are performed in the human parasite *P. falciparum*, but also in rodent parasites such as *P. berghei* and *P. yoelii*[209]. However, genetic manipulation of *P. falciparum* is challenging; routine culture of merozoites requires a ready supply of human red blood cells, transfection

efficiencies are comparatively low, and generation of stable clones requires extended periods of time (often one or more months) under drug selection[210]. Despite these difficulties, conditional systems for genetic manipulation such as DiCre[211] and targeted genome editing via CRISPR/CAS9[212,213] have been adapted for use in *Plasmodium* (detailed descriptions of both systems in *T. gondii* are provided in section 1.5.4), allowing for powerful genetic studies.

More recently, *Cryptosporidium* has become an increasingly tractable model for genetic manipulation and phenotypic characterization. Continuous *in vitro* culture has not been described; hence, new cultures must be derived from sporulated oocysts. Although historically lasting for only a few days[214], recent systems have been developed that allow *in vitro* culture for several months[215,216]. Similar to other apicomplexan systems, the recent adoption of CRISPR/CAS9 technology has made genetic manipulation tractable[217]. Given the unique intracellular niche of *Cryptosporidium*, wherein the parasite separates itself from the host cell cytosol via an electron-dense band and adjacent "feeder organelle"[218], and the basal position of *Cryptosporidium* in the Apicomplexa[219] (Figure 1.3), further development of this system is expected to yield great insights.

However, the majority of apicomplexans remain either refractory to culture and genetic manipulation or have not yet been the target of such attempts[220]. A recent study has shown that CRISPR/CAS9 constructs designed for *T. gondii* are also effective in *Neospora caninum*[221], yet *N. caninum* has not yet been extensively developed as a model system. The vast majority of research into coccidians (cyst-forming apicomplexans) is carried out in the model organism *Toxoplasma gondii*.

1.5 *Toxoplasma gondii*: the model apicomplexan

T. gondii represents an excellent model system within the Apicomplexa for genetic manipulation and phenotypic characterization. This is due to a number of factors, as discussed below.

1.5.1 *T. gondii* ultrastructure

T. gondii possesses a complex lifecycle (discussed in section 1.5.2), with multiple distinct cell types present at different stages and in different hosts. The most commonly studied cell type is the tachyzoite, the asexual form of the parasite that undergoes a lytic cycle within an intermediate host and is responsible for pathology. As such, this section will only describe the tachyzoite in detail, with occasional reference to other forms of *T. gondii*, when appropriate.

Compared to *Plasmodium* merozoites, which are exceptionally small eukaryotic cells (less than 2µm in diameter[222]), *T. gondii* tachyzoites are large (~2-3x6-7µm) cells, allowing better discrimination of intracellular structures in both fluorescence and electron microscopy[223]. Tachyzoites are polarized cells, with a larger, rounded base, and the cell body becoming increasingly narrower towards the apical tip. A combination of fluorescence microscopy with defined markers, as well as transmission and cryo-immunogold electron microscopy (TEM/iEM) have resulted in a good understanding of the main features of the parasite (Figure 1.4).

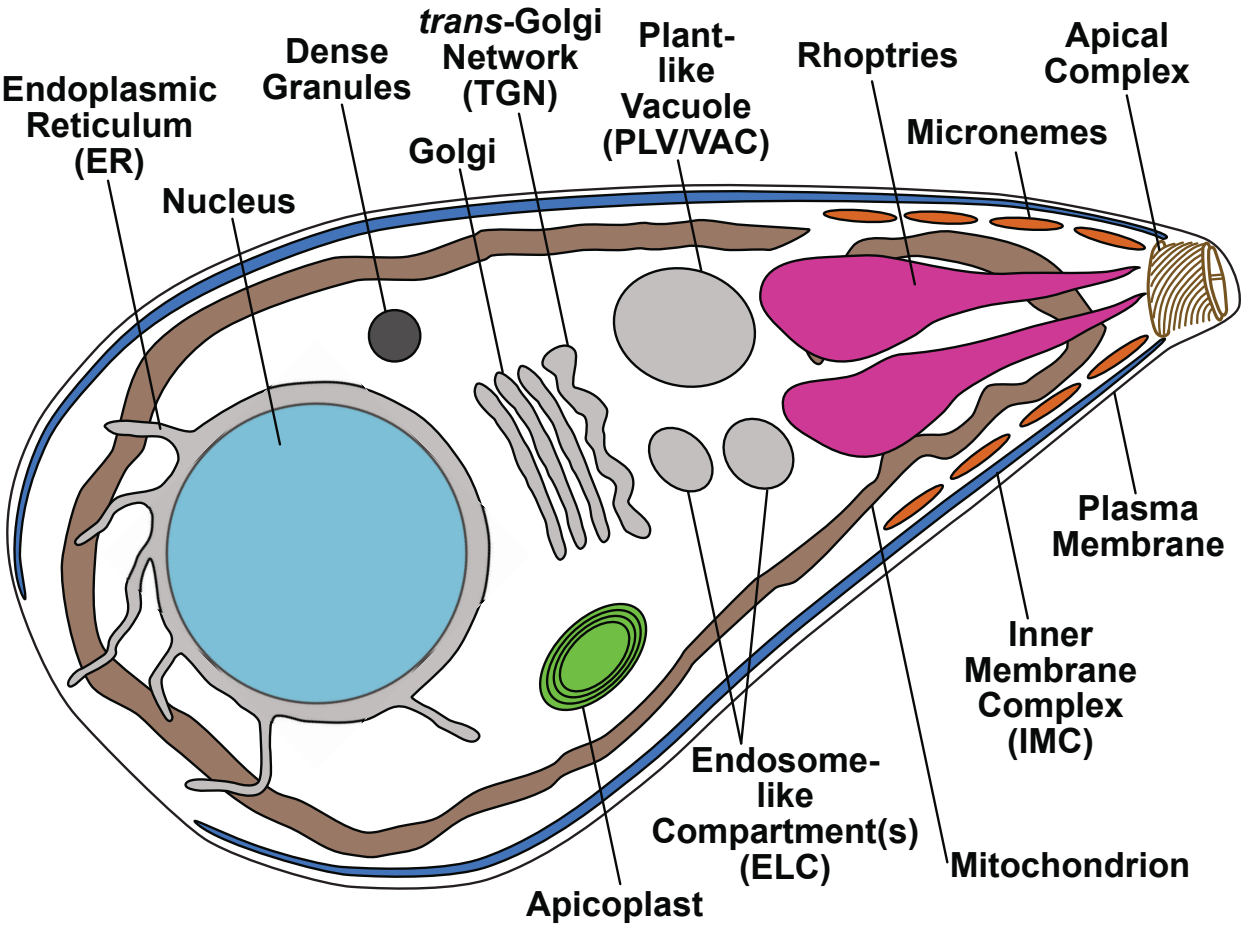
1.5.1.1 The apical complex and Inner Membrane Complex

The apical complex refers to a unique cytoskeletal structure and associated regulated secretory organelles (micronemes and rhoptries, discussed in section 1.5.1.4 and 1.5.1.5) at the apical end of apicomplexan zoites[224], although it is not present in an identical manner in all lifecycle stages of all apicomplexans[223]. The cytoskeletal portion comprises a conoid, a series of ribbon-like polymers of tubulin arrayed in a left-handed spiral, which is capped on the apical end by two preconoidal rings and on the basal end by an apical polar ring (APR)[225–227]. The APR is thought to be a microtubule-organizing centre (MTOC), from which the 22 cortical ("subpellicular") microtubules originate (discussed below)[228,229]. In addition, two microtubules attach to the anterior apical polar ring which run through, and extend below, the conoid[225]. The cytoskeletal portion of the apical complex is formed *de novo* during endodyogeny (cell division; discussed in section 1.5.2) and is thought to nucleate daughter cell formation[227]. Interestingly, incompletely closed conoid-like structures (termed "pseudoconoids") are found in related alveolates, including chromerids, colpodellids, and perkinsids; combined with the involvement of proteins otherwise associated with algal flagella in apical complex biogenesis, this has led to the hypothesis that the apical complex evolved from flagella in a free-living algal ancestor[230–233].

The Inner Membrane Complex (IMC) is a network of flattened membranous sacs connected together by "sutures" that lies just below the plasma membrane[234,235]. These two closely apposing structures together form what is referred to as the pellicle, with a small gap in between the two sets of membranes ~20nm wide; the IMC is not entirely continuous however, with gaps at both the apical and the basal ends[223]. The IMC is homologous to the canonical alveoli of other alveolates[236] and is formed from a combination of *de novo* and recycled maternal material during endodyogeny[237]. Firm attachment of the cortical microtubules to the IMC maintains the size and shape of zoites[238], and the IMC also harbors additional key apicomplexan proteins such as components of the "glideosome" (discussed in

Figure 1.4 Schematic of a *T. gondii* tachyzoite

This figure provides an overview of the basic ultrastructure of the asexual tachyzoite stage of *T. gondii*. The specialized secretory organelles are coloured: micronemes (orange), rhoptries (magenta), and dense granules (dark grey), as are the endosymbiotic organelles, the apicoplast (green) and mitochondrion (brown), and other organelles including the Inner Membrane Complex (IMC, dark blue), and the nucleus (light blue). The cytoskeletal portion of the apical complex (including the conoid) is depicted at the apical end in light brown. All other organelles, including the ER, Golgi, various ELCs, and PLV/VAC, are coloured light grey. Full descriptions of all organelles can be found in the main text.


Endoplasmic Reticulum (ER)
Dense Granules
trans-Golgi Network (TGN)
Plant-like Vacuole (PLV/VAC)
Rhoptries
Apical Complex
Micronemes
Golgi
Nucleus
Plasma Membrane
Inner Membrane Complex (IMC)
Endosome-like Compartment(s) (ELC)
Mitochondrion
Apicoplast

section 1.5.2). Additionally, gaps in the pellicle, termed "micropores", have been described through which internalization of material has been observed[239].

1.5.1.2 The nucleus, endoplasmic reticulum, and Golgi complex

The nucleus is the most prominent structure within *T. gondii* tachyzoites, measuring ~1μm in diameter and positioned either centrally, or towards the basal end of the cell[223] (Figure 1.4). The ER adopts a branched morphology, with extensions to the medial and basal ends of the nucleus; no distinct ER is discernable directly apical to the nucleus, with secretory transport structures budding directly from the outer nuclear envelope membrane[240].

The lack of a morphologically canonical Golgi complex in *Plasmodium* (i.e. not appearing as a series of closely apposed cisternae; not "stacked") led to questions regarding the nature of the Golgi in Apicomplexa[241,242]. However, subsequent studies confirmed that *T. gondii* possesses a morphologically stacked Golgi with between three and six cisternae and apparently distinct *cis/medial*- as well as *trans*-Golgi compartments[243–245].

1.5.1.3 The endosomal system

The organization of the *T. gondii* endosomal system, along with the identity of various described endosomal organelles is poorly understood. As will be discussed in section 1.6.2.5, the morphology of post-Golgi compartments is highly variable and their identity remains problematic, being mostly defined through localization of a series of partially overlapping marker proteins. Broadly, tubular extensions surrounding the Golgi, combined with one or more electron-lucent vesicular structures, appear to comprise the endosomal system; the poorly defined nature of these structures has prompted the extensive use of the generic term endosome-like compartment(s) (ELC) throughout the literature[245,246]. In addition, a putative vacuolar compartment, alternatively referred to as the *T. gondii* vacuole (VAC) or plant-like vacuole (PLV), is present (referred to herein as the VAC). The VAC is large and easily visible in extracellular tachyzoites, even by differential interference contrast microscopy, and appears as a large lucent vacuole in EM, occasionally with apparent internal membranous material. Following host cell invasion, the VAC fragments, and is discernable from other lucent structures in intracellular tachyzoites only through the application of immunolabeling techniques[247,248].

1.5.1.4 Micronemes

Micronemes represent one of three classes of specialized secretory organelles present in apicomplexans (the other two, rhoptries and dense granules, are discussed below), and

are localized at the apical tip of the parasite in close proximity to the apical complex[249]. Although microneme morphology varies between organisms and lifecycle stages[250], tachyzoite micronemes are "rod-shaped" organelles ~50nm wide by ~180nm long[251], and are moderately electron dense by TEM. Microneme biogenesis is known to involve the endosomal system (discussed in section 1.6.2), but is poorly described at the morphological level; in intraerythrocytic *P. falciparum*, immature micronemes have been described to bud from the Golgi and transit along microtubules to reach the apex of the developing merozoites[252]. It is therefore interesting to note that micronemes appear arrayed along sub-pellicular microtubules by super-resolution microscopy, and disruption of microtubules also disrupts the normal apical localization of micronemes, suggesting that they may be tethered to these microtubuless[229]. Although traditionally thought of as one population, recent work suggests at least two independent subsets of micronemes are present in *T. gondii*[253].

1.5.1.5 Rhoptries

Rhoptries are large (~2-3μm long) organelles that display a characteristic "club"- or "gourd"-shaped morphology, with a wide bulb at the basal end that tapers to a thin apical neck; these distinct regions also house different classes of proteins – rhoptry bulb (ROP) and rhoptry neck (RON) proteins[254]. The thin neck is almost uniformly electron dense by conventional TEM, with a less electron dense intermediate region that finally gives way to the bulb, which is heterogeneous with a "honeycombed" appearance[255]. Rhoptries have been described to form *de novo* during cell division, with highly acidic "pro-rhoptry" precursor organelles formed from the fusion of multiple post-Golgi compartments, which elongate and mature before being tethered at the apical end of the cell[256–258].

1.5.1.6 Dense granules

The dense granules are aptly named; they are uniformly electron dense spherical organelles approximately 300nm in diameter distributed throughout the cytosol[223]. Dense granule biogenesis has not been extensively studied, but they are believed to form directly from the Golgi apparatus (discussed in section 1.6.2.4). Recently, it has been postulated that, like micronemes, multiple distinct populations of dense granules might exist in *T. gondii*[259].

1.5.1.7 Endosymbiotic organelles: apicoplast and mitochondrion

The apicoplast, a four membrane-bound relict plastid of red algal origin that derived from a putative endosymbiotic event of secondary or higher level, is typically located close to both the ER and Golgi[204,260–262]. It can be recognized in conventional fluorescence microscopy

by DNA staining (e.g. with 4',6-diamidino-2-phenylindole (DAPI), Hoescht) as a distinct signal from the nucleus, but also by specific markers (discussed in section 1.6.2.2). Unlike the complex network of mitochondria observed in other cells, *T. gondii* possesses a single mitochondrion, whose morphology appears plastic throughout the lytic cycle but that commonly adopts an oval ("lasso") conformation in intracellular tachyzoites[263].

1.5.2 Lifecycle of *T. gondii*

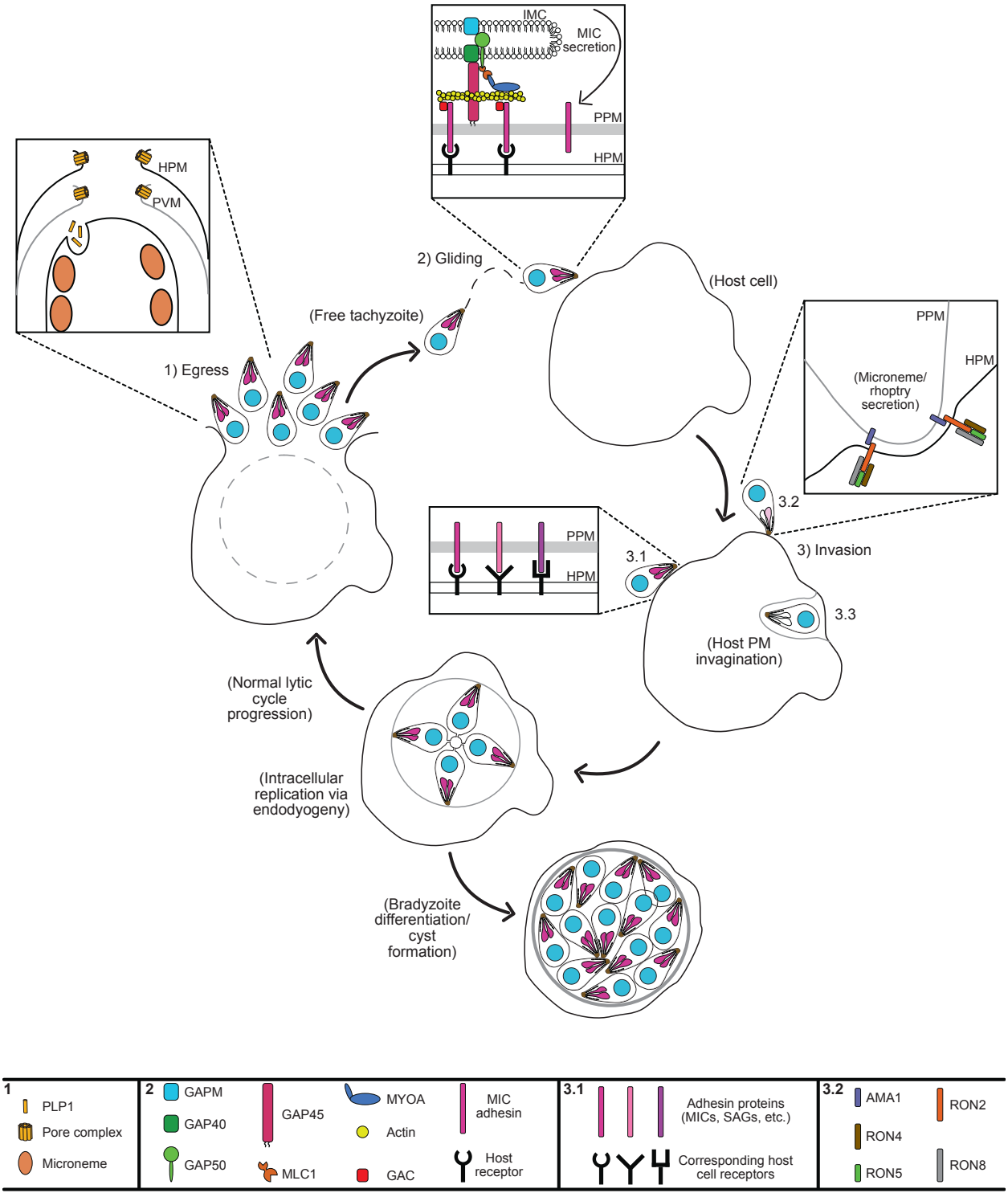
T. gondii possesses a complicated lifecycle involving both a definitive host, members of the *Felidae* genus (i.e. cats), and any number of intermediate hosts, generally any nucleated cell within a warm-blooded animal[264,265]. Although host infection is technically possible with any invasive form of the parasite (tachyzoite, bradyzoite, or sporozoite, discussed below), the majority of infections occur either through ingestion of tissue cysts in contaminated meat or other animal products, or oocysts, often in drinking water or soil[266–268].

Infection of the definitive host (cats) generally begins with the ingestion of infected prey, e.g. rodents, resulting in the rupture of tissue cysts within the intestinal tract. The released bradyzoites infect intestinal epithelial cells and proceed through five types (A through E) of morphologically distinct schizonts, each giving rise to individual merozoites[267,269]. Host cell rupture and parasite egress theoretically allows for further infection of intestinal cells by merozoites, although this has never been observed[223]. By a poorly understood mechanism, some infections result in the production of sexually differentiated forms, referred to as micro- and macrogametes[270,271]. Fertilization of an intracellular macrogamete by a microgamete results in a zygote, which subsequently becomes an immature oocyst[272]. Once shed, the maturation process proceeds to completion, resulting in an infectious sporulated oocyst containing two sporoblasts of four sporozoites each[273–275]. Sporulated oocysts are both highly infectious and extremely resilient, capable of surviving in the external environment for extended periods, and are resistant to ultraviolet radiation, osmotic imbalances, and many commonly used treatment agents such chlorine and ozone[276–278].

In addition to sexual development, the asexual lytic cycle and subsequent establishment of latent chronic infection can occur in any appropriate host. The lytic cycle is mediated by the rapidly growing tachyzoite stage and comprises three distinct steps: egress, motility, and invasion[279] (Figure 1.5). The process begins with egress, involving calcium-dependent activation and secretion of microneme proteins such as the perforin-like protein TgPLP1 to lyse both the parasitophorus vacuole (discussed below) and the host cell membrane[280–282]. Parasites concurrently activate gliding motility to traverse the extracellular space until a new suitable host cell is encountered. This process was originally proposed to

Figure 1.5 The asexual lytic cycle of *T. gondii*

This figure provides a cartoon overview of the basic steps within the lytic cycle; more detail for each can be found in the main text. The lytic cycle begins when intracellular parasites egress from infected host cells (1; host can typically be any nucleated cell in a warm-blooded organism), lysing the parasitophorus vacuole membrane (PVM) and host cell plasma membrane (HPM) and activating gliding motility. Membrane lysis is shown in the inset panel, whereby microneme secretion releases perforin-like proteins (e.g. PLP1), which subsequently form holes in the PVM/HPM, facilitating parasite exit. Parasites then move by gliding motility until they encounter a new host cell (2). Microneme secretion during gliding motility is shown in the inset panel, whereby trans-membrane MIC proteins in the parasite plasma membrane (PPM) form a bridge between molecules (e.g. sialic acid, chondroitin sulfate, some toll-like receptors) on the HPM (and/or proteins in the extracellular matrix) and the underlying parasite molecular machinery involved in gliding motility. Invasion (3) occurs in three distinct stages: initial (weak) attachment (3.1), microneme/rhoptry discharge to mediate tight junction formation and (strong) apical attachment (3.2), and invasion through the moving junction (3.3). Inset panel for 3.1 shows the initial weak attachment between one or more adhesin proteins on the PPM to corresponding molecules in the HPM. Inset panel for 3.2 shows the formation of the tight junction facilitating host cell invasion. Rhoptry secretion releases RON2/4/5/8 proteins into the host cell, which act as a specific receptor complex for the microneme protein AMA1 present on the parasite surface. Once inside the host cell, parasites replicate by endodyogeny, remaining connected to each other within the vacuole by the "filamentous network" (including the residual body). In some cases, parasites will differentiate into bradyzoites and modify the PVM to produce a tissue cyst. In most cases, parasites continue through the normal lytic cycle and eventually egress to begin another round of the cycle (1). For non-inset figures, black lines represent the host cell plasma membrane while grey lines represent the PV membrane; dotted grey lines represents PV membrane lysis. Symbols used in each inset panel are defined in the legend at the bottom of the figure. Abbreviations: HPM, host cell plasma membrane; PVM, parasitophorus vacuole membrane; PPM, parasite plasma membrane; IMC, inner membrane complex; PLP, perforin-like protein; GAP, glideosome-associated protein; MLC, myosin light chain; MYOA, myosin A; GAC, glideosome-associated connector; MIC, microneme protein; SAG, surface antigen group; AMA1, apical membrane antigen 1; RON, rhoptry neck protein.

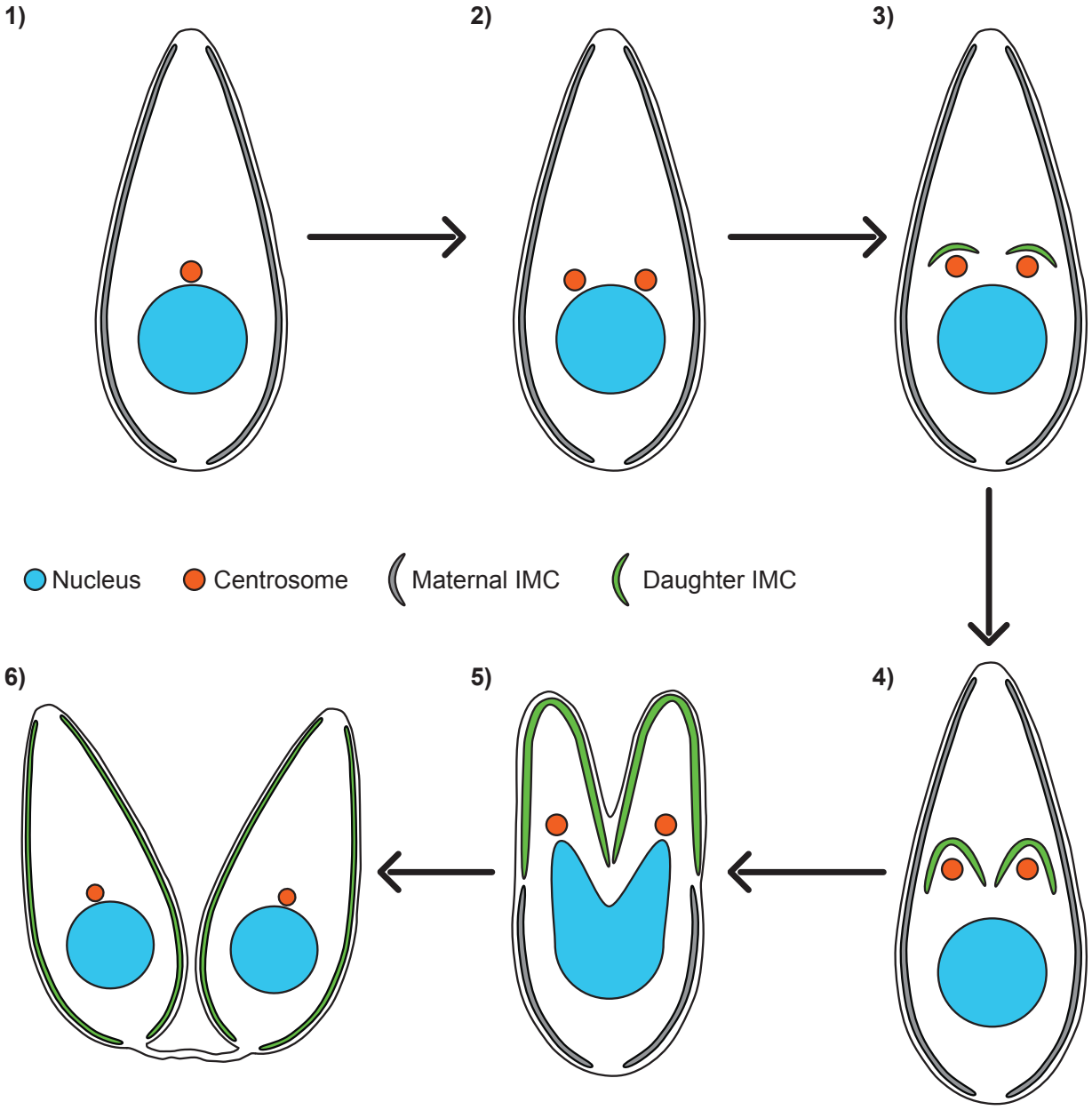

IMC
MIC secretion
PPM
HPM
HPM
PVM
2) Gliding
(Host cell)
(Free tachyzoite)
1) Egress
PPM
HPM
(Microneme/ rhoptry secretion)
3.2
3) Invasion
PPM
HPM
3.1
3.3
(Host PM invagination)
(Normal lytic cycle progression)
(Intracellular replication via endodyogeny)
(Bradyzoite differentiation/ cyst formation)
1
PLP1
Pore complex
Microneme
2
GAPM
GAP40
GAP50
GAP45
MLC1
MYOA
Actin
GAC
MIC adhesin
Host receptor
3.1
Adhesin proteins (MICs, SAGs, etc.)
Corresponding host cell receptors
3.2
AMA1
RON4
RON5
RON2
RON8

involve the "glideosome", a molecular complex anchored in the IMC that coupled myosin-dependent rearward translocation of transmembrane adhesin proteins to forward motion[283–289]. However, the glideosome model does not easily account for all observations and alternative models have been proposed[290–294]. Importantly though, all models proposed to date recognize the contribution of microneme secretion to motility; in the glideosome model, trans-membrane adhesins secreted from micronemes form the necessary connections between extracellular substrates and the molecular motor[295], while in the fountain flow model, microneme secretion adds membrane to the apical tip of the parasites, which establishes a retrograde membrane flow associated with forward movement[294]. Finally, the parasite engages with, and invades into, a new host cell. Invasion can be broken down into several steps. The parasite first attaches through low affinity interactions mediated by surface proteins (Figure 1.5, 3.1)[296]. Next, the parasite secretes a complex of rhoptry neck (RON) proteins (RON2/4/5/8) into the host cell, which associate with the host cell plasma membrane and act as an anchorage point for parasite plasma membrane-localized Apical Membrane Antigen 1 (TgAMA1, a microneme protein) to mediate tight apical attachment with the host cell (Figure 1.5, 3.2) and allow active entry into the host cell through a ring-like "moving junction" (Figure 1.5, 3.3)[297–303]. Additionally, there is evidence that the host cell is not entirely passive in this process, and that other *T. gondii* rhoptry proteins secreted during invasion, such as toxofilin, act to modulate host cell actin and further facilitate invasion[303–307]. Once inside the host cell, the parasite extensively remodels the parasitophorus vacuole by the release of dense granules, inducing the formation of an intra-vacuolar network that is thought to facilitate uptake of material from the host cell[308,309].

Once intracellular, the parasite then undergoes replication, with a characteristic doubling time of ~6-8 hours[310], through a process of endodyogeny in which two daughter zoites emerge within, and eventually consume, the maternal body (Figure 1.6)[311]. During endodyogeny, the parasite's centromeres duplicate first and reorient themselves to a position apical to the nucleus. In a series of coordinated steps, the parasite then replicates its autogenous organelles, including the Golgi, which are subsequently partitioned into the newly forming daughter cells. These daughters form within the mother cell, elongating towards the apical tip of the mother concurrent with both recycling and *de novo* formation of plasma membrane and IMC to encapsulate the newly forming daughters[223,237,256]. Other organelles, including the micronemes and rhoptries, are believed to be synthesized *de novo* during each round of replication[256]. Interestingly, although poorly defined, it has recently been proposed that micronemes and rhoptries may also be recycled into daughter cells during division, either directly or through break down and recycling of their constituents[312].

Figure 1.6 Replication by endodyogeny

This figure provides a highly simplified cartoon overview of asexual *T. gondii* replication via the process of endodyogeny. In resting parasites, a single centrosome is present in close association with the nucleus (1), which duplicates early in cell division (2). Early in endodyogeny, the IMC of each daughter cell begins to form (3) and the organelles begin to be partitioned into the developing daughter buds (not shown here for simplicity). As endodyogeny continues, the buds elongate toward the apical end of the mother cell (4), eventually deforming the maternal plasma membrane (5). Elongation continues as the mother cell's organelles are either recycled or degraded; completion of endodyogeny results in two daughter cells, which remain connected by a bounding membrane up until the intravacuolar network is disassembled prior to egress.


1)
2)
3)
4)
5)
6)
Nucleus
Centrosome
Maternal IMC
Daughter IMC

Following division, the daughter cells remain connected via membranous extensions from the basal end of the cell, which coalesce in a "residual body" containing the remnants of maternal cells not incorporated during division[313]. These connections can be visualized using camel-derived nanobodies (single-chain antibodies) directed against filamentous actin (F-actin) fused to a variety of tags or fluorescent proteins[314]. Live imaging of these markers showed that the "filamentous network" between parasites remained extensive throughout the intracellular stage, and was subsequently disassembled just prior to egress[314]. Artificially severing this network resulted in vacuoles becoming disorganized and impaired synchronous division cycles, suggesting that parasites were able to communicate via this network[315,316]. More recent work has shown movement of proteins between parasites through this network, confirming that it contributes to the proper development of intracellular parasites (Javier Periz, personal communication).

Under normal circumstances, parasites within a PV undergo synchronous rounds of endodyogeny, resulting in vacuoles containing multiples of two parasites (e.g. 2, 4, 8, 16, etc). Eventually, parasites egress (as above), and the lytic cycle continues anew.

Although the signals involved are not elucidated, it is believed that stress, related to the host immune response, triggers some tachyzoites to develop into bradyzoites (a slow growing form of the parasite responsible for chronic infection) *in vivo*[317]. Bradyzoite differentiation is accompanied by morphological changes, such as posterior migration of the nucleus, formation of uniformly electron dense rhoptries, and increase in number of amylopectin granules[318], and involves formation of tissue cysts with a thin (~500nm) wall[319], primarily in the brain and skeletal muscle. Tissue cysts can occasionally rupture, even in immunocompetent hosts[320], and may be involved in recurring infection, for example in ocular toxoplasmosis, as well as a source of protective immunity[321]. It is commonly believed that tissue cysts last "for the life of the host", including in humans as no treatment methods have been devised to clear latent infections[322], yet this has never definitively been shown and remains an open question[321]. Ingestion of tissue cysts, usually through carnivory, represents an important transmission mode.

1.5.3 *T. gondii* prevalence and pathogenesis

Current estimates suggest that *T. gondii* infects approximately one third of the global population, although seroprevalence varies significantly by region[266,323]. In the United States, a recent estimate has placed *T. gondii* seroprevalence at 11.14%[324]; equivalent data are not available for the Canadian population[325], but one study estimated *T. gondii* infection at 28.10 cases per 100,000[326]. Studies have investigated *T. gondii* seroprevalence among communities

in the northern part of Canada, including Inuit communities, where up to 65% of the population can be seropositive for *T. gondii*[327,328]. It is thought that the primary route of human infection is through ingestion of contaminated food and/or water[325].

Under normal circumstances, *T. gondii* infection in immunocompetent individuals is usually asymptomatic, although some patients present with a wide range of symptoms including "flu-like" symptoms (fever, fatigue, lymphadenopathy), as well as more specific symptoms such as myocarditis, myositis, hepatosplenomegaly, and chorioretinis[322]. However, as discussed above, acute infection can lead to formation of tissue cysts and establishment of latent chronic infection, which is problematic in a number of situations. HIV-infected individuals with $CD4^+$ cell counts < 100/mm^3 can present with toxoplasmic encephalitis, with multiple parenchymal brain lesions[329]. In comparison, ocular toxoplasmosis can occur in both immuno-competent and -deficient patients, and is associated with frequent reactivation of bradyzoite cyst forms into tachyzoites and the possibility of blindness due to retinal scarring[330]. Additionally, acute maternal infection during pregnancy carries the possibility of vertical transmission to the foetus; congenital toxoplasmosis may result in premature births, stillbirths, or spontaneous abortion, and even after birth is associated with a range of symptoms in the newborn including chorioretinitis and hydrocephaly[331].

Early studies investigating virulent versus avirulent *T. gondii* strains from North America and Europe identified that virulent strains comprised a single clonal lineage[332]. Subsequently, three clonal lineages were established in the Northern hemisphere[333], followed later by the identification of a fourth[334]. In comparison to the highly clonal population structure of Northern strains, strains from South America show evidence of frequent recombination; overall, current views of *T. gondii* population structure reveal six major clades with approximately 16 haplotypes[335]. Importantly, these strains vary in virulence, which has thus far been mapped mainly to the polymorphic ROP5 and ROP18 loci, encoding, respectively, a pseudokinase and serine/threonine kinase[336]. The importance of this population structure to understanding virulence is evident in the frequent fatal cases of acute toxoplasmosis associated with atypical strains in South America[337,338].

1.5.4 Genetic manipulation of *T. gondii*

Initial development of *T. gondii* as a model genetic system began in the 1970's with the use of chemical mutagenesis (mainly using N-methyl-N'-nitro-N-nitrosoguanidine or ethylnitrosourea) to generate temperature-sensitive mutants for phenotypic and biochemical analysis[339,340]. Drug resistance mutations generated through these efforts allowed for genetic

crosses in infected kittens and demonstrated both the haploid nature of all lifecycle stages except for sporulating oocysts, as well as the Mendelian nature of meiotic inheritance[340,341].

Identification of conditions suitable for *T. gondii* electroporation[342] permitted expression of transgenes, but only for ~72 hours post-transfection. However, a multitude of selectable markers were soon identified permitting stable transformation via positive selection: a mutated form of dihydrofolate reductase-thymidylate synthase (DHFR-TS) conferring pyrimethamine resistance[343], the chloramphenicol acetyl transferase (CAT) gene product conferring resistance to chloramphenicol[342,344], as well as complementation of *T. gondii* tryptophan auxotrophy[345] with *Escherichia coli trpB*[346]. Two additional markers, uracil phosphoribosyl transferase (UPRT) and hypoxanthine-xanthine-guanine phosphoribosyl transferase (HXGPRT) are unique among selectable markers for *T. gondii*, as they can be used both for positive and negative selection[347–350].

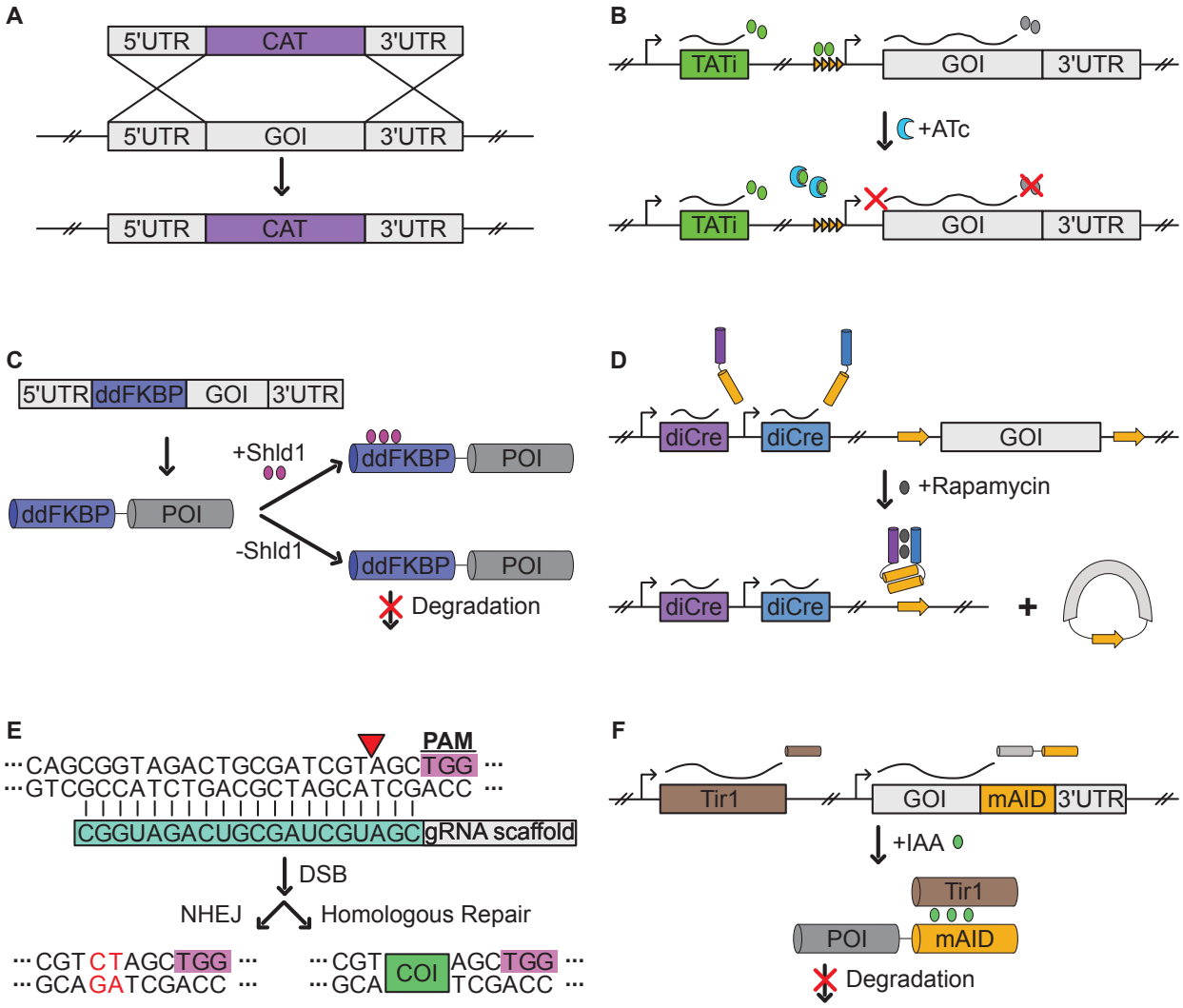
Insertional mutagenesis using vectors containing either genomically- or cDNA-derived DHFR-TS demonstrated the propensity for *T. gondii* to integrate exogenous DNA via nonhomologous recombination, and the necessity of long (several kilobases in length) segments of homologous sequence to drive homologous integration[351]. Forward genetic studies in *T. gondii* based on chemical and/or insertional mutagenesis has assisted in identifying stage conversion factors[352] and aided in the dissection of specific cellular systems (for example, cell division[353]). More recently, the sequencing of a complete *T. gondii* genome combined with the advent of CRISPR/CAS9 in *T. gondii* (discussed below) has allowed for large-scale genomic screens[354].

In addition to diverse forward genetic methods, multiple reverse genetic approaches have been developed to investigate specific gene products in *T. gondii*. The ability of homologous regions to drive gene integration/replacement in *T. gondii* (Figure 1.7A) led to the first complete knockout through transfection of a construct comprising the 5′ and 3′ regions of the *Rop1* locus surrounding CAT, producing a *Δrop1* strain[344]. The (prohibitive) requirement for long (>1-2 kb) flanking regions for efficient targeting was ameliorated through development of *Δku80* strains which lack the Ku80 product of the non-homologous end-joining (NHEJ) pathway and increase the frequency of homologous integration, even with short homologous regions[355,356]. Despite this, straight knockout approaches do not allow characterization of essential genes, as *T. gondii* is haploid throughout most of its lifecycle, including in the tachyzoite stage[341].

The first conditional system adapted for use in *T. gondii* was the tetracycline-repressor system, based off the tetracycline repressor (TetR) fused to a trans-activating domain. Although this study established anhydrotetracycline (ATc) as the optimal tetracycline

Figure 1.7 Overview of genetic manipulation techniques in *T. gondii*

This figure provides a cartoon overview of methods for genetic manipulation as discussed in the main text. A) straight gene knockouts by homologous recombination; a construct is prepared containing a selectable marker (here, CAT) flanked by sequences homologous to the locus to be disrupted. Drug selection favours replacement of the endogenous gene with the selectable marker. B) the TATi system. The target gene is placed under the control of a minimal tetracycline-responsive promoter (containing tetO repeats, orange triangles) in a parasite line expressing TetR fused to the trans-activating TATi peptide. The TetR-TATi fusion binds to the minimal promoter and induces transcription. Addition of anhydrotetracycline (ATc) sequesters TetR and prevents target gene transcription. C) Protein regulation via the destabilization domain (dd) system. A ddFKBP domain is placed within the coding sequence of the target gene (usually at the N-terminus), producing a fusion protein (blue and grey cylinders). The protein is degraded by the proteasome under normal conditions (-Shld1), while addition of Shld1 (+Shld1) prevents this degradation. D) the diCre system. The target gene is floxed (loxP sites placed on either side) in a line expressing N- and C-terminal halves of Cre recombinase bound to FRB and FKBP domains ("diCre"). Addition of rapamycin induces diCre dimerization, reconstituting Cre activity and recombining the loxP sites, resulting in removal of the target gene. E) genome editing by CRISPR-CAS9. The 20 nucleotide guide RNA (gRNA, shown in teal), is complementary to a target sequence adjacent to a three nucleotide PAM sequence (here conforming to NGG, shown in purple). CAS9 induces a double-strand break (red arrow), which can be repaired by two main pathways. Non-homologous end-joining (NHEJ) is error-prone and frequently results in indels (red nucleotides), whereas homologous recombination can be used to introduce cassettes (green box). F) the auxin-degron system. A tag (mAID) is added to a target gene in a line expressing Tir1. Addition of auxin (IAA) causes Tir1 association and subsequent proteasomal degradation. Abbreviations: UTR, untranslated region; GOI, gene of interest; POI, protein of interest; COI, cassette of interest.


A
5'UTR
CAT
3'UTR
5'UTR
GOI
3'UTR
5'UTR
CAT
3'UTR
B
TATi
GOI
3'UTR
+ATc
TATi
GOI
3'UTR
C
5'UTR
ddFKBP
GOI
3'UTR
+Shld1
ddFKBP
POI
ddFKBP
POI
-Shld1
ddFKBP
POI
Degradation
D
diCre
diCre
GOI
+Rapamycin
diCre
diCre
+
E
PAM
···CAGCGGTAGACTGCGATCGTAGCTGG···
···GTCGCCATCTGACGCTAGCATCGACC···
CGGUAGACUGCGAUCGUAGC
gRNA scaffold
DSB
NHEJ
Homologous Repair
···CGTCTAGCTGG···
···GCAGATCGACC···
···CGT
COI
AGCTGG···
···GCA
TCGACC···
F
Tir1
GOI
mAID
3'UTR
+IAA
Tir1
POI
mAID
Degradation

analogue for use with *T. gondii* and confirmed ATc-dependent binding of TetR to *tet*O repeats, no trans-activation was observed, and hence the system was appropriate only for conditional expression of dominant negative mutants and/or overexpression[357]. Insertional mutagenesis screening identified an endogenous trans-activation sequence (referred to as TATi), allowing the proper establishment of a tetracycline-inducible transactivator system (Figure 1.7B) and robust control of gene expression in an ATc-dependent manner[285]. Originally based on a two-step process involving introduction of a regulatable second copy of the gene-of-interest (GOI) followed by removal of the endogenous GOI, this was later modified by replacement or displacement of the endogenous promoter[358] (see also Chapter 5).

Although powerful, tetracycline-based systems rely on regulation of transcription, which is too slow for some applications. Rapid regulation of protein stability was found to be achieved through fusion of a mutated FKBP12 sequence ("ddFKBP") to a GOI (ddFKBP-GOI), which results in rapid degradation of the resulting fusion protein in the absence of rapamycin analogues (the most commonly used is referred to as Shld-1, Figure 1.7C)[359]. This system was subsequently adapted to Apicomplexa[360,361]. Although rapid (maximal protein levels are often reached within 6h of adding Shld-1), this system can exhibit leaky expression and requires the regulated protein have access to the proteasome; nevertheless, it is useful for overexpression/expression of dominant negative proteins[361–363].

A system based on dimerizable Cre recombinase (DiCre) was developed to produce true conditional knockouts (i.e. those that lack a GOI completely, Figure 1.7D). Based on the activity of Cre to induce recombination of LoxP sites, two functional halves of Cre are fused alternatively to the FKBP and FRB domains of mTOR, which dimerize in the presence of rapamycin and reconstitute the functional Cre recombinase to excise floxed (containing a LoxP site on either side) sequences from the genome[291]. The main issue with the DiCre system is low excision rates for some genes/genomic regions (for example, ~12% for TgAPμ1[258]). Similar to the tetracycline-based systems described above, this was originally achieved through a two-step process, but the advent of CRISPR/CAS9 technology in Apicomplexa (discussed below) has allowed specific insertion of LoxP sites into endogenous loci.

CRISPR/CAS9 is a genomic editing tool derived from the CRISPR (clustered regularly interspaced short palindromic repeats) bacterial/archaeal immune system (Figure 1.7E)[364]. A CRISPR RNA and a transactivating RNA (expressed as a single unit in the system referred to as a "guide RNA") complex with the CAS9 endonuclease and induce a double-strand DNA break (DSB) at a specific position that is homologous to the CRISPR RNA; in *T. gondii* this is typically repaired by the prevalent, error-prone, NHEJ pathway, leading to introduction of indels and disruption of a coding sequence[365,366]. The proclivity of *T. gondii* to use NHEJ

combined with the power of CRISPR/CAS9 allowed for the first whole-genome screen of essential genes in the tachyzoite stage[354]. In addition, DSBs were found to allow for efficient specific integration of exogenous DNA with small (~40bp) homology flanks[366,367], allowing for targeted homologous integration and improved editing of endogenous loci.

Finally, a recently adapted auxin-inducible degron (AID) system allows for rapid regulation at the protein level of genes tagged with the mini-AID amino acid sequence in parasites expressing TIR1 with addition of the plant hormone auxin (Figure 1.7F)[368]. Although extremely rapid (drastic decrease of tagged protein level within minutes), this system, like the ddFKBP system, requires access of the tagged protein to the proteasome and is therefore not appropriate for all candidates.

1.6 Membrane trafficking in Apicomplexa

The apicomplexan MTS remains poorly understood, despite increased efforts to characterize this system in the last ~20 years. This section will outline the coding complement of the apicomplexan MTS as it stands prior to Chapter 4 of this thesis; additional discussion on coding complement is provided in Chapter 2. Next, the function of studied components will be discussed, and a model presented to orient the reader for subsequent chapters. This model will be expanded in Chapter 5 to include data presented therein.

1.6.1 Coding Complement of Apicomplexa

1.6.1.1 Rab and Arf GTPases

Arf and Rab GTPases were introduced in sections 1.3.1.1 and 1.3.1.4, respectively. Langsley et al. (2008) reported a coding complement of 15 Rabs for *T. gondii*, compared to 11 in yeast and ~70 in mammalian cells[369,370]. In mammalian and yeast cells, Rab5 localizes to early endosomes, where it recruits effectors involved in the tethering of endocytic vesicles, maturation of the compartment and the switch to a "late" Rab7-positive compartment, as well as homo/heterotypic fusion of early and late endosomes[124,150,371–373]. In *T. gondii* there are three Rab5 paralogues, referred to as TgRab5A, TgRab5B, and TgRab5C[253]. The *Plasmodium* orthologue of Rab5B was noted for the fact that it is N-terminally myristoylated, bearing some similarity to *Arabidopsis thaliana* ARA6[374]. In *T. gondii*, all three localize to a sub-apical region housing the Golgi and endosomal compartments, and TgRab5B appears to localize to the plasma membrane as well[253] and interacts with the retromer subunit TgVps26[375].

In other systems, Rab6 localizes primarily to the Golgi/TGN and plays diverse roles, including in trafficking through the Golgi, Golgi organization, autophagy, and cytokinesis[376–

[379]. TgRab6 was found to localize to the Golgi, TGN, and somewhat to dense granules. Golgi localization was further supported by ER-like staining of TgRab6 following brefeldin A (BFA, a fungal metabolite that inhibits a subset of ArfGEF proteins and disrupts Golgi morphology and trafficking[35,380]) treatment. Single nucleotide mutants predicted to be either dominantly active or inactive altered bacterial alkaline phosphatase (BAP) and dense granule protein 1 (GRA1) staining (both dense granule markers), causing more to be retained in the Golgi, suggesting that Rab6 is involved in dense granule formation[381].

In opisthokonts, Rab7 primarily localizes to late endosomal compartments, and is responsible for fusion of these compartments as well as recruitment of retromer for cargo recycling prior to terminal degradation[107,382–384]. In *T. gondii* Rab7 localizes to compartments that stain with TgSORTLR (a Vps10/sortilin homologue), the propeptide of the MIC2-associated protein (proM2AP), a vacuolar-H+-pyrophosphatase (TgVP1), and a cathepsin L-like protease (TgCPL), all markers of endolysosomal structures[248,253]. Over-expression of TgRab7 shows no obvious defects[253], and over-expression of either constitutively active or inactive forms in *Plasmodium falciparum* show no apparent defects either[385]; as such, Rab7 function in Apicomplexa is currently unclear.

Rab11 is perhaps best known for its role in endosome to plasma membrane recycling [386], but is involved in diverse processes including cytokinesis, ciliogenesis, and autophagy[387–389]. *T. gondii* possess two paralogues, the pan-eukaryotic TgRab11A and the alveolate-specific TgRab11B[390]. TgRab11B localizes apical to the Golgi, interacts with the retromer subunit TgVps26, and is involved in delivery of early and late stage components to the IMC during endodyogeny[375,390]. TgRab11A also localizes apical to the Golgi and co-localizes with ROP5; disruption of TgRab11A results in aberrant localization of the surface protein TgSAG1 in the Golgi region, suggesting TgRab11A mediates trafficking and/or recycling of cell surface material[391], and may have other roles as well[392].

Other Rab GTPases in Apicomplexa are less well-characterized; an over-expression screen localized the majority to distinct locations within the cell, including the ER/Golgi (TgRab1B, TgRab2, TgRab18) and the Golgi/TGN (TgRab4). Of these, TgRab2 and TgRab4 over-expression blocked growth, although in a CRISPR-based genome-wide disruption screen, TgRab2 was predicted to be essential while TgRab4 was not[253,354]. Thus, further studies are required to elucidate Rab function in Apicomplexa.

The only other characterized GTPase in *T. gondii* is the ADP-ribosylation factor (ARF) homologue, TgARF1. This protein localizes to the Golgi complex, and expression of a dominant negative version appears to affect dense granule formation[393].

1.6.1.2 Adaptor Proteins and Cargo Adaptors

Adaptor protein complexes were introduced in section 1.3.1.2. *T. gondii* encodes orthologues for all subunits of all five complexes, though, like other eukaryotes outside of certain opisthokonts, the beta subunit for the AP-1 and AP-2 complexes is shared[394,395] (see also Chapter 2).

AP-1 was one of the first complexes to be studied in *T. gondii*, and was localized to the Golgi and post-Golgi compartments[396]. Genetic disruption and yeast-two-hybrid interaction data suggested that AP-1 could interact directly with TgROP2 and mediate its transport to the rhoptries, which was later called into question by the discovery that ROP2 was not actually a transmembrane protein, and hence could not bind AP-1 in a physiological manner[396,397]. The discovery of TgSORTLR appeared to answer the question of how this interaction could be mediated, as TgSORTLR binds both AP-1 and TgROP2, as well as numerous other microneme/rhoptry proteins[398]. Further studies in both *P. falciparum* and *T. gondii* implicated AP-1 in rhoptry trafficking and, at least in *T. gondii*, microneme trafficking and cell division[258,399].

No extensive characterization of other adaptor protein complexes has been performed in *T. gondii* to date. Endogenous tagging of AP-2 alpha reveals localization at the plasma membrane, and intracellular puncta[294]. AP-3 has not been characterized, but a class of parasitistatic compounds (N-benzoyl-2-hydroxybenzamides) inhibiting *T. gondii* target AP-3β and result in substantial mis-localization of markers for the VAC/PLV, formation of aberrant "empty" dense granules, and less severe defects in micronemes and rhoptries[400].

1.6.1.3 Coats

The various coat complexes, including COPI, COPII, and clathrin, were introduced in section 1.3.1.3. In *T. gondii* only the beta subunit of COPI has been characterized, which localizes primarily to the Golgi complex[240,401]. Conversely, TgSec23 staining was reported to be mostly cytosolic[240]. However, various COPII components (Sar1, Sec12, Sec13, and Sec24) have been localized to the ER in *Plasmodium falciparum*, suggesting that COPII may localize and function as in other eukaryotes[402–404]. Clathrin has been comparatively better-studied. Endogenously tagged clathrin heavy chain (TgCHC) localizes to the TGN, and over-expression of a dominant negative fragment ("CHC-Hub") resulted in pleiotropic defects consistent with a generalized role in Golgi maintenance and post-Golgi trafficking, consistent with its role in other systems[405].

1.6.1.4 Tethers

Tethers were discussed in section 1.3.1.5. With the exception of the VpsC core complex subunits of Vps11, Vps16, Vps18, and Vps33, no MTC complex is universally conserved in Apicomplexa, despite almost complete retention of most subunits in their free-living chromerid ancestors[127,395] (see Chapter 2).

The only characterized tether to date in Apicomplexa is the VpsC complex. This core complex additionally binds either Vps3/8 or Vps39/41 on its terminal edges to produce either the early endosomal CORVET or late endosomal HOPS complexes[406]. It appears that the only complex present in canonical form is the CORVET complex; Morlon-Guyot et al. (2015) describe the lone TgVps3/39-like subunit as Vps39, but robust phylogenetic analysis clearly demonstrates that this subunit is a Vps3 orthologue[127,407]. This study also reported the absence of Vps8[407], despite informatic prediction of a canonical Vps8 orthologue previously[127], and which was subsequently confirmed as a *bona fide* VpsC complex subunit by pulling down TgVps11[408]. An additional subunit, TgBDCP, may represent an analogous replacement for the canonical Vps39 subunit in these parasites, as it interacts with the VpsC complex and knockdown of the protein results in aberrant vacuolar morphology[408]. Determining Rab interactions between these subunits may further assist in confirming their identities, although interpretation of such results should be performed carefully, as yeast HOPS directly binds Rab7 while mammalian HOPS appears to do so only through intermediates[382,384].

1.6.1.5 SNAREs and ESCRTs

SNAREs and ESCRT complexes were discussed in sections 1.3.1.6 and 1.3.1.7, respectively. To date, only a Syntaxin-6 (TgStx6) homologue has been characterized. It localizes to the Golgi/TGN and is potentially involved in fusion of retrograde vesicular traffic from the endosomal system with the TGN, as over-expression causes fragmentation of Golgi and post-Golgi compartments[409]. Apicomplexa retain few ESCRT subunits; the apicomplexan ESCRT coding complement is discussed in Chapter 2.

1.6.2 Organization, Markers, and Function of the Apicomplexan MTS

1.6.2.1 Overview

Despite early skepticism regarding the putative nature of the apicomplexan MTS, based largely on the presence of unique organelles in addition to the apparent absence of others (for example, and as discussed in section 1.5.1.2, the absence of a stacked Golgi in *Plasmodium*[241,242]), numerous studies have elucidated a surprisingly conserved organellar

complement for the apicomplexan MTS. For simplicity, these are discussed here in terms of their evolutionary origin (endosymbiotic versus autogenous).

1.6.2.2 Endosymbiotic organelles – apicoplast and mitochondrion

The apicoplast, as discussed in section 1.5.1.7, is bound by four membranes, and can be recognized by DNA staining (e.g. with DAPI, Hoescht), but also by specific markers including the chaperone protein CPN60[410], and fluorescent protein fusions to the ferredoxin-NADP reductase, e.g. FNR-RFP[411]. Due to the location of the apicoplast outside of the ER, the presence of a fourth bounding membrane, and the paucity of defined import machinery, it is hypothesized that the initial step in delivery of nuclear-encoded proteins to the apicoplast involves either transient fusion of the ER with the apicoplast outer membrane or vesicular trafficking[236]. Initial studies in both *T. gondii* and *P. falciparum* suggested a direct route from ER to apicoplast, based mainly on the insensitivity of intra-apicoplast propeptide processing to treatment with BFA and the lack of effect of XDEL-based ER retention signals on the localization of apicoplast resident proteins[412,413]. However, a recent study in *P. falciparum* has used similar lines of evidence to suggest that apicoplast trafficking does in fact transit the Golgi[414]. In the latter study, lack of apicoplast protein processing upon addition of ER retention signals was interpreted as being due to continual cycling of the fusion protein between ER and Golgi, preventing its post-Golgi transport to the apicoplast, rather than by a direct block in an ER to apicoplast trafficking route, which was the interpretation favoured by the previous studies. It is also possible that different subsets of apicoplast proteins follow different routes or that there is some partial redundancy involved.

Unlike the apicoplast, nuclear-encoded proteins are thought to directly enter the mitochondrion from the cytosol via the translocase of the outer/inner mitochondrial membrane (TOM/TIM) translocation machinery[415]. Interestingly, mutation of the mitochondrial targeting sequence of TgSODB2 reroutes the protein from the mitochondria to the apicoplast[416]. Additionally, proteins targeted to both the apicoplast and mitochondria have been described[417], and are suggested to take an ER-Golgi route[418]. Hence, it is unclear how proteins are trafficked to endosymbiotic organelles in Apicomplexa. It is also possible that multiple routes exist and that proteins destined for different organelles, or even different compartments within the same organelle, might be trafficked differently.

1.6.2.3 ER, Golgi, and IMC

The irregular and branched ER of *T. gondii* may be visualized by fusions with the HDEL motif (for example, the P30-GFP-HDEL fusion[240]). Even more markers exist for the Golgi; for

the *cis*- and *medial*-Golgi, fusions of exogenous proteins to the human low density lipoprotein receptor, including BAP-LDLR[419], fluorescence fusions of the *T. gondii* homologue of ER-retention deficient (TgERD2)[401], and similar fusions to the mammalian GRASP55 protein[243] act as markers. In addition, identification of an O-linked glycosylation factor, UDP-GalNAc:polypeptide N-acetylgalactosaminyl-transferase (GalNAc or NAGTI) yielded a marker for the *trans*-Golgi network[244,245], and maybe for further post-Golgi compartments and/or TGN subdomains as well[420].

The most frequently used markers for the IMC are the various alveolin/IMC (e.g. IMC1) and gliding-associated (GAP, e.g. GAP45) proteins, which, as discussed above, rely on both TgRab11A and TgRab11B[390,391] to traffic to the IMC during endodyogeny. Far from a simple structure, the IMC has been shown to have multiple distinct subdomains[234,421]. Additionally, although it was previously thought that the IMC was rather impermeable, recent work has shown "gaps" in this structure[238]. These are potentially important for trafficking, including in delivering dense granule and surface material to the plasma membrane, and for internalization and subsequent trafficking of up taken material.

1.6.2.4 The Dense Granules

Although dense granules are known to move along actin tracks[422], little is known about their biogenesis. Early studies suggested that dense granules represent a form of constitutive secretion in the parasite[423], and this may account for mis-localization of apical organelle components with disrupted trafficking, as discussed below. Additionally, although poorly defined, disruption of TgRab6, TgARF1, TgDrpB (a dynamin-related protein), TgStx6, retromer, or the NSF/SNAP machinery (required for SNARE mediated fusion) adversely affected dense granules[375,381,393,409,424,425]; in the latter case, subsequent studies have shown inhibition of the SNAP machinery to have broad effects on the Golgi, secretory organelles, and apicoplast[426]. Processing of dense granule proteins has been shown to be mediated, at least in part, by the Golgi/TGN-localized TgASP5[427,428]. Combined with the observations regarding mis-localization of microneme/rhoptry proteins, which are known to transit the Golgi/TGN, it is likely that dense granule formation occurs from, or downstream of, this compartment.

1.6.2.5 The Endosomal System

The organization of the endosomal system is poorly described in *T. gondii.* Early studies localizing a paralogue of the early endosomal Rab5 GTPase (then referred to as TgRab51 but now known as TgRab5A) revealed its presence at the *trans* face of the Golgi and in tubulovesicular extensions and electron-lucent vesicular structures[246]. This marker was

subsequently found to co-localize with TgVP1, proM2AP, and other markers in a structure that was presumed to be something similar to an early endosome[245]. However, several observations, including partial co-localization with GalNAc-YFP and partial dispersal of TgRab5A signal upon treatment with BFA[245,246,258], suggests that at least some of this TgRab5A signal is associated to the Golgi/TGN.

However, these "early endosome" markers also co-localize with other markers, including the atypical TgRab5B and late endosomal TgRab7[253], TgStx6[409], TgSORTLR[398], the retromer complex[375], AP-1[258], and the Na+/H+ exchanger TgNHE3[429]. Additionally, the VAC is marked by cathepsin B and L proteases (TgCPB or "toxopain-1" and TgCPL, respectively), an aquaporin (TgAQP1), TgVP1, and TgRab7. This organelle is highly dynamic though, and fragments following host cell invasion such that TgCPL and TgVP1 no longer extensively co-localize[247,248]. In addition, despite being used as markers of the PLV/VAC, TgCPB is also found in rhoptries[430], and TgCPL in micronemes[248], suggesting that these compartments are part of the same trafficking system.

In terms of their evolutionary origins, micronemes and rhoptries are squarely placed as endo-lysosomal in origin. Early studies, on the basis of numerous lines of evidence suggested that rhoptries may be secretory lysosomes[431], or "lysosome-related organelles" (LROs). Subsequent studies focussed on molecular characterization of factors classically associated with the endosomal system, including TgDrpB[425], endocytic Rabs (TgRab5 and TgRab7)[253], the clathrin coat[405], TgSORTLR[398], AP-1[258], the retromer complex[375], and the VpsC tethering complexes[407,408]. In all of these studies, disruption of these components led to defects in the localization of microneme and rhoptry components, often with the absence of morphologically recognizable organelles. These results supported the LRO hypothesis and also suggested that micronemes represent yet another class of apicomplexan LRO[236].

In addition, these observations also lead to a hypothesis that Apicomplexa "repurposed" their endosomal systems to facilitate secretion, and downplayed a potential role for the same system in internalization/endocytosis[432]. However, due to limited metabolic capabilities, as well as the need for recycling during extracellular gliding motility and intracellular replication, it was subsequently recognized that parasites must be able to perform internalization/recycling. This was subsequently shown, first for intracellular[420,433] and later extracellular[294] stages of *T. gondii*. Hence endo/exocytic events appear intricately intertwined, and it appears that the organization of the endosomal system in *T. gondii* is likely more complex than previously envisioned.

1.6.2.6 A Model for the Apicomplexan MTS

This section introduces a new working model for post-Golgi trafficking and secretory organelle biogenesis which attempts to reconcile the available data from diverse studies (Figure 1.8). In this model, the TGN provides material to either a single "endosome-like compartment" or an intricately interconnected network of compartments (either way this is referred to using the common abbreviation ELC), and likely to nascent dense granules as well. As such, the TGN acts as a sorting station for exo/endocytic cargo, in which this cargo either traverses the system to terminal secretory organelles or degradative compartments or is trafficked/recycled to other subcellular locations. It is likely that some compartments mature and either fuse, or exchange material, with the terminal VAC. This model was envisioned to explain the following observations (also discussed further below):

1) Propeptide processing of MIC and ROP proteins is blocked by BFA, suggesting it occurs in a post-Golgi compartment[245,434–437].
2) TgSORLTR localizes primarily throughout the ER, Golgi, and endosomes, and binds microneme and rhoptry resident proteins, as well as AP-1 and retromer. Disruption prevents biogenesis of micronemes and rhoptries (and their resident proteins are mislocalized), but has no apparent effect on dense granules, the IMC, TgSAG1, or other organelles[398].
3) Overexpression of either TgRab5A or TgRab5C causes mis-localization of rhoptry bulb proteins but not microneme proteins; over-expression of dominant negative versions of either Rab additionally causes mis-localization of TgMIC3 and TgMIC8, but not TgM2AP or TgMIC2[253].
4) Retromer binds TgRab5B and TgRab11B, and co-localizes with diverse endosomal markers (TgSORTLR, proM2AP, and TgVP1). Disruption causes mis-localization of microneme, dense granule, and rhoptry components, abrogates the processing of TgROP1, TgROP2, TgROP4, TgM2AP, and TgMIC5, shifts localization of TgSORTLR from the Golgi to Rab5/7-positive compartments, and is defective in recycling of trans-membrane plasma membrane proteins. The processing defect likely reflects the arrest of proteins in the Golgi/TGN, prior to a compartment in which they can be cleaved, likely due to the re-distribution of TgSORTLR[375,398].
5) AP-1 localizes primarily to the TGN (co-localizes with GalNAc-YFP, TgSORTLR, proMIC3, and proROP4, but not proM2AP) and is partially dispersed by BFA treatment. AP-1 knockout has no effect on dense granules but re-routes TgMIC3/MIC8 and proROP4/ROP2,3,4 to the vacuolar space, possibly through dense granules. Over-expression has no effect on microneme or rhoptry protein processing or microneme protein localization but results in ROP signal in puncta

throughout the cytosol. This is likely because forming rhoptries require fusion of multiple AP-1-derived vesicles, whereas forming micronemes do not, and hence, over-expression of AP-1 leads to increased vesiculation and prevents fusion of ROP-positive vesicles and hence rhoptry maturation[258,396]. Additionally, similar to TgRab11A, AP-1 disruption results in defective membrane delivery/recycling and aberrant endodyogeny[258,391].

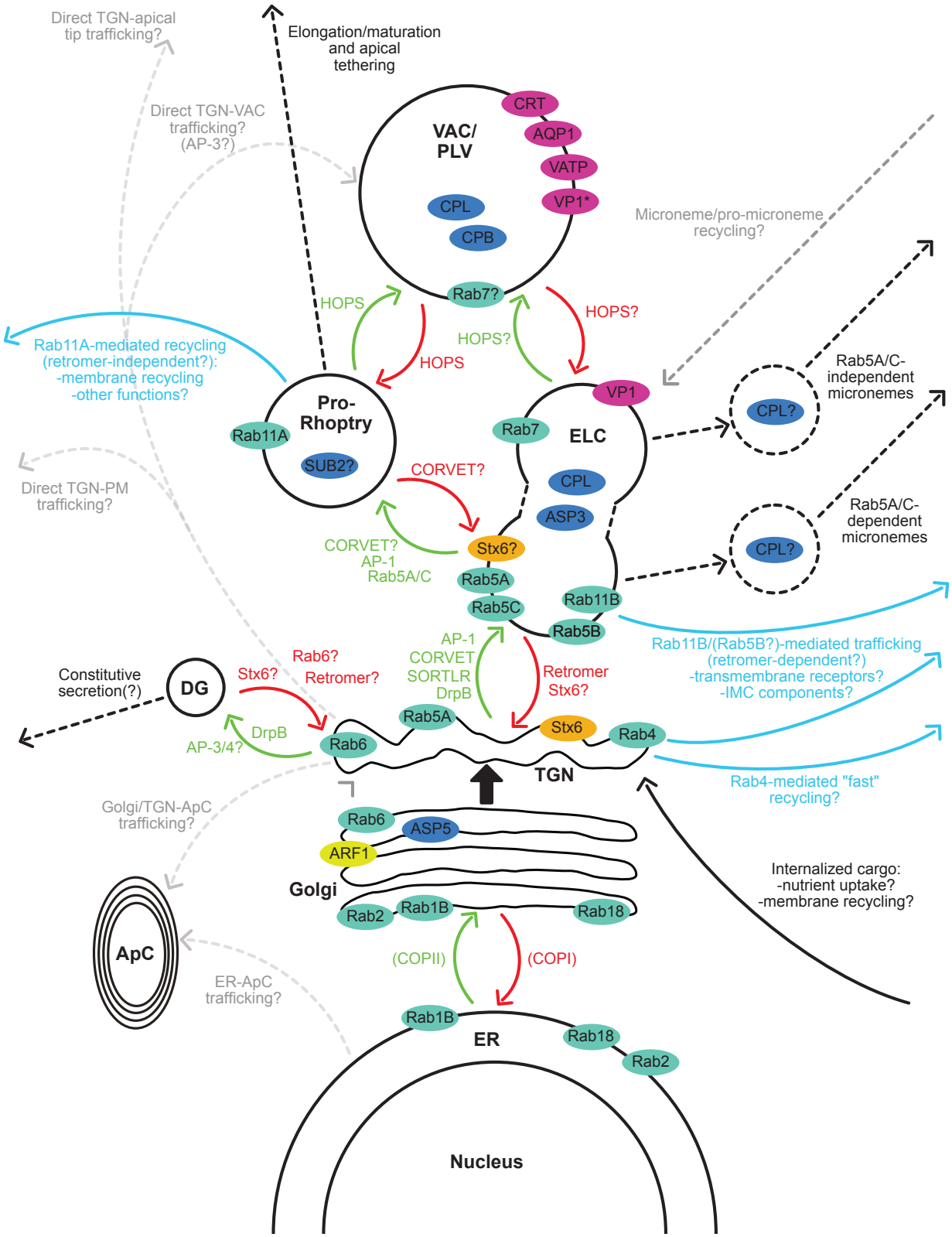
6) A class of parasitistatic compounds that appear to specifically interact with and/or inhibit the beta subunit of the AP-3 complex in *T. gondii* result in vacuolar fragmentation, formation of "empty" DGs, and some mislocalization of both microneme and rhoptry proteins[400].
7) Ingested protein co-localizes with GalNAc-YFP, dd-GFP-DrpB, proM2AP, proMIC5, and TgCPB/L, but not with TgNHE3 or proRON4. As uptake appears to occur throughout the intracellular lifecycle, this suggests that material is taken up either into a sub-domain of the TGN that lacks TgNHE3 or a separate Rab5-positive compartment that contains proMIC, but not proRON, components, prior to trafficking to the VAC[420,433].
8) In extracellular parasites, uptake of exogenous material occurs concurrently with gliding motility (possibly allowing for membrane recycling). In these parasites, a substantial amount of internalized lipid is redistributed to the ER, Golgi, and VAC, but less is found co-localized with proM2AP-positive compartments. Additionally, some internalized material incapable of being sorted (nano-gold particles) localizes to the rhoptries[294].
9) TgRab11A localizes to rhoptries and mediates trafficking (and possibly recycling) of TgSAG1, and potentially other surface proteins. TgRab11B, alternatively, localizes in the vicinity of the Golgi and subsequently to the IMC, and mediates delivery of early and late stage IMC components during endodyogeny[390,391].
10) TgCPB is found both in the VAC and in rhoptries and TgCPB inhibition leads to decreased processing, as assessed with the ROP2,3,4 antibody[430]. Conversely, CPL is found in the PLV/VAC, compartments containing proM2AP, and even in mature micronemes as well. CPL knockout parasites fail to process proM2AP/MIC3 but not MIC5, MIC6, or AMA1[248]. Additionally, M2AP signal in trafficking-deficient parasites localizes differently than MIC5 signal, suggesting that different maturases function at different stages or that different trafficking pathways from the site of maturation to the apical end of the parasite exist[438].

11) TgVps11, a core VpsC subunit, co-localizes with GalNAc-YFP, TgCPL, TgRab5B, TgRab7, proM2AP, and proROP4. Vps11 knockdown causes mis-localization of diverse microneme, rhoptry, and dense granule markers without affecting endosymbiotic organelles or the IMC/plasma membrane[407]. TgBDCP, a putative VpsC interactor in *T. gondii,* co-localizes with GalNAc-YFP, TgRab5B, proROP4, and TgCPL, but not with proM2AP or TgRab7; knockdown causes a change in TgARO (a rhoptry surface protein involved in apical rhoptry positioning[439]) and TgCPL staining, and increases overlap between TgROP7 and TgCPL. TgVps8, a CORVET-specific subunit, co-localizes well with TgRab5B and somewhat with TgRab7. KD of TgVps8 causes similarly broad trafficking defects as TgVps11[408].
12) TgStx6 localizes mainly to the TGN and proM2AP-positive compartments, and occasionally at the cell surface. Over-expression of a dominant negative version causes the membranes in the Golgi region to fragment/vesiculate, as assessed by electron microscopy, and causes proM2AP and TgVP1 staining in fluorescence microscopy to increase as well as become fragmented. Additionally, it disrupts normal cell morphology similar to TgRab11A and likely has an effect on dense granule biogenesis[409]. This suggests that this factor acts to tether endosomally-derived vesicles at the TGN, and possibly also mediates some homo/heterotypic endosomal fusion.
13) Disruption of some factors, including TgVps11[407] and AP-1[258], have resulted in extreme apical staining of a subset of microneme proteins (e.g. MIC1, MIC2, M2AP, MIC4, MIC6). This was suggested to represent a hitherto unrecognized direct TGN-apical microneme trafficking pathway. However, this phenotype is also observed in parasites lacking TgSORTLR (for example, see Figure 4 in Sloves et al. (2012)[398]). Hence, here it is considered more likely that this phenotype represents aberrant inclusion of a subset of microneme proteins into a different pathway, potentially for delivery of components to the apical complex, as they likely accumulate in the TGN in these mutants.
14) A dynamin-related protein, TgDrpB, localizes in the vicinity of the Golgi, and over-expression of a dominant negative form causes broad defects on micronemes, rhoptries, and dense granules[425].

Almost all models proposed to date involve microneme and rhoptry trafficking proceeding through a post-Golgi/TGN compartment, with micronemes distinct from the endosomal

Figure 1.8 Proposed model of *T. gondii* endosomal trafficking

Organelles are depicted with associated molecular markers. Potential trafficking routes between organelles are also shown, along with known or hypothesized machinery for each step; for simplicity, not all organelles and markers are shown. Dashed lines represent uncertainty in organelle identity or trafficking step. Solid green arrows represent anterograde, and red arrows retrograde, trafficking steps while black arrows represent internalization and light blue arrows putative recycling steps. The thick black arrow between the Golgi and TGN represents cisternal maturation. Dashed grey lines represent trafficking steps that are not described, but that may exist on the basis of other evidence. Bold text denotes organelle labels while all other text denotes pathways or trafficking machinery. Markers are shown as filled ovals, with the colour corresponding to the type of marker: magenta, various trans-membrane proteins; deep blue, proteases/maturases; teal, Rabs; orange, SNAREs; yellow, ARFs. Abbreviations: CRT, chloroquine resistance transporter; AQP, aquaporin; VATP, vacuolar ATPase; VP, vacuolar-H+-pyrophosphatase; CPL/B, cathepsin protease L/B; SUB, subtilisin-like protease; ASP, aspartyl protease; VAC/PLV, plant-like vacuolar compartment; ELC, endosome-like compartment; ApC, apicoplast; DG, dense granule; TGN, *trans*-Golgi network; ER, endoplasmic reticulum.


Direct TGN-apical tip trafficking?
Elongation/maturation and apical tethering
Direct TGN-VAC trafficking? (AP-3?)
VAC/ PLV
CRT
AQP1
VATP
VP1*
CPL
CPB
Microneme/pro-microneme recycling?
Rab7?
HOPS
HOPS
HOPS?
HOPS?
Rab11A-mediated recycling (retromer-independent?): -membrane recycling -other functions?
Pro-Rhoptry
Rab11A
SUB2?
VP1
Rab7
ELC
CPL?
Rab5A/C-independent micronemes
Direct TGN-PM trafficking?
CORVET?
CPL
ASP3
CORVET? AP-1 Rab5A/C
Stx6?
Rab5A
Rab5C
Rab11B
Rab5B
CPL?
Rab5A/C-dependent micronemes
Rab11B/(Rab5B?)-mediated trafficking (retromer-dependent?) -transmembrane receptors? -IMC components?
AP-1 CORVET SORTLR DrpB
Retromer Stx6?
Rab6? Retromer?
Constitutive secretion(?)
DG
Stx6?
DrpB
AP-3/4?
Rab6
Rab5A
Stx6
Rab4
TGN
Rab4-mediated "fast" recycling?
Golgi/TGN-ApC trafficking?
Rab6
ASP5
ARF1
Golgi
Rab2
Rab1B
Rab18
Internalized cargo: -nutrient uptake? -membrane recycling?
ApC
(COPII)
(COPI)
ER-ApC trafficking?
Rab1B
ER
Rab18
Rab2
Nucleus

organelles themselves (see for example Venugopal and Marion (2018)[392]); this view is maintained here.

The emergence of proROP proteins from a Rab5A-positive compartment(s) has been recently described in detail[258]. The identity of Rab5- and Rab7-compartments has been difficult to pin down precisely, based largely on partially overlapping TGN/endosomal markers, such as TgRab5A/B/C, TgRab7, TgVP1, proMIC proteins (e.g. proM2AP, proMIC3), and the partial disruption of TgRab5A/AP-1 upon BFA treatment, which suggests that at least some of this "ELC" marker is present connected with the Golgi/TGN. This model presents either a single organelle that is positive for both markers, but possesses distinct subdomains, or a single maturing organelle/network that changes from being Rab5- to Rab7-positive. The former is similar to the situation described in *Arabidopsis*[440,441], and more recently, yeast[442], whereas the latter is more similar to mammalian endosomal systems[443]. Either way, the TGN is envisioned as the first stop for internalized material (i.e. more similar to yeast/plant systems). This is based on several lines of evidence, such as the co-localization of internalized material in both intracellular and extracellular tachyzoites with Golgi/TGN markers[294,420,433], the presence of TgRab4, which in mammalian cells mediates "fast" recycling of cargo from early endosomes to the plasma membrane[444], at the TGN[253], and the apparent cytokinetic defect observed in AP-1 deficient parasites, which is consistent with a reliance for membrane recycling on an early TGN-ELC step[258].

Even with the limited data currently available, it is clear that many differences exist between trafficking in Apicomplexa versus other well-studied systems such as human cells and yeast. This includes the presence of specialized organelles such as the micronemes and rhoptries, which, as argued above, appear to represent additional endosomal compartments. The question of how newly distinct organelles emerge in eukaryotes will be considered in the next section.

1.7 Mechanisms of MTS evolution

Despite the presence of internal compartmentalization in some bacteria, such as planctomycetes[445], the MTS is considered a defining hallmark of eukaryotic life[446]. Despite a paucity of possible homologues in closely related archaeal taxa[446,447], parsimony-based reconstruction of multiple MTS families suggest that the last eukaryotic common ancestor (LECA) possessed a complement of MTS genes similar to that in extant eukaryotes[448]. Given that a large number of MTS genes represent paralogous gene families, and that combinatorial interactions between paralogues of each family are capable of encoding the specificity of each trafficking step, a model was derived to explain both the complexity of protein-coding genes

in the eukaryotic MTS, but also the diversification of autogenously derived organelles (i.e. not derived by endosymbiosis like plastids or mitochondria, Figure 1.9).

The organelle paralogy hypothesis (OPH) posits that individual ancestral homologues of the main paralogous gene families (e.g. Rabs, SNAREs, etc.) duplicated to give rise to multiple paralogues. Gene duplication as a mechanism of generating biological complexity was first formally described by Ohno (1970)[449], who described a neofunctionalization mechanism whereby the additional copy was free of selective pressure and hence could acquire new function. A competing model to explain retention of duplicated genes is subfunctionalization, in which deleterious mutations occur in both the original and duplicated gene such that both are required to maintain ancestral function[450]. Other models posit that gene duplication is followed first by a period of subfunctionalization and then subsequently by neofunctionalization of one or more gene copies[451,452]. For simplicity, I focus on the role of possible eventual neofunctionalization throughout this thesis, without explicit reference to the existence of subfunctionalization during the process of gene duplication giving rise to functionally distinct paralogues.

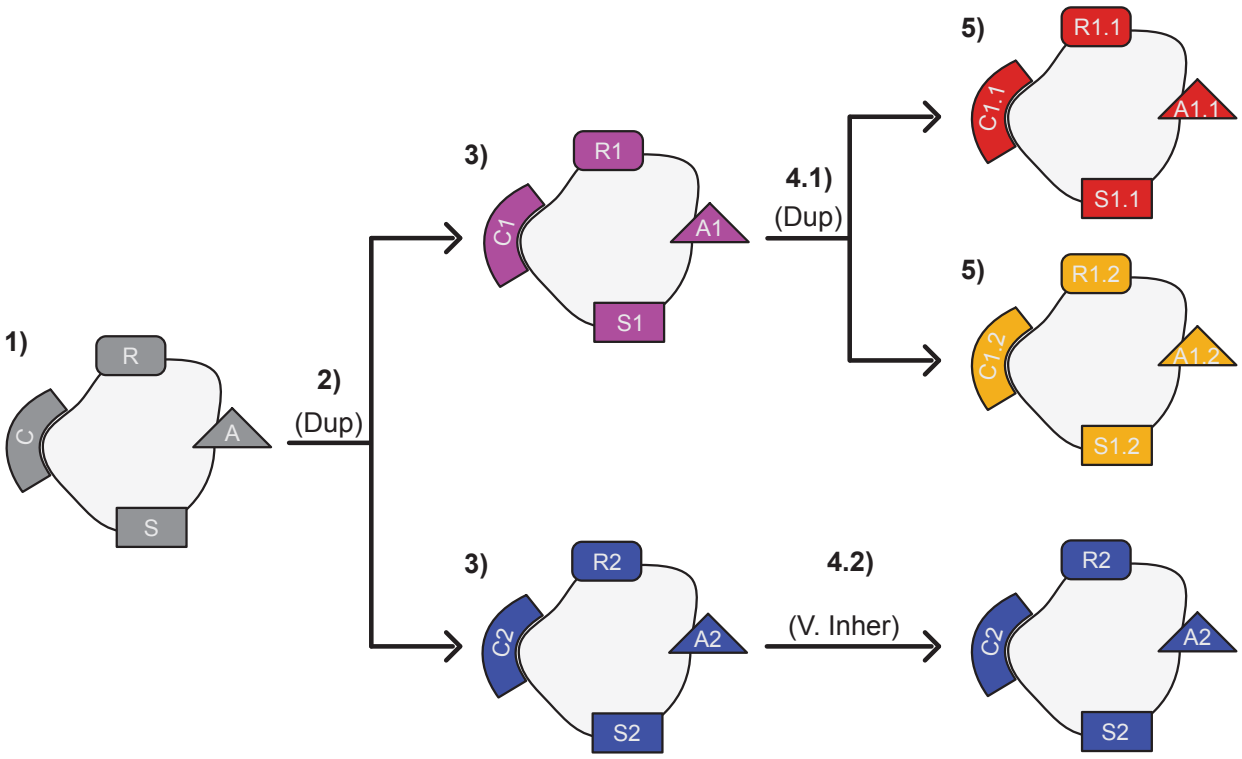
Provided the initial paralogue provided its required function, or if that function was no longer necessary, then the additional paralogues were unconstrained to accumulate mutations or other novel sequence features, and, importantly, new functions. Novel autogenous organelles could therefore arise over time, with these additional paralogues available to provide specific trafficking required to maintain organelle identity[453]. This model is supported by data from more recent paralogous duplications such as the split of a single beta subunit of both AP-1 and AP-2 complexes into individual β1 and β2 subunits in some taxa[394], and from simulations regarding the link between gene complement and organelle composition[454].

1.8 Rationale and hypothesis

As discussed in section 1.5.1.1, the apical complex, together with the associated micronemes and rhoptries, is similar to structures found in other myzozoan taxa[232]. Basal myzozoan taxa, such as *Colpodella vorax*, are flagellated heterotrophs that feed by a unique method known as myzocytosis: they anchor via their apical pole to prey and "suck" the contents into their own cell body prior to fusion with a digestive compartment. Although unclear, it has been described that this process is coupled to secretion of microneme and rhoptry-like organelles[455]. It has been hypothesized that this intimate association, mediated through the apical complex, eventually "flipped" from one of extracellular digestion to intracellular parasitism[456]. As was discussed in section 1.6, micronemes and rhoptries are thought to be derived from the endosomal system[236], although they are not a universal

Figure 1.9 Model for the role of gene duplication in organellogenesis

This figure provides an overview of the organelle paralogy hypothesis (OPH), whereby gene duplication and divergence are coupled to the increase in complexity of organelle complement. 1) In this example, there is a single initial endomembrane compartment, with exemplar specificity-encoding machinery: a protein coat (C), Rab GTPase (R), Arf family GTPase (A), and SNARE (S). These machineries duplicate (2), giving rise to paralogues (e.g. C1 and C2), denoted here by different colours, and, as a result, two separate endomembrane compartments (3). One set of machinery repeats the process of duplication (4.1), giving rise to further paralogues (e.g. C1.1 and C1.2) and an additional compartment (5), while the other set of machinery is maintained unchanged through vertical inheritance (4.2). In this toy example, a single ancestral compartment eventually gives rise to three compartments, based on the mechanism proposed by the OPH. Note that, for simplicity, not all possible machinery is shown (e.g. GEFs and GAPs), and all machinery duplicates in each case. Abbreviations: Dup, duplication; V. Inher, vertical inheritance. Figure adapted from Dacks and Field (2007)[453].


1)
R
C
A
S
2)
(Dup)
3)
R1
C1
A1
S1
4.1)
(Dup)
5)
R1.1
C1.1
A1.1
S1.1
5)
R1.2
C1.2
A1.2
S1.2
3)
R2
C2
A2
S2
4.2)
(V. Inher)
R2
C2
A2
S2

eukaryotic feature. Hence, the advent of the apical complex and its associated organelles in the myzozoan ancestor represents an organellogenesis event crucial to the development of intracellular parasitism.

As was discussed above in section 1.7, new organelles can theoretically be generated through duplication and diversification of MTS machinery. However, this is also contingent on several factors, all of which centre around the fundamental question of functional homology – whether gene function is maintained, and hence can be reliably predicted, across orthologous genes from diverse taxa. If functional homology is indeed maintained, then the presence of an orthologue of known function in the genome of an organism is unlikely to be associated with a novel organelle, except in the instance that the original function of that gene (and hence, likely the associated organelle) are absent. A conserved complement is therefore unlikely to be associated with novel organelles; in the case of gene loss, this is even more unlikely. However, additional MTS paralogues arising in extant taxa could be associated with the advent of novel organelles, as predicted by the OPH[453].

The aim of this thesis is to examine the extent to which Apicomplexa, with specific reference to *T. gondii*, can inform the hypothesis that an OPH-like mechanism of duplication and diversification has operated since the LECA to contribute to the diversity of organelles found in extant eukaryotes. I hypothesized that the advent of novel apicomplexan organelles, specifically micronemes and rhoptries, is associated with additional novel paralogues of MTS machinery in Apicomplexa and their closely associated sister taxa.

In order to explore this, Chapter 2 reports the coding complement of Apicomplexa in more detail for several MTS families and reveals a common pattern associated with parasite evolution: loss of non-essential genes. Chapter 3 explores the relationship between gene orthology and function across eukaryotes using a variety of model taxa, strongly arguing for the existence of functional homology, at least within the trafficking system. Having established that Apicomplexa have lost numerous components, and that functional homology exists to a large extent in the MTS, Chapters 4 and 5 then explore a phylogenetic workflow to identify novel paralogues in subsets of eukaryotic genomes and details the molecular characterization of three novel Arl proteins in *T. gondii*. Chapter 6 then discusses these results in detail to present a view as to the extent that paralagous expansion and diversification within the MTS can explain the diversity of organelles across eukaryotes.

2. Chromerid genomes and the evolution of Apicomplexa

2.1 Chromerids as free-living apicomplexan relatives

Approximately 20 years ago, the identity of an enigmatic multiple membrane-bound compartment in Apicomplexa was elucidated – this was a relic plastid (now known as the apicoplast), suggestive of a photosynthetic apicomplexan ancestor[262]. These results are consistent with the presence of plastids in the apicomplexan sister clade of dinoflagellates, and also suggest that the apparent absence of a plastid in both *Cryptosporidium*[457,458] and at least some gregarines[219] is a secondarily derived character. However, for this latter point to be fully supported, it would be predicted that additional photosynthetic basal taxa to Apicomplexa would exist, which would make multiple independent acquisitions of a plastid in the myzozoan tree increasingly unlikely (Figure 1.3). However, for a long time all described basal taxa, such as colpodellids, were known as free-living heterotrophic flagellates rather than photosynthesizers[455,459].

Evidence for such photosynthetic relatives came from environmental sequencing of plastid 16S rRNA genes, which revealed the presence of numerous undescribed lineages globally associated with coral reefs[206]. Two of these were described at the ultrastructural level and shown to have fully photosynthetic plastids primarily pigmented by chlorophyll a, but lacking chlorophyll c, unlike dinoflagellates and most chromalveolates[207,208]. The first of these, *Chromera velia*, was described in 2008[207] and followed by a description of *Vitrella brassicaformis* in 2012[208]. Although these two organisms are typically referred to as "chromerid algae" or simply "chomerids", this group is likely paraphyletic and these two representatives encompass a comparatively small sampling of the diversity of basal apicomplexan-related lineages[206]. Also of note, the plastid in both *C. velia* and *V. brassicaformis* is bound by four membranes, similar to the apicoplast but with one more membrane than surrounds the ancestral peridinin dinoflagellate plastid, suggesting that the ancestral myzozoan plastid was bound by four membranes[208].

Chromerids represent a key group to understand the evolution of Apicomplexa. As the closest free-living photosynthetic apicomplexan relatives known to date, they represent a key outgroup for comparative studies. In general, the evolution of parasitism in eukaryotes is associated with reduction in the number of encoded genes and pathways, referred to as "stripping down" or "streamlining". The basic assumption is that parasites are able to rely to some extent on their host, especially for key metabolites, and are therefore able to lose biosynthetic and other cellular machinery so long as the function of this machinery is compensated for[460]. As absence in comparative genomic analysis is not informative, it is useful to have one or more free-living outgroups to the parasitic group under study to time gene loss. In the case of the Apicomplexa, the closest outgroups up until recently for which genomic

data was available are the dinoflagellates and their basal sister taxa such as *Perkinsus marinus*[461]. However, to date, dinoflagellate genomes are only available for several species within the *Symbiodinium* genus, and all such assemblies are incomplete and have some aberrant gene models. A more distant outgroup, the ciliates, have comparatively more genomic data but frequently have massively expanded gene families (due in part to the occurrence of whole genome duplications in some lineages), and also possess a number of unique morphological and genetic features[462,463].

Chromerids represent a key outgroup and a logical target for sequencing, and indeed were sequenced at high coverage to provide robust assemblies. These assemblies are crucial to the understanding of apicomplexan MTS evolution as described in this thesis. Not only do they provide a more accurate estimate of the timing of gene loss (described both in this chapter and in Chapter 4), but they also represent a key outgroup to understanding the timing of emergence of novel features (some examples are described in Chapter 4). This chapter presents an analysis of MTS machinery performed for the initial description of the chromerid genomes (section 2.2); for brevity, and because much of the work was not that of the author, the entire paper is not included, but is cited as Woo et al. (2015)[395].

2.2 Endomembrane trafficking system

2.2.1 Materials and Methods

The predicted proteomes of 26 species have been searched for endomembrane trafficking components. Initial homology searching was carried out using BLAST[464]. Known sequences from human (*Homo sapiens*) and yeast (*Saccharomyces cerevisiae*) were used to search the proteomes of each organism including *Chromera* and *Vitrella* to identify potential homologues of proteins implicated in endomembrane trafficking. Any sequences scoring an initial E value of 0.05 or lower were subjected to confirmation by reciprocal BLASTP. This involved the use of candidate homologous sequences as queries against the relevant *H. sapiens* or *S. cerevisiae* genome. Sequences that retrieved the query sequence, or named homologues/paralogues/isoforms thereof, first with an E value of 0.05 or lower were considered true homologues.

Additional searches were carried out using HMMER[465]. The HMMs for the initial queries were built and used to search each proteome. Top hits based on BLASTP results with E values less than 0.05 were considered confirmed homologues, and not subjected to further analysis. Subunits with significant HMMER hits were further investigated by reciprocal BLASTP as described above. Further HMMER searches were carried out with the addition of homologous

sequences from *Bigelowiella natans*, *Phytophthora infestans*, and *T. gondii* to the original HMMs. Results were analyzed identically to the first round. All identified endomembrane components are listed in Online Appendix Table 2.1.

To identify homologous proteins not predicted by the gene prediction software, TBLASTN was used with the homologous protein from the closest related organism in our data set against scaffolds and contigs; E value cut-off was identical to BLASTP analysis. Additionally, BLASTP was used to search either genome with an identified homologue from the other, if it was present. The final results are summarized in Figure 2.1 using the Coulson Plot Generator software[466].

2.2.2 Results and discussion

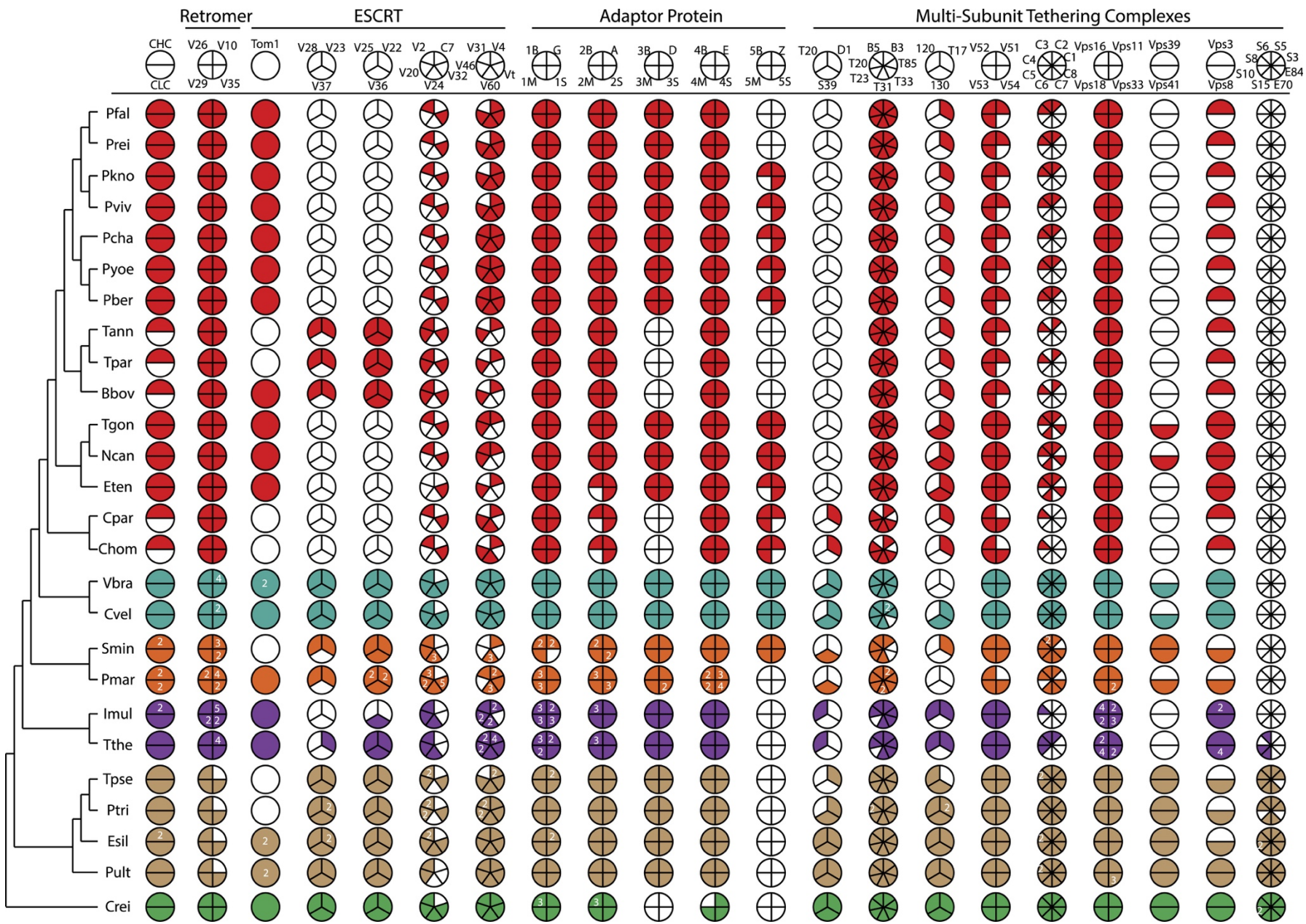
Apicomplexa possess unusual features in their membrane trafficking systems. Non-canonical membranous inclusions such as the invasion organelles, the micronemes, rhoptries, and dense granules are present[467]. Though canonical, stacked, Golgi bodies are present in *T. gondii*[243], other apicomplexan species possess Golgi bodies with aberrant morphology and unusual characteristics[403]. Combined with other organelle destinations such as mitochondria, digestive vacuoles involved in hemoglobin catabolism in *P. falciparum*, and plant-like lytic vacuoles in *T. gondii*[247], specificity of protein and lipid components of these various organelles suggest a need for unique trafficking pathways mediated by distinct protein machinery.

Interestingly, previous studies demonstrated the loss of trafficking machinery in Apicomplexa, including three key sets of proteins in the ESCRT machinery[182], adaptor protein complex (AP) families[48,468], and multi-subunit tethering complexes (MTCs)[126,127]. Several of the aforementioned families are involved in trafficking within the late endosomal system in opisthokont models and so may be associated with the evolution of the rhoptries and micronemes within the apicomplexan or myzozoan lineage. Consistent with this idea, some cases of reduction were not limited to Apicomplexa, and could be observed in the sister phyla of the ciliates and dinoflagellates.

This pattern of loss raises the question of what losses correlate with the transition to parasitism and which are pre-adaptive, arising more deeply in the lineage. The unique phylogenetic position of chromerids[202,206,208] allows finer dissection of the patterns of retention/loss observed previously. Hence, this section focusses on detailed characterization of the three previously studied sets of membrane trafficking machinery in the predicted proteomes of *Chromera* and *Vitrella*, together with 24 closely related organisms for comparison.

Fig. 2.1. An overview of endomembrane trafficking components

Coulson plot representation of the retention/loss of genes encoding trafficking gene complement of the Retromer, Clathrin, ESCRT, AP, and MTC family proteins amongst the 26 species. The fill colors indicate different phyla, for example, red Coulson plots for apicomplexans. Legends at the top of each column denote subunit components of complexes. For each organism, filled sectors of the pie represent presence of the corresponding protein, whereas empty sectors represent a failure to identify the corresponding protein in the genome. In cases where multiple copies of the protein are present, and can confidently be ascribed to unique genes, numbers indicate relevant paralog counts. The 26 species are shown on the left side with a phylogenetic tree. For simplicity, all subunits are listed as per yeast nomenclature, and only revert to human nomenclature when no homologous yeast gene exists. Taxon abbreviations: Pfal, *Plasmodium falciparum*; Prei, *P. reichenowi*; Pkno, *P. knowlesi*; Pviv, *P. vivax*; Pcha, *P. chabaudi*; Pyoe, *P. yoelii*; Pber, *P. berghei*; Tann, *Theileria annulate*; Tpar, *T. parva*; Bbov, *Babesia bovis*; Tgon, *Toxoplasma gondii*; Ncan, *Neospora caninum*; Eten, *Eimeria tenella*; Cpar, *Cryptosporidium parvum*; Chom, *C. hominis*; Vbra, *Vitrella brassicaformis*; Cvel, *Chromera velia*; Smin, *Symbiodinium minutum*; Pmar, *Perkinsus marinus*; Imul, *Icthyophthirius multifiliis*; Tthe, *Tetrahymena thermophila*; Tpse, *Thalassiosira pseudonana*; Ptri, *Phaeodactylum tricornutum*; Esil, *Ectocarpus siliculosus*; Pult, *Pythium ultimum*; Crei, *Chlamydomonas reinhardtii*. Abbreviations: CHC, Clathrin heavy chain; CLC, Clathrin light chain; V, Vps; C, CHMP; Vt, Vta1; B, Beta, M, Mu; S, Sigma, G, Gamma; A, Alpha; D, Delta; E, Epsilon; Z, Zeta; T20, Tip20; D1, Dsl1; S39, Sec39; T, Trs; T17, Tca17; C, COG; S, Sec; E, Exo; ESCRT, Endosomal Sorting Complex Required for Transport; MTC, multi-subunit tethering complex; AP, Adaptor Protein. IDs of genes encoding the components are listed in Online Appendix Table 2.1.


Retromer
ESCRT
Adaptor Protein
Multi-Subunit Tethering Complexes
CHC
CLC
V26 V10
V29 V35
Tom1
V28 V23
V37
V25 V22
V36
V2 C7
V20 V32
V46
V24
V31 V4
Vt
V60
1B G
1M 1S
2B A
2M 2S
3B D
3M 3S
4B E
4M 4S
5B Z
5M 5S
T20 D1
S39
B5 B3
T20 T85
T23 T33
T31
120 T17
130
V52 V51
V53 V54
C3 C2
C4 C1
C5 C8
C6 C7
Vps16 Vps11
Vps18 Vps33
Vps39
Vps41
Vps3
Vps8
S6 S5
S8 S3
S10 E84
S15 E70
Pfal
Prei
Pkno
Pviv
Pcha
Pyoe
Pber
Tann
Tpar
Bbov
Tgon
Ncan
Eten
Cpar
Chom
Vbra
Cvel
Smin
Pmar
Imul
Tthe
Tpse
Ptri
Esil
Pult
Crei

2.2.2.1 ESCRT machinery

The ESCRT machinery is a set of five sub-complexes involved in recognition of ubiquitylated proteins and recruitment to the multi-vesicular body (MVB)/late endosome for degradation[182]. Most eukaryotes, including *Chlamydomonas reinhardtii* and the representative stramenopile taxa (*Thalassiosira pseudonana, Phaeodactylum tricornutum, Ectocarpus siliculosus*, and *Pythium ultimum*), have a complete set of the ESCRT machinery, suggesting that the ancestor of alveolates, and indeed the Last Eukaryotic Common Ancestor (LECA) likely had it. Though this ancestral complement appears to have been reduced in ciliates in the ESCRTI and III complexes, and a few components are missing from dinoflagellate taxa, numerous gene duplications have occurred as well, suggesting sculpting of the machinery. By comparison, apicomplexan parasites exhibit significant reductions in their ESCRT machinery[182]. Cryptosporidia, coccidia, and plasmodia appear to lack any subunits of the ESCRTI and II complexes. ESCRTIII conservation is better, though no apicomplexan encodes Vps24, and multiple taxa have lost Vps20 as well. A similar pattern is seen for the ESCRTIII-a machinery, with piroplasmids encoding only Vps46 and Vps4. Coccidia additionally encode Vps31, and cryptosporidia Vps60, whereas plasmodia encode all subunits (rodent parasites like *Plasmodium chabaudi*), or lack Vps31 (human or simian parasites like *P. falciparum*). *Chromera* and *Vitrella* possess all ESCRT subunits except for the ESCRT-III component CHMP7, which is rarely found outside the opisthokont supergroup[182]. This observation suggests two conclusions regarding the evolution of the ESCRT machinery within alveolates: massive gene loss within the Apicomplexa occurred recently, after the split from the proto-apicomplexan ancestor, and some losses of machinery shared between apicomplexans and other alveolates are due to independent losses. An excellent example of this latter case is that of Vps37, which is present only in chromerids, but in no other alveolate included in the current study, suggesting its function was dispensable in a large number of lineages.

2.2.2.2 APs

The APs are heterotetrameric complexes that select cargo for inclusion into transport vesicles at organelles of the late secretory system and endocytic system. AP-1 and AP-3 are involved in the transport between the trans-Golgi network (TGN) and endosomes. AP-2 is involved in the transport from the cell surface. AP-4 is involved in TGN transport to either endosomes or the cell surface, while the recently described AP-5 complex is involved in the transport between late endosomes and early endosomes/TGN[52,469]. All five complexes are

ancient, having likely been present in the LECA[48,468]. However, the complexes have also been secondarily lost on multiple occasions as well. Outgroup taxa in our data set possess AP-1-4 complexes, with the exception of *C. reinhardtii* lacking AP-3, but only *Symbiodinium minutum* possesses an AP-5 complex.

Apicomplexa display higher variability in AP complex retention. With the exception of AP-2M in cryptosporidia, all taxa retain full AP-1, -2, and -4 complexes. Piroplasms lack all subunits of the AP-3 complex, and together with *P. falciparum* and *Plasmodium reichenowi*, lack AP-5 as well. Other plasmodia possess all AP-5 subunits with the exception of the mu subunit. This result was unexpected, based on the usual patterns of conservation seen across *Plasmodium* species. Presence of AP-5 in the majority of these organisms suggests the exciting possibility of a novel trafficking pathway absent from the comparatively well-studied human parasite *P. falciparum*. Additionally, our increased taxon sampling has suggested that AP-5 may be well conserved across Myzozoa, a result otherwise indeterminable from previous studies of this protein family[48]. Cryptosporidia also lack AP-3, but unlike piroplasmids, they possess almost a complete AP-5 complex, missing only the sigma subunit. Coccidia are the exception, possessing all five AP complexes in their entirety. Excitingly, *Chromera* and *Vitrella,* like coccidia, possess a complete complement of adaptin subunits, suggestive of a more complex set of trafficking pathways to endosomal organelles in these organisms.

2.2.2.3 MTCs

The MTCs are an assembly of heteromeric protein complexes involved in the first stage of vesicle fusion and delivery of contents from a transport vesicle to a destination organelle. Each one is specific to an organelle or transport pathway and all eight complexes have been deduced as present in the LECA, with some interesting cases of secondary loss. While *C. reinhardtii* and the stramenopiles encode a complete set of MTC machinery, several of these MTCs have interesting patterns of conservation, specifically in the Apicomplexa[127].

The conservation of the TRAPP I-II complexes is unclear through eukaryotes and clear patterns are difficult to draw. However, the apparent absence of the entire TRAPPII complex in *Vitrella* may be due to gaps/biases/absences in sequencing, protein prediction, or analysis, but has interesting ramifications if proven to be a real biological phenomenon.

Exocyst is involved in diverse processes, all of which involve polarized exocytosis[470]. *Tetrahymena* appears to encode only four of the Exocyst subunits. None of the eight subunits were identifiable in *Chromera*, *Vitrella*, nor in any of the Apicomplexa or dinoflagellates. This confirms, and extends, a previous result suggesting the absence of this complex within the

Myzozoa, suggesting a bona fide ancestral loss concurrent with the acquisition of an apical complex that could have served an analogous tethering function for secretory organelles.

Conserved oligomeric Golgi (COG) is an octameric complex involved in tethering at the Golgi body[127]. The COG complex is poorly conserved in Apicomplexa and *Tetrahymena thermophila* only encodes half of the COG subunits. In contrast, all eight COG subunits are present in *Chromera* and *Vitrella*. The retention of a complete COG complex in both *Chromera* and *Vitrella* contrasts with the substantial loss of subunits in Apicomplexa, especially outside the coccidians[127] (Figure 2.1, Online Appendix Table 2.1). Notably, this conservation is consistent with the presence of robust, stacked Golgi bodies in *Chromera*[471] and *T. gondii*[243], compared to aberrant morphology in other Apicomplexa.

Chromerids exhibit complex life cycles, from immotile vegetative cells to multi-cellular sporangia, and occasionally motile flagellated cells. Both lineages contain numerous potential locales for intracellular trafficking including mitochondria, plastid, starch granules, flagella, micronemes, and, in *Chromera*, the chromerosome. Additionally, vesicular traffic to the sporangial/cyst wall has been visualized in both lineages[208]. Our results indicate that chromerids possess an appropriately complex complement of membrane trafficking machinery to achieve these requirements.

Though MVBs have not been explicitly imaged or characterized in either lineage to date, both *Chromera* and *Vitrella* encode a complete set of ESCRT machinery, suggestive of the presence of functional MVBs. These may play a key role in modulating surface protein expression in various life cycle stages. Importantly, the close evolutionary position of *Chromera* and *Vitrella* to Apicomplexa suggests that the extensive decrease in ESCRT subunit conservation in Apicomplexa occurred in the immediate ancestor and is not an ancestral feature of a more inclusive group[182] (Figure 2.1, Online Appendix Table 2.1). Particularly, the lack of some ESCRT subunits such as Vps37 in ciliates and dinoflagellates is most parsimoniously attributed to multiple independent losses. Further evidence for a complete set of ESCRT machinery in the last common alveolate ancestor comes from the conservation of all subunits to the exclusion of CHMP7 in the outgroup stramenopile taxa and in *C. reinhardtii*. The absence of CHMP7 in all taxa is not unusual, as it is lost in numerous taxa across eukaryotes[182].

Conservation of adaptin subunits is striking, particularly the complete retention of AP-5 in chromerids. In an initial study of seven organisms from the SAR supergroup (the group in which chromerids belong to), only two (*B. natans* and *T. gondii*) were found to encode the complex; conservation across eukaryotes was similarly sparse[48]. The presence of a complete AP-5 complex in chromerids and coccidians may be indicative of a conserved function in both

lineages. Likewise, the retention of an almost complete AP-5 in cryptosporidia and plasmodia may have functional significance or may simply represent a reductive evolutionary process that has not yet reached completion. The complete lack of AP-5 in *P. falciparum* and *P. reichenowi* supports the latter view. As with the ESCRT complexes, the presence of AP-1-5 in chromerids suggests the loss of AP-3 and AP-5 observed in some Apicomplexa is secondary, as well as the loss of AP-5 in *Perkinsus marinus*, and in both ciliate lineages.

Presence of a complete VpsC core along with an additional CORVET subunit Vps3 in the majority of apicomplexan genomes suggests the potential for a modified HOPS/CORVET complex that interacts with Rab5 to direct tethering at the micronemes/rhoptries. This is in keeping with the view of rhoptries/micronemes as divergent endolysosomal organelles[236]. However, chromerids do not appear to possess rhoptries, although chromerids possess cellular components analogous to micronemes[208,471]. More HOPS/CORVET subunits were found to be conserved in *T. gondii*, which are the only apicomplexan to date to be described as possessing a canonical lysosome-like compartment[5], suggesting that complete complexes are retained in these lineages because they are required for trafficking to canonical lysosome-related organelles as well. Additionally, *Chromera* possesses the chromerosome, which often displays intralumenal vesicles similar to MVBs, suggesting it may also be derived from endosome-like organelles[471].

In conclusion, apicomplexan possess unusual endomembrane compartments including atypical Golgi and endosome-derived invasion organelles such as micronemes and rhoptries[236]. Modifications in the complement of membrane trafficking machinery, including the loss of key protein complexes found in most eukaryotes, have been observed in the apicomplexan lineage, potentially associated with the specialization of the endomembrane system. The absence of some components (Exocyst, Vps39, Trs120, Tip20) within *Chromera* and *Vitrella* suggests pre-adaptation to parasitism deeper in the apicomplexan lineage. By contrast, the presence of near complete complements of key machinery (AP-1-5, ESCRTs, COG) absent in many apicomplexans, pinpoints the timing of the losses at the colpodellid/apicomplexan transition.

2.3 Additional discussion

The above study clearly shows the value of chromerid algae for timing gene loss within the Apicomplexa and their close outgroups. As mentioned in section 2.1, both dinoflagellates and ciliates are problematic to use as outgroups in comparative analysis due to their unusual cell biology and/or to the low quality of the assemblies currently available. This is evident in Figure 2.1, where subunits present in both stramenopiles and chromerids are absent from

ciliates and/or dinoflagellates. For example, all four ciliate/dinoflagellate taxa were described as missing the ESCRT subunit Vps37, despite its complete retention in both chromerids and all included stramenopiles (Figure 2.1). Although this loss could be genuine, the main issue is that, without the chromerid outgroup, the comparable loss of Vps37 in all Apicomplexa would be placed as a single event at the base of alveolates rather than three separate events, on the branches leading to ciliates, dinoflagellates, and apicomplexans.

Further enforcement of the utility of chromerids for comparative studies is the high level of conservation of machinery in their genomes; not including the exocyst complex, which appears to have been lost at the base of Myzozoa, 72 of 76 genes studied are present in at least one chromerid, allowing the subsequent loss of a substantial amount of machinery in Apicomplexa to be interpreted as having occurred in the "proto-apicomplexan" (the most recent common apicomplexan ancestor) or deeper in one or more apicomplexan lineages. The presence of the chromerids therefore clarifies that, as is generally suggested for parasitic eukaryotes, machinery was lost following the split with close free-living relatives. Some of this lost machinery deserves further discussion, especially in light of results in the literature since the publication of the chromerid genomes.

The loss of much of the ESCRT machinery in Apicomplexa, especially outside of the piroplasmids, is notable for several reasons. Although loss of some subunits, and the complete absence of the canonical ESCRT-0 complex outside of opisthokonts, is common across eukaryotes, no other organism included for study by Leung et al. (2008) had lost both ESCRT-I and ESCRT-II completely[182]. ESCRT-0 binds ubiquitylated cargo to initiate intralumenal vesicle (ILV) formation on MVBs[472], although this function appears to be compensated for by Tom1 in organisms lacking the canonical ESCRT-0[473]. Both ESCRT-I and ESCRT-II additionally bind ubiquitin[474,475], and form a bridge of interactions that eventually link cargo-dependent and ESCRT-0-mediated nucleation to recruitment of ESCRT-III[476–480]. ESCRT-I and ESCRT-II are thought to be responsible for the initial budding, with ESCRT-III involved in scission at the bud neck[481–484]. Hence, it would appear that apicomplexans, with the possible exception of piroplasmids, do not possess a sufficient complement of ESCRTs to mediate the formation of ILVs or other structures requiring membrane budding away from the cytosol. Yet, overexpression of PfVps4 in *T. gondii* results in the formation of large MVB-like structures[485] and ILVs are occasionally observed in the *T. gondii* VAC[248]. It is possible that the minimal machinery that remains is capable of ILV formation through noncanonical mechanisms, or through the use of analogous components undetected by homology searching. ESCRTs play multiple other roles in the cell, such as in autophagosome formation[486] and during cytokinetic abscission[487]. It is therefore possible that the remaining ESCRT subunits play a role in one or

more specific cellular processes. Full understanding of the significance of widespread loss of ESCRT subunits in Apicomplexa will require in depth characterization of the remaining subunits.

Similarly, the pattern of adaptor protein loss in Apicomplexa is also intriguing. Multiple independent losses of the AP-3 complex had been described previously in Apicomplexa[468], and can now be extended to the AP-5 complex as well, which was lost independently in piroplasmids and in a subset of malaria parasites (represented here by *P. falciparum* and *P. reichenowi*, Figure 2.1). AP-5 is known to be sparsely conserved across eukaryotes[48], and was similarly absent from almost all outgroup taxa apart from chromerids and *S. minutum* (Figure 2.1). As the most recently described AP complex, little is known about the exact function of AP-5, although recent studies in HeLa cells suggest that AP-5 localizes to late endosomal structures and is responsible for early to late endosome, as well as late endosome to Golgi, trafficking[52,469]. These results suggest a possible role for this complex in endosomal trafficking in Apicomplexa as well. Although all four subunits are expressed in *T. gondii* (https://toxodb.org/toxo/), none are predicted to be essential based on a genome-wide CRISPR-based screen (the lowest phenotypic score is for AP-5μ, -2.04, i.e. higher than the cutoff of ~-2.5 for essential genes)[354]. Furthermore, previous attempts to tag and disrupt these proteins have thus far proved unsuccessful (C.M. Klinger, unpublished, and M. Meissner, personal communication). The role of AP-5 in Apicomplexa remains unknown, as does why it has been lost in some apicomplexans. Similarly, although AP-3 has been lost multiple times its function remains obscure in the lineages in which it is still present; the only functional report in *T. gondii* is in the context of resistance to parasitistatic compounds, and suggest a possible role for TgAP-3β in endosomal trafficking[400].

Patterns of conservation and loss among tethering complexes in Apicomplexa was previously discussed by Klinger et al. (2013)[127]. The addition of chromerid genomes clarified some previously noted patterns, such as the independent loss of multiple COG subunits in both apicomplexans and ciliates, as well as the apparent loss of the exocyst complex at the base of Myzozoa (Figure 2.1). The almost complete lack of the Dsl1 complex in Apicomplexa is also intriguing; Dsl1 is a MTC functioning to tether vesicles at the ER[130]. In *T. gondii*, for example, the ER is known to possess an unusual organization, being essentially absent from the apical portion of the nucleus at which vesicles bud directly from the outer membrane of the nuclear envelope to reach the Golgi, while showing extended and branched morphology medial/basal to the nucleus[240]. It is possible that this unique organization arose concurrently with the loss of Dsl1, although the ER morphology has not been intensely studied in chromerids to date, so this remains speculative. Despite this, it is still curious to consider

what machinery functions to tether vesicles at the ER. Although retrograde trafficking to the ER has not been extensively studied in Apicomplexa, the presence of machinery such as an ER-retention deficient (ERD) homologue[401] suggests that, at least, Golgi-to-ER trafficking occurs. In addition, as discussed in Klinger et al. (2013)[127], loss of Dsl1 subunits appears to correlate with the absence of peroxisomes across eukaryotes; peroxisomes appear to be completely absent in Apicomplexa, along with homologues of the PEX machinery mediating their biogenesis[488,489].

Finally, as discussed in the results, the loss of the exocyst complex in all myzozoans studied to date is a potentially important synapomorphic feature. Although the exocyst has numerous functions across eukaryotes including in regulating cytokinesis[490] and autophagy[491], one of its roles, and the one for which it was discovered, is its ability to regulate fusion of secretory organelles during exocytosis[143,144,492]. Given that Apicomplexa rely heavily on regulated secretion of micronemes and rhoptries to progress through a lytic cycle and establish infection in a host[493], the absence of exocyst is striking. One notable feature of micronemes, which was noted in Section 1.5.1.4, is the relationship between sub-pellicular microtubules and micronemes. In *T. gondii*, this was recently demonstrated through the use of genetic mutants which destabilized the sub-pellicular microtubules and caused dispersal of the otherwise organized fluorescence signal for the resident microneme protein TgMIC2[229]. EM studies in *Plasmodium* almost two decades ago suggested that micronemes formed from the Golgi during intracellular development and traveled along a microtubular track to reach the apical end of the cell[252]. Both of these results suggest an intimate relationship between micronemes and microtubules, which suggest that micronemes may be tethered prior to secretion via this interaction.

Additionally, other recent studies have elucidated machinery required for secretion of micronemes and/or rhoptries. In the case of micronemes, a guanylate cyclase activates PKG, which, through a series of downstream steps, eventually causes release of calcium from intracellular stores to regulate microneme exocytosis through calcium-dependent protein kinases (CPDKs; in *T. gondii* TgCDPK1)[368,494,495]. One important calcium-dependent mediator of exocytosis is the tandem C2 domain protein DOC2.1, which may promote membrane fusion[496]. The signaling cascade that results in calcium release also triggers plasma membrane remodeling, specifically the creation of phosphatidic acid, which interacts with the microneme surface protein APH to mediate microneme exocytosis[497]. Although less is known regarding the regulation of rhoptry secretion, a multiple C2 domain-containing protein, FER2, was recently described to be essential for rhoptry secretion[498]. Hence, it appears that, although the exocyst complex is absent in Apicomplexa, the crucial function of tethering regulated

secretory organelles has been provided by additional, non-homologous, machinery. The presence of similar regulated secretory organelles in other myzozoan taxa[232] suggests that the appearance of apical secretory organelles may have arisen concurrently with the loss of the exocyst complex in this lineage.

The small subset of MTS machinery presented in this chapter demonstrates the utility of chromerid genomes to understand the patterns of reduction frequently observed in apicomplexan genomes. Importantly, although not discussed here for trafficking machinery, the gain of genes and even whole gene families was described in Woo et al. (2015)[395]; see for example the simplified schematic in Figure 2A and Figure 3 within that publication. Hence, including chromerids in comparative analyses can help to understand three basic evolutionary modes: 1) loss of machinery in Apicomplexa following the chomerid/apicomplexan split, 2) loss of machinery prior to the chromerid/apicomplexan split, and 3) gain of machinery in Apicomplexa and their outgroups. The ability to distinguish between 1) and 2) is only possible with the addition of genomic data for chromerids, as prior to the advent of these genomes, the most recent apicomplexan ancestor that could be constructed was the myzozoan ancestor. As will be explored in Chapter 4, chromerids are also useful to distinguish timing of gains as well, as, for example, there is at least one trafficking paralogue that is present in Apicomplexa and not in chromerids. Even for machinery that was gained in the common ancestor of alveolates or myzozoans, the presence of two additional sampling points closer to the base of Myzozoa improves the confidence of any inference regarding taxonomic spread, and hence timing of, novel MTS components.

The study described in this chapter was relatively small, focussing on only five sets of machinery: clathrin, retromer, ESCRTs, MTCs, and APs. However, the pattern observed – general retention of machinery in free-living chromerid relatives followed by extensive lineage-specific loss within extant apicomplexan lineages – represents a common paradigm for other MTS families as well (see Chapter 4). Hence, this chapter established that Apicomplexa have lost MTS machinery during the transition from a free-living to parasitic lifestyle. As discussed in the introduction, this is important for two key reasons: 1) gene loss is inconsistent with proposed mechanisms of organellogenesis, and 2) gene loss suggests that any additional organelles would evolve concurrently with additional paralogues, rather than through co-opting existing machinery. However, both of these notions hinge on the idea that existing machinery would be constrained to perform an expected function, and hence would be unable to participate in additional trafficking pathways. This notion, referred to as "functional homology", is key to determining whether the above assumptions are valid in the

case of Apicomplexa (and are vital to understanding similar patterns across eukaryotes), and will be the subject of the next chapter.

3. Exploring the relationship between gene homology and function

3.1 Preface

The aim of this chapter is to explore the relationship between gene homology and function in order to gain insight into the reliability of functional predictions based solely on the encoded gene complement of an organism. Generally speaking, one aim of comparative genomic analyses is to map information derived in one or more experimental model systems to organisms that are either comparatively understudied at the molecular level, or not studied at all. These understudied models may exist due to their novelty, i.e. that there has not been sufficient time between their description and the time of investigation such that a model system could have been developed, due to their refractory nature towards culturing, genetic manipulation, or some other factor, or due to a lack of resources available for their study. Other reasons why an organism may be understudied exist, and it is likely that a large amount of eukaryotic diversity will remain known only at a shallow level, as the number of described species far exceeds the capacity of relevant investigators to study them.

Compared to traditional molecular methods, such as genetic manipulation, microscopy, and phenotypic analysis, which are time and resource-intensive, the speed of informatic sequence analysis means that entire systems can be studied *in silico* for any organism for which the necessary sequence data exists (or can be obtained). Identification of homologous genes is frequently performed using sequence-to-sequence analysis, such as BLAST[464], profile-to-sequence analysis, using hidden Markov models (HMMs) and programs such as HMMer[499], and profile-to-profile analysis, such as employed by the HHSuite of software tools[500]. Broadly speaking, profile-based methods incorporate information from multiple sequences to generate distributions of expected bases at each homologous position within an alignment[501], which provides greater sensitivity. However, all of these methods provide the same basic information: a list of putative homologues for any given gene of interest in the dataset(s) under investigation.

By searching for all components of a given system, as identified and defined through molecular investigations in a model system(s), it is possible to "reconstruct" that system for a given organism based only on genomic/transcriptomic information. Although absence is not informative, as genome sequences are often at least partially incomplete and transcriptomic analysis is not anticipated to capture the full coding depth of a genome unless all possible conditions for gene expression can be met[502], the presence of a given gene suggests that it may be expressed and functional. If enough genes for a given complex and/or system are present, it is frequently taken to mean that the organism possesses the function associated with that complex/system. Hence, organism function can be inferred from genomic data.

This has caveats, the most obvious of which is the presumed correlation between the presence of a gene in genomic/transcriptomic data and the presence of a functional protein in a living cell. Some genes are retained as non-transcribed and/or non-functional pseudogenes and, even if the gene is functional, gene transcription and/or translation may be specific to a given cell type and/or lifecycle stage. Finally, the gene product may not actually perform the same function as the homologous gene in the model organism for which a function was described. This notion, that the relationship between genes (homology) and their function are correlated strongly enough that functional inferences can safely be made, is termed "functional-homology", and is discussed at length in section 3.2. This homology-function relationship is further complicated by the presence, in some cases, of multiple paralogues (either in the organism for which function is being ascribed, or in the model organism(s) from which functional data is derived).

Homologues are genes with a common evolutionary origin, whether by means of duplication, vertical descent, or horizontal transfer; any genes that can be traced to a common ancestor at some point during their evolutionary history are homologous. Orthologues and paralogues are subsets of homologues, such that, for example, all orthologues are homologues but not all homologues are orthologues. Paralogues arise by gene duplication, generating genes that are initially identical or highly similar, but which may diverge over time to acquire new functions. Orthologues arise by vertical descent and can be thought of as "the same gene" in diverse organisms[503].

Paralogues pose a challenge to functional homology. For example, when an orthologue in one lineage duplicates, one or both of those genes may diverge and acquire new function. If the original function of the gene is required, it is likely that at least one such paralogue (regardless of the number of duplications that occur) will be constrained to maintain the ancestral function. However, all other paralogues will be comparatively unconstrained. As described in section 1.7, this mechanistic process of gene expansion and diversification is considered to have been key to the organization of the MTS in the LECA[453]. In this thesis, it is further argued that the process has continued since the LECA and may be important for the generation of novel autogenous organelles in extant eukaryotes as well.

Considering this, the concept of functional homology becomes key for two main reasons. 1) Functional homology suggests that homologues in diverse organisms will generally perform the "same" function. If this is indeed true, then, as long as that function is required by a given organism, that gene product will be unlikely to participate in additional/novel functions. This is not a strict rule; for example, in *T. gondii*, the Vps10 homologue sortilin (TgSORTLR) mediates anterograde trafficking of both microneme and

rhoptry proteins, which is thought to be possible due to a slight shift in expression of these two sets of proteins during division[310,398]. Although this is a superficial example, mechanisms such as changes in expression could theoretically allow a gene product to perform functions above and beyond what is predicted by functional homology. 2) If gene products indeed maintain function across organisms, this would suggest that the most straight-forward mechanism to obtain cellular machinery capable of performing novel functions is through duplication and diversification. Again, pre-existing machinery could be recruited, especially if the original function is no longer required, but for systems such as the MTS that perform essential functions for all eukaryotic cells, this scenario is comparatively less likely.

In order to understand the patterns of gene loss (as described in Chapter 2, and as will be described in Chapter 4) and gain in Apicomplexa (as will be described in Chapter 4), it is therefore crucial to test the validity of functional homology in extant eukaryotes. If functional homology does indeed exist in the MTS across eukaryotes, it is safer to assume that pan-eukaryotic homologues conserved in apicomplexan genomes are already performing a known function. In addition, it would suggest that additional paralogues, which arose more recently in apicomplexan evolution, might possess novel functions. To answer this question, I previously led an effort to perform a large literature analysis on model systems across eukaryotes to compare MTS proteins which had associated cell biological data for a majority of these model systems, which has been published as Klinger et al. (2016)[504]. This analysis is presented in the next section.

3.2 Resolving the homology-function relationship through comparative genomics of membrane-trafficking machinery and parasite cell biology

3.2.1 Introduction

Genomics, the sequencing and analysis of genomes, has empowered tremendous advances. Possessing a genome sequence for an organism, particularly one difficult to culture or genetically manipulate, allows the prediction of cellular organization, metabolism, gene expression mechanisms, and organellar complement, through *in silico* analysis of the corresponding predicted proteome.

This is essentially a comparative analysis, which at its heart relies on robust evidence of function in one or more organisms. Comparative genomics allows reconstruction of pan-eukaryotic complements of cellular components, including the cytoskeleton, nuclear transport, metabolism, and mitochondrion ([505], *inter alia*), providing evidence for the general or core aspects of cellular systems and which aspects are lineage-specific. This evidence is an

important basis for understanding evolutionary mechanisms behind emergence of cellular complexity. Furthermore, the acceleration in understanding gained by the annotation of thousands of genes is invaluable, by producing initial hypotheses for expected interactions, pathways, and organellar roles that can be tested.

Inherent in comparative genomic studies is the assumption of functional homology, i.e. that orthologous genes retain equivalent function. Orthology is the relationship between two genes in distinct taxa that are directly related by vertical descent[506], and which may be considered as the "same gene"; the expectation is that such gene pairs retain equivalent properties and roles within the cell[503]. This assumption has been generally regarded as safe, based on a model of conservation of function rather than the widespread gain of novel functions or neofunctionalization and based on experimental validation of enzymes assayed heterologously or *in vitro*, where 'function' can be relatively readily defined. However, much of our understanding of eukaryotic cell biology is based on evidence from a small sample of true eukaryotic diversity and frequently from a restricted region of the eukaryotic tree. Given this sampling bias, to what extent can 'function' be reliably predicted across eukaryotic diversity based on sequence similarity alone?

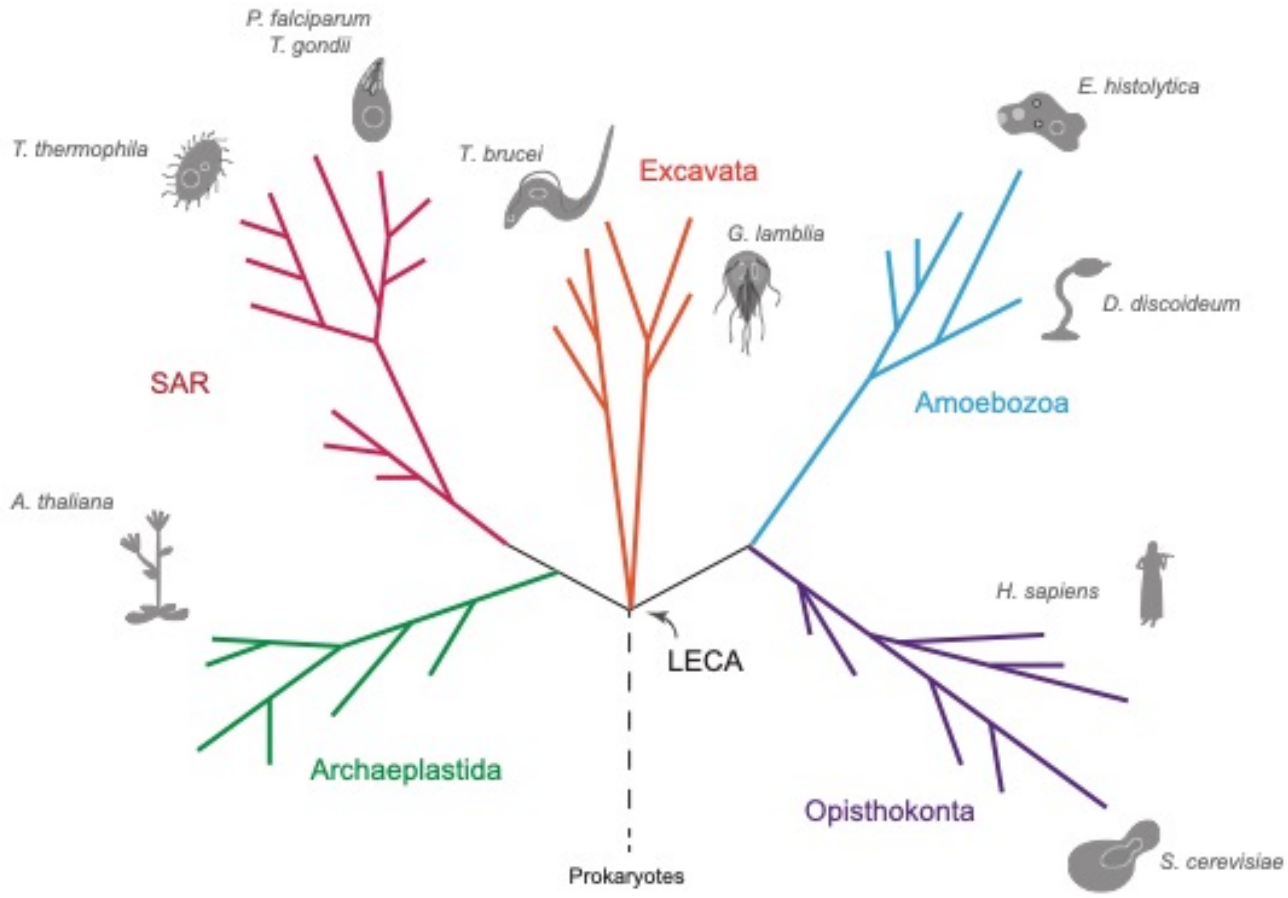
Testing the assumption of functional homology requires experimental evidence from organisms across a full taxonomic range of eukaryotes, and there are now fortunately tractable organisms from each of the major eukaryotic divisions or Supergroups (Figure 3.1). Here we have chosen a subset of non-metazoan organisms and assessed comparative data available for genes of the membrane trafficking system, a crucial cellular system underpinning pathogenic mechanisms in many parasitic protists, and which has been well studied. We not only assess the validity of the core assumption of functional homology in comparative studies of membrane trafficking genes, but also begin to identify the manner in which the endomembrane system is modified in individual parasitic lineages and which speaks directly to mechanisms of disease and the origins of parasitism.

3.2.1.1 The membrane-trafficking system: a modern molecular view

Membrane trafficking is the process by which proteins and other macromolecules are distributed throughout organelles of the endomembrane system, and released into, or internalized from, the extracellular environment. Trafficking is vital for metabolism, signaling, and interacting with the external environment. Transport vesicles act to transfer cargo molecules between the organelles of the endomembrane system, which possess discrete morphology, localization, and functions[21].

Figure 3.1 Model Organisms Across Eukaryotes

This figure demonstrates the distribution of model organisms across eukaryotic diversity. Colour-coded branches and corresponding labels denote eukaryotic Supergroups, with the branching order roughly corresponding to the organization of known diversity within each group. Model organisms are represented by greyscale illustrations and corresponding labels in italics. The position of the Last Eukaryotic Common Ancestor (LECA) is indicated. Though additional model organisms exist for each of these groups, they are excluded from this figure for simplicity.


P. falciparum
T. gondii
T. thermophila
T. brucei
Excavata
E. histolytica
G. lamblia
D. discoideum
SAR
Amoebozoa
A. thaliana
H. sapiens
LECA
Archaeplastida
Opisthokonta
S. cerevisiae
Prokaryotes

Anterograde trafficking involves movement from the endoplasmic reticulum (ER) through the Golgi complex, the *trans*-Golgi network (TGN), and on to the plasma membrane[507], whilst endocytosis begins at the plasma membrane where cargo is sorted by endosomes before recycling or targeting to acidic terminal organelles. During endocytosis organelles acidify, may acquire intralumenal vesicles (present in multi-vesicular bodies or MVBs), and modify their compositions[443]. In all trafficking pathways retrograde transport steps recycle selected components back to previous organelles for use in future rounds of trafficking.

Specialized protein complexes controlling vesicle budding, tethering, and fusion, many of which are large paralagous families, regulate transport. Arf/Sar family small GTPases and their regulators, cargo adaptors, and coat protein complexes are involved in vesicle formation/fission. Rab GTPases are involved in vesicle targeting, whilst coiled coil SNARE proteins are central to vesicle fusion[21]. Importantly, members of these multiple families act at discrete locations or trafficking pathways; the specificity of trafficking is in part encoded in the combinatorial interactions of these various players[508]. For example, COPII-coated vesicles mediate anterograde transport from the ER to the Golgi, while the corresponding retrograde transport step requires COPI vesicle formation[509]; clathrin-coated vesicles mediate multiple post-Golgi transport routes[510].

Our view of membrane trafficking is dominated by studies in animal and yeast cells. However, membrane trafficking is a central process underpinning growth, cell surface presentation and secretion and thus critical to pathogenic mechanisms of many parasitic protists, for example by mediating host cell invasion[312] and immune system evasion[511]. It is therefore reasonable to ask what complement of membrane trafficking proteins is present across the broad diversity of eukaryotes and what we can infer about both evolution of the membrane trafficking system and the conserved set of eukaryotic membrane trafficking machinery, and how this has been modified in parasitic protists.

3.2.1.2 Evolution of membrane-trafficking: LECA complement and modern innovations

Comparative studies have allowed reconstruction of the gene complement of the last eukaryotic common ancestor, or LECA. The rationale is simple and powerful: if orthologues of a gene are identified in organisms covering the breadth of eukaryotic diversity, then parsimony dictates that gene was present in the LECA[505].

Three general patterns are observed. Some families, such as clathrin, retromer, COPI, and COPII are widely conserved and inferred present in the LECA; though few deviations from

the ancestral complement of core machinery exist in extant organisms, some variability is seen in retention of accessory components[33,88,96,512]. Other families are more variable, for example the heterotetrameric adaptor protein complexes. The adaptor protein (AP) complexes 1 and 2 are well conserved, but AP-3 through AP-5 and TSET, a recently described member, while found in widely diverse taxa are frequently absent[48,49]. This is interpreted as ancestral presence in LECA and frequent subsequent loss of the latter complexes in extant eukaryotes. The third pattern, lineage-specific expansion, is exemplified by the Rab family, which reveals a patchy distribution in extant eukaryotes, but critically with new clades and paralogous expansion of conserved subfamilies arising in some lineages[110,111,513].

Hence, extant eukaryotes have gained and lost membrane trafficking machinery since diverging from LECA. Paralogous expansion and other lineage-specific features certainly provide machinery theoretically required for novel function and endomembrane specialization, but loss of machinery may also be involved in this process, and a full understanding necessitates comparison across eukaryotic diversity.

3.2.2 Emerging model organisms

Phylogenetics has resolved eukaryotic diversity into five Supergroups, creating the necessary framework for comparative analyses (Figure 3.1). Despite increased knowledge of the taxonomic affiliation and cell biology of diverse eukaryotes, cell biological models remain biased towards the Supergroup Opisthokonta, namely humans and yeast (Figure 3.1, purple). Nonetheless, model organisms have been established across eukaryotes, including parasites, and many possess endomembrane features (proteins and organelles) not present in canonical models.

The multicellular plant *Arabidopsis thaliana* (Figure 3.1, green – Supergroup Archaeplastida) encodes a large genome with multiple paralogues for many membrane trafficking genes. *A. thaliana* has an endomembrane system largely similar to model opisthokonts. However, a key difference is the lack of a discrete early endosomal compartment, as internalized material is distributed to the TGN before being recycled or transiting the endosomal system for degradation in the vacuole[440,441,514].

The ciliate *Tetrahymena thermophila* (Figure 3.1, red – Supergroup SAR) is a ciliated heterotroph that engulfs prey in phagosomes that subsequently mature and undergo fission/fusion with other intracellular compartments before releasing their remaining contents. A prominent contractile vacuole is present for osmoregulation and dense core secretory granules underlie the plasma membrane. Canonical endomembrane compartments are

present, though their intracellular location and arrangement differ from yeast and mammalian cells[515].

Also within the SAR (stramenopiles, alveolates, and rhizarians) Supergroup are the apicomplexan parasites *Toxoplasma gondii* and *Plasmodium falciparum*, causative agents of toxoplasmosis and malaria, respectively (Figure 3.1, red). These organisms possess a polarized endomembrane system including apical or "invasion" organelles, micronemes and rhoptries, to mediate host cell invasion and egress[467]. Apical organelles are likely divergent endo-lysosomes and other endo-lysosomal compartments, including an endosome-like compartment and vacuole, are also present, though the organization and identity of the endosomal system remains poorly understood[236,312,431].

Giardia lamblia, causative agent of giardiasis, is a member of the Supergroup Excavata possessing a reduced endomembrane system (Figure 3.1, brown). *Giardia* cells are bilaterally symmetric, possessing two diploid nuclei and four pairs of flagella. Aside from Golgi-like encystation-specific vesicles in encysting cells, the organism maintains only an ER and peripheral vacuoles, which perform functions associated with endo-lysosomes in model systems[516].

Another intensely studied group of excavates are the trypanosomatids (Figure 3.1, brown). Trypanosomes cause disease in humans, wild and domestic animals, insects, plants, and fish, as well as having free-living relatives, and hence have provided a wealth of data on genome evolution, cell biology, and mechanisms of interaction with, and adaptation to, their hosts[517]. *Trypanosoma brucei* is the organism of choice for dissection of tryapnosomatid cell biology, owing to the application of RNA interference and other technologies. Trypanosomes possess an endomembrane system similar to that in mammalian model systems, but differ in some aspects, such as restricting all endocytic uptake to a cellular region known as the flagellar pocket[518].

Entamoeba histolytica is a member of the Supergroup Amoebozoa (Figure 3.1, blue) with an unusual tubulovesicular endomembrane organization[519]. Consistent with its name, *histolytica*, this organism combines secreted virulence factors with cell killing via a specialized phagocytic process (trogocytosis) to induce host tissue damage and necrosis in the intestinal tract and liver[520]. Additionally, *E. histolytica* is capable of efficient whole-cell phagocytosis, but the exact mechanism is slightly different than in mammalian cells, involving fusion of nascent phagosomes with a pre-existing pre-phagosomal vacuole[521].

Dictyostelium discoideum (Figure 3.1, blue – Supergroup Amoebozoa) has a complex life cycle, encompassing unicellular amoebae that aggregate under starvation conditions to form transiently multicellular entities, first a bulbous slug, which then forms an elongated

stalk structure known as a fruiting body from which to release spores[522]. The endomembrane system of *D. discoideum* is reminiscent of model organisms but also features non-acidic post-lysosomes and a prominent contractile vacuole[523]. Owing to ease of genetic manipulation, *D. discoideum* has contributed understanding to cellular processes including cell-cell adhesion, chemotactic signaling, cytoskeleton-dependent locomotion, cytokinesis, and, as a professional phagocyte, the formation and maturation of phagosomes as well[524].

3.2.3 Examining the case for functional homology

Prior to assessing functional homology it is worth defining our criteria, which we have divided into three categories of evidence.

(i) Localization. Functional homology implies the gene product in question localizes to organelles or structures that are homologous in the respective cells.

(ii) Interactions. Functional homology implies that gene products should interact with homologous proteins, or in the case of other molecules, those of the same or similar molecular composition such as binding specific phosphoinositides or ions.

(iii) Genetic disruption. Functional homology implies that disruption should result in a similar phenotype between taxa. However, differences in cell physiology can make phenotypes difficult to directly compare and hence require careful interpretation.

3.2.4 Functional homology in trafficking machinery between divergent organisms

We have focused on proteins where broadly equivalent evidence from multiple organisms permits comparison of function in a relevant manner, including the adaptor proteins, ESCRT and retromer complexes, and finally select Rab GTPases.

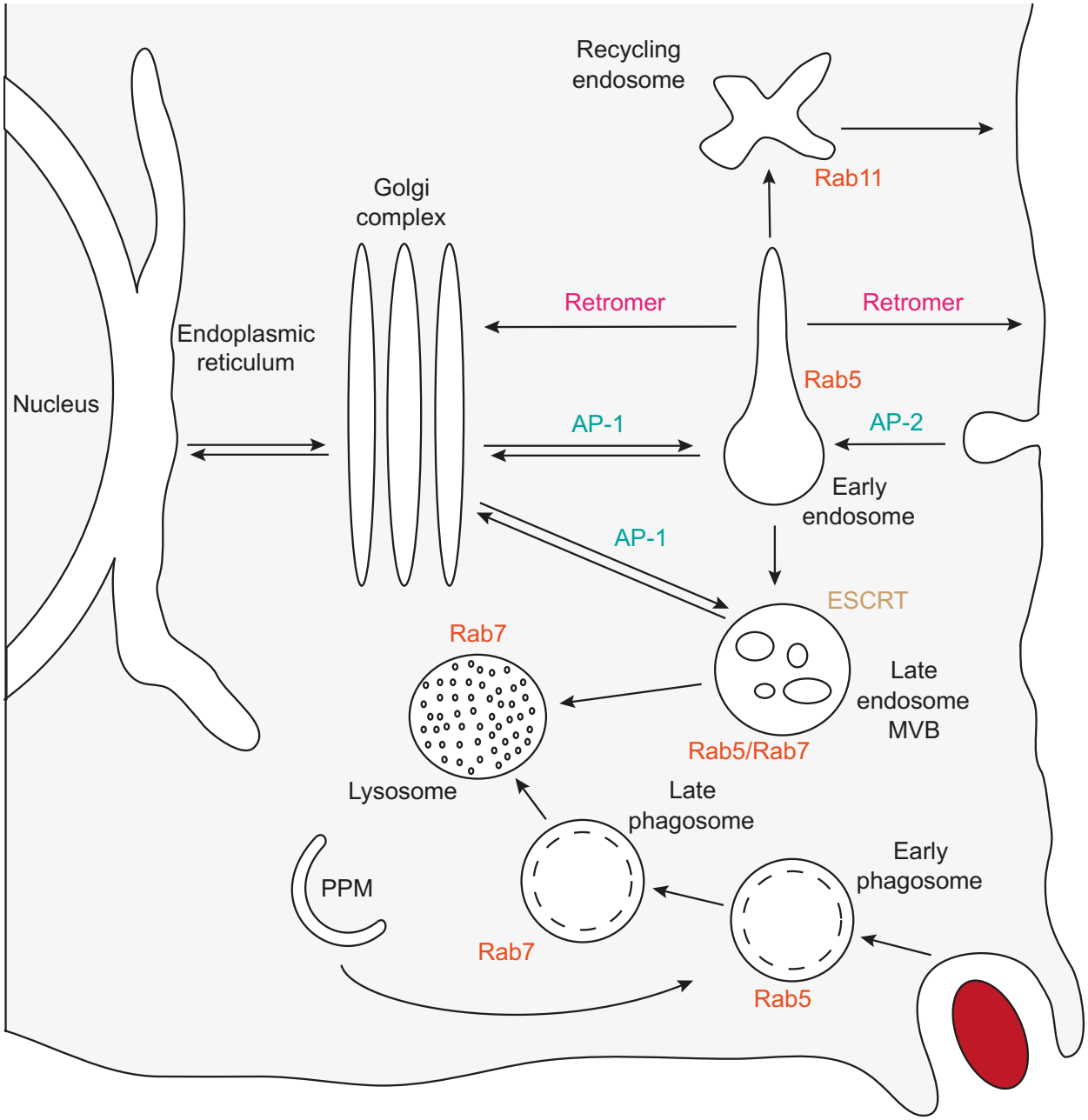
3.2.4.1 Adaptor proteins

The adaptor protein complexes bind cargo proteins for inclusions into vesicular carriers that are then formed by the action of membrane-deforming coat proteins such as clathrin. There are five heterotetrameric adaptor complexes (AP-1 through AP-5) composed of two large (γ, α, δ, ε, ζ and β1-5), one medium (μ1-5), and one small subunit (σ1-5). They are related to other such complexes, including the coat-like TSET complex and the COPI coat[49]. We focus on AP-1 and AP-2, as the role of these complexes in mediating specific intracellular trafficking events together with clathrin is well established in model systems[510,525], and they are similarly the best-studied adaptor proteins in other organisms.

3.2.4.1.1 AP-1

Figure 3.2 Function of select membrane-trafficking machinery in a model endomembrane system

This figure depicts roles for membrane-trafficking system machinery under discussion in a generalized eukaryotic cell, based on studies primarily in yeast and mammalian systems. Components are colour-coded, with adaptor proteins (AP, teal), ESCRT (brown), retromer (magenta), and Rab GTPases (orange). Organelles are depicted based on common morphology and labeled in plain text. Arrows, including the directionality of each step, indicate trafficking between organelles. The presence of a dotted line in the interior of phagosomes represents the presence of either a single bounding membrane (phagosomes), or two bounding membranes (autophagosomes). The red oval represents a particle to be phagocytosed. Additional machinery is required for each trafficking event shown, but for simplicity is not included in this diagram. Note that not all organisms perform the illustrated trafficking events, and other events occur that are not depicted in this diagram.


Recycling
endosome
Rab11
Golgi
complex
Retromer
Retromer
Endoplasmic
reticulum
Nucleus
Rab5
AP-1
AP-2
Early
endosome
AP-1
ESCRT
Rab7
Late
endosome
MVB
Rab5/Rab7
Lysosome
Late
phagosome
Early
phagosome
PPM
Rab7
Rab5

In opisthokonts, the AP-1 complex is primarily localized to the TGN and early endosomes. It mediates transport between these organelles in both directions, but also mediates some trafficking between these organelles and the PM[526]. AP-1 interacts with clathrin and various monomeric adaptors, as well as trans-membrane receptors important for sorting biosynthetic endo-lysosomal cargo[527].

In *A. thaliana* AP-1 is primarily associated with the TGN/early endosome, as evidenced by co-localization with various markers for this organelle and correspondingly poor co-localization with markers of the Golgi or MVBs[528–530]. AP-1 subunits interact with clathrin heavy chain[528], the adaptor EPSIN1[531], and two vacuolar sorting receptors[528,532]. Genetic disruption of AP-1 subunits results in defects in both vacuolar trafficking and TGN/early endosome to plasma membrane recycling[528–530].

Little is known about adaptor protein function in *T. thermophila,* but both AP-1μ subunit paralogues localize to distinct intracellular locations[533]. Early studies in *T. gondii* localized AP-1μ at the Golgi, endosome-like compartment, and rhoptries[396]. This is consistent with a recent study in *P. falciparum* showing the dynamic localization of tagged AP-1μ in puncta adjacent to the Golgi and rhoptries throughout the intracellular life cycle[399]. Expression of a dominant negative mu subunit in *T. gondii* causes mis-localization of the rhoptry protein ROP2 and impairs rhoptry formation, and AP-1μ both co-localizes, as well as interacts with, the vacuolar receptor TgSORTLR[396,398,534].

In *G. lamblia*, AP-1μ localizes to perinuclear regions and the cell periphery, in the latter case co-localizing with peripheral vacuole proteins, and can interact with clathrin[535]. AP-1μ also binds the vacuolar receptor Vps, and its knockdown by dsRNA induces degradation of Vps; this is specific to AP-1, as AP-2μ does not bind Vps[536]. Knockdown of AP-1μ also results in mis-localization of two peripheral vacuole proteins[535].

None of the AP complexes have been successfully localized in trypanosomes, and it is unclear why this may be so. AP-1 is involved in lysosomal delivery of p67, the major lysosomal glycoprotein, in *T. brucei* and there is evidence that this is developmentally regulated[537,538]. More recently AP-1 was implicated in sensitivity of *T. brucei* to suramin, an important frontline drug, and this appears to synergize with endocytosis of surface components, presumably to "condition" the lysosome in some manner to maintain sensitivity to the drug[537].

Though AP-1γ was identified in *E. histolytica* by proteomics to be associated with phagosomes, little else is currently known about its function[539]. In *D. discoideum*, AP-1γ localizes to phagosomes as well as multiple distinct intracellular puncta, some of which co-localize with the Golgi marker comitin[540,541]. Time course isolation of phagosomal membranes shows that AP-1 associates early and is subsequently lost over time[541]. As in model systems,

AP-1 interacts with clathrin[540], but also the contractile vacuole protein Rh50[542]. Consistent with these observations, knock out of AP-1μ results in secretion of unprocessed lysosomal enzymes, defects in phagocytosis and fluid phase uptake, and mis-localization of contractile vacuole markers[540,541].

3.2.4.1.2 AP-2

In animals and fungi, the AP-2 complex has a well-defined role in mediating clathrin-dependent endocytic uptake of specific cargo at the plasma membrane, often through interaction with other cargo adaptors[543].

The *A. thaliana* AP-2 complex is dynamically associated with the plasma membrane, as evidenced by a multitude of studies using tagged AP-2 subunits or specific antibodies[544–549]. Consistent with studies in model systems, various approaches indicate co-localization[544–547], and physical interactions[545–548], of AP-2 subunits with clathrin. In addition, AP-2α can interact with the C-terminal region of the monomeric clathrin adaptor AP180[550]. Genetic disruption of AP-2 subunits, or use of chemical inhibitors of clathrin-mediated endocytosis, results in decreased endocytic uptake of specific plasma membrane cargo[544–547,549]. The severity of the resulting phenotype varies depending on the method of disruption, and this may be due to the role of the TPLATE/TSET complex in endocytosis in this lineage[549,551].

D. discoideum AP-2 localizes to distinct puncta near the cell surface which co-localize with clathrin; both AP-2 and clathrin also partially localize to the contractile vacuole network[552]. Similarly, the single beta subunit involved in both AP-1 and AP-2 complexes in *D. discoideum* localizes to the plasma membrane and also to intracellular structures[553]. Consistent with a role in endocytosis, AP-2 interacts with an Eps15-related protein[552], but also with the SNARE VAMP7[554], which is known to associate with the contractile vacuole[555,556]. Oddly, knockout of AP-2 subunits does not affect the internalization of the contractile vacuole marker dajumin[552], or the localization of p25 or p80 endosomal markers[557]. Comparatively, knockout of the lone AP-1/2β subunit results in pleiomorphic defects, including impaired osmotic stress response[553], likely due to its function in both complexes.

Little is currently known regarding AP-2 function in other systems. *T. thermophila* AP-2μ co-localizes with a dynamin-related protein known to be important for endocytosis at the plasma membrane, as well as to contractile vacuole pores[533]. *E. histolytica* AP-2β was identified on isolated phagosomes by proteomics[539]. In *G. lamblia*, AP-2μ co-localizes with LysoTracker Red, which labels acidic organelles such as lysosomes, and also clathrin heavy chain, at peripheral vacuoles. Knockdown using dsRNA does not affect fluid phase uptake, but does impair receptor-mediated endocytosis[558]. AP-2 is absent in trypanosomatids that express

the variant surface glycoprotein, which may represent an adaptation connected with very rapid endocytosis seen in African trypanosomes and critical for antigenic variation[511,559].

3.2.4.1.3 Functional homology in adaptor proteins

AP-1 mediates trafficking events between the Golgi, endosomes, and the PM, while AP-2 mediates endocytic uptake at the PM. Localization of these components in diverse eukaryotes is consistent with these roles: AP-1 and AP-2 in *G. lamblia* localize to peripheral vacuoles, which are thought to serve the function of endo-lysosomes, and potentially also the Golgi, and in both *T. gondii* and *P. falciparum* AP-1 localizes to the Golgi and endosomes. A role for AP-1 in phagosome function has been reported previously in murine macrophages[560], and this function may also be present in *D. discoideum* and *E. histolytica*. AP-1 and AP-2 in *G. lamblia* mediate trafficking to peripheral vacuoles from the ER and plasma membrane, respectively. Furthermore, interaction between *Toxoplasma* AP-1 and a vacuolar receptor, as well as a direct effect of AP-1 disruption on trafficking of rhoptry proteins, suggests AP-1 retains homologous function in Apicomplexa as well. AP-1 and AP-2 localize as expected in *A. thaliana*, and possess conserved roles in vacuolar trafficking and recycling, and endocytosis, respectively.

3.2.4.2 The ESCRT complexes

The endosomal sorting complexes required for transport (ESCRT) machinery mediate diverse processes from sorting of ubiquitylated cargo into intralumenal vesicles at MVBs to mediating cytokinesis and autophagy[561,562]. Of the five sub-complexes (ESCRTs 0,I,II,II,and IIIa), 0 is known to be opisthokont-specific while the others are found across eukaryotic diversity[182,563].

A. thaliana encodes all canonical ESCRT subunits, including multiple paralogues in many cases[564,565]. Specific antibodies against, or fluorescent fusions of, ESCRT-I[514,564] and ESCRT-II[514] components reveal primarily TGN/early endosome localization. C-terminal YFP fusions of ESCRT-III components partially co-localize with an MVB marker[566] and, although these fusions may not act in a physiological manner[566,567], additional work confirms an MVB localization for the ESCRT-IIIa component SKD1/Vps4[566,568,569]. Hence, ESCRT components appear to be recruited sequentially during endosomal maturation. Functional disruption of ESCRT components results in aberrant vacuolar morphology, failure to degrade transmembrane vacuolar cargo, enlarged MVBs, impaired intralumenal vesicle formation, and impaired autophagy[566,568–572]. Additional plant-specific ESCRT components have been

described[567,573–577], the presence of which suggests that lineage-specific functional innovations are also present.

A lack of detailed characterization makes it unclear how a reduced ESCRT complement functions in Apicomplexa[182,395]. When expressed in either *T. gondii* or *P. falciparum*, the *Plasmodium* Vps4 orthologue is primarily cytosolic. Vps4 mutants predicted to be blocked in ATP binding or hydrolysis instead localize to distinct puncta, which co-localize with markers of the endosome-like compartment. Electron microscopy of these mutants reveal enlarged structures reminiscent of MVBs that are not observed in wild-type parasites[485].

G. lamblia encodes two paralogues of Vps46, one of which, Vps46A, localizes to the cytoplasm and shows intense signal near the plasma membrane, consistent with a possible role at peripheral vacuoles[578,579]. Furthermore, either paralogue is capable of restoring vacuolar sorting of carboxypeptidase S in a yeast Vps46 knockout[579], suggesting at least partial conservation of function between yeast and *Giardia.*

ESCRT components have been localized in trypanosomes, and as expected appear to be present at late endosomal compartments. This is consistent with the importance of ubiquitylation for turnover of surface molecules in *T. brucei*[182,580]. Whilst knockdowns suggest a role in trafficking of surface proteins in *T. brucei*, the impact is not strong, albeit this poor penetrance has also been observed in other eukaryotes. Although the absence of an endocytic blockade has been interpreted in trypanosomes as evidence for a divergent pathway for surface protein turnover[581], the paucity of data and clear soft phenotype obtained by knockdown at present make any firm conclusions unsafe.

In *E. histolytica* Vps4 localizes to small cytoplasmic puncta under normal conditions, but also surrounds ingested red blood cells following phagocytosis. An ATPase assay confirmed Vps4 ATPase activity, and overexpression of an enzymatically dead mutant impairs phagocytosis and the organism's ability to cause hepatic abscesses in hamsters[582]. Three *E. histolytica* proteins contain a Bro1 domain, and thus may be homologues of Bro1/Vps31: ADH112, ADH112-like 1 and ADH112-like 2. Overexpressed ADH112 localizes to the plasma membrane and cytoplasmic vesicles and accumulates on MVBs, and can interact with the ESCRT subunit Vps32. Expression of exogenous Bro1 has a dominant negative effect on red blood cell phagocytosis[583], suggesting a possible role for ESCRT machinery in this process.

Tom1 has been proposed as an analogue of ESCRT 0 outside of opisthokonts, and in *D. discoideum* localizes to intracellular puncta distinct from p25 or p80 positive endosomes, and co-localizes with ubiquitin. It does interact with another ESCRT component Vps23/Tsg101, but also with ubiquitin, an Eps15-related protein, and clathrin[183]. Whereas Bro1/ALIX knockout cells cannot form spores or fruiting bodies[584], suggesting a possible

function in differentiation or cytokinesis, Tom1 knockout cells do not show these defects, and display only mildly impaired fluid-phase uptake[183]. As such, the exact function of the ESCRT complexes in *D. discoideum* is currently unclear.

3.2.4.2.1 Functional homology in ESCRT complexes

Localization of Vps46 at peripheral vacuoles in *Giardia* is consistent with their putative homologous relationship to endo-lysosomes, and endo-lysosomal localization of ESCRT components in trypanosomes and *A. thaliana* has also been shown. The function of both *Giardia* Vps46 paralogues is sufficiently conserved to complement a yeast knockout, and ESCRT machinery in trypanosomes also appears to be functionally conserved. Functional conservation in *A. thaliana* has been convincingly demonstrated, as mutants fail to properly sort cargo and accumulate intralumenal vesicles that remain contiguous with the MVB bounding membrane. Localization of *Entamoeba* subunits Vps4 and ADH112 to both early and late phagosomal structures suggests some difference between *E. histolytica* and model systems, likely due to the unusual endomembrane organization in *E. histolytica*. Although alteration of *Entamoeba* Vps4 activity, or expression of exogenous Bro1, leads to defects in phagocytosis and pathogenicity, the exact function of the *E. histolytica* ESCRT machinery remains unclear. Further investigation into non-endocytic functions of ESCRT across eukaryotes may provide further insight into the patterns of subunit retention, for example in the Apicomplexa where conservation of ESCRT-III components may be due to a need for accurate cytokinesis and not be related to MVB formation.

3.2.4.3 Retromer

The retromer complex consists of a trimeric cargo-selective complex, comprising Vps26, Vps29, and Vps35, which interacts with sorting nexin (SNX) family proteins and other factors including Rab7 to mediate endosome-to-TGN and endosome-to-plasma membrane trafficking pathways[106,585]. One of the best-known functions of retromer, and that for which it was discovered, is recycling of the Vps10 cargo receptor[100].

A. thaliana encodes three copies of Vps35, two of Vps26, and a single copy of Vps29, together with SNX1, SNX2A, and SNX2B sorting nexins. The exact localization of retromer components has been disputed. VPS35, VPS29, and SNX2 co-localize with MVB/vacuole markers[586–592], while one study reported a primarily TGN localization for both VPS35 and SNX2A[593]. Vps26 double mutants lack VPS35[589,594] while vps29 mutants have reduced levels of VPS35[595], suggesting VPS35 stability is dependent on its presence in a complex. All three VPS35 genes can be disrupted, but triple null mutants are not viable; mutants in vps35a show

different phenotypes from those in vps35b, suggesting sub-complexes exist with distinct functions[588,596,597]. Disruption of retromer function results in fragmented vacuoles, accumulation of vacuolar cargo precursors, and secretion of vacuolar cargo into the extracellular space, which in *Arabidopsis* constitutes a default pathway[588–592,594,595,598]. Despite similarity in retromer trafficking compared to model systems, *A. thaliana* appears to possess a number of differences related to mechanisms of retromer subunit recruitment[589,592,594].

In *T. thermophila* only the Vps10 receptor has been investigated. Four Vps10/sortilin-like proteins, Sor1 through Sor4, are present. Sor4 stains cytoplasmic puncta distinct from secretory granules, but interacts with the secretory granule protein Grt1. Knockout of Sor4 causes mis-localization of two resident secretory granule proteins, as well as the aspartyl cathepsin protease CTH3, which is capable of processing secretory granule protein pro-domains[599,600].

The trimeric retromer complex in *T. gondii* co-localizes and interacts with the Vps10-like receptor TgSORTLR[375,398,534], and is involved in recycling between the endosome-like compartment and both the TGN and plasma membrane. In *P. falciparum* Vps29 and Vps35 localize to punctae throughout the intracellular lifecycle that are distinct from markers for the ER, Golgi, plastid, mitochondria, rhoptries, and micronemes[385]. Conversely, PfSORTLR co-localizes with the Golgi marker ERD2, indicating that the receptor is primarily present at the Golgi. Attempts to knockout retromer subunits in *P. falciparum* failed, suggesting the gene product is essential in intracellular parasites[385].

In *G. lamblia* Vps35 localizes to the cell periphery, consistent with peripheral vacuole localization, while Vps26 and Vps29 co-localize with the ER marker BiP; some partial co-localization between subunits is observed in a subset of peripheral vacuoles, and the observed localization patterns are further supported by sub-cellular fractionation. Vps35 co-localizes and interacts with the Vps10-like receptor Vps, and additionally interacts strongly with both Vps26 and Vps29[601,602].

T. brucei encodes single orthologues of Vps26, Vps29, and Vps35, as well as a single SNX protein. Vps5 and Vps26 localize to the region between the nucleus and kinetoplast, consistent with endosomal localization. Additionally, Vps26 co-localizes with early endosomal markers including clathrin, Rab5A, Rab11, and EpsinR, and closely apposes signals for the MVB and lysosome. RNAi-mediated knockdown of these components exhibits mild defects in trafficking of p67 (lysosome) and ISG75 (plasma membrane), as well as Golgi fragmentation, suggesting a similar function of trypanosome retromer to that in mammalian and yeast systems[96].

In *E. histolytica* proteomic studies have identified Vps34, a PI-3-kinase known to regulate retromer function through generation of the phosphoinositide PI3P, on phagosomes[539]. Additionally, Vps26, Vps29, and Vps35 form a complex *in vivo*[603], and, together with Rab7A, retromer is likely involved in the maintenance of the pre-phagosomal vacuole[521,603]. These data point to a primary role for *Entamoeba* retromer in phagocytosis. Despite that *D. discoideum* possesses all retromer subunits[96], no functional data exist yet for retromer in this organism.

3.2.4.3.1 Functional homology in retromer

The localization of retromer across systems corresponds to its function in model organisms. Localization to pre-phagosomal vacuoles and phagosomes is consistent with their endo-lysosomal nature. However, differences in the localization of *G. lamblia* Vps35 and Vps26/Vps29 is at odds with their strong interaction and suggests a dynamic localization. Despite some studies indicating a primarily TGN localization of *A. thaliana* components, the bulk of evidence places retromer primarily at late endosomal compartments. The majority of evidence for retromer function in other non-model organisms is indirect, through characterization of the well-known retromer cargo Vps10/sortilin. Vps10 homologues mediate trafficking to secretory organelles in *T. thermophila* and Apicomplexa, and the *G. lamblia* homologue directly interacts with Vps35. Additionally, there is evidence for Vps10 homologues interacting with AP-1 in both *G. lamblia* and in *T. gondii*. This likely reflects AP-1 and retromer mediating distinct Vps10-dependent trafficking events, potentially anterograde and retrograde Golgi-endosome transport, respectively. In *A. thaliana*, where retromer has been better characterized, it appears to be important for vacuolar trafficking, as mutants secrete vacuolar cargo into the extracellular space via a default constitutive pathway.

3.2.4.4 Rab GTPases

While the above machinery is involved in vesicle formation, vesicle fusion machinery can similarly be assessed, perhaps most tractably the Rab GTPases. Like other GTPases Rabs cycle between GTP- and GDP-bound states. The state of the bound nucleotide has a direct effect on the conformation of the GTPase and regulates the ability of the GTPase to bind specific effector proteins[604]. Additional factors, e.g. guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs), regulate the switch between bound nucleotide state, and can precisely regulate the intracellular location and concentration of GTP- and GDP-bound forms of specific GTPase proteins. Hence, Rabs are often referred to as "master

regulators" or switches of processes, including membrane trafficking[112]. Three Rabs are well studied in many systems and have well-defined functions: Rab5, Rab7, and Rab11.

3.2.4.4.1 Rab5

In opisthokonts, Rab5 is present on early endosomal compartments and mediates the recruitment of effectors involved in the Rab5 to Rab7 switch important in endosome maturation[119,605]. Despite putative orthologues being present in their genomes, we could not find relevant characterization of Rab5 in either *T. thermophila* or *D. discoideum*, and a Rab5 orthologue has yet to be identified in *G. lamblia*[110].

A. thaliana encodes three Rab5 family proteins: RHA1/RABF2a, ARA7/RABF2b, and ARA6/RABF1. All three paralogues label endosomes, with RHA1 and ARA7 co-localizing, while ARA6 shows variable overlap with either RABF2 protein[606–612]. These likely represent endosomal populations, with RABF2 variants acting at MVBs and RABF1 at a variant of recycling endosomes. Constitutively active ARA6 localizes to the plasma membrane[606,608], and ARA6 co-localizes with endocytosed plasma membrane proteins[608], and yet, unlike RHA1 and ARA7, is not associated with vacuolar targeting of soluble cargo[608,613].

T. gondii and *P. falciparum* both encode three Rab5 paralogues, Rab5A, Rab5B, and Rab5C. Tagged versions of each paralogue in *T. gondii* revealed localization consistent with the endosome-like compartment[245,246,248,253], and overexpression of all three paralogues ablate parasite growth. However, only functional disruption of Rab5A or Rab5C result in mis-localization of a subset of microneme and rhoptry proteins[253]. Though Rab5B function is unknown, it is present in a retromer interactome[375].

In contrast, *P. falciparum* Rab5A is localized to haemoglobin-containing structures[374,614]. Expression of a constitutively active Rab5A increases haemoglobin uptake and food vacuole size, consistent with a role in endocytic uptake[614]. Rab5B, localizes to the plasma membrane and food vacuole of intracellular parasites[374]. Though Rab5B localization is consistent with an endocytic role, its function is currently unclear; it is essential in *Plasmodium*, despite the presence of both Rab5A and Rab5C paralogs, suggesting these paralogues do not possess redundant function[374].

All trypanosomatids encode two Rab5 paralogues, which represent a lineage-specific duplication. Both are essential, and critical for endocytosis of surface components in *T. brucei*[615,616]. Significantly, these two paralogues apparently mediate the trafficking of distinct cargo proteins[617], but the basis for the targeting of a molecule to a Rab5A or Rab5B-specific route, or the functional need for such a division, has remained elusive.

In *E. histolytica* Rab5 was identified on phagosomal membranes, albeit only at different time points and dependent on the material taken up[618,619], suggesting a similar association of Rab5 with phagosomes as seen in model systems, but also a potential for complex and dynamic regulation. Additionally, Rab5 associates with Rab7 in pre-phagosomal vacuoles in resting cells. Different from a model view of Rab5 localization though, assays using the fluid-phase marker FITC-dextran suggest that Rab5 does not localize to early endosomal structures in *E. histolytica*[521], in contrast to what has been observed in mammalian cells[620].

3.2.4.4.2 Rab7

Rab7 is present on mature endosomes, MVBs, and lysosomes, as well as on phagosomes. It is involved in recruitment of the HOPS tethering complex to ensure regulated fusion with the degradative compartment[151,621], as well as the retromer complex to ensure recycling of components prior to terminal degradation[106].

A. thaliana encodes eight putative Rab7 family proteins belonging to the RABG3 group, suggesting the potential for redundancy and/or novel functions. RABG3f primarily co-localizes with MVB and vacuole markers, and expression of a dominant negative version causes fragmentation of the vacuole and inhibits vacuolar trafficking[622]. RABG3b is involved in autophagic processes such as cell death and differentiation during growth and pathogen response[623,624]. Some functional redundancy likely exists, as various quintuple and sextuple mutants show phenotypic defects but remain viable[611].

Rab7 has not been extensively characterized in *T. thermophila,* but is present in a phagosome proteome[625], and tagged Rab7 is present both as bright puncta on phagosomes, as well as structures containing LysoTracker Red[626].

In *T. gondii* Rab7 localizes in the late secretory system of the parasite, and partially co-localizes with various markers of the endosome-like compartment and vacuole, but is distinct from both Rab5A and the Golgi protein GRASP[247,248,253]. Parasites overexpressing Rab7, or expressing constitutively active or dominant negative versions of Rab7, exhibit growth defects but no obvious trafficking defects[253]; this is at odds with an interaction between active Rab7 and the retromer component Vps26[375]. Hence, the function of Rab7 in *T. gondii* is unclear.

P. falciparum Rab7 localizes primarily to distinct puncta throughout the intracellular life cycle that partially co-localize with the retromer component Vps35 but is distinct from Golgi-associated Rab6. Expression of constitutively active or dominant negative versions, similar to *T. gondii*, showed no appreciable trafficking defect[385].

As with Rab5, we could not find evidence of a Rab7 orthologue in *G. lamblia*. Trypanosomes retain a single Rab7 paralogue, which closely juxtaposes to the lysosome. Knockdown of TbRab7 impairs uptake of a subset of endocytic cargo, but does not appear to affect the delivery of biosynthetic lysosomal cargo[627].

E. histolytica has multiple Rab7 paralogues. Rab7A through Rab7E are present by proteomic analysis on phagosomal membranes at multiple time points[618], and, as previously mentioned, Rab7 associates with Rab5 at pre-phagosomal vacuoles and interacts with the retromer complex[521,603]. Overexpression of Rab7 results in enlarged intracellular vesicles, and an overall increase in cell acidity, but no apparent defect in phagocytosis or endocytosis[603]. Though four Rab7 paralogues are present in a cell surface proteome their localization and function has yet to be fully elucidated[628].

In *D. discoideum* Rab7 has been localized to phagosomes by proteomics of isolated organelles[555,629,630]. By microscopy, Rab7 localizes to phagosomes, macropinosomes, lysosomes, and post-lysosomes[631–633]. Expression of a dominant negative Rab7 inhibits macropinocytosis and phagocytosis[631,633], and prevents delivery of endo-lysosomal components, yet enhances the delivery of unprocessed proteases and sugar-linked proteins, to maturing phagosomes[632].

3.2.4.4.3 Rab11

In opisthokonts Rab11 is primarily involved in recycling of cell surface proteins, but also plays a role in other cellular processes including innate immune responses, delivery of components to the cleavage furrow during cytokinesis, and ciliogenesis, at least in mammalian cells[388,634].

The Rab11 subfamily is highly expanded in *A. thaliana*, with 26 putative members divided into six sub-groups, RABA1 through RABA6. RABA1 members display dynamic localization between the TGN, endosomes, and plasma membrane[635–637], suggestive of a possible recycling function; consistent with this, RABA1b mutants show hypersensitive intracellular aggregation of plasma membrane proteins in response to Brefeldin A[635], and the RABA1 quadruple mutant is sensitive to salinity stress[637,638]. All RABA2 and RABA3 members appear to localize to the same compartment, which is distinct from the Golgi and late endosomes, but does overlap with markers of the TGN and other Rab11 members[635,639]. During cell division, various RABA members re-locate to the cell plate, where they co-localize with KNOLLE, a SNARE involved in cytokinesis[639]. Consistent with this, cell wall analysis revealed a decrease in specific constituents in rabA2b, rabA2d, and three rabA4 mutants[640]. Additionally, RABA4 members localize to the tip area of growing cells[641–644], where they

interact with PI-4-kinases and phosphatases[641–645] to mediate polarized growth; RABA4c also plays a role in recycling of plasma membrane receptors[636].

T. thermophila encodes multiple Rab11 paralogues, one of which, Rab11A, labels posterior to anteriorly directed vesicles, which may represent recycling endosomes, and also partially labels the contractile vacuole[626].

A proteomic study of isolated rhoptries in *T. gondii* revealed the presence of Rab11A in this compartment[254]. Confirming this, Rab11A partially co-localizes with the rhoptry protein ROP5, but also with endosome-like compartment markers. Expression of a dominant negative Rab11A does not affect invasion organelles, endosymbiotic organelles, or the Golgi, but prevents delivery of late stage components of a plasma membrane-associated complex termed the IMC, and results in defective cell division[391]. Rab11B, the other Rab11 paralogue, co-localizes with a Golgi marker in resting parasites, but relocates to the IMC in developing daughter cells. Expression of a dominant negative Rab11B shows a similar defect in cell division as Rab11A, albeit due to distinct trafficking pathways with different timing[390], and Rab11B is also present in a retromer interactome[375].

Similar to *T. gondii*, Rab11A was found to localize in discrete puncta throughout the intracellular lifecycle of *P. falciparum*, some of which co-localize with the resident rhoptry protein Rhop2 and the IMC protein GAP45[391].

The single Rab11 in *G. lamblia* is present in puncta or stacks in cells preparing to encyst, and at the cell periphery in mature cysts, where it co-localizes with the cyst wall protein CWP1. Ribozyme-mediated knockdown results in a decrease in CWP1 present in encystation-specific vesicles, instead being present in numerous cytoplasmic puncta, suggesting a trafficking defect[646].

Rab11 is a major regulator of recycling pathways in African trypanosomes. Turnover of surface proteins in *T. brucei* is strongly influenced by Rab11, while extensive disruption of endocytic pathways follows Rab11 knockdown. Furthermore the underlying interactome for Rab11 is divergent between trypanosomes and mammalian cells; FIP proteins that mediate Rab11 function in mammalian cells are absent, and at least one trypanosome-specific interacting protein has been identified[647]. In *T cruzi* Rab11 mediates an unusual pathway that traffics the critical *trans*-sialidase surface protein family to the surface, but which is via the contractile vacuole[648]. This suggests that the diversification of function within trypanosomes is often cryptic, and as discussed above, can depend on the precise cellular configuration.

In *E. histolytica* Rab11 is enriched in endosomal fractions[649], but microscopy revealed localization in small cytoplasmic vesicles, and a lack of co-localization with phagocytosed *E. coli*, endocytosed transferrin, or markers of the ER or Golgi[650]. Similarly, Rab11B is associated

with non-acidified compartments that are distinct from the ER, early endosomes, and lysosomes. Rab11B overexpression enhances exocytosis of fluid phase markers, intracellular and secreted cysteine protease activity, and improves killing efficiency, suggesting a potential role in recycling and release of pathogenesis factors[651].

Multiple Rab11 paralogues exist in *D. discoideum*. Rab11A localizes to the contractile vacuole network, and also co-localizes, as well as interacts with, the contractile vacuole-associated ion channel P2XA[652]. A previous study identified Rab11 in contractile vacuole-associated fractions by blotting, and co-localized Rab11 with other markers of the contractile vacuole network[653]. Overexpression, or expression of a dominant negative version, of Rab11 results in aberrant contractile vacuole morphology and impaired osmotic stress response[652,653]. Correlative data suggests that Rab11A and Rab11C may be involved in delivery of a V-ATPase to phagosomes[654], which is consistent with their identification in a proteomic analysis of purified phagosomes[630].

3.2.4.4.4 Functional homology in Rab GTPases

Rab5 and Rab7 have well defined localisations and functions in model systems, and the Rab5 to Rab7 switch is a paradigm for dynamic protein association during organelle maturation. The localization and function of Rab5 in trypanosomes is consistent with a canonical role, while the role of Rab5A and Rab5C in trafficking to *T. gondii* apical organelles is conserved when these organelles are viewed as derived endo-lysosomes. Similarly, Rab7 performs the expected function in trypanosomes, and its localization in Apicomplexa to compartments homologous to late endosomes/lysosomes, is also consistent with model systems. Paralogous expansion of both Rab5 and Rab7 in *A. thaliana* complicates assessment of functional homology, including the role of ARA6 in recycling traffic, though overall localization and function imply conservation. Studies in *D. discoideum* and *E. histolytica* suggest that Rab5 and Rab7 maintain a conserved role in the function and maturation of compartments derived from internalization of extracellular material

Rab11 primarily mediates trafficking through recycling endosomes. *Entamoeba* Rab11 is present at compartments distinct from early and late endosomes, potentially in a recycling endosome, which is consistent with the increased exocytosis noted in cells overexpressing Rab11B. Similarly, *T. brucei* Rab11 is important for recycling traffic. The primary role of Rab11 in *G. lamblia*, *T. gondii*, and *P. falciparum* can generally be described as delivery of cargo to structures adjacent to the plasma membrane. The unique mechanisms by which apicomplexan parasites undergo cell division (endodyogeny in *T. gondii* and schizogeny in *P. falciparum*) are important when assessing functional homology. In these organisms progeny emerge from

within the mother cell, mediated in part through the specific and timely IMC formation[237,655,656], which is mediated by both Rab11 paralogues. This is reminiscent of the regulatory role for Rab11 in animal cell cytokinesis, together with exocyst[657]. The extensive diversification of the Rab11 family in *A. thaliana* is unprecedented in other eukaryotes, but some members possess functions such as recycling and trafficking of plasma membrane and cell wall constituents during cell division and polarized cell growth.

Rab11 may be involved in contractile vacuole function in both *D. discoideum* and *T. thermophila.* The contractile vacuole is an enigmatic organelle present in a subset of organisms across eukaryotic diversity though it is not yet established whether these are homologous or analogous. A role for Rab11 in the function of this compartment is consistent with exocyst involvement in the contractile vacuole of *D. discoideum*, as well as the unicellular archaeplastid *Chlamydomonas reinhardtii*[658–660]. Additionally, Rab11 has been identified in proteomic studies of the contractile vacuole in *T. cruzi*[661], and recycling traffic appears to transit this organelle[648]. Finally, though current evidence is limited, Rab11 also appears to play a role in trafficking to phagosomes in *D. discoideum*. This is consistent with recent studies suggesting such a role for Rab11 and exocyst in phagosome maturation in endothelial cells[662].

3.2.5 Discussion

3.2.5.1 Overview

With the increasing ease and prevalence of comparative genomics, the validity of assuming functional homology is both critical to assess and fruitful to explore. First and foremost, the simple conclusion from our comparative survey is that yes, orthology does appear to translate into functional homology. However, this is complicated by many factors, and needs to be taken as a first foray into this kind of assessment, and not a question laid to rest.

Firstly, despite considerable efforts to expand experimental investigation into non-model eukaryotes, there are still large gaps in our knowledge base, as evidenced by the fact that we were only able to find comparable molecular cell biological data for a small set of membrane-trafficking genes, essentially all within the endocytic system. Future studies expanding into the secretory system and encompassing machinery identified in diverse eukaryotes but that is absent or diverged in opisthokont taxa, for example the TSET complex and novel ArfGAP subfamily ArfGAPC2[32,49], will aid in correcting the asymmetrical bias on opisthokonts in our models of membrane-trafficking.

Table 3.1 Functional homology across model systems

This table provides a brief summary of the evidence for functional homology for select membrane trafficking components across discussed model organisms. Trafficking machinery is listed by row and organisms by column. For each component listed, the major localization and presumed function are listed, with appropriate references for each; for more extensive description of the underlying evidence please see the relevant main text section(s). Abbreviations: Com, component; Evi, evidence; Des, description; Ref, references; Loc, localization; Fxn, function; PM, plasma membrane; TGN, trans-Golgi network; MVB, multi-vesicular body; LE, late endosome; CV, contractile vacuole; DCG, dense core granule; ELC, endosome-like compartment; Mic, microneme; Rhop, rhoptry; VAC, vacuolar compartment; DV, digestive vacuole; IMC, inner membrane complex; PPV, pre-phagosomal vacuole. Blank cells are present where components are either unknown or no evidence exists.

		A. thaliana		*T. thermophila*		*T. gondii* & *P. falciparum*		*G. lamblia*		*T. brucei*		*E. histolytica*		*D. discoideum*	
Com	**Evi**	***Des***	***Ref***	***Des***	***Ref***	***Des***	***Ref***	***Des***	***Ref***	***Des***	***Ref***	***Des***	***Ref***	***Des***	***Ref***
AP-1	***Loc***	TGN, endosomes	[528-530]	Puncta	[533]	Golgi, ELC	[396,399]	PV	[535]			Phagosomes	[539]	Phagosomes, Golgi	[540, 541]
	Fxn	Vacuolar delivery, PM recycling	[528-530]			Mic/Rhop biogenesis	[396]	PV trafficking	[535, 536]	Lysosomal delivery	[537, 538]			Phagocytosis, CV, lysosomal delivery	[540-542]
AP-2	**Loc**	PM, puncta	[544-549]	CV, basal bodies	[533]			PV	[558]			Phagosomes	[539]	PM, CV, puncta	[552, 553]
	Fxn	Endocytosis	[544-547, 549]					Endocytosis, cyst formation	[558]						
ESCRT	**Loc**	TGN, endosomes, MVBs	[514,564, 566-569]			Vps4 cytosolic	[485]	PV	[578, 579]	LE/MVB	[182]	Phagosomes, MVBs	[582, 583]	Intracellular puncta	[183]
	Fxn	Vacuolar delivery, autophagy	[566,568-572]							Vacuolar delivery	[182]	Phagocytosis	[583]	Differentiation	[584]
Retromer	**Loc**	TGN, endosomes	[586-593]	Vps10-like puncta	[599]	TGN,ELC	[375, 385]	PV, ER	[601, 602]	Endosomes	[96]				
	Fxn	Vacuolar delivery	[588-592,594, 595,598]	DCG biogenesis	[599, 600]	ELC to TGN and PM recycling	[375]			Vacuolar delivery	[96]	PPV maintenance	[521, 603]		
Rab5	**Loc**	Endosomes	[606-612]			ELC, PM	[245,246, 248,253, 374,614]			Endosomes	[615-617]	Phagosomes, PPVs	[521, 618, 619]		
	Fxn	Vacuolar delivery, recycling	[606,608, 613]			Mic/Rhop and DV trafficking	[253, 374, 614]			Endocytosis	[615-617]	PPV maintenance	[521]		
Rab7	**Loc**	Endosomes, vacuole	[622]	Phagosomes	[625, 626]	ELC, VAC	[247,248, 253,375, 385]			LE/MVB	[627]	Phagosomes, PPVs	[521, 618]	Phagsosomes, late endocytic	[555, 628-633]
	Fxn	Vacuolar delivery, autophagy	[622-624]							Vacuolar delivery	[627]	PPV maintenance	[521, 603]	Phagocytosis, lysosomal delivery	[631-633]
Rab 11	**Loc**	PM,TGN, endosomes	[635-637,639, 641-644]	Endosomes	[626]	Rhops, PM (IMC)	[254,390, 391]	Puncta, PM	[646]	Endosomes	[647, 648]	Endosomes, puncta	[649-651]	CV, phagosomes	[630, 652, 653]
	Fxn	Cytokinesis, PM trafficking	[635-645]			Cell division	[390, 391]	Cyst formation	[646]	Recycling	[647, 648]	Recycling	[651]	CV function, osmotic stress	[652, 653]

Nonetheless, this basic position of functional homology enables hypotheses to be generated and tested to better understand the effect of paralogous expansion and accretion of novel factors. Additionally, our comparative analysis indicates that considering differences in endomembrane organization and trafficking pathways (e.g. the presence of unique organelles or expanded trafficking pathways), is essential to assessment of both functional homology and novelty among lineages.

3.2.5.2 Functional homology of trafficking machinery in diverse eukaryotes

Our pan-eukaryotic comparisons highlight the plasticity of the endomembrane system, not only in parasites, which possess modifications concurrent with their unique pathogenic mechanisms, but also in free-living taxa, and this plasticity must be considered in order to properly assess functional homology.

Perhaps the best example is *G. lamblia*, where the peripheral vacuoles correspond to, and encompass the function of, diverse endo-lysosomes present in model systems. Hence, localization of a plethora of machinery, including AP-1, AP-2, ESCRT, and retromer to these structures is consistent with conserved function, though coincident localization of all these factors in other cells would be unusual.

Understanding trafficking in higher plants requires consideration of the unique organization of their endocytic system, namely that of a combined TGN/early endosome. Some phenotypes, such as aggregation of plasma membrane receptors in response to Brefeldin A, make sense only in the context of this feature. Additionally, the endosomal system in these organisms is likely more complex than has been fully appreciated in previous studies: MVBs appear to bud directly from the TGN/early endosome[514], incomplete co-localization of endosomal markers suggests existence of sub-populations, and a recent study has suggested at least three distinct pathways exist for the movement of cargo from the TGN/early endosome to the vacuole[611].

The organization of the apicomplexan endomembrane system shows significant lineage-specific divergence. The role of a Vps10-like receptor, Rab5A and Rab5C, AP-1, and retromer in mediating apical organelle biogenesis appears at odds with canonical functions for these proteins. However, apical organelles are homologous to endo-lysosomes, and some evidence points to a plant-like organization for the TGN/endosome-like compartment. Hence, these factors can be understood to mediate both anterograde and retrograde transport through an intermediate compartment within the endosomal system, and their function is thus conserved.

In *E. histolytica,* as in humans, Rab5 and Rab7 are involved in phagocytosis, yet Rab5 does not appear to be involved in endocytosis. Subunits of the AP-1 and AP-2 complexes, as well as retromer, are found at phagosomes, and, while this may seem superficially like a case of neofunctionalization, is consistent with a role for AP-1 in phagocytosis in murine macrophages[560], and evidence for roles for both AP-2 and retromer in phagocytic clearance of apoptotically killed cells in *Caenorhabditis elegans*[663,664]. Therefore, many seemingly non-canonical functions of trafficking factors in *E. histolytica* may represent specialization common to professional phagocytic cells.

3.2.5.3 Evolutionary precedent of conserved and novel features

The cell biological complement in the LECA served as initial building blocks for environmental adaptation during eukaryotic radiation, including in parasites. It is likely that drastic alterations from an established state would be selected against, unless the environment was radically different than that encountered by previous generations. This both explains the gross underlying pattern of functional homology and provides a precedent for trafficking system modification.

In many cases, such as *Giardia*, apicomplexans and to a lesser extent kinetoplastids, parasites have reduced their membrane-trafficking gene complements[395,517,665], often interpreted as jettisoning unnecessary or redundant pathways. Further experimental characterization will be needed to determine the extent to which this interpretation bears out. By contrast, other taxa, such as *Entamoeba*, *Dictyostelium* and *Tetrahymena*, have expanded their complements. In cases where multiple paralogues exist, some may possess a similar basic function, but may do so only in specific life cycle stages, or only in a restricted region of the cell, allowing for polarized trafficking and specialization.

We argue that this latter mode of innovation in the trafficking system is best viewed as an extension of the Organelle Paralogy Hypothesis[453]. Just as the process of gene duplication and co-evolution of identity encoding machinery is proposed to have given rise to the basic set of membrane-trafficking organelles prior to the LECA[394], the same process should continue to act in extant eukaryotes. Hence new organelles may arise from an ancestral compartment through concurrent duplication and co-evolution of the underlying identity-encoding trafficking machinery, such that the machinery acquires specific features for this role. This may include specific trafficking signals, the ability to bind to specific proteins or phosphoinositides, and additionally they may be further regulated by specific factors such as GEF and GAP proteins.

By extending this to descendants of a lineage in which the organelle arose, particularly when the homologous organelle is present and its function required, some paralogues that arose concurrently with it would be maintained and constrained to performing required functions for organelle biogenesis and/or maintenance, and hence will be functionally homologous. However, in descendants no longer possessing the organelle or its required function, or in cases of further expansion, regardless of the presence or absence of a homologous organelle, paralogues are unconstrained and may acquire new function. Hence, despite a conserved set of organelles and machinery inferred in the LECA, extant eukaryotes display an array of unique features. This not only applies to the endomembrane system, as we have described here, but also likely extends across cellular systems.

Although we can be relatively optimistic in assuming functional homology within the membrane trafficking system, equivalent assessments may or may not show the same thing in other cellular systems; the question is certainly worth asking.

3.2.5.4 Conclusions and future perspectives

In conclusion, despite considerable divergence in cellular systems among diverse eukaryotes since the LECA, efforts to map function on the basis of comparative genomic data appear to be well founded. Our literature review revealed that functional homology is present in membrane trafficking system machinery in several taxa spanning eukaryotic diversity and encompassing both free-living and parasitic organisms. This allows for some further degree of confidence in continued molecular evolutionary and comparative genomic analysis as well as providing a lens through which to view the unique cell biological adaptation present in each organism in order to fully appreciate how these systems may differ. In particular, expanding this analysis across systems between parasites and their hosts can be expected to provide valuable insight into the complex interactions between them.

3.3 Additional discussion

The overall conclusion reached by this analysis was that functional homology is broadly present across eukaryotic diversity, at least for select components of the MTS, but that lineage-specific differences in morphology and trafficking pathways had to be considered to arrive at this conclusion. Overall, this restriction is additionally consistent with the concept of functional homology, whereby "function" may be defined by considering homologous structures, which themselves may not retain the same morphology and/or organization.

For example, within the endosomal system, the mammalian model of discreet endosomal compartments with at least one early endosomal structure separating material

taken up from the extracellular milieu from the Golgi/TGN, forms the basis for the "textbook model" of endocytic trafficking. However, this appears not to be the case with plants such as *A. thaliana*[441], and has recently been suggested to also not explain the organization of the endosomal system in a more closely related organism, the budding yeast *S. cerevisiae*[442,666]. In these organisms, uptake results in direct distribution to the Golgi/TGN, which itself acts more like an early-endosome-like sorting station. However, similar functions of Rab5 homologues, ScVps21p/ScYpt52p/ScYpt53p and AtRHA1/AtARA6/AtARA7, suggest that within the confines of the different organization present in each system, similar functions must still be fulfilled and that it is still the homologous proteins that (by and large) perform these functions.

In terms of the apicomplexan MTS, these results suggest two things: 1) the reduced complement of MTS proteins encoded in Apicomplexa as compared to the LECA (Chapters 2 and 4) suggest that some functions became dispensable over time, allowing loss of the corresponding trafficking factors, and 2) that the presence of novel organelles in these parasites may correlate with the advent of additional machinery. As will be discussed more fully in Chapter 4, these organelles include the micronemes, rhoptries, and dense granules, as well as the IMC and the apicoplast. Not all trafficking to these organelles might necessarily require additional trafficking factors; for example, the dense granules appear to form directly from the Golgi/TGN in *T. gondii* and superficially resemble dense core secretory granules of specialized secretory cells such as endocrine and neuroendocrine cells[667]. Although the case for homology between these structures requires further study of the morphology and biogenesis of dense granules, it is possible that their formation and budding might use otherwise conserved resident Golgi/TGN proteins.

In the case of the micronemes and rhoptries, as was discussed in section 1.6, these organelles are present within the endosomal system, although their exact identity remains to be established. Although this might suggest that they also do not require additional MTS machinery, the presence of other endosomal compartments (one or more endosomal structures and a terminal degradative compartment)[245,247,248], suggest that further machinery might be required for their biogenesis and maintenance. The rationale behind these conclusions will be revisited in Chapters 4 and 5, and then discussed at length in Chapter 6.

There is also the question of the implications of functional homology for eukaryotic organisms in general. At a surface level, it suggests that broad comparative studies that aim to reconstruct ancestral states based on the conservation of factors in extant descendants of that ancestor are justified in their approach. For example, numerous studies have reconstructed the LECA gene complement for systems as diverse as trafficking, cell division,

and the cytoskeleton[505]. Our ability to infer the lifestyle of this ancestral organism entirely hinges on whether it is a valid assumption that the homologues inferred as present in LECA would perform a similar function as those homologues do in extant eukaryotes. Similarly, our assumptions regarding the LECA organelle complement also ride, to a certain extent, on functional homology; this latter point is slightly undercut by the finding in this analysis that gene function was conserved to a higher extent than gene localization (i.e. than the retention of morphologically comparable structures).

However, functional homology is not just important for understanding evolutionary history. Throughout this thesis it is argued that the concept of functional homology is inextricably linked to the OPH and the gain and loss of autogenously generated organelles. As discussed above, the transition from the first eukaryotic common ancestor (FECA) to LECA, and therefore the array of organelles and functions within LECA itself, is dependent on the ability to confidently infer function from homology. But the underlying OPH mechanism is simplistic and straight-forward[453], and not conceptually unique to the ancestral eukaryotic lineage; the OPH would be expected to operate in extant eukaryotes as well. Hence, novel organelles across eukaryotes may be associated with paralogous expansion and diversification of MTS machinery. This represents not only an intriguing hypothesis, but also one which can be tested. At a surface level, gene complement can be inferred from sequence data and paralogous expansions identified (this will be discussed at length in Chapter 4). Although expansions alone provide circumstantial evidence, in models in which access to traditional cell biological methods of genetic manipulation and phenotypic characterization are available, the function of these additional paralogues can be tested to confirm or refute the informatically generated hypotheses (for example, as performed in Chapter 5).

Although comparative genomic studies have been carried out for over two decades, the implicit assumptions underlying these studies have not been explicitly tested. This chapter has provided evidence for the existence of functional homology, which is arguably key to understanding both the advent of canonical eukaryotic organelles deep in evolutionary time as well as to understanding the diverse array of cellular features in modern eukaryotes. Continued studies across diverse eukaryotes will provide additional information to either strengthen or weaken the concept of functional homology. Regardless, this simple concept, that the evolutionary relationship between genes is linked to their function, represents a powerful framework that should be considered by evolutionary and molecular biologists alike. In the next two chapters, the relationship between functional homology, paralogous expansion, and generation of autogenous organelles in extant eukaryotes is explored for the Apicomplexa, and specifically for the model organism *T. gondii*. Chapter 4 provides evidence

for paralogous expansion of MTS machinery during apicomplexan evolution, while Chapter 5 describes the characterization of three such additional paralogues, with special emphasis on the role of one such paralogue in mediating trafficking to micronemes and rhoptries.

4. An Informatics Screen to Identify Novel Paralogues in Apicomplexa

4.1 An argument for novel paralogues in Apicomplexa

This chapter focusses on the potential for Apicomplexa, and possibly their close relatives as well, to encode additional MTS paralogues beyond those present in the LECA. To begin, Chapters 2 and 3 are briefly summarized and used to provide the basis for the hypothesis. The overall strategy used is explained and then results shown for selected MTS families. Possible ramifications of these data are then discussed.

4.1.1 Loss: the common pattern observed previously

As parasites of global importance, Apicomplexa were among the first eukaryotes to be fully sequenced, with completed genomes for *P. falciparum* in 2002[668], *T. gondii* in 2003[669], and *C. parvum* in 2004[458]. A large quantity of other genome data (some of it unpublished), is provided to the community through EuPathDB (https://eupathdb.org/eupathdb/).

Due in part to this early availability, Apicomplexa were frequently included in early studies investigating the spread of MTS components known from molecular studies in humans and yeast. These included broad studies into endocytic components[88], as well as deeper investigations into gene families including the Adaptor proteins[48,49,468], ESCRTs[182], MTCs[126,127], retromer[96], and others. However, in the vast majority of these studies, only one or a few apicomplexan genomes were included, frequently *P. falciparum* and *T. gondii*, as these remain arguably the best-known members of the phylum.

Chapter 2 described a study on several of these components, Clathrin, retromer, Adaptins, ESCRTs, and MTCs, across a larger and more focussed apicomplexan-centric dataset. Crucially, this was also the first study to include the important outgroup taxa *C. velia* and *V. brassicaformis*, which provided unprecedented resolution in the timing of changes leading up to, and following the divergence of, Apicomplexa themselves (section 2.2).

The overall pattern observed across all of these studies was that of loss of machinery determined to be present in the LECA; i.e., secondary loss within Apicomplexa, or within a larger group that includes Apicomplexa. A loss of machinery in parasites is often interpreted as "jettisoning" components that a parasitic lifestyle has rendered non-essential, and is a common trope in the literature[460]. Although attractive in its simplicity, this interpretation does not hold in all cases, as will be discussed throughout this chapter and throughout the remainder of this thesis.

4.1.2 Apicomplexa possess organelles not found in other eukaryotes

As discussed in Chapter 3, a combination of functional homology and the OPH provide a basis for understanding the patterns of gene gain and loss across eukaryotes. The FECA-

LECA transition is hypothesized to have involved the expansion of a primordial set of MTS machinery, concurrently with an expansion from a primordial, to a more modern, complement of autogenous organelles[446,453]. Hence, machinery inferred to be present in the LECA based on comparative studies was likely already involved in the maintenance of organelle identity for the "canonical" eukaryotic organelles, such as the ER, Golgi, and various endosomes. The advent of new organelles in a given lineage, especially in cases where canonical organelles are still present, would therefore likely involve the advent of new MTS factors. This could theoretically be achieved either through paralogous expansion of existing factors, or through accretion of non-homologous machinery. The latter case has been demonstrated in the literature, for example in the gain of novel clathrin-interacting proteins in trypanosomes[670]; however, in general, identification of these factors is only achievable through direct biochemical means. The former case is attractive from the perspective of *in silico* analyses, as homology provides a basis for identification using existing algorithms.

As discussed in section 3.3, Apicomplexa possess several organelles without obvious homologues in other eukaryotes, including the dense granules, micronemes, and rhoptries, as well as the IMC and the apicoplast. Of these organelles, the need for additional machinery is less clear for the dense granules and the apicoplast. Although it has yet to be demonstrated convincingly, available evidence suggests that dense granules bud from the Golgi/TGN and form the "constitutive" arm of the apicomplexan secretory pathway[423,427,428]. Disruption of conventional trafficking machinery such as ARF, Rab6, Stx6, and α-SNAP[375,381,393,409,424,425] affect dense granule biogenesis and/or trafficking, which are all consistent with an origin at the Golgi/TGN. Additionally, although immature versions of dense granules (in analogy to immature secretory granules) have not been identified to date, dense granules are superficially similar to dense core secretory granules of certain mammalian cell types[671]. As discussed in section 1.6.2.2, trafficking to the apicoplast is both poorly studied and poorly understood. Unlike most other secondary (or higher) plastids, the apicoplast is bound by four membranes and is present outside the ER[260,672]. Additionally, it is thought that the outermost membrane is traversed by cargo through the act of vesicular fusion[236] rather than by any form of channel. No trafficking machinery has been unambiguously determined to date to affect apicoplast trafficking, and the exact route of the trafficking, either ER-to-apicoplast or Golgi-to-apicoplast, remains contentious[412–414].

The IMC is an enigmatic structure subtending the plasma membrane in Apicomplexa consisting of either one large continuous compartment or a series of distinct compartments held together by proteinaceous "sutures"[234,235]. Although the IMC is unambiguously homologous to the alveoli of other alveolates, such as ciliates[236], the case for homology of

these structures to other eukaryotic organelles is relatively unclear. Imaging-based approaches in *P. falciparum* showed that the IMC component PfGAP50 is present first in the ER and then at the IMC, suggesting ER-IMC trafficking during intracellular development[673]. In *T. gondii*, TgIMC1 has been observed in close proximity to the nucleus during early endodyogeny, consistent with the ER/Golgi of the parasite[674]. Two subsequent studies demonstrated the role of the pan-eukaryotic TgRab11A and alveolate-specific TgRab11B in IMC biogenesis[390,391], albeit both had different localization patterns and specific roles. Hence, the IMC appears to form from the early secretory pathway (ER/Golgi), but involves machinery typically functioning in the endosomal system (Rab11).

Comparatively, trafficking to micronemes and rhoptries is much better understood. A number of microneme and rhoptry resident proteins are synthesized as apoproteins and processed during trafficking. Localization and biochemical studies have convincingly demonstrated that processing occurs in a compartment within the late Golgi/TGN or beyond[245,434–437], suggesting that the trafficking is a post-Golgi event. Additionally, studies of machinery traditionally associated with the endosomal system in *T. gondii*, such as AP-1[258,396,399], Rab5A and Rab5C[253], DrpB[425], clathrin[405], retromer[375], and the VpsC complexes[407], have shown effects on micronemes and rhoptries. Based on these results, it has been hypothesized that micronemes and rhoptries represent some form of endosomal compartment, although the exact nature of the relationships are difficult to ascertain[236,431]. However, Apicomplexa possess other endosomal compartments; a poorly defined structure that might represent one or a collection of endosome-like structures (the ELC)[245], as well as a vacuolar structure (PLV/VAC)[247,248]. In particular, the presence of more canonical endolysosomal compartments in addition to micronemes and rhoptries suggests that they truly represent additional compartments, which would be consistent with an OPH-like expansion of the underlying trafficking machinery.

4.1.3 Rationale and hypothesis

The presence of these aforementioned additional organelles would suggest that Apicomplexa encode additional MTS machinery to mediate the additional trafficking steps that are necessary to maintain organelle identity. However, as discussed in section 4.1.1, previous studies into the apicomplexan MTS identified only examples of gene loss. Although some of these losses may be associated simply with a shift in the emphasis of function for a given system or complex, for example the likely emphasis for a role of apicomplexan ESCRT in cell division (discussed in section 2.3), the pattern is overall inconsistent with organelle expansion under the OPH.

Hence, I hypothesized that there must be additional machinery that mediates trafficking to at least a subset of the additional organelles in Apicomplexa. Importantly, this hypothesis is already supported by key examples in the literature. The alveolate-specific Rab11B[390] is involved in trafficking to the IMC (an organelle found only in alveolates), and, although not described as such at the time, the myzozoan-specific Rab5C (see section 4.4.4) is involved in trafficking to rhoptries and a subset of micronemes[253]. Given that microneme and rhoptry-like organelles are present across the Myzozoa[232], these two examples represent the advent of paralogues in a lineage concurrently with the advent of novel organelles, exactly as functional homology and the OPH would predict.

Additional machinery might be present either as novel factors not homologous to any known MTS component, or represent additional paralogues within known MTS families. As the former would be difficult, or even impossible, to screen for using informatic methods, I chose to focus on the latter. Additionally, as the two previous examples in the literature represent additional paralogues within known MTS families, I expected this approach to yield additional candidates. The workflow is described in section 4.3.

4.2 Materials and Methods

Predicted proteomes of all organisms under study were downloaded from relevant public databases; information regarding all datasets is found in Supplementary Table S4.1. Initial identification of homologues was performed using HMMer v3.1b1[499] followed by reciprocal BLASTp (v2.8.1)[464] searches against the *Homo sapiens* predicted proteome to check for false positives. All homology searching employed an e-value cut off of 0.05; discrimination between positive and negative hits in reciprocal BLAST used a two order of magnitude cut off, whereby true homologues were considered to hit a relevant homologue in *H. sapiens* with an e-value at least two orders of magnitude greater than the first non-homologous hit. In some cases, additional homologues were identified by reciprocal BLASTp[464] analysis using an identified homologue from the most closely related taxon within the dataset. BLAST and HMMer searches were run using Goat, a graphical user interface (GUI)-based wrapper to the underlying search algorithms I wrote using Python3 (https://github.com/chris-klinger/Goat). Goat was tested by searching for all homologues identified in a previous study wherein I performed all homology searching manually[127]; Goat identified all homologues through reciprocal BLASTp analysis apart from those hits that were originally identified by manually searching genome scaffolds (Online Appendix Table 4.1). In addition, all homology searches were confirmed through manual inspection and additional phylogenetic analyses (see below). Domain prediction used PfamScan v1.6[675] with an e-value cut off of 0.01; reported start and

stop positions represent the domain "envelope". All results of homology searching and domain prediction analysis can be found in Online Appendix Tables 4.2-4.11.

Structural modeling was carried out using the "intensive" mode on the Phyre2 web server[676]. Additional confirmative modeling of the β-propeller structure of *T. gondii* Rab1K used HHpred[500] and MODELLER[677] through the MPI Bioinformatics Toolkit[678]. Both web servers were accessed in July 2019.

All alignments were carried out using MAFFT v7.407[679]; for alignments less than ~250 sequences, the slow and accurate L-INS-i method was used, while larger alignments used the ––auto option. Alignments were manually inspected and trimmed by hand. Phylogenetic analysis was carried out using one of three methods. IQ-TREE v1.6.11[680] was used for rapid inference of large datasets, under the best model as inferred by each program run and performing 1000 replicates for both single-branch SH-like likelihood ratio testing and ultra-fast bootstrapping. RAxML v8.2.12[681] was used both for initial phylogenies during sequence classification, employing maximum-likelihood tree inference and rapid bootstrapping with 100 replicates for each run, and for final bootstrapping of datasets, using the autoMRE criterion to determine a sufficient number of bootstrap replicates for each dataset[682]. In all cases, model selection was performed by RAxML during each program run. Bayesian phylogenies were performed using MrBayes v3.2.7a[683]. Four independent runs of four chains were run for 1,000,000 MCMC generations, sampling every 500 generations under a mixed amino acid model. The consensus tree and statistics were calculated following removal of the first 20% of samples from each run as burn-in. For all phylogenetic inference, rate variation among sites was modeled using a discrete gamma distribution with four rate categories. All phylogenetic trees in this chapter are provided in Newick format in the Online Appendix.

All trees were viewed using FigTree v1.4.4 (https://github.com/rambaut/figtree) and tree figures manually modified using Affinity Designer v1.6.1 for Mac. Dot plot representations of homologue conservation among study taxa were generated using a custom Python script (included under the PyCPGen folder in the Online Appendix).

4.3 An informatics screen to identify novel paralogues

In this thesis novel paralogues, subsequently referred to as lineage-specific paralogues (LSPs), are defined as paralogues that are present in a restricted subset of taxa, but not across all eukaryotes. Although some examples are mentioned wherein duplications within one or more apicomplexan lineages were identified, LSPs specifically refer to those paralogues that are present across Apicomplexa, or present in both Apicomplexa and a subset of outgroup taxa, as these represent paralogues predicted to be present in the common apicomplexan

ancestor. Given this definition, identification of LSPs necessarily requires comparative outgroup taxa and the use of phylogenetic analysis.

The workflow used is outlined in Figure 4.1. A starting query (1), either an HMM model of a group of known homologues, or one representing a defining domain of a specific MTS family, is used to identify all putative homologues across the dataset (3), as described in the materials and methods (2). For initial phylogenetic analysis (4), putative homologues within the dataset are partitioned into smaller groups (in this thesis, the groups are Apicomplexa, chromerids, dinoflagellates, ciliates, stramenopiles, a single group comprising *Bigelowiella natans*, *Emiliania huxleyi*, and *Guillardia theta*, and archaeplastids), and marker sequences corresponding to known pan-eukaryotic paralogues added (separate blue bars in Figure 4.1). The small nature of these datasets allows for robust classification of pan-eukaryotic paralogues from among all identified homologues in each group (5). However, some sequences within each group remain unclassified, i.e. fail to group with a marker sequence with statistically significant node support (6). All such sequences from each of the separate groups are then combined and run as a single dataset in further phylogenetic analysis (7). Some of these previously unclassified sequences may group together with significant support, forming a "novel" clade; these sequences correspond to LSPs (8). Other sequences fail to form groups, or form groups within individual species or restricted sets of lineages; these sequences remain "unclassified" (9). Finally, sequences corresponding to LSPs are combined with each set of classified sequences (from step 5) to form a final (single) dataset that can be used to investigate the relationship of each LSP to pan-eukaryotic paralogues.

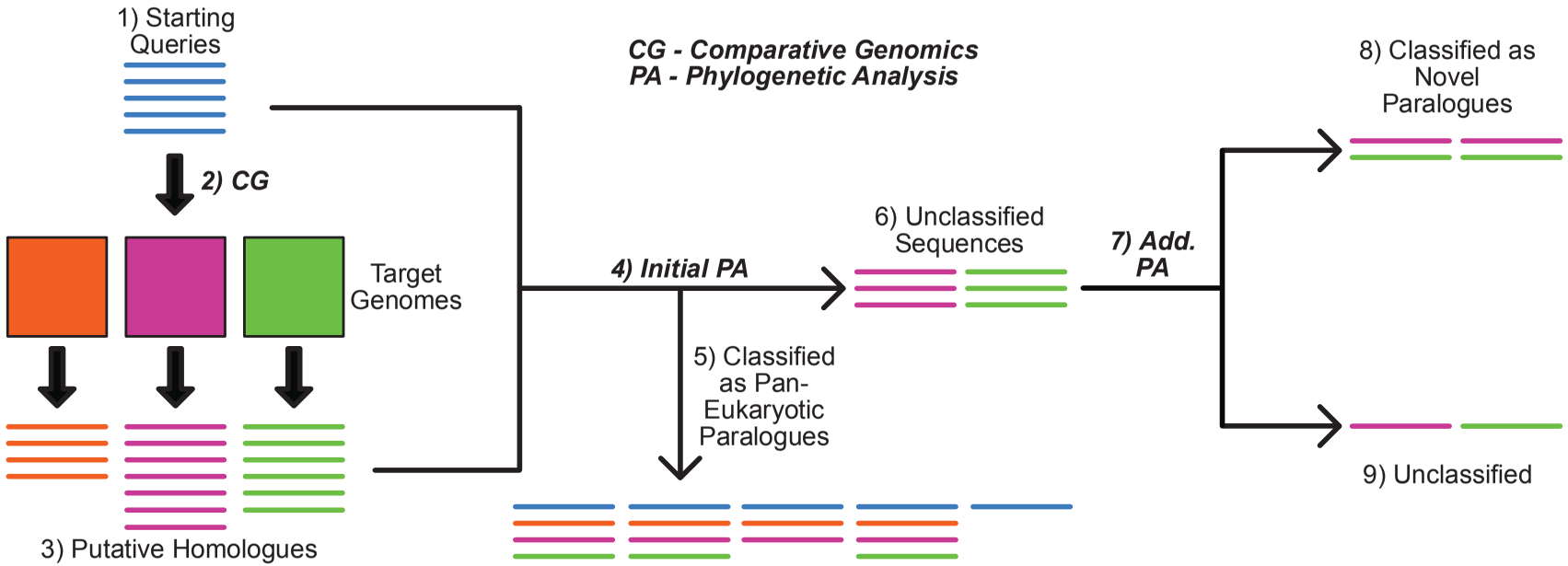
Although simple, this approach has the benefit of allowing "step-by-step" reconstruction of clades while minimizing the effect of divergent sequences and truly unclassifiable sequences, many of which are divergent. These divergent sequences can disrupt tree topologies, even when a sufficiently good evolutionary model is used, through long-branch artefacts[684], and can make classification difficult. Hence, datasets are initially split and run separately to classify sequences, and only combined in the end when divergent sequences have been removed through careful analysis of each smaller dataset.

4.4 Results

This thesis presents data for seven paralogous MTS families across a broad dataset of 50 genomes, including 27 Apicomplexa and 23 outgroup taxa (materials and methods, Supplementary Table S4.1). These taxa can be broadly divided into nine groups: the Apicomplexa, including the gregarine *Gregarina niphandrodes*, the two chromerid algae *C.*

Figure 4.1. Phylogenetic workflow to identify lineage-specific paralogues

This figure provides an overview of the workflow employed to separate conserved from lineage-specific paralogues. Homology searching (2) using a starting query/set of queries for each family (1) results in identification of all putative homologues in the dataset (3). Phylogenetic analysis of each sub-group (e.g. Apicomplexa, stramenopiles) is performed using known markers of pan-eukaryotic paralogues (4) to classify all such paralogues amongst the putative homologues (5). Remaining unclassified sequences (6) from all sub-groups are gathered and run in additional analyses (7) to identify novel paralogues (8). These novel paralogues are subsequently combined with the previously classified pan-eukaryotic paralogues from step 5 and phylogenies constructed to identify the most likely origin of each lineage-specific paralogue (not shown in the figure). The phylogenies shown in this chapter are the result of these final analyses including both pan-eukaryotic and lineage-specific paralogues. Abbreviations: CG, comparative genomics; PA, phylogenetic analysis; Add., additional.


1) Starting Queries
2) CG
Target Genomes
3) Putative Homologues
CG - Comparative Genomics
PA - Phylogenetic Analysis
4) Initial PA
5) Classified as Pan-Eukaryotic Paralogues
6) Unclassified Sequences
7) Add. PA
8) Classified as Novel Paralogues
9) Unclassified

velia and *V. brassicaformis*, the dinoflagellates, including the basal group *Perkinsus marinus*, the ciliates, the stramenopiles, the rhizarian *Bigelowiella natans,* haptophyte *Emiliania huxleyi*, cryptophyte *Guillardia theta*, and finally the blue-green (glaucophyte), green, and red algae along with *A. thaliana* (colour-coded as in Figure 4.2 below as red, teal, olive, purple, brown, orange, sky blue, and green, respectively).

For each family, a representative HMM model was built using either characterized homologues from model systems or an alignment corresponding to the Pfam model for a domain common to all members (e.g. the TBC domain for RabGAPs; see below). All putative hits were subjected to reverse BLASTp analysis against the *H. sapiens* proteome and domain prediction via Pfam. Some homologues were additionally identified via further reciprocal BLAST searches using an orthologue from the closest taxa found to encode one. All initial homologues were further classified by maximum-likelihood phylogenetic analysis using RAxML, wherein orthology was inferred based on a minimum bootstrap support of 50 uniting the sequence in question with a known marker sequence. Any sequences not classified via phylogenetics were assigned as "unclassified". Specific details of the analysis are provided in the materials and methods section.

The results section begins by discussing three families in which no clear LSPs (barring the TBS proteins containing an ArfGEF domain, see section 4.4.1) were identified, the ArfGEFs, ArfGAPs, and SM proteins. Following this, results for the other four families, in which LSPs were confidently identified, are presented by first discussing SNARE proteins, then following up with Rab GTPases and their TBC GAP proteins, and finally discussing Arf/Arl proteins, which are also the subject of Chapter 5.

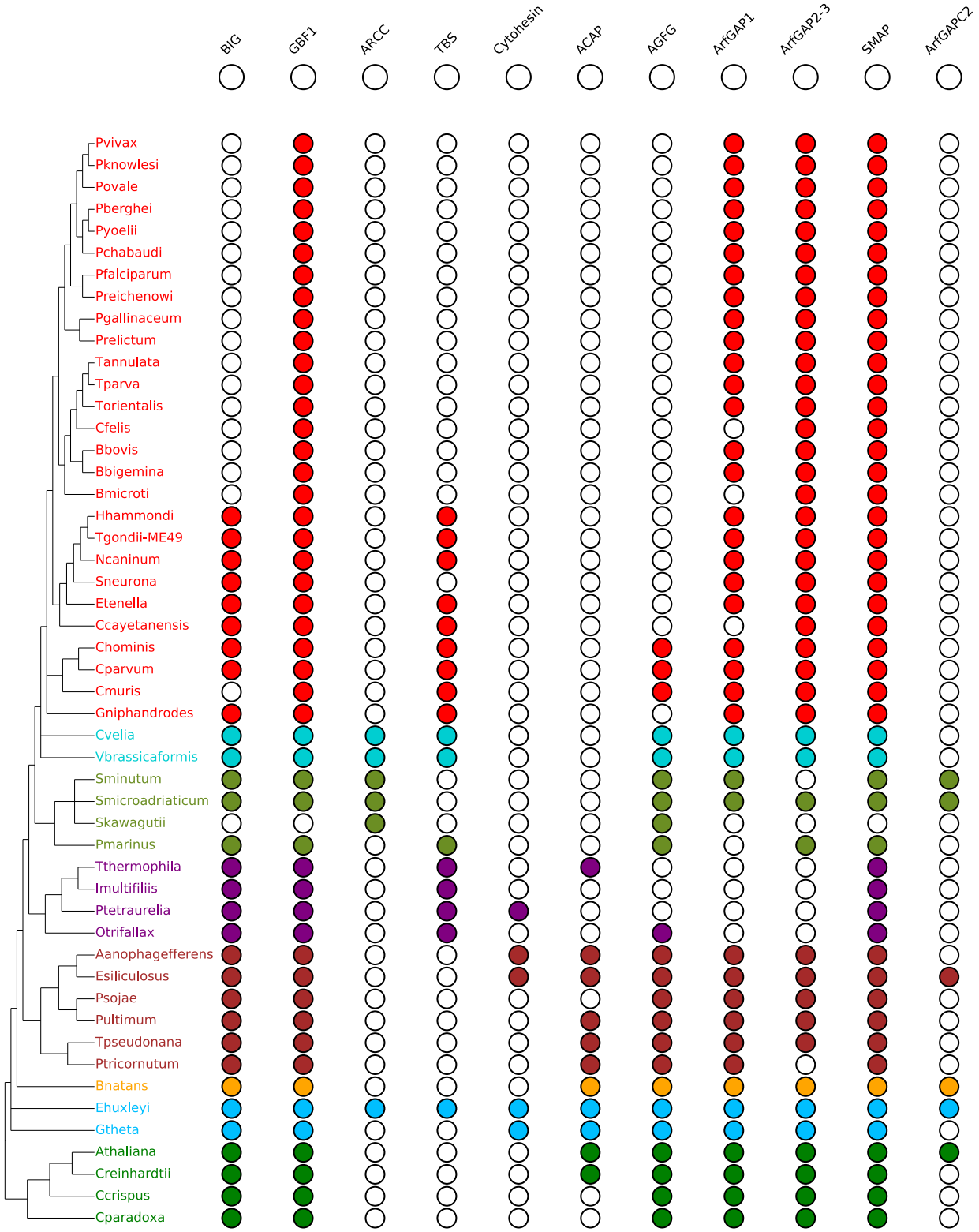
4.4.1 Arf regulators: ArfGEFs and ArfGAPs

Regulators for Arf family G proteins, the ArfGEFs and ArfGAPs, were introduced in section 1.3.1.1. Three ArfGEF families, all sharing the defining Sec7 domain, are presumed to be ancient: BIG, GBF1, and Cytohesin; an additional family, defined by the presence of N-terminal ankyrin repeats, ARCC, is found across eukaryotes but it is unclear whether this represents an ancient family or a case of convergent domain architecture[36]. Similarly, six ArfGAP families, all sharing a common ArfGAP domain, are presumed ancient: ArfGAP1, ArfGAP2/3, SMAP, ACAP, AGFG, and ArfGAPC2[32].

HMMs built from the Sec7 (Pfam ID PF01369) and ArfGAP (Pfam ID PF01412) domains were used to search for putative homologues in the dataset. A total of 200 putative ArfGEFs (Online Appendix Table 4.2) and 358 putative ArfGAPs (Online Appendix Table 4.3) were

Figure 4.2 ArfGEF and ArfGAP conservation

This figure shows the conservation of select ArfGEF/GAPs in study taxa. Each column represents an ArfGEF/GAP, as indicated at the top, while each row represents a different taxon. Filled circles represent the presence of a confidently inferred homologue while empty circles represent the failure to identify a homologue. Cladogram on the left represents the relationships between study taxa. Colour-coding is as follows: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids.


BIG
GBF1
ARCC
TBS
Cytohesin
ACAP
AGFG
ArfGAP1
ArfGAP2-3
SMAP
ArfGAPC2
Pvivax
Pknowlesi
Povale
Pberghei
Pyoelii
Pchabaudi
Pfalciparum
Preichenowi
Pgallinaceum
Prelictum
Tannulata
Tparva
Torientalis
Cfelis
Bbovis
Bbigemina
Bmicroti
Hhammondi
Tgondii-ME49
Ncaninum
Sneurona
Etenella
Ccayetanensis
Chominis
Cparvum
Cmuris
Gniphandrodes
Cvelia
Vbrassicaformis
Sminutum
Smicroadriaticum
Skawagutii
Pmarinus
Tthermophila
Imultifiliis
Ptetraurelia
Otrifallax
Aanophagefferens
Esiliculosus
Psojae
Pultimum
Tpseudonana
Ptricornutum
Bnatans
Ehuxleyi
Gtheta
Athaliana
Creinhardtii
Ccrispus
Cparadoxa

identified. A combination of BLAST searches, domain prediction, and phylogenetic analysis (materials and methods) classified the majority of sequences (Figure 4.2); 24 ArfGEFs and 45 ArfGAPs remained unclassified. As expected based on previous studies demonstrating patchy distribution of both families across eukaryotes[32,36], conservation among family members was low. In fact, only the ArfGEF GBF1 and ArfGAP SMAP were universally conserved, apart from in the dinoflagellate *Symbiodinium kawagutii*, whose genome is only ~80% complete[685] (Figure 4.2).

Surprisingly, piroplasmids and *Plasmodium spp.* encode only a single ArfGEF, GBF1, suggesting that they require a paucity of regulation for ARF activation (Figure 4.2). Coccidia and *Cryptosporidium spp.* additionally encode a homologue of the other large (~2000 aa) ArfGEF BIG, as well as homologues of TBS, which encodes both an ArfGEF and TBC RabGAP domain[686], and which will also be discussed in section 4.4.5.

Conservation among the ArfGAPs in Apicomplexa is better, with almost all taxa encoding homologues of ArfGAP1, ArfGAP2/3, and SMAP (Figure 4.2). *Cryptosporidium spp.* additionally encode AGFG; chromerids encode homologues of all of these ArfGAPs. Conversely, ACAP was restricted to outgroup taxa, suggesting it was lost during apicomplexan evolution, possibly in the myzozoan ancestor. ArfGAPC2 was the most sparsely conserved ArfGAP, consistent with previous reports[32], although its presence in *S. minutum* and *S. microadriaticum* as well as several outgroup taxa suggests it was lost numerous times within the diaphoretickes.

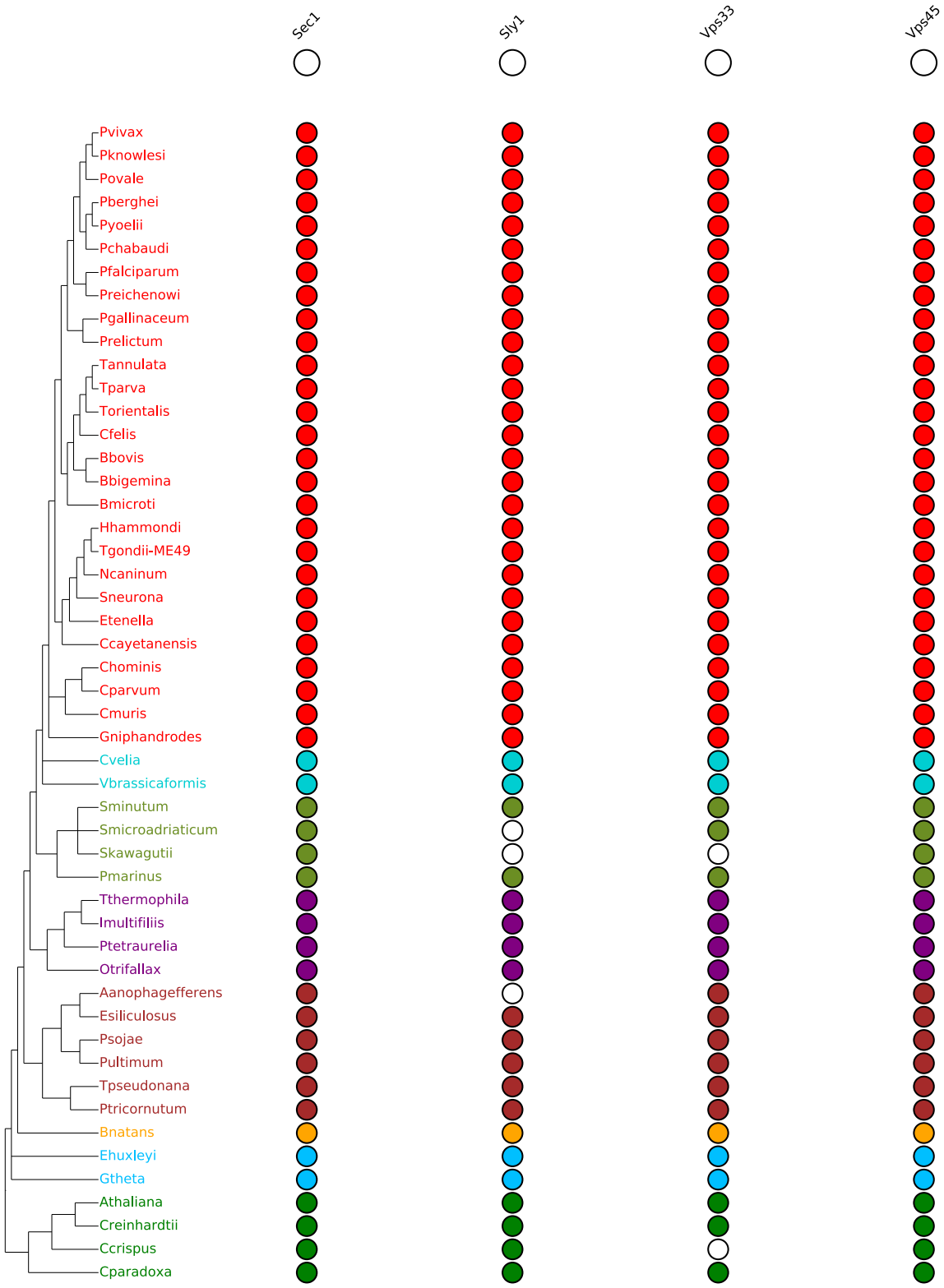
Despite detailed studies into both of these families, no obvious LSPs were identified, suggesting that ancestral ArfGEF and ArfGAP families can sufficiently regulate existing Arf family members in Apicomplexa.

4.4.2 SM proteins

SM proteins were introduced in section 1.3.1.6 as key regulators of SNARE function. Four SM proteins, Sec1, Sly1, Vps33, and Vps45, are known, and all are well-conserved across eukaryotes[126]. An HMM derived from the Sec1 domain (Pfam ID PF00995) was used to identify all putative SM protein homologues in the dataset (materials and methods). In total, 262 putative homologues were identified, all of which were classified as one of the four known SM proteins (Online Appendix Table 4.4). As a result, SM protein conservation was excellent, with the few instances of missing homologues possibly resulting from issues with gene prediction and/or genome assembly in specific taxa (Figure 4.3). No obvious LSPs were identified, suggesting that no expansions of SM proteins accompanied the observed expansions in SNARE complement within the Myzozoa (described in the next section).

Figure 4.3 SM protein conservation

This figure shows the conservation of select SM proteins in study taxa. Each column represents a SM protein paralogue, as indicated at the top, while each row represents a different taxon. Filled circles represent the presence of a confidently inferred homologue while empty circles represent the failure to identify a homologue. Cladogram on the left represents the relationships between study taxa. Colour-coding is as follows: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids.


Sec1
Sly1
Vps33
Vps45
Pvivax
Pknowlesi
Povale
Pberghei
Pyoelii
Pchabaudi
Pfalciparum
Preichenowi
Pgallinaceum
Prelictum
Tannulata
Tparva
Torientalis
Cfelis
Bbovis
Bbigemina
Bmicroti
Hhammondi
Tgondii-ME49
Ncaninum
Sneurona
Etenella
Ccayetanensis
Chominis
Cparvum
Cmuris
Gniphandrodes
Cvelia
Vbrassicaformis
Sminutum
Smicroadriaticum
Skawagutii
Pmarinus
Tthermophila
Imultifiliis
Ptetraurelia
Otrifallax
Aanophagefferens
Esiliculosus
Psojae
Pultimum
Tpseudonana
Ptricornutum
Bnatans
Ehuxleyi
Gtheta
Athaliana
Creinhardtii
Ccrispus
Cparadoxa

4.4.3 SNAREs

As discussed in section 1.3.1.6, the SNAREs are divided into four subfamilies, the Qa, Qb, Qc, and R families, and appear to be broadly conserved across eukaryotes[167,168]. For each SNARE subfamily, an HMM was built combining known homologues from *H. sapiens*, *A. thaliana*, and *S. cerevisiae*, and used to search the dataset for all putative homologues.

4.4.3.1 Qa-SNAREs

Qa-SNAREs include the Syntaxins (abbreviated here Stx), Stx5, Stx16, Stx17, Stx18, as well as those of the endosomes (StxE, in humans comprising Stx7 and Stx12 and herein referred to by Stx12) and the plasma membrane (SynPM, in humans comprising Stx1-4 and herein referred to by Stx2).

In total, 407 putative Qa-SNARE homologues were identified (Online Appendix Table 4.5), of which 353 could be confidently classified into pan-eukaryotic clades following initial phylogenetic analysis (Figure 4.4). Stx2, Stx5, Stx12, Stx16, and Stx18 were well-conserved across the dataset, with at least one member of each of the major taxonomic groups encoding a homologue of each Qa SNARE (Figure 4.4). This conservation extended to the Apicomplexa as well, with very few failures to identify a putative homologue outside of the apparent lack of Stx12 in *Cryptosporidium*. Conversely, some small expansions were noted in Apicomplexa, such as a duplication of Stx2 in the common ancestor of piroplasmids and *Plasmodium spp.* (Figure 4.4, Online Appendix Table 4.5).

Of the remaining 54 sequences that could not be confidently classified in initial phylogenetic analysis, 23 remained unclassified following further combined phylogenetic analysis. The other 31 were initially identified as additional Stx12 homologues but grouped separately within this clade (see below). These LSP sequences were combined with all pan-eukaryotic sequences and run in a large-scale phylogenetic analysis to confirm the initial classifications based on taxon subsets (Figure 4.5). Each group of putative homologues resolved into the expected monophyletic group with strong support: Stx2 (97 Shimodaira-Hasegawa approximate likelihood ratio test, SH-aLRT; 100 ultra-fast bootstrap, UF-bb), Stx5 (83 SH-aLRT, 98 UF-bb), a single group containing all Stx12-related sequences ("Stx12-like", 96.5 SH-aLRT, 96 UF-bb), Stx16 (97.9 SH-aLRT, 98 UF-bb), and Stx18 (99.3 SH-aLRT, 100 UF-bb). Additionally, Stx5 and Stx18 were each other's closest neighbor (96.6 SH-aLRT; 98 UF-bb), and both Stx2 and Stx12 clades grouped as well (92.9 SH-aLRT, 96 UF-bb); Stx16 branched in between these two groups.

To explore the nature of the putative LSP further, a robust phylogenetic analysis was carried out on all Stx2- and Stx12-like sequences (Figure 4.6). As expected, the additional

Figure 4.4 Qa-SNARE conservation

This figure shows the conservation of select Qa-SNAREs in study taxa. Each column represents a Qa-SNARE paralogue, as indicated at the top, while each row represents a different taxon. Filled circles represent the presence of a confidently inferred homologue while empty circles represent the failure to identify a homologue. Cladogram on the left represents the relationships between study taxa. Colour-coding is as follows: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids.

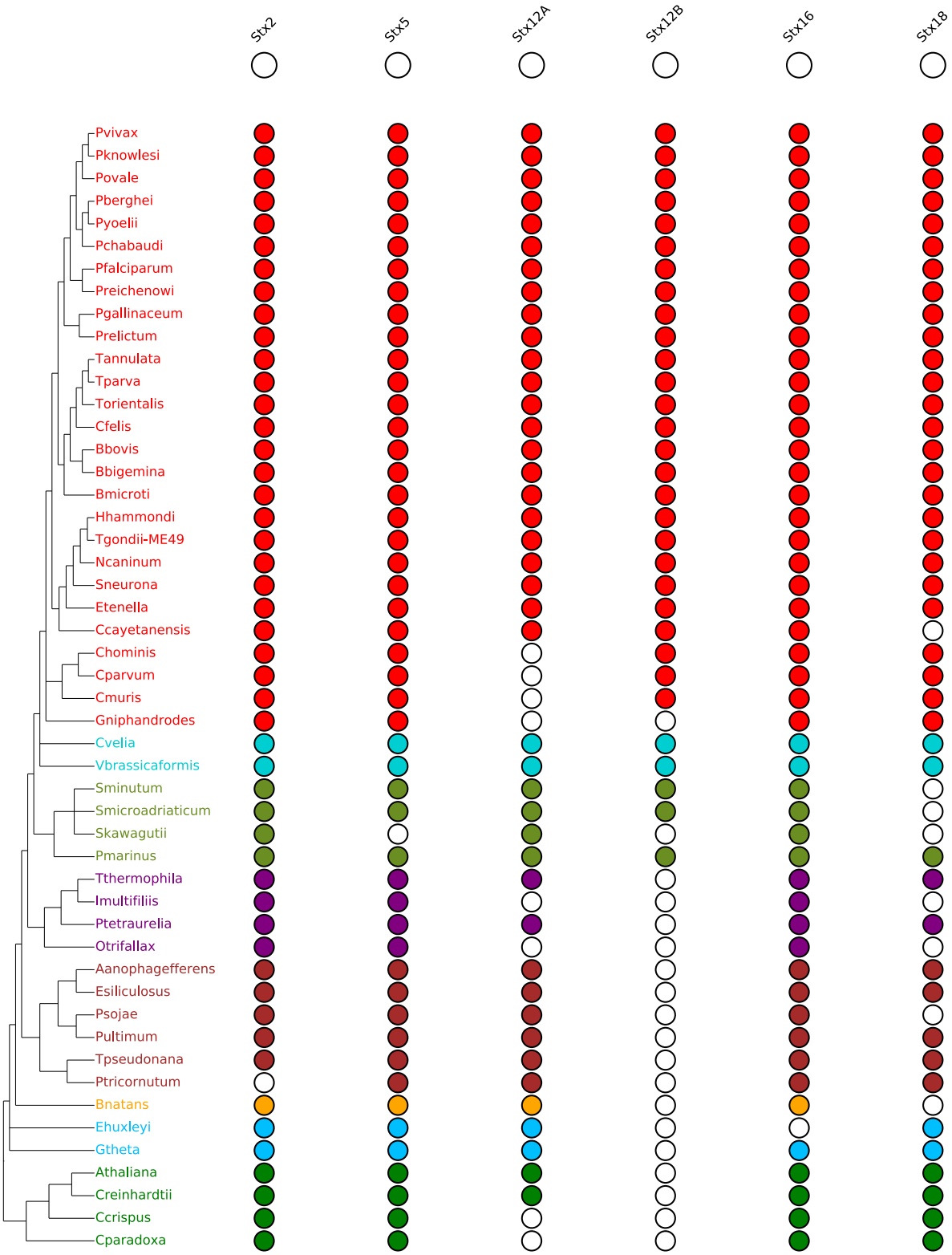

Stx2
Stx5
Stx12A
Stx12B
Stx16
Stx18
Pvivax
Pknowlesi
Povale
Pberghei
Pyoelii
Pchabaudi
Pfalciparum
Preichenowi
Pgallinaceum
Prelictum
Tannulata
Tparva
Torientalis
Cfelis
Bbovis
Bbigemina
Bmicroti
Hhammondi
Tgondii-ME49
Ncaninum
Sneurona
Etenella
Ccayetanensis
Chominis
Cparvum
Cmuris
Gniphandrodes
Cvelia
Vbrassicaformis
Sminutum
Smicroadriaticum
Skawagutii
Pmarinus
Tthermophila
Imultifiliis
Ptetraurelia
Otrifallax
Aanophagefferens
Esiliculosus
Psojae
Pultimum
Tpseudonana
Ptricornutum
Bnatans
Ehuxleyi
Gtheta
Athaliana
Creinhardtii
Ccrispus
Cparadoxa

Figure 4.5 Qa-SNARE phylogeny

This figure shows an IQ-TREE phylogeny obtained from a 362 taxon by 150 position alignment of Qa-SNAREs. Clades are collapsed for simplicity and support values listed for all nodes in order of Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT)/ultra-fast bootstrap (UF-bb). Reliable clades are those with >= 80% SH-aLRT and >= 95% UF-bb support. Scale bar represents number of substitutions per site.

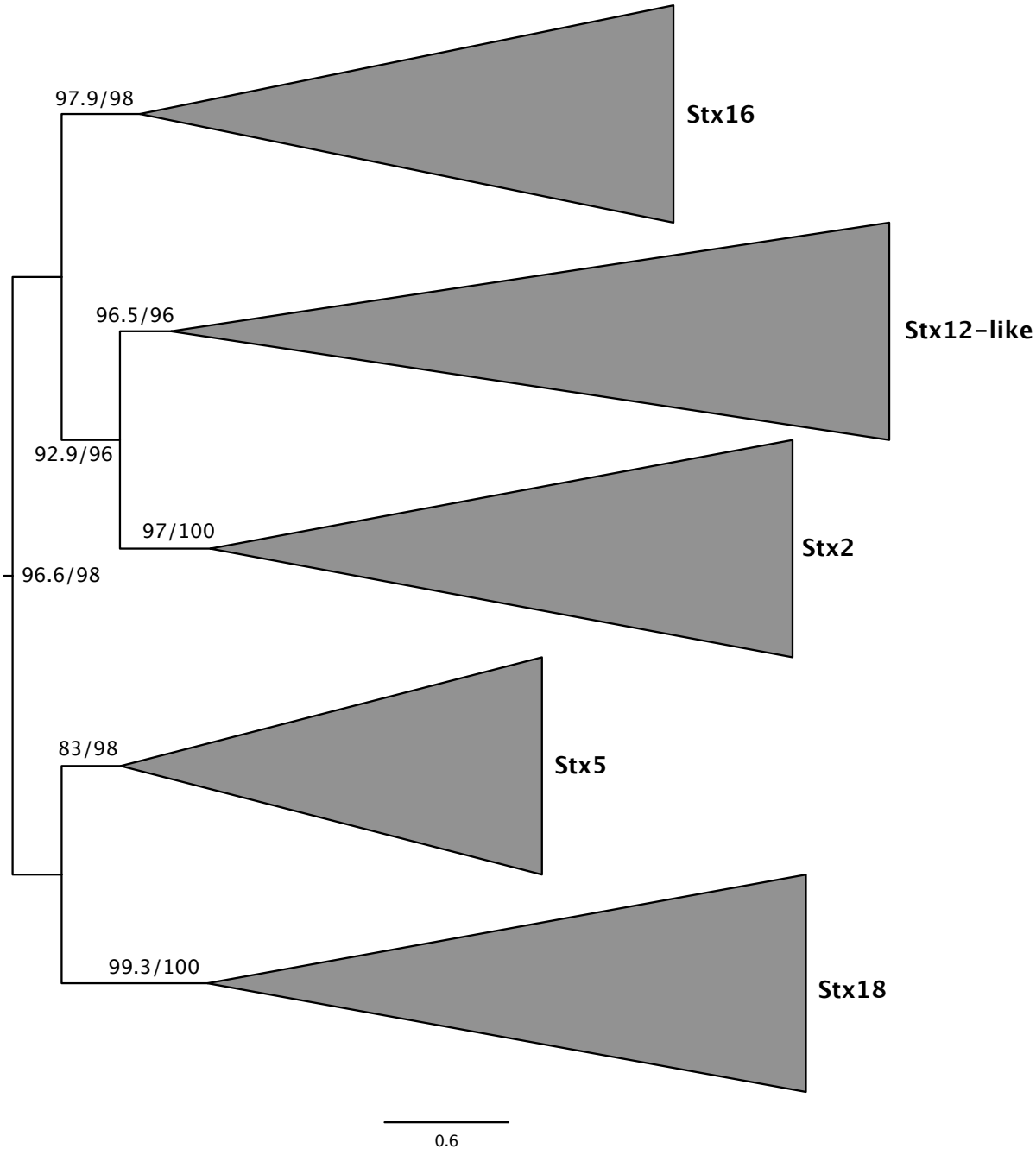
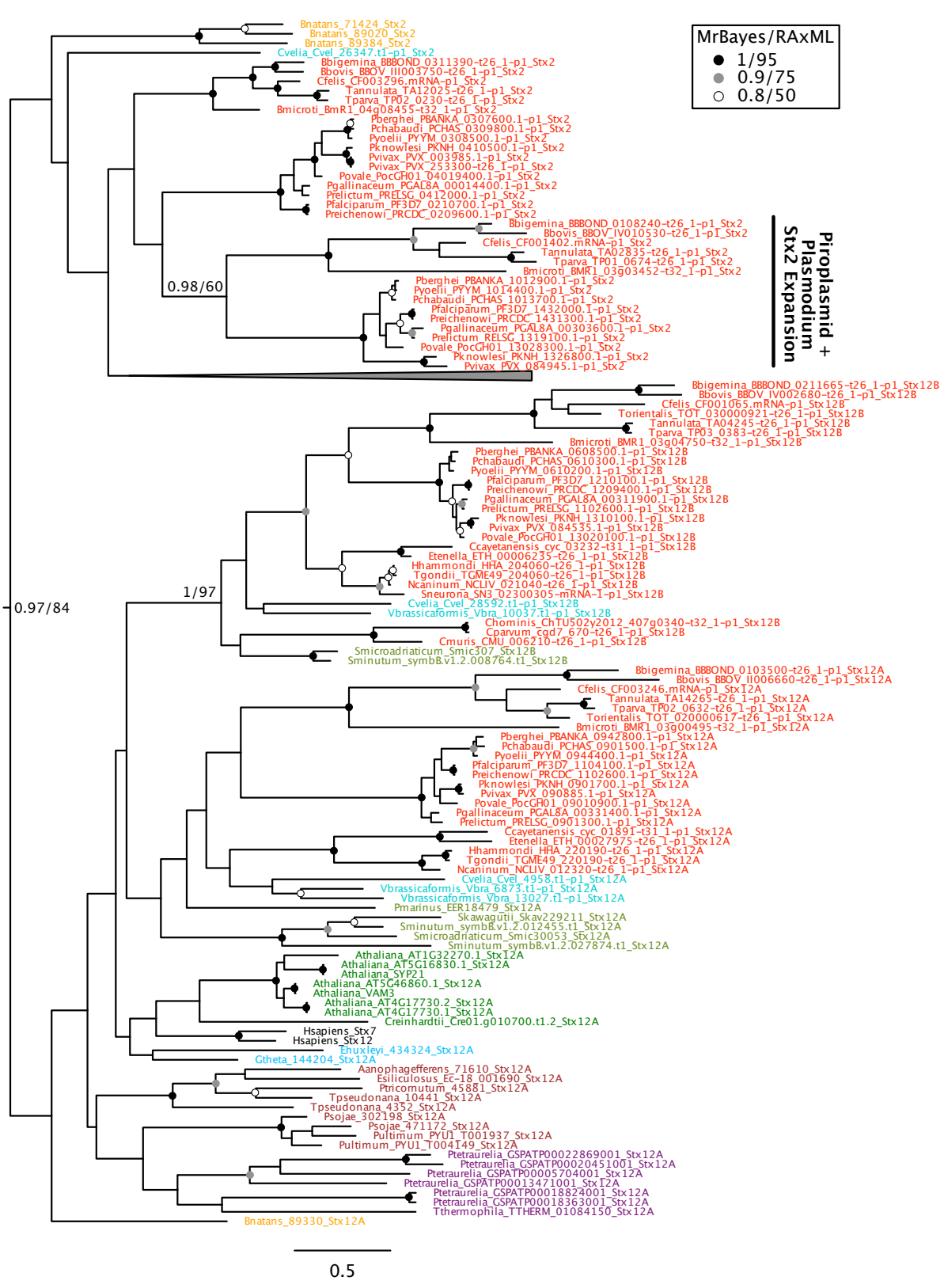

97.9/98
Stx16
96.5/96
Stx12-like
92.9/96
97/100
Stx2
96.6/98
83/98
Stx5
99.3/100
Stx18
0.6

Figure 4.6 Detailed phylogeny of Stx2 and Stx12

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 190 taxon by 150 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.


MrBayes/RAxML
1/95
0.9/75
0.8/50
Stx2
Piroplasmid + Plasmodium Stx2 Expansion
Stx12B
Stx12A
0.98/60
1/97
0.97/84
0.5
Bnatans_71424_Stx2
Bnatans_89020_Stx2
Bnatans_89384_Stx2
Cvelia_Cvel_26347.t1-p1_Stx2
Bbigemina_BBBOND_0311390-t26_1-p1_Stx2
Bbovis_BBOV_III003750-t26_1-p1_Stx2
Cfelis_CF003296.mRNA-p1_Stx2
Tannulata_TA12025-t26_1-p1_Stx2
Tparva_TP02_0230-t26_1-p1_Stx2
Bmicroti_BmR1_04g08455-t32_1-p1_Stx2
Pberghei_PBANKA_0307600.1-p1_Stx2
Pchabaudi_PCHAS_0309800.1-p1_Stx2
Pyoelii_PYYM_0308500.1-p1_Stx2
Pknowlesi_PKNH_0410500.1-p1_Stx2
Pvivax_PVX_003985.1-p1_Stx2
Pvivax_PVX_253300-t26_1-p1_Stx2
Povale_PocGH01_04019400.1-p1_Stx2
Pgallinaceum_PGAL8A_00014400.1-p1_Stx2
Prelictum_PRELSG_0412000.1-p1_Stx2
Pfalciparum_PF3D7_0210700.1-p1_Stx2
Preichenowi_PRCDC_0209600.1-p1_Stx2
Bbigemina_BBBOND_0108240-t26_1-p1_Stx2
Bbovis_BBOV_IV010530-t26_1-p1_Stx2
Cfelis_CF001402.mRNA-p1_Stx2
Tannulata_TA02835-t26_1-p1_Stx2
Tparva_TP01_0674-t26_1-p1_Stx2
Bmicroti_BMR1_03g03452-t32_1-p1_Stx2
Pberghei_PBANKA_1012900.1-p1_Stx2
Pyoelii_PYYM_1014400.1-p1_Stx2
Pchabaudi_PCHAS_1013700.1-p1_Stx2
Pfalciparum_PF3D7_1432000.1-p1_Stx2
Preichenowi_PRCDC_1431300.1-p1_Stx2
Pgallinaceum_PGAL8A_00303600.1-p1_Stx2
Prelictum_RELSG_1319100.1-p1_Stx2
Povale_PocGH01_13028300.1-p1_Stx2
Pknowlesi_PKNH_1326800.1-p1_Stx2
Pvivax_PVX_084945.1-p1_Stx2
Bbigemina_BBBOND_0211665-t26_1-p1_Stx12B
Bbovis_BBOV_IV002680-t26_1-p1_Stx12B
Cfelis_CF001065.mRNA-p1_Stx12B
Torientalis_TOT_030000921-t26_1-p1_Stx12B
Tannulata_TA04245-t26_1-p1_Stx12B
Tparva_TP03_0383-t26_1-p1_Stx12B
Bmicroti_BMR1_03g04750-t32_1-p1_Stx12B
Pberghei_PBANKA_0608500.1-p1_Stx12B
Pchabaudi_PCHAS_0610300.1-p1_Stx12B
Pyoelii_PYYM_0610200.1-p1_Stx12B
Pfalciparum_PF3D7_1210100.1-p1_Stx12B
Preichenowi_PRCDC_1209400.1-p1_Stx12B
Pgallinaceum_PGAL8A_00311900.1-p1_Stx12B
Prelictum_PRELSG_1102600.1-p1_Stx12B
Pknowlesi_PKNH_1310100.1-p1_Stx12B
Pvivax_PVX_084535.1-p1_Stx12B
Povale_PocGH01_13020100.1-p1_Stx12B
Ccayetanensis_cyc_03232-t31_1-p1_Stx12B
Etenella_ETH_00006235-t26_1-p1_Stx12B
Hhammondi_HHA_204060-t26_1-p1_Stx12B
Tgondii_TGME49_204060-t26_1-p1_Stx12B
Ncaninum_NCLIV_021040-t26_1-p1_Stx12B
Sneurona_SN3_02300305-mRNA-1-p1_Stx12B
Cvelia_Cvel_28592.t1-p1_Stx12B
Vbrassicaformis_Vbra_10037.t1-p1_Stx12B
Chominis_ChTU502y2012_407g0340-t32_1-p1_Stx12B
Cparvum_Cgd7_670-t26_1-p1_Stx12B
Cmuris_CMU_006210-t26_1-p1_Stx12B
Smicroadriaticum_Smic307_Stx12B
Sminutum_symbB.v1.2.008764.t1_Stx12B
Bbigemina_BBBOND_0103500-t26_1-p1_Stx12A
Bbovis_BBOV_II006660-t26_1-p1_Stx12A
Cfelis_CF003246.mRNA-p1_Stx12A
Tannulata_TA14265-t26_1-p1_Stx12A
Tparva_TP02_0632-t26_1-p1_Stx12A
Torientalis_TOT_020000617-t26_1-p1_Stx12A
Bmicroti_BMR1_03g00495-t32_1-p1_Stx12A
Pberghei_PBANKA_0942800.1-p1_Stx12A
Pchabaudi_PCHAS_0901500.1-p1_Stx12A
Pyoelii_PYYM_0944400.1-p1_Stx12A
Pfalciparum_PF3D7_1104100.1-p1_Stx12A
Preichenowi_PRCDC_1102600.1-p1_Stx12A
Pknowlesi_PKNH_0901700.1-p1_Stx12A
Pvivax_PVX_090885.1-p1_Stx12A
Povale_PocGH01_09010900.1-p1_Stx12A
Pgallinaceum_PGAL8A_00331400.1-p1_Stx12A
Prelictum_PRELSG_0901300.1-p1_Stx12A
Ccayetanensis_cyc_01891-t31_1-p1_Stx12A
Etenella_ETH_00027975-t26_1-p1_Stx12A
Hhammondi_HHA_220190-t26_1-p1_Stx12A
Tgondii_TGME49_220190-t26_1-p1_Stx12A
Ncaninum_NCLIV_012320-t26_1-p1_Stx12A
Cvelia_Cvel_4958.t1-p1_Stx12A
Vbrassicaformis_Vbra_6873.t1-p1_Stx12A
Vbrassicaformis_Vbra_13027.t1-p1_Stx12A
Pmarinus_EER18479_Stx12A
Skawagutii_Skav229211_Stx12A
Sminutum_symbB.v1.2.012455.t1_Stx12A
Smicroadriaticum_Smic30053_Stx12A
Sminutum_symbB.v1.2.027874.t1_Stx12A
Athaliana_AT1G32270.1_Stx12A
Athaliana_AT5G16830.1_Stx12A
Athaliana_SYP21
Athaliana_AT5G46860.1_Stx12A
Athaliana_VAM3
Athaliana_AT4G17730.2_Stx12A
Athaliana_AT4G17730.1_Stx12A
Creinhardtii_Cre01.g010700.t1.2_Stx12A
Hsapiens_Stx7
Hsapiens_Stx12
Ehuxleyi_434324_Stx12A
Gtheta_144204_Stx12A
Aanophagefferens_71610_Stx12A
Esiliculosus_Ec-18_001690_Stx12A
Ptricornutum_45881_Stx12A
Tpseudonana_10441_Stx12A
Tpseudonana_4352_Stx12A
Psojae_302198_Stx12A
Psojae_471172_Stx12A
Pultimum_PYU1_T001937_Stx12A
Pultimum_PYU1_T004149_Stx12A
Ptetraurelia_GSPATP00022869001_Stx12A
Ptetraurelia_GSPATP00020451001_Stx12A
Ptetraurelia_GSPATP00005704001_Stx12A
Ptetraurelia_GSPATP00013471001_Stx12A
Ptetraurelia_GSPATP00018824001_Stx12A
Ptetraurelia_GSPATP00018363001_Stx12A
Tthermophila_TTHERM_01084150_Stx12A
Bnatans_89330_Stx12A

Stx2 homologues in piroplasmids and *Plasmodium spp.* grouped within the larger Stx2 clade (0.98 Bayesian posterior probability, pp; 60 bootstrap support, bb). This suggests that Stx2 duplicated in the common ancestor of these two groups. Additionally, a clade of Stx12-related sequences corresponding to the previously identified LSP was recovered with strong support (1 pp, 97 bb, Figure 4.6). This group of sequences branches within the larger pan-eukaryotic Stx12 clade and are referred to as Stx12B to delineate it from the pan-eukaryotic Stx12A clade (Figures 4.4, 4.6). Stx12B is conserved among Apicomplexa, chromerids, and dinoflagellates, suggesting that it arose in the myzozoan ancestor.

Hence, although both Stx2 and Stx12 have expanded within the Apicomplexa, only Stx12B represents a true LSP, as defined in section 4.3, within the Q-SNARE family.

4.4.3.2 Qb- and Qc-SNAREs

Qb-SNAREs include the members GOSR1, GOSR2, NPSN11, Sec20, and Vti1, while the Qc-SNAREs include Bet1, Stx6, Stx8, SYP71, and Use1. In total, 384 putative Qb-SNARE, and 317 putative Qc-SNARE, homologues were identified (Online Appendix Tables 4.6 and 4.7). Following initial phylogenetic analyses, 332 Qb-SNARE and 256 Qc-SNARE homologues could be confidently classified as members of pan-eukaryotic clades (Figure 4.7).

Similar to the Qa-SNAREs, conservation of both subfamilies was good across the dataset, although some members had a patchy distribution within certain taxonomic groups (Figure 4.7). Among the Apicomplexa, notable absences include that of GOSR2 in the piroplasmids with the sole exception of the basal taxon *Babesia microti,* as well as the absence of Use1 homologues in all cryptosporidians. Although expansions were infrequent, most taxa possessed multiple Vti1 homologues; for example, *T. gondii* encodes three putative Vti1 homologues (Online Appendix Table 4.6).

Of the remaining unclassified sequences, further phylogenetic analysis revealed that 36 sequences possess two SNARE domains, which were considered putative Qbc SNAREs[168], and investigated further below. The remaining sequences, 16 Qb- and 25 Qc-SNAREs, remained unclassified.

Comparisons of the putative Qbc proteins to both Qb- and Qc-SNAREs within the dataset demonstrated that the N-terminal domain was most similar to Qb-SNAREs while the C-terminal domain was most similar to Qc-SNAREs, as expected. Hence, and to confirm all initial classifications, the separate N-terminal Qb- and C-terminal Qc-domains of each Qbc protein were combined with all pan-eukaryotic sequences from the relevant subfamily and run in large-scale phylogenetic analyses (Figures 4.8 and 4.9). Moderate to strong support for all clades was observed; although Vti1 homologues did group together in initial analyses,

Figure 4.7 Qb- and Qc-SNARE conservation

This figure shows the conservation of select Qb- and Qc-SNAREs in study taxa. Each column represents a Qb- or Qc-SNARE paralogue (or Qbc-SNARE), as indicated at the top, while each row represents a different taxon. Filled circles represent the presence of a confidently inferred homologue while empty circles represent the failure to identify a homologue. Cladogram on the left represents the relationships between study taxa. Colour-coding is as follows: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids.

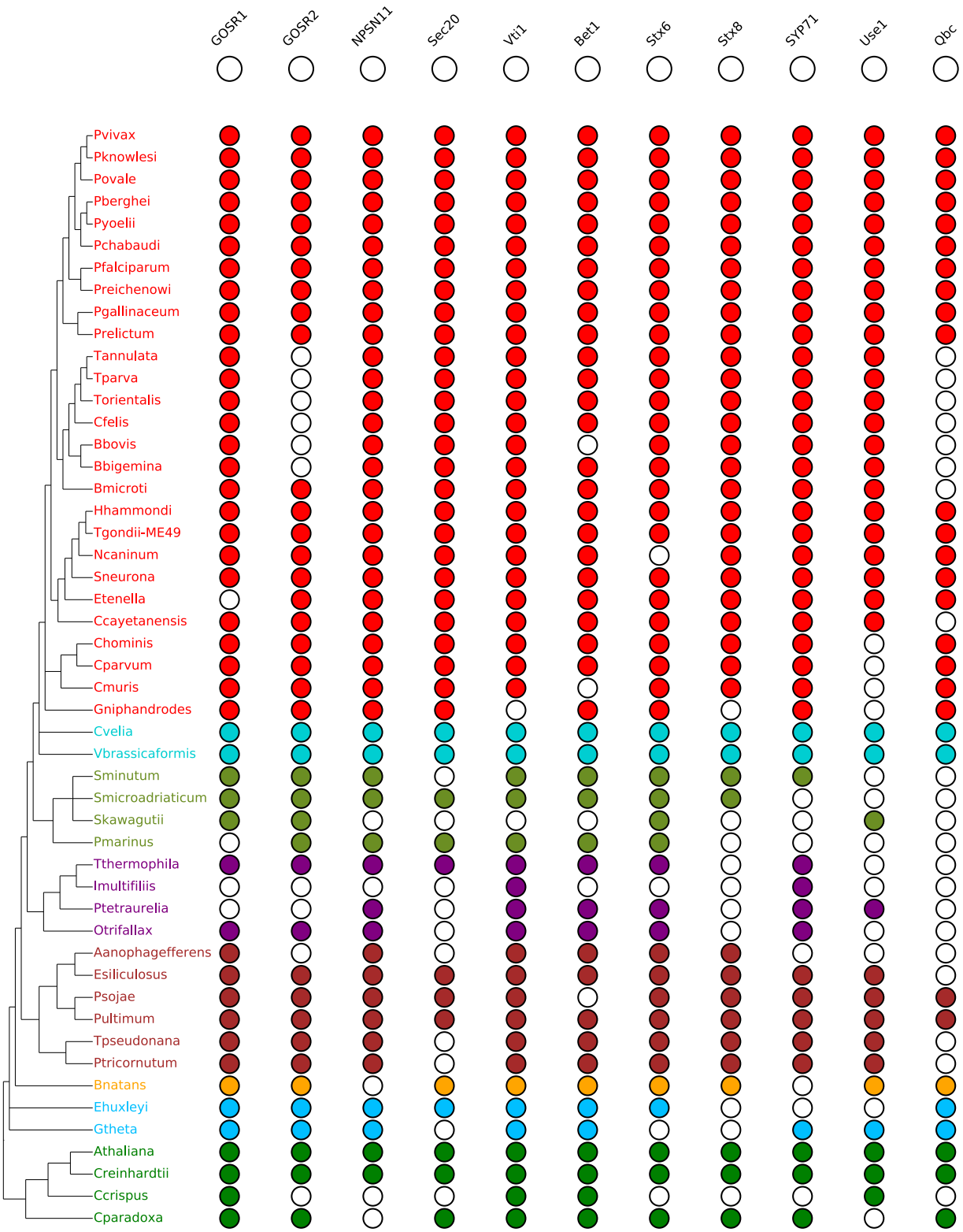

GOSR1
GOSR2
NPSN11
Sec20
Vti1
Bet1
Stx6
Stx8
SYP71
Use1
Qbc
Pvivax
Pknowlesi
Povale
Pberghei
Pyoelii
Pchabaudi
Pfalciparum
Preichenowi
Pgallinaceum
Prelictum
Tannulata
Tparva
Torientalis
Cfelis
Bbovis
Bbigemina
Bmicroti
Hhammondi
Tgondii-ME49
Ncaninum
Sneurona
Etenella
Ccayetanensis
Chominis
Cparvum
Cmuris
Gniphandrodes
Cvelia
Vbrassicaformis
Sminutum
Smicroadriaticum
Skawagutii
Pmarinus
Tthermophila
Imultifiliis
Ptetraurelia
Otrifallax
Aanophagefferens
Esiliculosus
Psojae
Pultimum
Tpseudonana
Ptricornutum
Bnatans
Ehuxleyi
Gtheta
Athaliana
Creinhardtii
Ccrispus
Cparadoxa

Figure 4.8 Qb-SNARE phylogeny

This figure shows an IQ-TREE phylogeny obtained from a 228 taxon by 128 position alignment of Qb-SNAREs. Clades are collapsed for simplicity and support values listed for all nodes in order of Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT)/ultra-fast bootstrap (UF-bb). Reliable clades are those with >= 80% SH-aLRT and >= 95% UF-bb support. Scale bar represents number of substitutions per site.

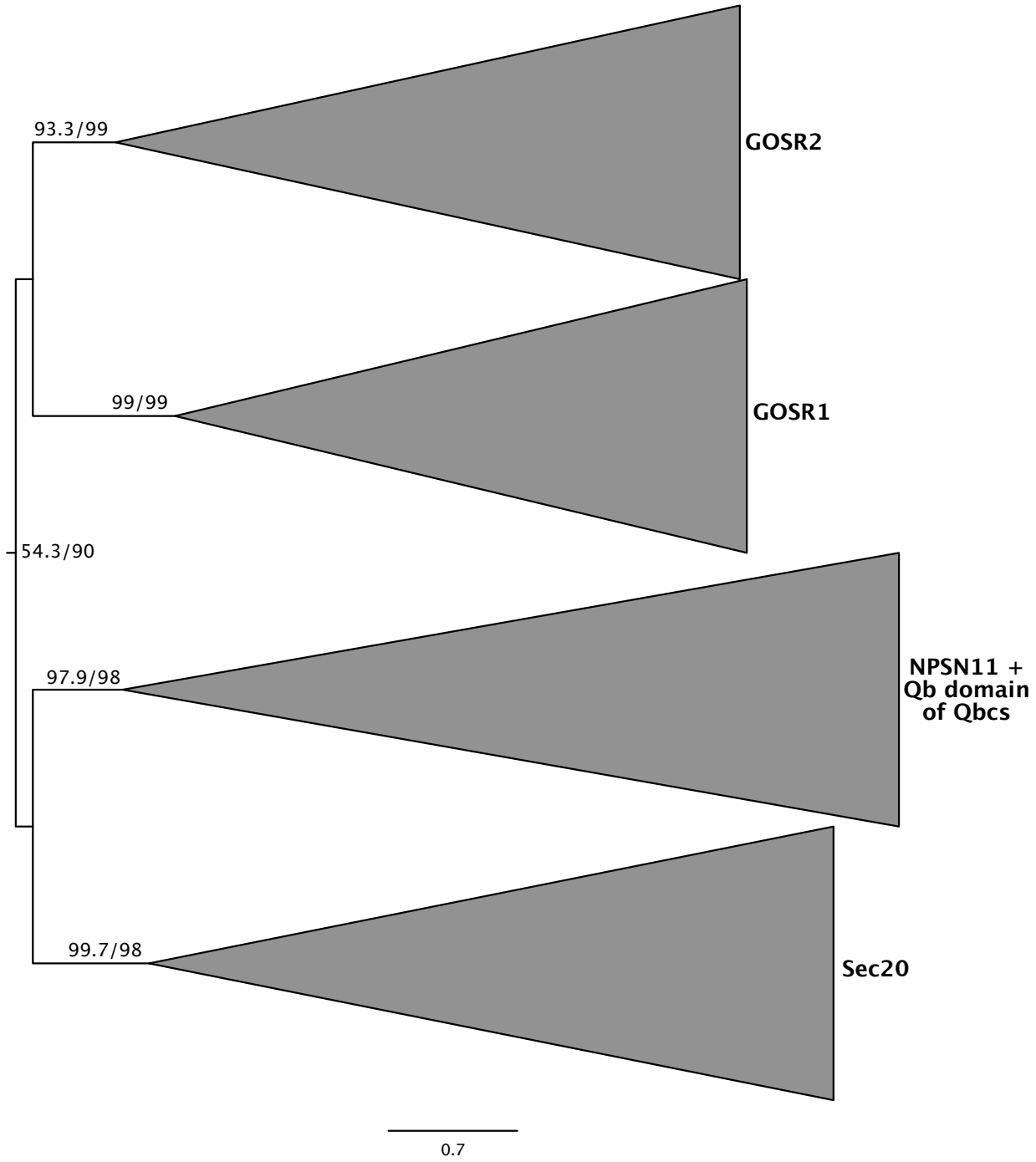
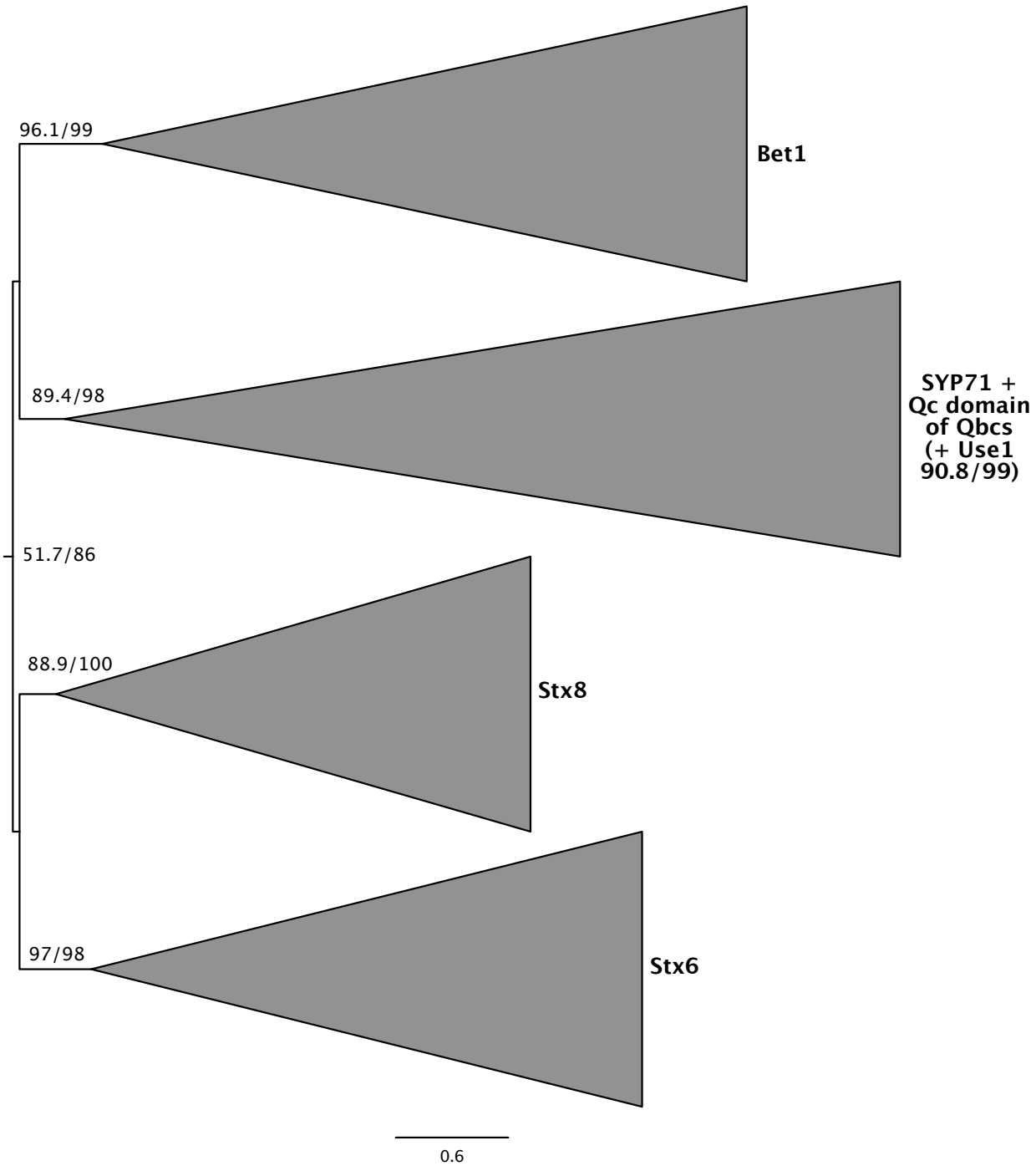

93.3/99
GOSR2
99/99
GOSR1
54.3/90
97.9/98
NPSN11 +
Qb domain
of Qbcs
99.7/98
Sec20
0.7

Figure 4.9 Qc-SNARE phylogeny

This figure shows an IQ-TREE phylogeny obtained from a 272 taxon by 164 position alignment of Qc-SNAREs. Clades are collapsed for simplicity and support values listed for all nodes in order of Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT)/ultra-fast bootstrap (UF-bb). Reliable clades are those with >= 80% SH-aLRT and >= 95% UF-bb support. Scale bar represents number of substitutions per site.


96.1/99
Bet1
89.4/98
SYP71 +
Qc domain
of Qbcs
(+ Use1
90.8/99)
51.7/86
88.9/100
Stx8
97/98
Stx6
0.6

bootstrap support for this clade was relatively low (~50 in various analyses using RAxML), and so this clade was not included in the final large-scale analysis depicted in Figure 4.8. Separate analyses aimed at resolving the internal structure of the Vti1 clade to understand the pattern of the aforementioned expansions also suffered from unstable topologies and low bootstrap support; further work is required to understand the nature of Vti1 expansion in these taxa. The Qb domain of Qbc homologues strongly grouped together with NPSN11 (97.9 SH-aLRT, 98 UF-bb, Figure 4.8), suggesting, as previously described[168], that the Qb domain of Qbc proteins is derived from NPSN11. Similarly, the Qc domain of Qbc homologues grouped together with SYP71 (89.4 SH-aLRT, 98 UF-bb, Figure 4.9), although Use1 also branched from within this clade (90.8 SH-aLRT, 99 UF-bb). In other analyses though, Use1 branched separately, suggesting that the Qc domain of Qbc proteins is derived from SYP71.

Qbc homologues were the least well-conserved among the Qb- and Qc-SNAREs, being present only in select outgroup taxa, though not in ciliates or dinoflagellates (Figure 4.7), and in all apicomplexans except piroplasmids. It is clear from the pattern of conservation that the loss in piroplasmids is secondary; the significance of this observation awaits functional evaluation of the function of Qbc SNARE homologues in Apicomplexa.

4.4.3.3 R-SNAREs

R-SNAREs include Sec22 and Ykt6, along with the VAMP proteins, including the well-conserved VAMP7 and other VAMP sequences (see below). Initial homology searching identified 382 putative R-SNARE homologues (Online Appendix Table 4.8). Following initial classification, only 212 sequences could be confidently classified as belonging within a pan-eukaryotic clade. Some, like the VAMP7- and Stx6-related LSPs described below, clearly grouped within these larger clades, while others, totaling 97 sequences**,** appeared related to other VAMP sequences in *H. sapiens* but lacked support for a clear association with any single paralogue (referred to as "VAMPX", Figure 4.10); 27 sequences remained unclassified following all analyses.

Similar to the Q-SNAREs, R-SNAREs were found to be well-conserved across the dataset, with at least one homologue of the Sec22, VAMP7, and Ykt6 families encoded in almost all taxa (Figure 4.10). Additional R-SNARE homologues, which were classified here simply as "VAMPX", were also present in almost all taxa (Figure 4.10); additional analyses will be required to resolve these sequences.

As mentioned above, large expansions were noted in both VAMP7 and Ykt6 (Online Appendix Table 4.8) during initial phylogenetic analyses. Additional robust analyses of each group, including only Sec22 as an outgroup, allowed for detailed investigation of their internal

Figure 4.10 R-SNARE conservation

This figure shows the conservation of select R-SNAREs in study taxa. Each column represents a R-SNARE paralogue, as indicated at the top, while each row represents a different taxon. Filled circles represent the presence of a confidently inferred homologue while empty circles represent the failure to identify a homologue. Cladogram on the left represents the relationships between study taxa. Colour-coding is as follows: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids.

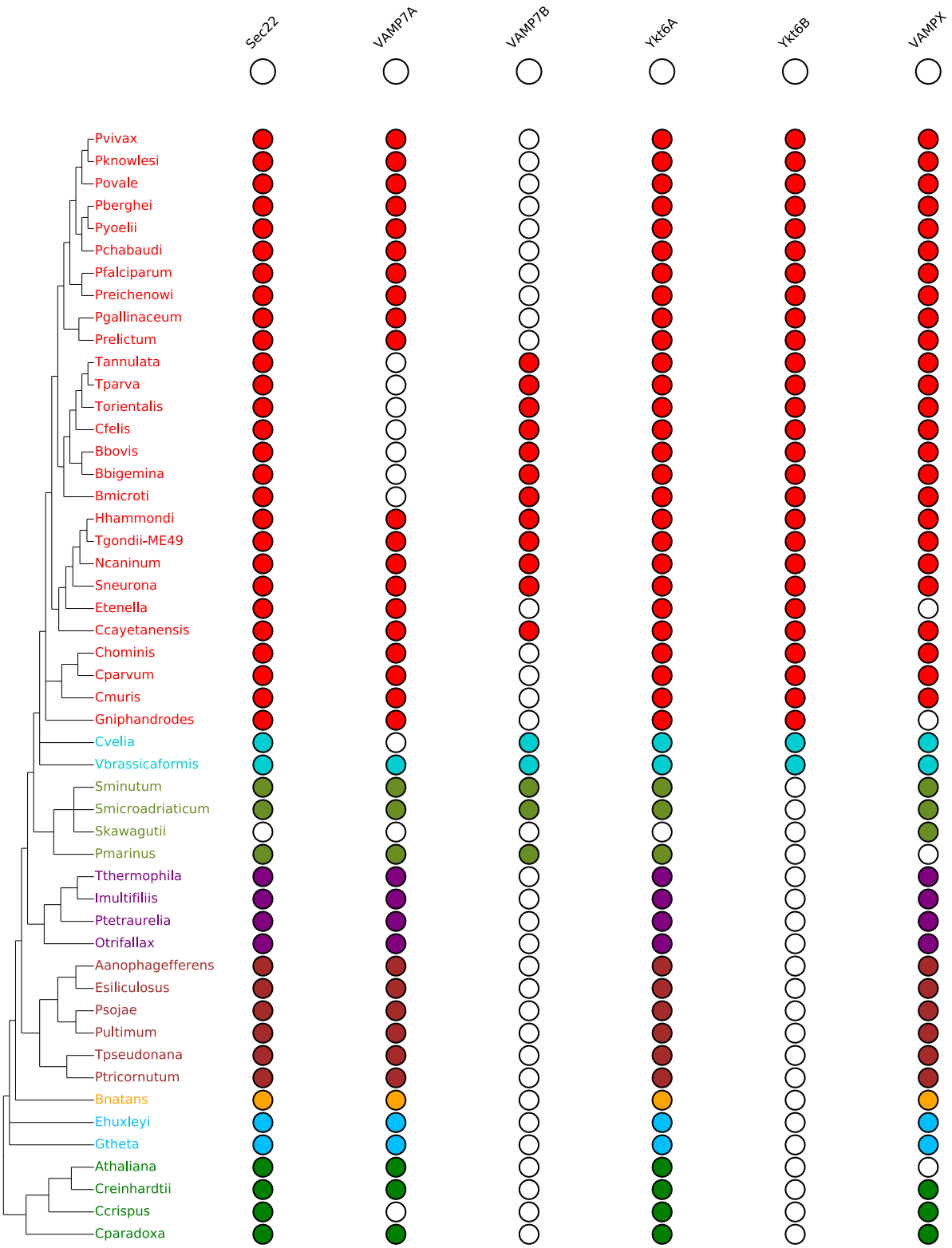

Sec22
VAMP7A
VAMP7B
Ykt6A
Ykt6B
VAMPX
Pvivax
Pknowlesi
Povale
Pberghei
Pyoelii
Pchabaudi
Pfalciparum
Preichenowi
Pgallinaceum
Prelictum
Tannulata
Tparva
Torientalis
Cfelis
Bbovis
Bbigemina
Bmicroti
Hhammondi
Tgondii-ME49
Ncaninum
Sneurona
Etenella
Ccayetanensis
Chominis
Cparvum
Cmuris
Gniphandrodes
Cvelia
Vbrassicaformis
Sminutum
Smicroadriaticum
Skawagutii
Pmarinus
Tthermophila
Imultifiliis
Ptetraurelia
Otrifallax
Aanophagefferens
Esiliculosus
Psojae
Pultimum
Tpseudonana
Ptricornutum
Bnatans
Ehuxleyi
Gtheta
Athaliana
Creinhardtii
Ccrispus
Cparadoxa

Figure 4.11 Detailed phylogeny of VAMP sequences

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 151 taxon by 158 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.

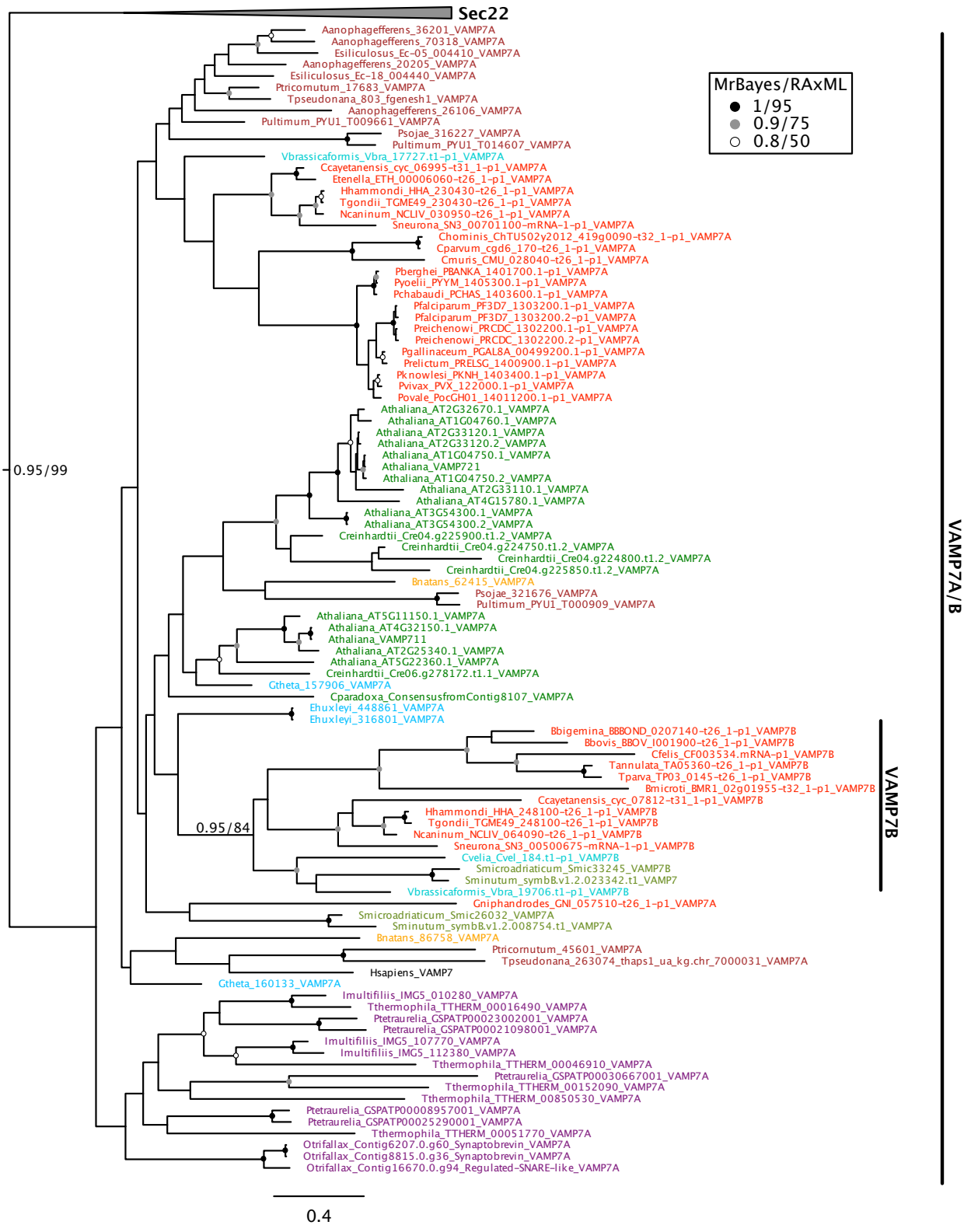
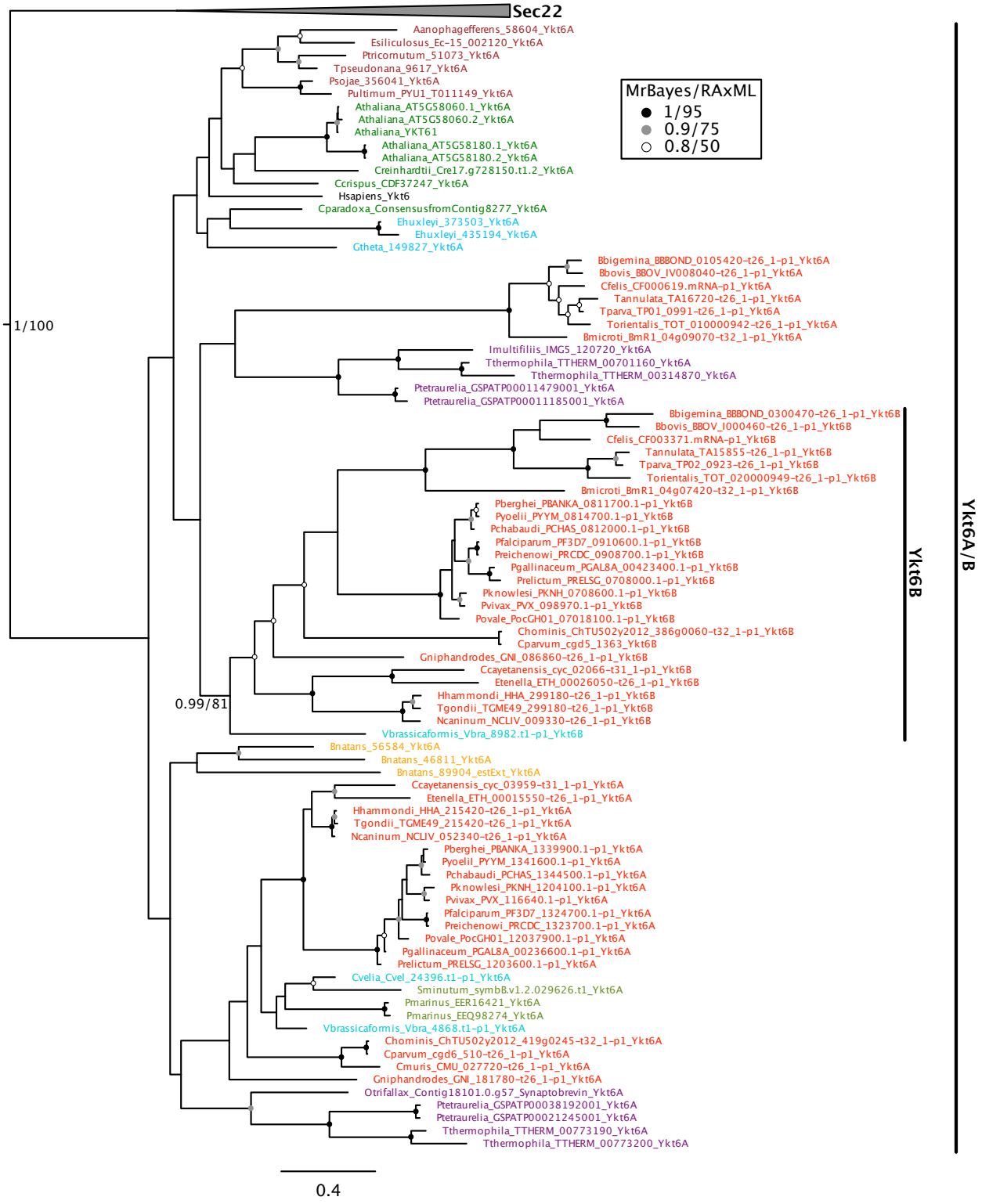

Sec22
MrBayes/RAxML
1/95
0.9/75
0.8/50
0.95/99
0.95/84
VAMP7A/B
VAMP7B
0.4
Aanophagefferens_36201_VAMP7A
Aanophagefferens_70318_VAMP7A
Esiliculosus_Ec-05_004410_VAMP7A
Aanophagefferens_20205_VAMP7A
Esiliculosus_Ec-18_004440_VAMP7A
Ptricornutum_17683_VAMP7A
Tpseudonana_803_fgenesh1_VAMP7A
Aanophagefferens_26106_VAMP7A
Pultimum_PYU1_T009661_VAMP7A
Psojae_316227_VAMP7A
Pultimum_PYU1_T014607_VAMP7A
Vbrassicaformis_Vbra_17727.t1-p1_VAMP7A
Ccayetanensis_cyc_06995-t31_1-p1_VAMP7A
Etenella_ETH_00006060-t26_1-p1_VAMP7A
Hhammondi_HHA_230430-t26_1-p1_VAMP7A
Tgondii_TGME49_230430-t26_1-p1_VAMP7A
Ncaninum_NCLIV_030950-t26_1-p1_VAMP7A
Sneurona_SN3_00701100-mRNA-1-p1_VAMP7A
Chominis_ChTU502y2012_419g0090-t32_1-p1_VAMP7A
Cparvum_cgd6_170-t26_1-p1_VAMP7A
Cmuris_CMU_028040-t26_1-p1_VAMP7A
Pberghei_PBANKA_1401700.1-p1_VAMP7A
Pyoelii_PYYM_1405300.1-p1_VAMP7A
Pchabaudi_PCHAS_1403600.1-p1_VAMP7A
Pfalciparum_PF3D7_1303200.1-p1_VAMP7A
Pfalciparum_PF3D7_1303200.2-p1_VAMP7A
Preichenowi_PRCDC_1302200.1-p1_VAMP7A
Preichenowi_PRCDC_1302200.2-p1_VAMP7A
Pgallinaceum_PGAL8A_00499200.1-p1_VAMP7A
Prelictum_PRELSG_1400900.1-p1_VAMP7A
Pknowlesi_PKNH_1403400.1-p1_VAMP7A
Pvivax_PVX_122000.1-p1_VAMP7A
Povale_PocGH01_14011200.1-p1_VAMP7A
Athaliana_AT2G32670.1_VAMP7A
Athaliana_AT1G04760.1_VAMP7A
Athaliana_AT2G33120.1_VAMP7A
Athaliana_AT2G33120.2_VAMP7A
Athaliana_AT1G04750.1_VAMP7A
Athaliana_VAMP721
Athaliana_AT1G04750.2_VAMP7A
Athaliana_AT2G33110.1_VAMP7A
Athaliana_AT4G15780.1_VAMP7A
Athaliana_AT3G54300.1_VAMP7A
Athaliana_AT3G54300.2_VAMP7A
Creinhardtii_Cre04.g225900.t1.2_VAMP7A
Creinhardtii_Cre04.g224750.t1.2_VAMP7A
Creinhardtii_Cre04.g224800.t1.2_VAMP7A
Creinhardtii_Cre04.g225850.t1.2_VAMP7A
Bnatans_62415_VAMP7A
Psojae_321676_VAMP7A
Pultimum_PYU1_T000909_VAMP7A
Athaliana_AT5G11150.1_VAMP7A
Athaliana_AT4G32150.1_VAMP7A
Athaliana_VAMP711
Athaliana_AT2G25340.1_VAMP7A
Athaliana_AT5G22360.1_VAMP7A
Creinhardtii_Cre06.g278172.t1.1_VAMP7A
Gtheta_157906_VAMP7A
Cparadoxa_ConsensusfromContig8107_VAMP7A
Ehuxleyi_448861_VAMP7A
Ehuxleyi_316801_VAMP7A
Bbigemina_BBBOND_0207140-t26_1-p1_VAMP7B
Bbovis_BBOV_I001900-t26_1-p1_VAMP7B
Cfelis_CF003534.mRNA-p1_VAMP7B
Tannulata_TA05360-t26_1-p1_VAMP7B
Tparva_TP03_0145-t26_1-p1_VAMP7B
Bmicroti_BMR1_02g01955-t32_1-p1_VAMP7B
Ccayetanensis_cyc_07812-t31_1-p1_VAMP7B
Hhammondi_HHA_248100-t26_1-p1_VAMP7B
Tgondii_TGME49_248100-t26_1-p1_VAMP7B
Ncaninum_NCLIV_064090-t26_1-p1_VAMP7B
Sneurona_SN3_00500675-mRNA-1-p1_VAMP7B
Cvelia_Cvel_184.t1-p1_VAMP7B
Smicroadriaticum_Smic33245_VAMP7B
Sminutum_symbB.v1.2.023342.t1_VAMP7
Vbrassicaformis_Vbra_19706.t1-p1_VAMP7B
Gniphandrodes_GNI_057510-t26_1-p1_VAMP7A
Smicroadriaticum_Smic26032_VAMP7A
Sminutum_symbB.v1.2.008754.t1_VAMP7A
Bnatans_86758_VAMP7A
Ptricornutum_45601_VAMP7A
Tpseudonana_263074_thaps1_ua_kg.chr_7000031_VAMP7A
Hsapiens_VAMP7
Gtheta_160133_VAMP7A
Imultifiliis_IMG5_010280_VAMP7A
Tthermophila_TTHERM_00016490_VAMP7A
Ptetraurelia_GSPATP00023002001_VAMP7A
Ptetraurelia_GSPATP00021098001_VAMP7A
Imultifiliis_IMG5_107770_VAMP7A
Imultifiliis_IMG5_112380_VAMP7A
Tthermophila_TTHERM_00046910_VAMP7A
Ptetraurelia_GSPATP00030667001_VAMP7A
Tthermophila_TTHERM_00152090_VAMP7A
Tthermophila_TTHERM_00850530_VAMP7A
Ptetraurelia_GSPATP00008957001_VAMP7A
Ptetraurelia_GSPATP00025290001_VAMP7A
Tthermophila_TTHERM_00051770_VAMP7A
Otrifallax_Contig6207.0.g60_Synaptobrevin_VAMP7A
Otrifallax_Contig8815.0.g36_Synaptobrevin_VAMP7A
Otrifallax_Contig16670.0.g94_Regulated-SNARE-like_VAMP7A

Figure 4.12 Detailed phylogeny of Ykt6 sequences

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 139 taxon by 194 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.


Sec22
Aanophagefferens_58604_Ykt6A
Esiliculosus_Ec-15_002120_Ykt6A
Ptricornutum_51073_Ykt6A
Tpseudonana_9617_Ykt6A
Psojae_356041_Ykt6A
Pultimum_PYU1_T011149_Ykt6A
Athaliana_AT5G58060.1_Ykt6A
Athaliana_AT5G58060.2_Ykt6A
Athaliana_YKT61
Athaliana_AT5G58180.1_Ykt6A
Athaliana_AT5G58180.2_Ykt6A
Creinhardtii_Cre17.g728150.t1.2_Ykt6A
Ccrispus_CDF37247_Ykt6A
Hsapiens_Ykt6
Cparadoxa_ConsensusfromContig8277_Ykt6A
Ehuxleyi_373503_Ykt6A
Ehuxleyi_435194_Ykt6A
Gtheta_149827_Ykt6A
MrBayes/RAxML
1/95
0.9/75
0.8/50
1/100
Bbigemina_BBBOND_0105420-t26_1-p1_Ykt6A
Bbovis_BBOV_IV008040-t26_1-p1_Ykt6A
Cfelis_CF000619.mRNA-p1_Ykt6A
Tannulata_TA16720-t26_1-p1_Ykt6A
Tparva_TP01_0991-t26_1-p1_Ykt6A
Torientalis_TOT_010000942-t26_1-p1_Ykt6A
Bmicroti_BmR1_04g09070-t32_1-p1_Ykt6A
Imultifiliis_IMG5_120720_Ykt6A
Tthermophila_TTHERM_00701160_Ykt6A
Tthermophila_TTHERM_00314870_Ykt6A
Ptetraurelia_GSPATP00011479001_Ykt6A
Ptetraurelia_GSPATP00011185001_Ykt6A
Bbigemina_BBBOND_0300470-t26_1-p1_Ykt6B
Bbovis_BBOV_I000460-t26_1-p1_Ykt6B
Cfelis_CF003371.mRNA-p1_Ykt6B
Tannulata_TA15855-t26_1-p1_Ykt6B
Tparva_TP02_0923-t26_1-p1_Ykt6B
Torientalis_TOT_020000949-t26_1-p1_Ykt6B
Bmicroti_BmR1_04g07420-t32_1-p1_Ykt6B
Pberghei_PBANKA_0811700.1-p1_Ykt6B
Pyoelii_PYYM_0814700.1-p1_Ykt6B
Pchabaudi_PCHAS_0812000.1-p1_Ykt6B
Pfalciparum_PF3D7_0910600.1-p1_Ykt6B
Preichenowi_PRCDC_0908700.1-p1_Ykt6B
Pgallinaceum_PGAL8A_00423400.1-p1_Ykt6B
Prelictum_PRELSG_0708000.1-p1_Ykt6B
Pknowlesi_PKNH_0708600.1-p1_Ykt6B
Pvivax_PVX_098970.1-p1_Ykt6B
Povale_PocGH01_07018100.1-p1_Ykt6B
Chominis_ChTU502y2012_386g0060-t32_1-p1_Ykt6B
Cparvum_cgd5_1363_Ykt6B
Gniphandrodes_GNI_086860-t26_1-p1_Ykt6B
Ccayetanensis_cyc_02066-t31_1-p1_Ykt6B
Etenella_ETH_00026050-t26_1-p1_Ykt6B
0.99/81
Hhammondi_HHA_299180-t26_1-p1_Ykt6B
Tgondii_TGME49_299180-t26_1-p1_Ykt6B
Ncaninum_NCLIV_009330-t26_1-p1_Ykt6B
Vbrassicaformis_Vbra_8982.t1-p1_Ykt6B
Bnatans_56584_Ykt6A
Bnatans_46811_Ykt6A
Bnatans_89904_estExt_Ykt6A
Ccayetanensis_cyc_03959-t31_1-p1_Ykt6A
Etenella_ETH_00015550-t26_1-p1_Ykt6A
Hhammondi_HHA_215420-t26_1-p1_Ykt6A
Tgondii_TGME49_215420-t26_1-p1_Ykt6A
Ncaninum_NCLIV_052340-t26_1-p1_Ykt6A
Pberghei_PBANKA_1339900.1-p1_Ykt6A
Pyoelii_PYYM_1341600.1-p1_Ykt6A
Pchabaudi_PCHAS_1344500.1-p1_Ykt6A
Pknowlesi_PKNH_1204100.1-p1_Ykt6A
Pvivax_PVX_116640.1-p1_Ykt6A
Pfalciparum_PF3D7_1324700.1-p1_Ykt6A
Preichenowi_PRCDC_1323700.1-p1_Ykt6A
Povale_PocGH01_12037900.1-p1_Ykt6A
Pgallinaceum_PGAL8A_00236600.1-p1_Ykt6A
Prelictum_PRELSG_1203600.1-p1_Ykt6A
Cvelia_Cvel_24396.t1-p1_Ykt6A
Sminutum_symbB.v1.2.029626.t1_Ykt6A
Pmarinus_EER16421_Ykt6A
Pmarinus_EEQ98274_Ykt6A
Vbrassicaformis_Vbra_4868.t1-p1_Ykt6A
Chominis_ChTU502y2012_419g0245-t32_1-p1_Ykt6A
Cparvum_cgd6_510-t26_1-p1_Ykt6A
Cmuris_CMU_027720-t26_1-p1_Ykt6A
Gniphandrodes_GNI_181780-t26_1-p1_Ykt6A
Otrifallax_Contig18101.0.g57_Synaptobrevin_Ykt6A
Ptetraurelia_GSPATP00038192001_Ykt6A
Ptetraurelia_GSPATP00021245001_Ykt6A
Tthermophila_TTHERM_00773190_Ykt6A
Tthermophila_TTHERM_00773200_Ykt6A
0.4
Ykt6A/B
Ykt6B

structure. A single clade of VAMP7 sequences, referred to as VAMP7B, branched within the larger pan-eukaryotic VAMP7A clade with strong support (0.95 pp, 84 bb, Figure 4.11). Although clearly a myzozoan-specific paralogue, VAMP7B was absent from cryptosporidians and *G. niphandrodes*, suggesting it was secondarily lost in some taxa (Figure 4.10). Surprisingly, the lone VAMP7 homologue identified in piroplasmids was clearly identified as VAMP7B (Figures 4.10, 4.11), suggesting that in this lineage, the pan-eukaryotic paralogue has been lost instead.

Similar analysis of Ykt6 sequences (Figure 4.12) revealed a duplication of the pan-eukaryotic Yk6A to yield a lineage-specific Ykt6B paralogue (0.99 pp, 81 bb). However, unlike VAMP7B, Ykt6B was not identified in *P. marinus* or any of the *Symbiodinium* species, suggesting that it arose later in the chromerid-apicomplexan ancestor (Figure 4.10). Furthermore, Ykt6B homologues were identified in all chromerid and apicomplexan genomes, suggesting that it is well-conserved within this group (Figure 4.10).

These data demonstrate that at least two R-SNARE LSPs are present, a myzozoan-specific duplication of VAMP7 and chromerid/apicomplexan-specific duplication of Ykt6.

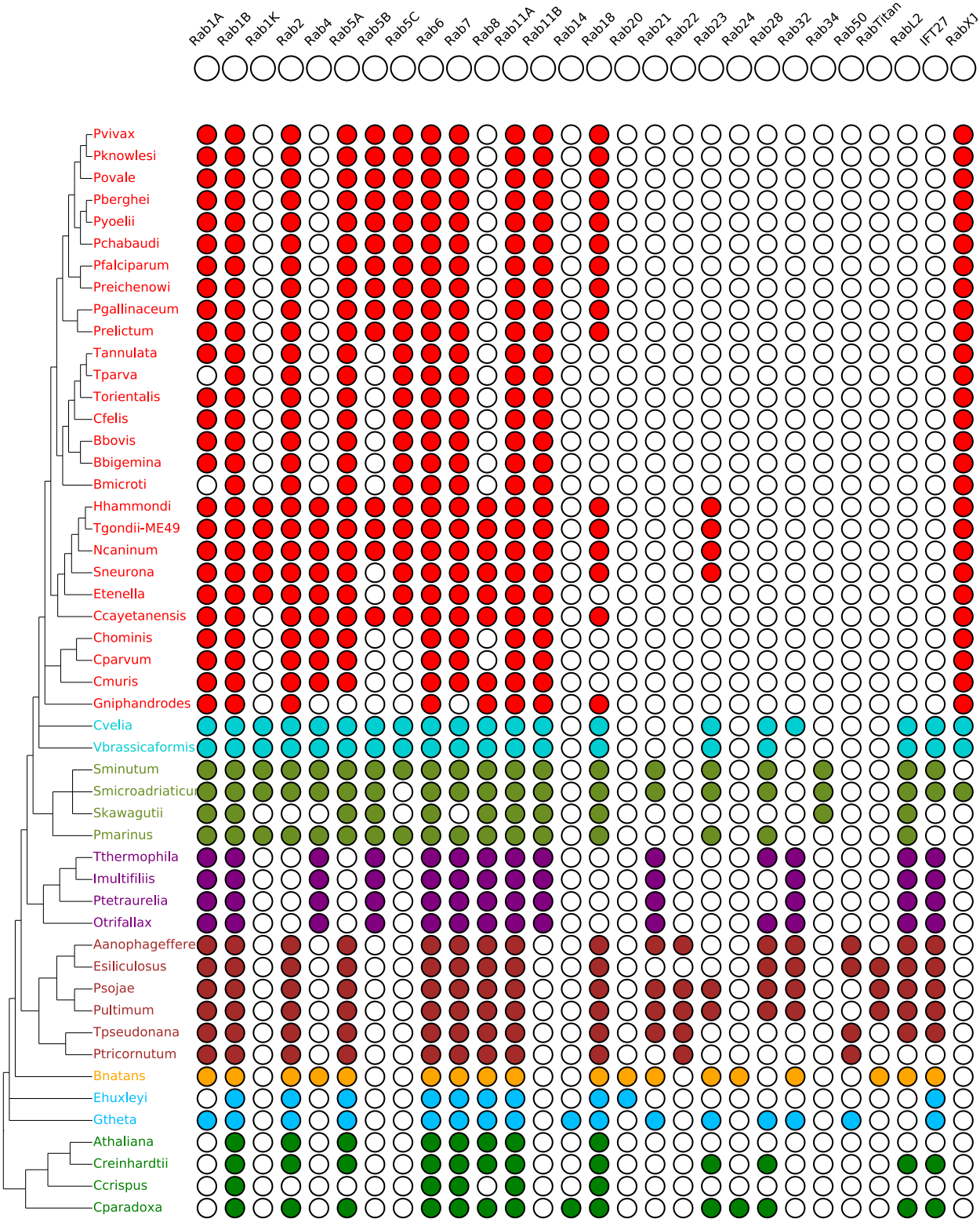
4.4.4 Rab GTPases

Rabs were introduced in section 1.3.1.4 as a group of GTPases within the Ras superfamily that are key regulators of membrane trafficking. Although the LECA has been reconstructed as possessing approximately 23 Rab paralogues, some eukaryotes have subsequently lost ancient paralogues, while others have massively expanded their Rab complement[110]. As similar GTPase domains exist between different members of the Ras superfamily, an HMM comprising all known Rab homologues from *H. sapiens, S. cerevisiae,* and the excavate *Naegleria gruberi* was used to search for Rab homologues instead.

Initial homology searching resulted in identification of 1589 putative Rab homologues in the dataset (Online Appendix Table 4.9). Of these, only about half (868) could be confidently ascribed to a pan-eukaryotic clade during initial phylogenetic analysis of individual groups. 79 sequences were identified as homologues of rab-like proteins, including DNAJ, SGP, RabL3, and RabL6, and were not further pursued. 57 sequences, all from ciliates, branched basally to the pan-eukaryotic Rab2 clade (referred to as basal Rab2, "bRab2", Online Appendix Table 4.9). Another 106 sequences appeared similar to pan-eukaryotic Rab paralogues, including Rab5, Rab6, and Rab11, but did not meet the significance cut-off for inclusion in these groups and were classified as "-like", e.g. Rab5-like. In the end, 227 sequences remained unclassified; the remainder fell into several LSPs, described below.

Figure 4.13 Rab conservation

This figure shows the conservation of select Rabs in study taxa. Each column represents a Rab paralogue, as indicated at the top, while each row represents a different taxon. Filled circles represent the presence of a confidently inferred homologue while empty circles represent the failure to identify a homologue. Cladogram on the left represents the relationships between study taxa. Colour-coding is as follows: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids.


Rab1A
Rab1B
Rab1K
Rab2
Rab4
Rab5A
Rab5B
Rab5C
Rab6
Rab7
Rab8
Rab11A
Rab11B
Rab14
Rab18
Rab20
Rab21
Rab22
Rab23
Rab24
Rab28
Rab32
Rab34
Rab50
RabTitan
RabL2
IFT27
RabX1
Pvivax
Pknowlesi
Povale
Pberghei
Pyoelii
Pchabaudi
Pfalciparum
Preichenowi
Pgallinaceum
Prelictum
Tannulata
Tparva
Torientalis
Cfelis
Bbovis
Bbigemina
Bmicroti
Hhammondi
Tgondii-ME49
Ncaninum
Sneurona
Etenella
Ccayetanensis
Chominis
Cparvum
Cmuris
Gniphandrodes
Cvelia
Vbrassicaformis
Sminutum
Smicroadriaticu
Skawagutii
Pmarinus
Tthermophila
Imultifiliis
Ptetraurelia
Otrifallax
Aanophageffere
Esiliculosus
Psojae
Pultimum
Tpseudonana
Ptricornutum
Bnatans
Ehuxleyi
Gtheta
Athaliana
Creinhardtii
Ccrispus
Cparadoxa

Within the Apicomplexa, only a subset of the LECA Rabs are conserved: Rab1, Rab2, Rab4, Rab5, Rab6, Rab7, Rab8, Rab11, Rab18, and, in coccidians at least, Rab23 (Figure 4.13). Chromerids additionally encode putative homologues of Rab28, Rab32, RabL2/RTW, and RabL4/IFT27, and dinoflagellates Rab21 and Rab34. No alveolate taxa were found to encode homologues of Rab14, Rab20, Rab22, Rab24, Rab50, or RabTitan. In fact, these paralogues were patchily distributed, being largely restricted to *B. natans*, *E. huxleyi*, and *G. theta*, as well as a subset of stramenopiles and archaeplastids. As well, none of the analyses performed could sufficiently separate the identified Rab32 homologues into Rab32A/B clades, and so these sequences are reported simply as Rab32. Importantly though, at least one taxon in the dataset was found to encode each of the inferred ancestral Rabs (Figure 4.13), supporting their proposed identity as pan-eukaryotic Rab paralogues.

A large-scale phylogenetic analysis of combined pan-eukaryotic and LSP sequences confirmed the general robustness of the initial classifications, as well as the high level of sequence conservation among Rab paralogues, as all clades could be reconstructed even with ~900 sequences in the analysis (Figure 4.14). Several paralogues were identified, both during the initial and subsequent phylogenetic classification steps using RAxML, as well as in the large IQ-TREE-based analysis, that branched within, or as sister to, pan-eukaryotic Rab clades, but which were restricted in their taxonomic spread. Several of these mirrored previously reported LSPs in the literature: a Rab1 paralogue restricted to SAR as well as the cryptophyte *G. theta*[513], the Rab5-like proteins Rab5B and Rab5C[253], and the alveolate-specific Rab11B paralogue[390]. However, an additional, previously unreported Rab1 paralogue was identified (see below), as well as a Rab-like paralogue that bore no obvious similarity to any other Rab. To confirm and expand on these putative LSPs, additional analyses were performed using robust phylogenetic methods and focussing on smaller subsets of the total Rab dataset.

4.4.4.1 Rab1-related GTPases

All putative Rab1-related sequences were run together with Rab18 sequences as an outgroup Figure 4.15). As expected, Rab18 sequences clustered together strongly to the exclusion of all Rab1-like sequences (1 pp, 100 bb). Within the Rab1 sequences, three clades emerged with varying levels of statistical support. The pan-eukaryotic paralogue (confusingly referred to as "Rab1B" to stay consistent with Elias et al. (2009)[513]) branched basally to a separate clade of Rab1A and Rab1K sequences (0.94 pp, 63 bb). These sequences represent the previously described Rab1A paralogue[513], which, as expected, was found across SAR and in the cryptophyte *G. theta*, as well as an additional clade of sequences, referred to as Rab1K

Figure 4.14 Rab phylogeny

This figure shows an IQ-TREE phylogeny obtained from a 916 taxon by 162 position alignment of Rabs. Clades are collapsed for simplicity and support values listed for all nodes in order of Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT)/ultra-fast bootstrap (UF-bb). Reliable clades are those with >= 80% SH-aLRT and >= 95% UF-bb support. Scale bar represents number of substitutions per site.

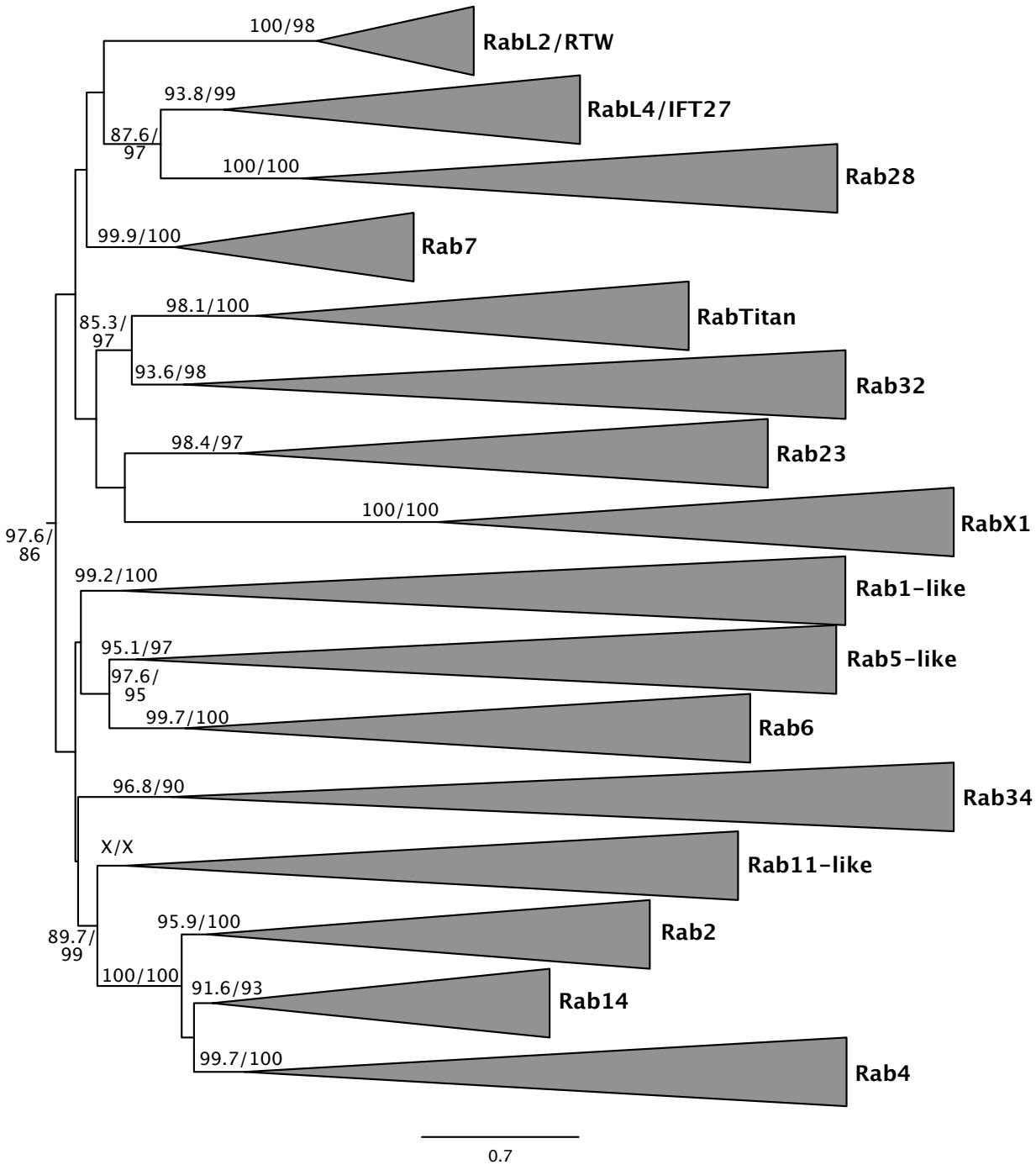
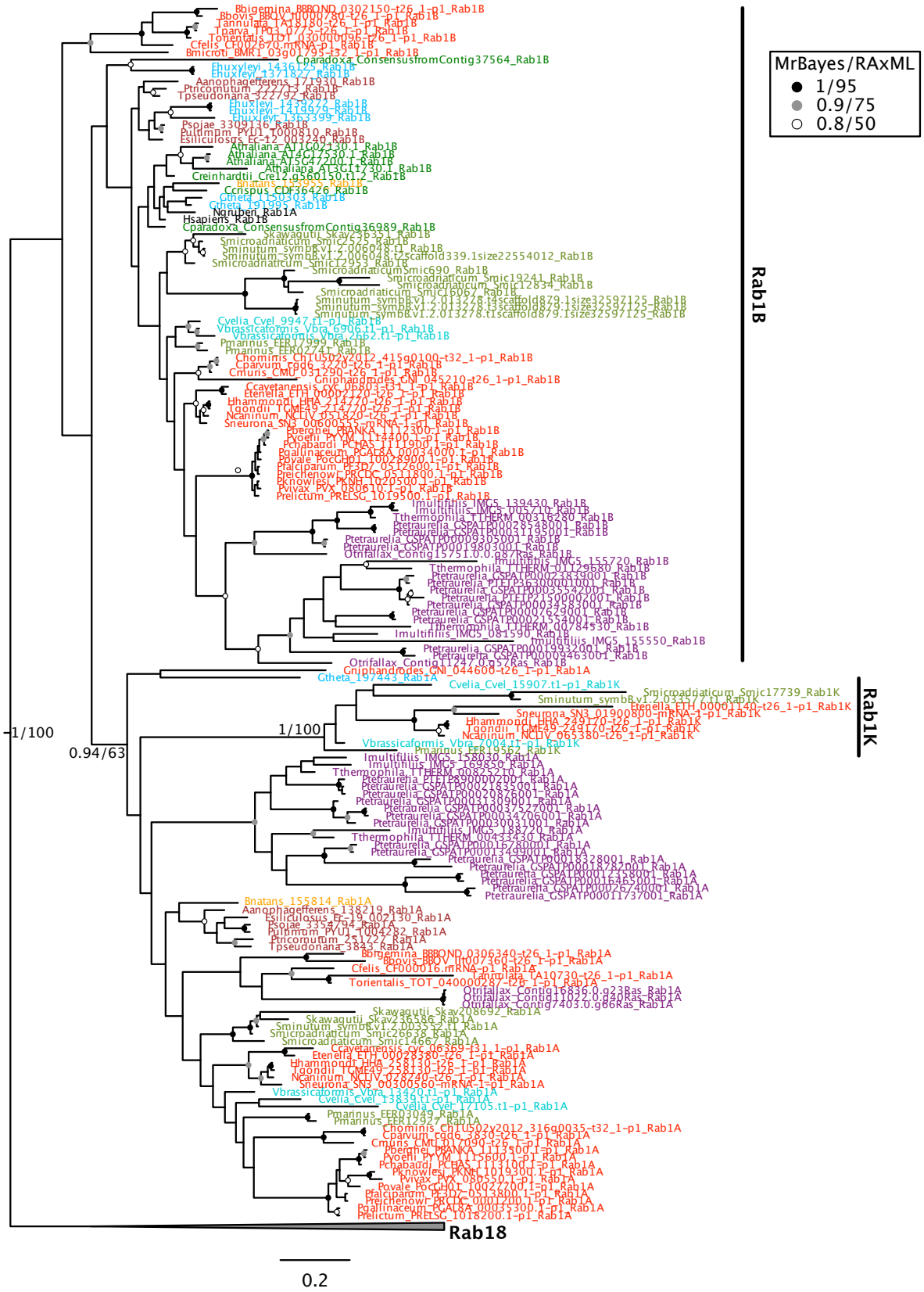

100/98
RabL2/RTW
93.8/99
RabL4/IFT27
87.6/
97
100/100
Rab28
99.9/100
Rab7
98.1/100
RabTitan
85.3/
97
93.6/98
Rab32
98.4/97
Rab23
100/100
RabX1
97.6/
86
99.2/100
Rab1-like
95.1/97
Rab5-like
97.6/
95
99.7/100
Rab6
96.8/90
Rab34
X/X
Rab11-like
95.9/100
Rab2
89.7/
99
100/100
91.6/93
Rab14
99.7/100
Rab4
0.7

Figure 4.15 Detailed phylogeny of Rab1-related sequences

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 204 taxon by 165 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.


MrBayes/RAxML
1/95
0.9/75
0.8/50
Rab1B
Rab1K
Rab1A/K
Rab18
1/100
1/100
0.94/63
0.2

("K" standing for "Kelch", as described below), which branched within the Rab1A clade with strong support (1 pp, 100 bb). Whereas Rab1A was found broadly across SAR, Rab1K was restricted only to coccidian apicomplexans, chromerids, and dinoflagellates (Figures 4.13, 4.15). Although large Rabs (greater than ~200 aa) have been reported previously in the literature, e.g. RabTitan[110], they are still uncommon; surprisingly, Rab1K sequences tend to be long (~600 aa) with an extended N-terminus followed by a C-terminal GTPase domain. Additional investigation of alignments, combined with domain predictions, revealed the presence of a variable number of Kelch motif repeats in the N-termini of Rab1K homologues (Online Appendix Table 4.9), which are predicted to form β-propellers known to be involved in protein-protein interactions[687]. Structural prediction of TgRab1K (materials and methods) supports this notion, as the extended N-terminus models onto known β-propeller structures (Figure 4.16).

Hence, in addition to the previously reported Rab1A paralogue in SAR and *G. theta*, an additional paralogue, Rab1K, likely arose in the myzozoan ancestor before being secondarily lost in multiple extant apicomplexans.

4.4.4.2 Rab5-related GTPases

Rab5B and Rab5C have been previously described in the literature[253], however, their exact identity as *bona fide* Rab5 paralogues has not previously been assessed by adequate phylogenetic analysis. Therefore, a phylogenetic analysis including all Rab5-related sequences (Rab20, Rab21, Rab22, Rab24, and Rab50, referred to as the "primordial endocytic" clade in Elias et al. (2012)[110] was performed with Rab6 included as an outgroup (Figure 4.17).

As expected, Rab6 sequences clustered together apart from all Rab5-related sequences with strong support (1 pp, 100 bb). Of the Rab5-related clades, few were reconstructed with overwhelmingly strong support; for example, Rab20 and Rab24 were not supported as separate clades in this analysis, but the single clade encompassing both met the threshold of statistical support (1 pp, 57 bb). Rab22 sequences, which never grouped together strongly in any of the independent RAxML runs performed during sequence classification, had similarly low support (0.77 pp, 56 bb). Importantly though, all sequences formed the expected monophyletic clades, and no previously classified sequences resolved within unexpected clades, suggesting that the phylogenetic reconstruction is accurate.

Homologues of Rab5A and Rab5C grouped within a single large clade (0.97 pp, 62 bb), and Rab5C sequences themselves branched within this assemblage with very strong support (1 pp, 98 bb); in other analyses performed during classification, these two clades branched

Figure 4.16 Secondary and tertiary structure of Rab1K

This figure provides an overview of the unique sequence features identified for Rab1K homologues. A) Domain structure of five exemplar Rab1K homologues, based on Pfam domain prediction. Black lines represent amino acid sequence, with green (Kelch) and blue (GTPase) cylinders corresponding to predicted domains. Scale bar indicates 100 amino acids. B) Predicted structure of the entire *T. gondii* Rab1K protein sequences, as assessed by Phyre2. Note the presence of an N-terminal β-propeller followed by a well-folded C-terminal region of largely alpha-helices corresponding to the GTPase domain. C) Confirmatory modeling of the N-terminus of TgRab1K using HHPred/MODELLER; note that the sequence still folds into a predicted β-propeller. For B and C, rainbow colouring is used, with the N-terminus in blue and C-terminus in red.

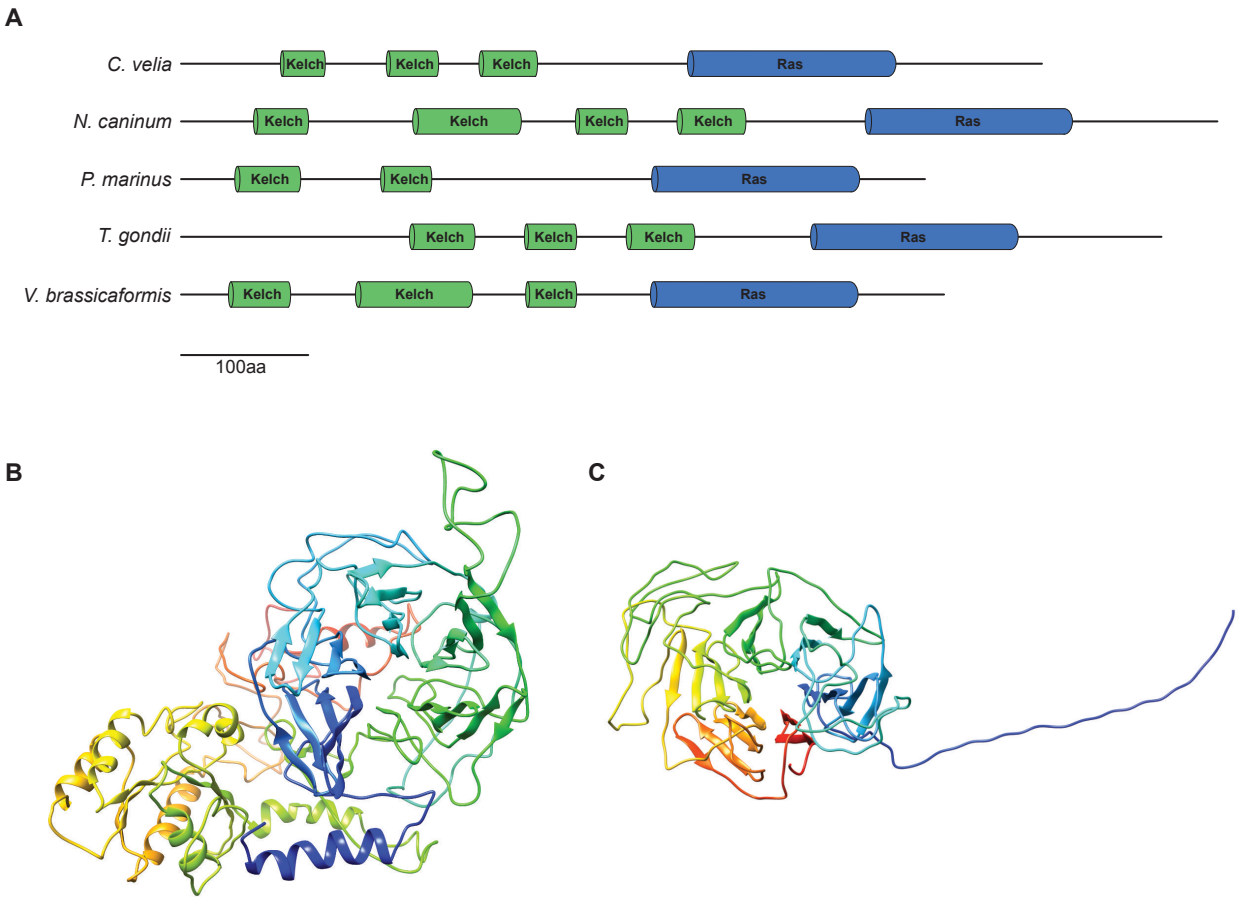
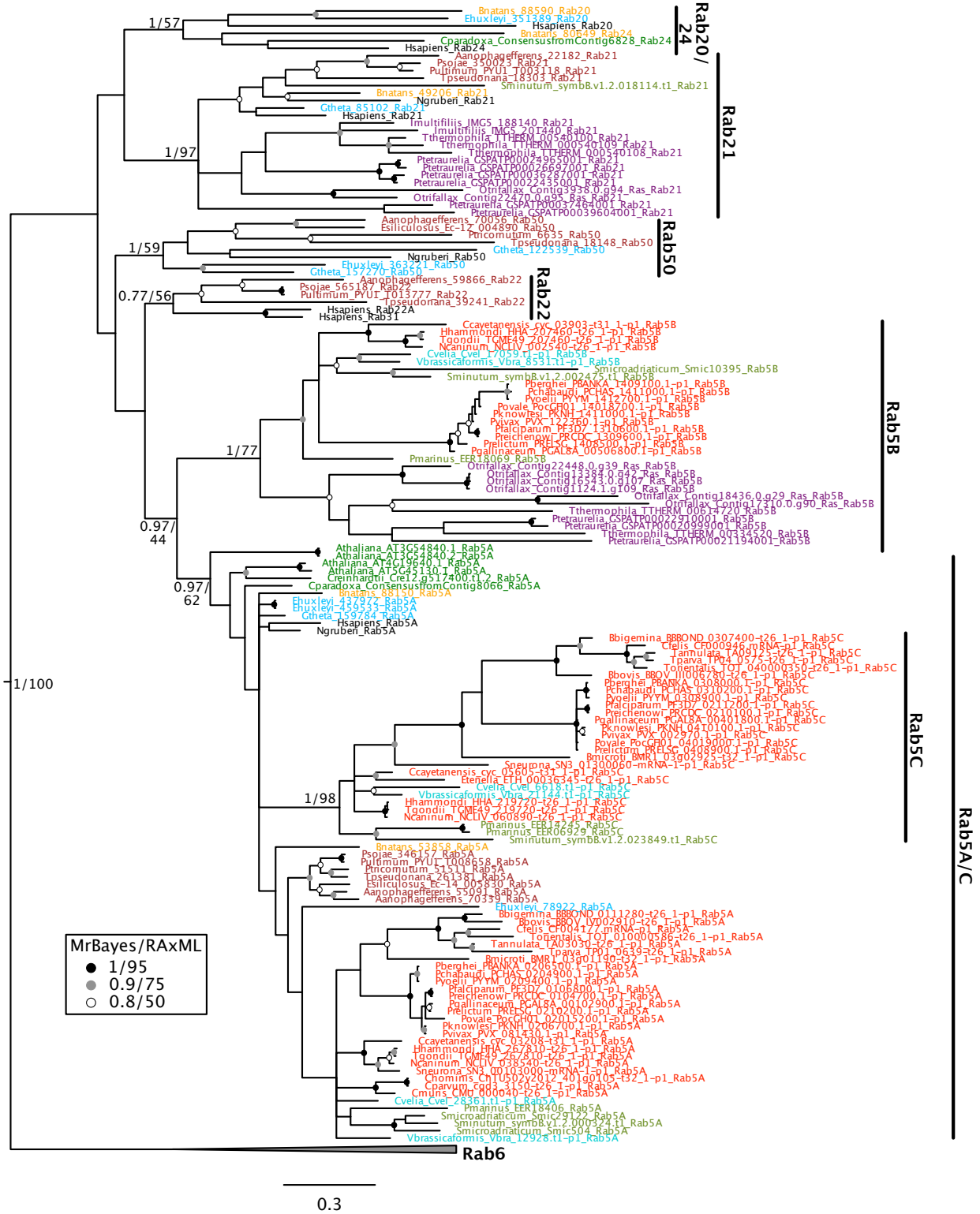

A
C. velia
Kelch
Kelch
Kelch
Ras
N. caninum
Kelch
Kelch
Kelch
Kelch
Ras
P. marinus
Kelch
Kelch
Ras
T. gondii
Kelch
Kelch
Kelch
Ras
V. brassicaformis
Kelch
Kelch
Kelch
Ras
100aa
B
C

Figure 4.17 Detailed phylogeny of Rab5-related sequences

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 202 taxon by 161 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.


Rab20/24
Rab21
Rab50
Rab22
Rab5B
Rab5C
Rab5A/C
Rab6
1/57
1/97
1/59
0.77/56
1/77
0.97/44
0.97/62
1/100
1/98
MrBayes/RAxML
1/95
0.9/75
0.8/50
0.3

as sisters. Overall, these results suggest that Rab5A and Rab5C represent each other's closest relative. Rab5B sequences grouped together with moderate support (1 pp, 77 bb), and formed an outgroup to the Rab5A/5C clade with good posterior probability but comparatively low bootstrap support (0.97 pp, 44 bb). The position of Rab5B within the Rab5-related clade was comparatively unstable; whereas Rab5A and Rab5C always grouped together, Rab5B either formed a basal branch to this combined clade or branched with low support as sister to another Rab5-related clade.

These results suggest that Rab5C represents a Rab5 paralogue arising in the myzozoan common ancestor, while Rab5B, although definitely restricted to alveolates (Figure 4.13), is either an additional divergent Rab5 paralogue, or represents another Rab5-related paralogue arising in the alveolate ancestor.

4.4.4.3 Rab11-related GTPases

Rab11B was previously described as an alveolate-specific Rab11 paralogue[390]. To confirm this, all Rab11-like sequences identified from initial classification were run together with Rab18 as an outgroup. As with the other analyses described thus far, the Rab18 sequences grouped together with strong support separate from all Rab11-like sequences (1 pp, 100 bb). Among the Rab11 sequences, Rab11A sequences formed a basal group from which a strongly supported Rab11B (1 pp, 94 bb) clade emerged (Figure 4.18). The exact topology within this combined Rab11 clade varied among analyses, with Rab11A and Rab11B occasionally forming sister clades, especially in larger analyses, but Rab11B sequences were always grouped to the exclusion of all Rab11A sequences with at least moderate support. Expansions within the Rab11A/Rab11B clades were frequently observed, especially in *A. thaliana* and in ciliates, but these were clearly restricted to one or a few lineages.

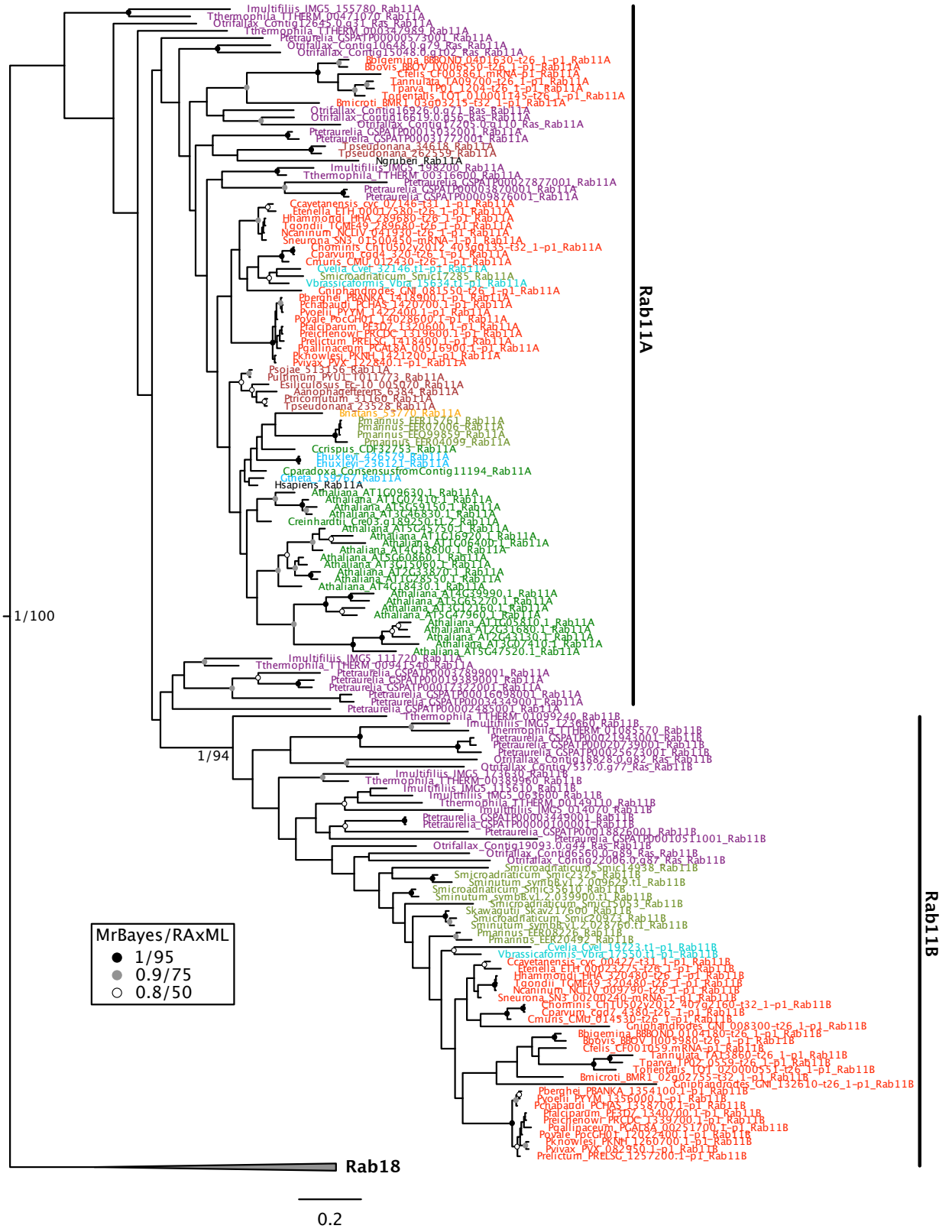
These results support the previous result reported in the literature and suggest that Rab11 duplicated in the alveolate ancestor to yield an additional Rab11B paralogue.

4.4.4.4 RabX1, an uncharacterized myzozoan-specific Rab

In addition to the Rab LSPs described above, which, on the basis of both homology searching and initial phylogenetic analyses, could be confidently ascribed as paralogues of known pan-eukaryotic Rab sequences, one additional putative LSP was detected among the Myzozoa. This paralogue, referred to as RabX1, did not confidently group with any pan-eukaryotic Rab paralogue in any of the initial RAxML-based reconstructions. However, RabX1 homologues consistently retrieved Rab homologues in reverse BLASTp searches (Online Appendix Table 4.9), suggesting that they do represent *bona fide* Rab paralogues. The top hit

Figure 4.18 Detailed phylogeny of Rab11-related sequences

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 197 taxon by 166 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.


Rab11A
Rab11B
Rab18
1/100
1/94
MrBayes/RAxML
1/95
0.9/75
0.8/50
0.2

in each case was different, precluding any consistent inference of similarity on the basis of BLAST analysis alone. However, RabX1 did consistently group within a larger sub-section of Rab paralogues, namely that of RabL2/RTW, RabL4/IFT27, Rab7, Rab23, Rab28, Rab32, and RabTitan (97.6 SH-aLRT, 86 UF-bb, Figure 4.14).

Therefore, to further investigate the putative identity of RabX1, these sequences, together with Rab34, which grouped with Rab23 in previous Rab trees[110], were run in a separate phylogenetic analysis (Figure 4.19). As expected, each putative clade was recovered with strong support, with the exception of Rab32 – in this case though, a more expansive Rab32/Titan clade did meet the threshold of statistical support (0.92 pp, 67 bb). RabX1 homologues grouped together with maximal support (1 pp, 100 bb), and, surprisingly, were most closely related to Rab28 (0.78 pp, 45 bb). Although just shy of the threshold considered to represent minimal support for credible clades (i.e. 0.8 pp, 50 bb), RabX1 was often found to group as sister to either Rab28, or to RabL4/IFT27, which formed an outgroup to the RabX1/Rab28 clade in Figure 4.19 with low support (0.75 pp, 26 bb). Rab28 and IFT27 were also found to be sister clades in a previous high-resolution analysis of Rabs[110], albeit again without significant support; the results presented here are consistent with Rab28 and IFT27 being closely related, and additionally suggest that RabX1 is also related to this group.

Hence, although it cannot be confidently asserted, it appears as though RabX1 may represent a divergent Rab28-like group. Critically though, chromerids and dinoflagellates also encode canonical Rab28 paralogues (Figure 4.13), suggesting that RabX1 represents a true LSP rather than a divergent Rab28 homologue.

4.4.5 RabGAPs (TBCs)

As discussed in section 1.3.1.4, almost all known RabGAPs are members of a large paralogous family defined by the presence of a Tre-2/Bub2/Cdc16 (TBC) domain. 13 TBC domain proteins, TBC-B, -D, -E, -F, -G, -H, -I, -K, -L, -M, -N, -Q, and -RootA, are inferred as ancient, although additional families were found in more restricted portions of the eukaryotic tree as well[123]. An HMM derived from the TBC domain (Pfam ID PF00566) was used to identify all putative TBC homologues within the dataset.

In total, 1049 putative TBC homologues were identified within the dataset (Online Appendix Table 4.10). Similar to the Rabs, initial phylogenetic analyses resulted in approximately half (547) being confidently classified as pan-eukaryotic paralogues. Further analyses of the remaining sequences placed roughly half into clades of LSPs (described below), while 275 sequences remained unclassified.

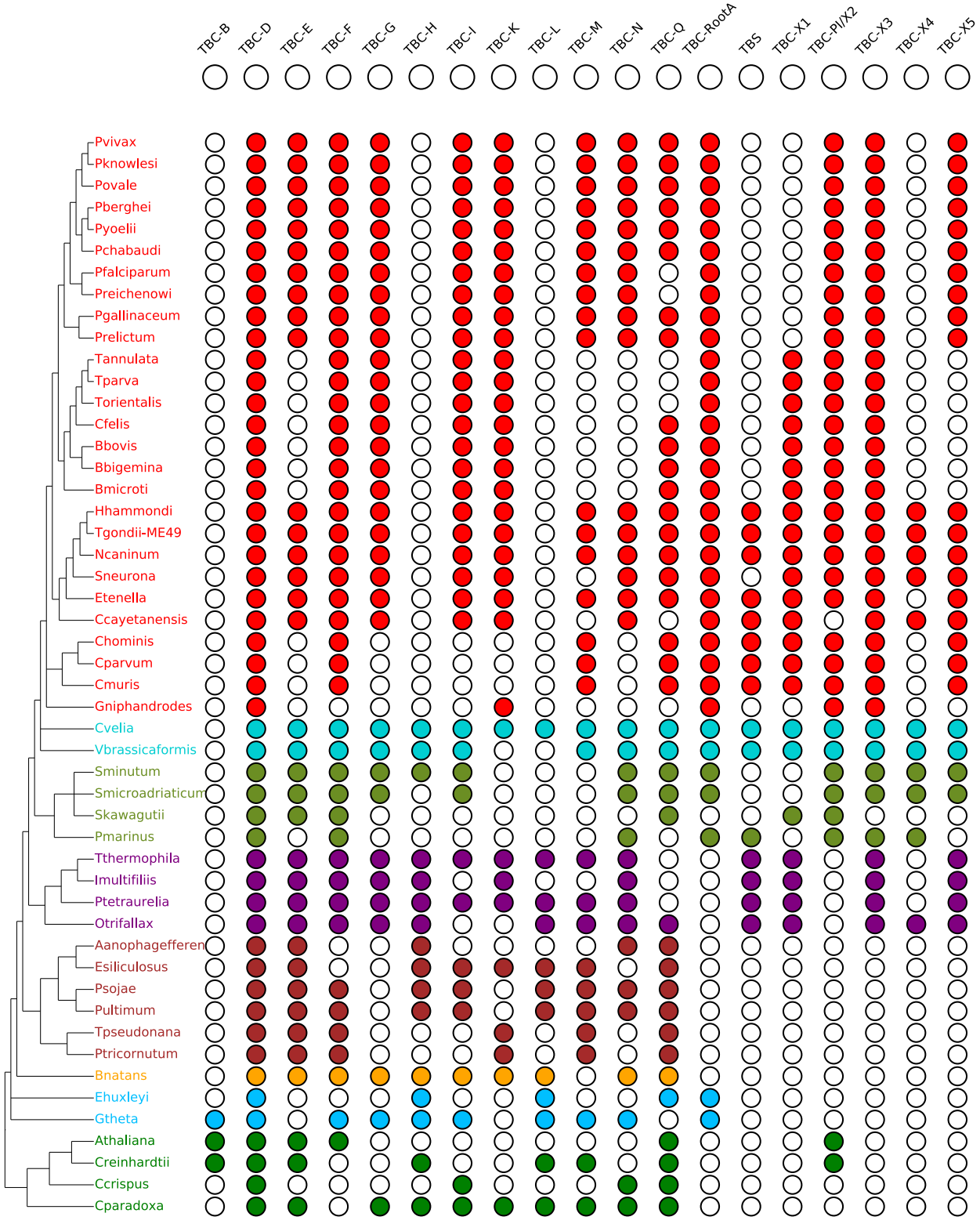
Figure 4.19 Detailed phylogeny of RabX1-related sequences

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 164 taxon by 155 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.


MrBayes/RAxML
1/95
0.9/75
0.8/50
Rab32/Titan
Rab23
Rab7
RabX1
Rab28
RabL4/IFT27
RabL2/RTW
Rab34
0.92/67
0.92/83
1/98
1/96
1/100
0.78/
45
0.96/
94
1/95
0.75/
26
1/91
1/99
0.4

Figure 4.20 TBC conservation

This figure shows the conservation of select TBCs in study taxa. Each column represents a TBC paralogue, as indicated at the top, while each row represents a different taxon. Filled circles represent the presence of a confidently inferred homologue while empty circles represent the failure to identify a homologue. Cladogram on the left represents the relationships between study taxa. Colour-coding is as follows: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids.


TBC-B
TBC-D
TBC-E
TBC-F
TBC-G
TBC-H
TBC-I
TBC-K
TBC-L
TBC-M
TBC-N
TBC-Q
TBC-RootA
TBS
TBC-X1
TBC-PI/X2
TBC-X3
TBC-X4
TBC-X5
Pvivax
Pknowlesi
Povale
Pberghei
Pyoelii
Pchabaudi
Pfalciparum
Preichenowi
Pgallinaceum
Prelictum
Tannulata
Tparva
Torientalis
Cfelis
Bbovis
Bbigemina
Bmicroti
Hhammondi
Tgondii-ME49
Ncaninum
Sneurona
Etenella
Ccayetanensis
Chominis
Cparvum
Cmuris
Gniphandrodes
Cvelia
Vbrassicaformis
Sminutum
Smicroadriaticum
Skawagutii
Pmarinus
Tthermophila
Imultifiliis
Ptetraurelia
Otrifallax
Aanophagefferen
Esiliculosus
Psojae
Pultimum
Tpseudonana
Ptricornutum
Bnatans
Ehuxleyi
Gtheta
Athaliana
Creinhardtii
Ccrispus
Cparadoxa

Apicomplexa encode 10 of the 13 ancestral TBC proteins, with the exception of TBC-B, which was only identified in archaeplastids and *G. theta*, TBC-H, which was present in both chromerids, and TBC-L, which was identified in *C. velia* but not *V. brassicaformis* (Figure 4.20). Hence, it is likely that the loss of TBC-H and TBC-L occurred in Apicomplexa after the split from the apicomplexan-chromerid ancestor. Multiple secondary losses were inferred in the Apicomplexa as well: TBC-E, -G, -I, -K, and -N in *Cryptosporidium spp.*, TBC-E, -M, and -N in piroplasmids, as well as TBC-Q in *Theileria spp.* but not other basal piroplasmids. Curiously, TBC-Q also appears to be absent from both *P. falciparum* and *P. reichenowi*, despite being conserved in all other included members of the genus (Figure 4.20). Interestingly, TBC-RootA, which is almost universally conserved in Myzozoa, was absent from all outgroup taxa with the exception of *E. huxleyi* and *G. theta*, suggesting it has been lost multiple times independently within the diaphoretickes.

These LSPs were combined with all pan-eukaryotic sequences and run in a large-scale phylogenetic analysis to confirm the initial classifications (Figure 4.21). Encouragingly, the observed tree topology was similar to those from previous analyses[123], with TBC-K, -N, -Q, and -RootA forming one group, TBC-B, -D, -E, and -F forming another, and the other TBCs placing between them. Combined results from homology searching and initial phylogenetic analyses suggested that several clades including apicomplexan homologues were present that may represent LSPs: the TBC portion of TBS proteins, which appeared to group with TBC-N homologues, a clade referred to as TBC-X2, which grouped with TBC-PI homologues from archaeplastids, and four other clades, which all appeared most similar to TBC-Q. Hence, further detailed phylogenetic analyses were performed for each group to confirm the identity of each putative LSP.

4.4.5.1 TBC-N and TBS

As discussed in section 4.4, TBS proteins possess both a Sec7 ArfGEF domain and a TBC RabGAP domain. As part of a larger study on pan-eukaryotic ArfGEF conservation, I previously reported results using a smaller sampling of alveolate taxa showing that the TBC domain of TBS proteins is derived from TBC-N[36]. The phylogenetic analysis here supports this view as it groups TBC-N and TBS proteins (separate from TBC-D, 1 pp, 100 bb, Figure 4.22), and further supports a single alveolate TBS clade within the larger TBC-N clade (0.79 pp, 50 bb). Hence, these results here confirm and extend the previous analysis.

Figure 4.21 TBC phylogeny

This figure shows an IQ-TREE phylogeny obtained from a 679 taxon by 153 position alignment of TBCs. Clades are collapsed for simplicity and support values listed for all nodes in order of Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT)/ultra-fast bootstrap (UF-bb). Reliable clades are those with >= 80% SH-aLRT and >= 95% UF-bb support. Scale bar represents number of substitutions per site.

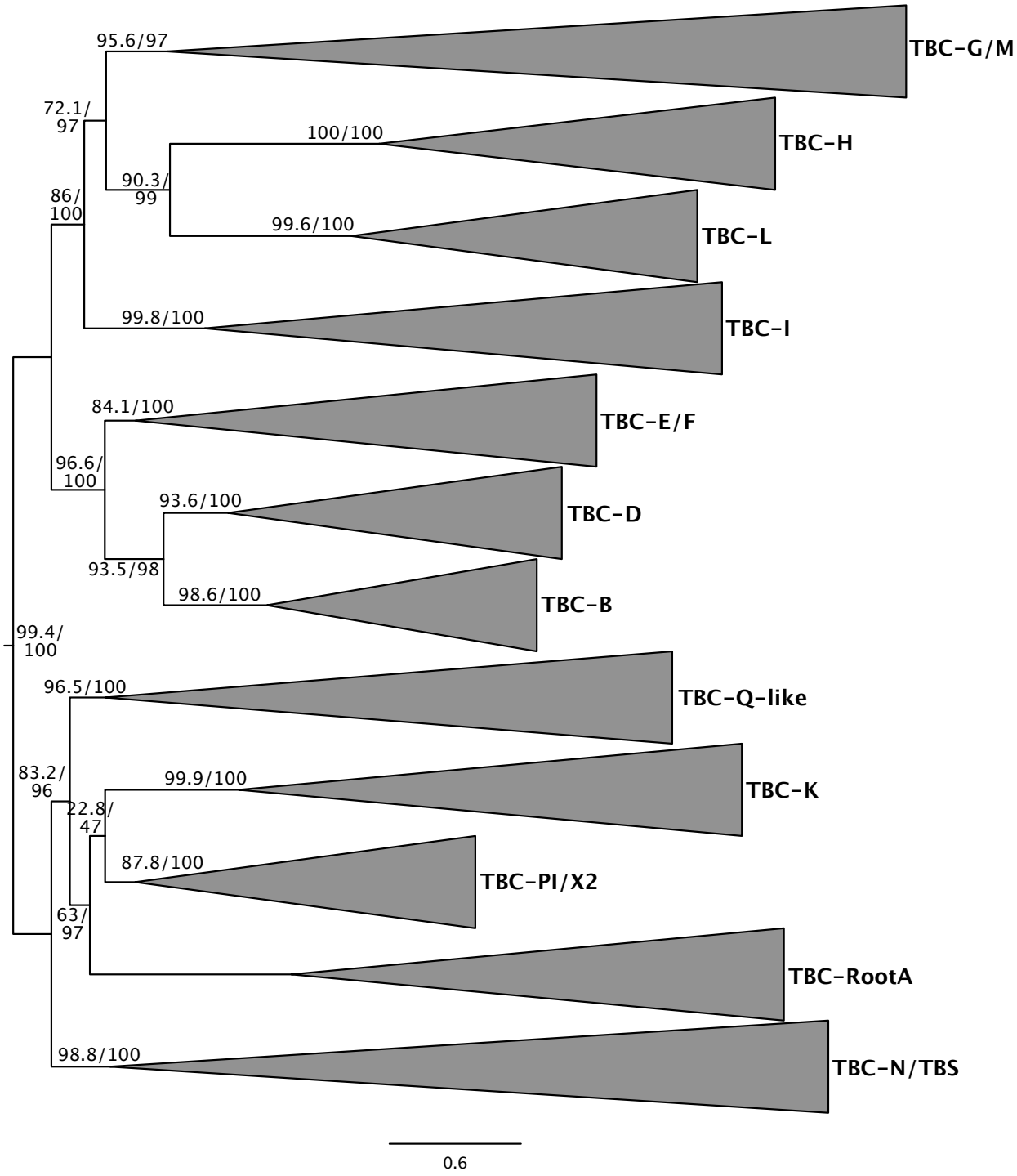
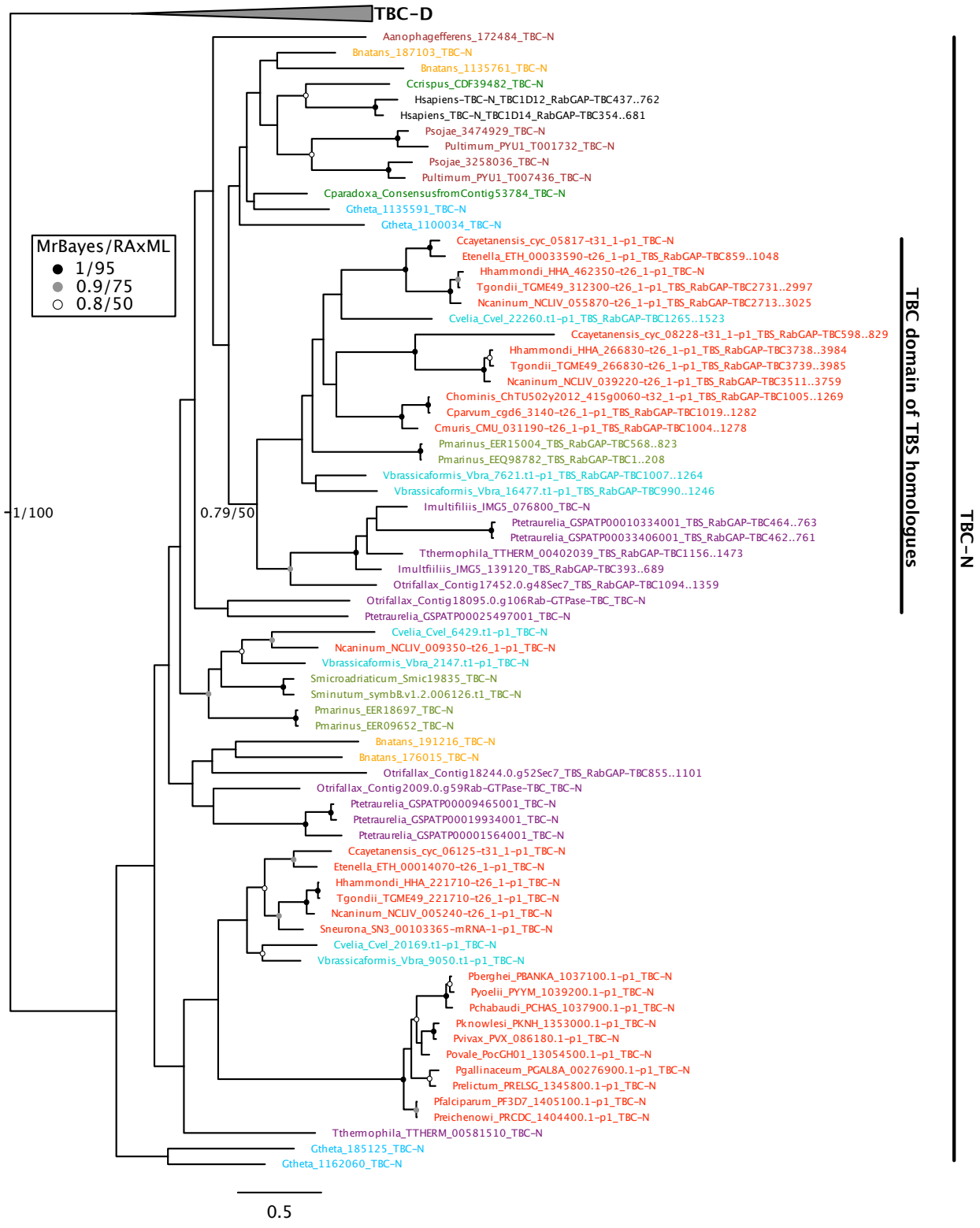

95.6/97
TBC-G/M
72.1/
97
100/100
TBC-H
90.3/
99
86/
100
99.6/100
TBC-L
99.8/100
TBC-I
84.1/100
TBC-E/F
96.6/
100
93.6/100
TBC-D
93.5/98
98.6/100
TBC-B
99.4/
100
96.5/100
TBC-Q-like
83.2/
96
99.9/100
TBC-K
22.8/
47
87.8/100
TBC-PI/X2
63/
97
TBC-RootA
98.8/100
TBC-N/TBS
0.6

Figure 4.22 Detailed phylogeny of TBC-N-related sequences

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 132 taxon by 172 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.


TBC-D
Aanophagefferens_172484_TBC-N
Bnatans_187103_TBC-N
Bnatans_1135761_TBC-N
Ccrispus_CDF39482_TBC-N
Hsapiens-TBC-N_TBC1D12_RabGAP-TBC437..762
Hsapiens_TBC-N_TBC1D14_RabGAP-TBC354..681
Psojae_3474929_TBC-N
Pultimum_PYU1_T001732_TBC-N
Psojae_3258036_TBC-N
Pultimum_PYU1_T007436_TBC-N
Cparadoxa_ConsensusfromContig53784_TBC-N
Gtheta_1135591_TBC-N
Gtheta_1100034_TBC-N
MrBayes/RAxML
1/95
0.9/75
0.8/50
Ccayetanensis_cyc_05817-t31_1-p1_TBC-N
Etenella_ETH_00033590-t26_1-p1_TBS_RabGAP-TBC859..1048
Hhammondi_HHA_462350-t26_1-p1_TBC-N
Tgondii_TGME49_312300-t26_1-p1_TBS_RabGAP-TBC2731..2997
Ncaninum_NCLIV_055870-t26_1-p1_TBS_RabGAP-TBC2713..3025
Cvelia_Cvel_22260.t1-p1_TBS_RabGAP-TBC1265..1523
Ccayetanensis_cyc_08228-t31_1-p1_TBS_RabGAP-TBC598..829
Hhammondi_HHA_266830-t26_1-p1_TBS_RabGAP-TBC3738..3984
Tgondii_TGME49_266830-t26_1-p1_TBS_RabGAP-TBC3739..3985
Ncaninum_NCLIV_039220-t26_1-p1_TBS_RabGAP-TBC3511..3759
Chominis_ChTU502y2012_415g0060-t32_1-p1_TBS_RabGAP-TBC1005..1269
Cparvum_cgd6_3140-t26_1-p1_TBS_RabGAP-TBC1019..1282
Cmuris_CMU_031190-t26_1-p1_TBS_RabGAP-TBC1004..1278
Pmarinus_EER15004_TBS_RabGAP-TBC568..823
Pmarinus_EEQ98782_TBS_RabGAP-TBC1..208
Vbrassicaformis_Vbra_7621.t1-p1_TBS_RabGAP-TBC1007..1264
Vbrassicaformis_Vbra_16477.t1-p1_TBS_RabGAP-TBC990..1246
1/100
0.79/50
Imultifiliis_IMG5_076800_TBC-N
Ptetraurelia_GSPATP00010334001_TBS_RabGAP-TBC464..763
Ptetraurelia_GSPATP00033406001_TBS_RabGAP-TBC462..761
Tthermophila_TTHERM_00402039_TBS_RabGAP-TBC1156..1473
Imultfiiliis_IMG5_139120_TBS_RabGAP-TBC393..689
Otrifallax_Contig17452.0.g48Sec7_TBS_RabGAP-TBC1094..1359
Otrifallax_Contig18095.0.g106Rab-GTPase-TBC_TBC-N
Ptetraurelia_GSPATP00025497001_TBC-N
Cvelia_Cvel_6429.t1-p1_TBC-N
Ncaninum_NCLIV_009350-t26_1-p1_TBC-N
Vbrassicaformis_Vbra_2147.t1-p1_TBC-N
Smicroadriaticum_Smic19835_TBC-N
Sminutum_symbB.v1.2.006126.t1_TBC-N
Pmarinus_EER18697_TBC-N
Pmarinus_EER09652_TBC-N
Bnatans_191216_TBC-N
Bnatans_176015_TBC-N
Otrifallax_Contig18244.0.g52Sec7_TBS_RabGAP-TBC855..1101
Otrifallax_Contig2009.0.g59Rab-GTPase-TBC_TBC-N
Ptetraurelia_GSPATP00009465001_TBC-N
Ptetraurelia_GSPATP00019934001_TBC-N
Ptetraurelia_GSPATP00001564001_TBC-N
Ccayetanensis_cyc_06125-t31_1-p1_TBC-N
Etenella_ETH_00014070-t26_1-p1_TBC-N
Hhammondi_HHA_221710-t26_1-p1_TBC-N
Tgondii_TGME49_221710-t26_1-p1_TBC-N
Ncaninum_NCLIV_005240-t26_1-p1_TBC-N
Sneurona_SN3_00103365-mRNA-1-p1_TBC-N
Cvelia_Cvel_20169.t1-p1_TBC-N
Vbrassicaformis_Vbra_9050.t1-p1_TBC-N
Pberghei_PBANKA_1037100.1-p1_TBC-N
Pyoelii_PYYM_1039200.1-p1_TBC-N
Pchabaudi_PCHAS_1037900.1-p1_TBC-N
Pknowlesi_PKNH_1353000.1-p1_TBC-N
Pvivax_PVX_086180.1-p1_TBC-N
Povale_PocGH01_13054500.1-p1_TBC-N
Pgallinaceum_PGAL8A_00276900.1-p1_TBC-N
Prelictum_PRELSG_1345800.1-p1_TBC-N
Pfalciparum_PF3D7_1405100.1-p1_TBC-N
Preichenowi_PRCDC_1404400.1-p1_TBC-N
Tthermophila_TTHERM_00581510_TBC-N
Gtheta_185125_TBC-N
Gtheta_1162060_TBC-N
0.5
TBC domain of TBS homologues
TBC-N

4.4.5.2 TBC-PI and TBC-X2

TBC-PI was previously reported as an archaeplastid-specific TBC protein present in plants and green algae, but absent from red algae, that could be further divided into two paralogues, TBC-PIA and TBC-PIB[123]. A myzozoan-specific group of sequences, originally referred to as TBC-X2 (Figure 4.20), was identified that appeared to group with TBC-PI sequences in multiple phylogenetic analyses (87.8 SH-aLRT, 100 UF-bb, Figure 4.21). A robust phylogenetic analysis focussing on these sequences strongly supported their grouping to the exclusion of the next closest paralogue, TBC-K (1 pp, 100 bb, Figure 4.23). Although the node uniting TBC-PIA, TBC-PIB, and TBC-X2 sequences resolved as a polytomy, the topology is consistent with all three groups representing a single clade, with a subsequent split into TBC-PIA and TBC-PIB paralogues in the Archaeplastida (0.99 pp, 60 bb for the node uniting these two clades, Figure 4.23). Hence, TBC-PI is more widespread than previously reported, being present in at least Archaeplastida and Myzozoa.

4.4.5.3 TBC-Q and other novel TBC LSPs

The remaining putative TBC LSPs grouped strongly with TBC-Q (96.5 SH-aLRT, 100 UF-bb, Figure 4.21). All four of these clades, referred to as TBC-X1, -X3, -X4, and -X5, are inferred to have arisen in the alveolate ancestor (Figure 4.20). TBC-X1 has been secondarily lost in *Plasmodium spp.*, TBC-X4 in *Plasmodium spp.*, piroplasmids, and *Cryptosporidium spp.*, and TBC-X5 in piroplasmids; all four paralogues are present in coccidians.

Robust phylogenetic analysis supported their close similarity (the main node resolving again as a polytomy) and grouped them to the exclusion of the closely related TBC-K with maximal support (1 pp, 100 bb, Figure 4.24). Although TBC-Q (0.98 pp, 18 bb) and TBC-X5 (0.62 pp, 23 bb) clades were not strongly supported, the other three clades, TBC-X1 (1 pp, 94 bb), TBC-X3 (1 pp, 89 bb), and TBC-X4 (1 pp, 79 bb) were. However, all clades were reconstructed as monophyletic in multiple analyses, suggesting that they are *bona fide* LSPs.

4.4.6 Arf and Arl G proteins

Arf family G proteins were introduced in section 1.3.1.1, as another member of the Ras superfamily (in addition to Rabs) mediating trafficking in eukaryotes. In addition, as discussed in section 1.3.1.1, the LECA complement of ARF and Arl proteins has not been sufficiently investigated, though a single ARF and, at least, Arl1, Arl2, Arl3, Arl5, Arl8 and ARFRP1, are expected to be found across eukaryotes[31]. An HMM derived from the Arf domain (Pfam ID PF00025) was used to identify all putative Arf family homologues within the dataset.

Figure 4.23 Detailed phylogeny of TBC-PI/X2 sequences

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 89 taxon by 213 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.

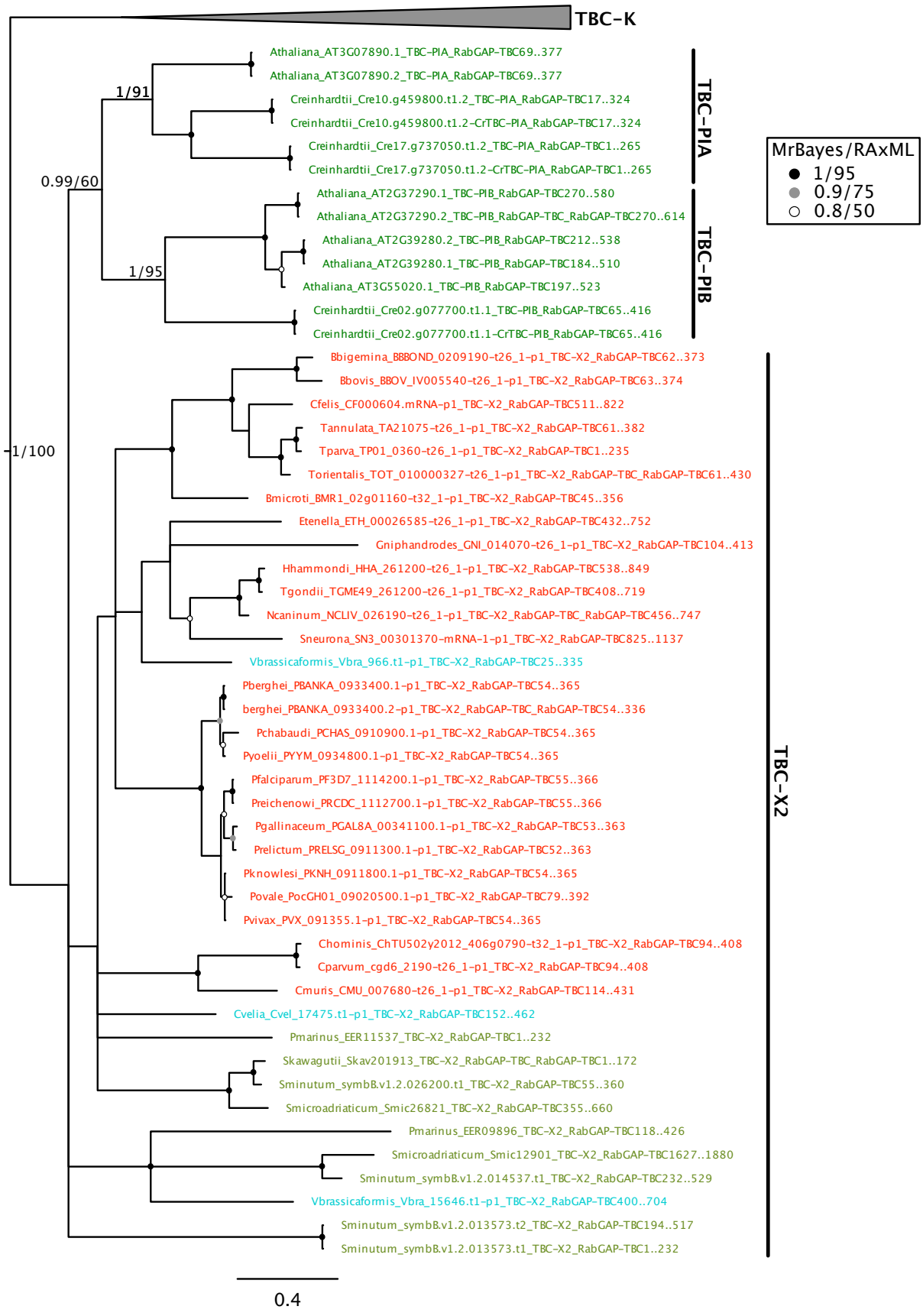
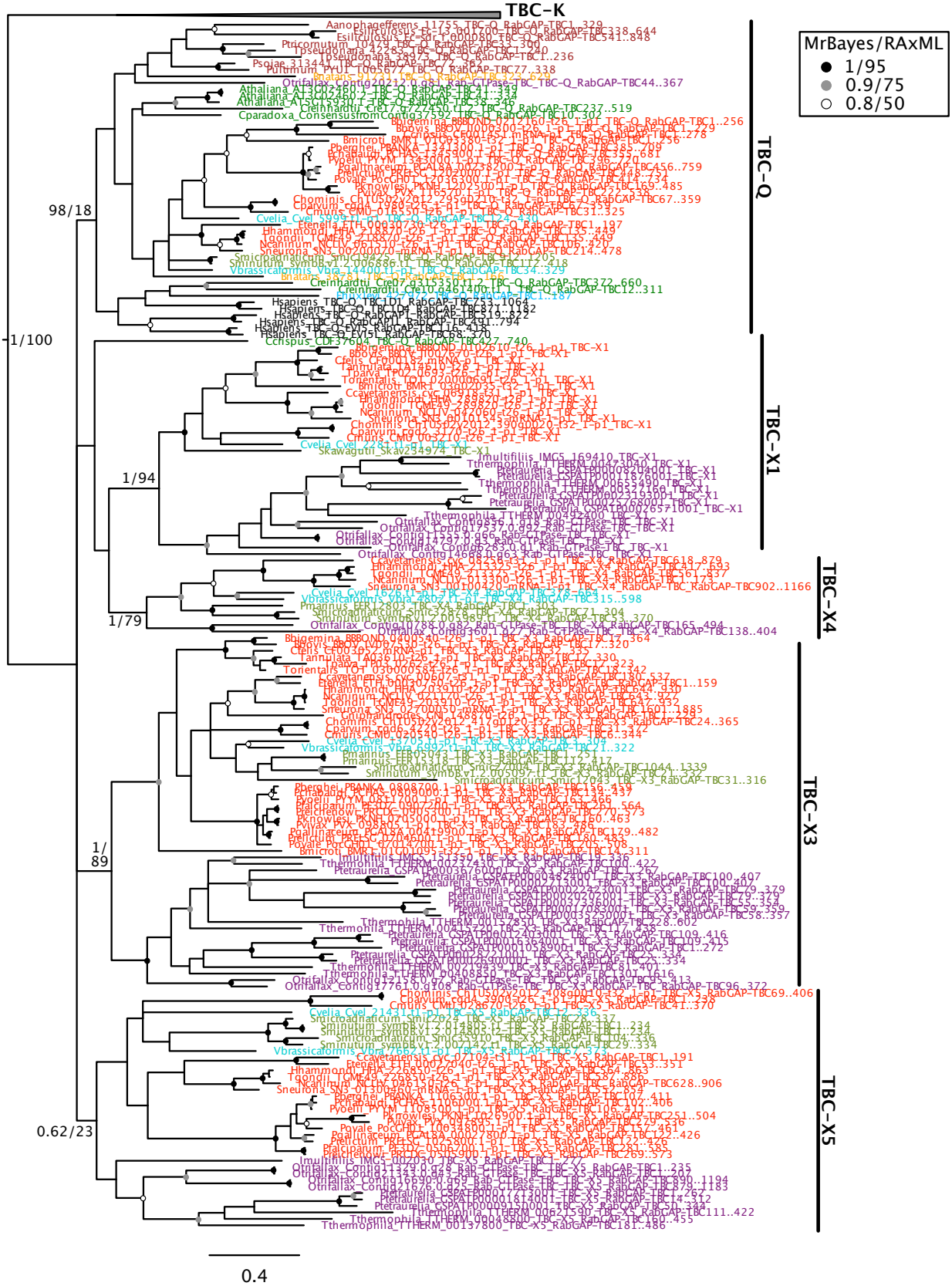

TBC-K
Athaliana_AT3G07890.1_TBC-PIA_RabGAP-TBC69..377
Athaliana_AT3G07890.2_TBC-PIA_RabGAP-TBC69..377
1/91
Creinhardtii_Cre10.g459800.t1.2_TBC-PIA_RabGAP-TBC17..324
Creinhardtii_Cre10.g459800.t1.2-CrTBC-PIA_RabGAP-TBC17..324
Creinhardtii_Cre17.g737050.t1.2_TBC-PIA_RabGAP-TBC1..265
Creinhardtii_Cre17.g737050.t1.2-CrTBC-PIA_RabGAP-TBC1..265
TBC-PIA
0.99/60
Athaliana_AT2G37290.1_TBC-PIB_RabGAP-TBC270..580
Athaliana_AT2G37290.2_TBC-PIB_RabGAP-TBC_RabGAP-TBC270..614
Athaliana_AT2G39280.2_TBC-PIB_RabGAP-TBC212..538
Athaliana_AT2G39280.1_TBC-PIB_RabGAP-TBC184..510
1/95
Athaliana_AT3G55020.1_TBC-PIB_RabGAP-TBC197..523
Creinhardtii_Cre02.g077700.t1.1_TBC-PIB_RabGAP-TBC65..416
Creinhardtii_Cre02.g077700.t1.1-CrTBC-PIB_RabGAP-TBC65..416
TBC-PIB
MrBayes/RAxML
1/95
0.9/75
0.8/50
Bbigemina_BBBOND_0209190-t26_1-p1_TBC-X2_RabGAP-TBC62..373
Bbovis_BBOV_IV005540-t26_1-p1_TBC-X2_RabGAP-TBC63..374
Cfelis_CF000604.mRNA-p1_TBC-X2_RabGAP-TBC511..822
Tannulata_TA21075-t26_1-p1_TBC-X2_RabGAP-TBC61..382
1/100
Tparva_TP01_0360-t26_1-p1_TBC-X2_RabGAP-TBC1..235
Torientalis_TOT_010000327-t26_1-p1_TBC-X2_RabGAP-TBC_RabGAP-TBC61..430
Bmicroti_BMR1_02g01160-t32_1-p1_TBC-X2_RabGAP-TBC45..356
Etenella_ETH_00026585-t26_1-p1_TBC-X2_RabGAP-TBC432..752
Gniphandrodes_GNI_014070-t26_1-p1_TBC-X2_RabGAP-TBC104..413
Hhammondi_HHA_261200-t26_1-p1_TBC-X2_RabGAP-TBC538..849
Tgondii_TGME49_261200-t26_1-p1_TBC-X2_RabGAP-TBC408..719
Ncaninum_NCLIV_026190-t26_1-p1_TBC-X2_RabGAP-TBC_RabGAP-TBC456..747
Sneurona_SN3_00301370-mRNA-1-p1_TBC-X2_RabGAP-TBC825..1137
Vbrassicaformis_Vbra_966.t1-p1_TBC-X2_RabGAP-TBC25..335
Pberghei_PBANKA_0933400.1-p1_TBC-X2_RabGAP-TBC54..365
berghei_PBANKA_0933400.2-p1_TBC-X2_RabGAP-TBC_RabGAP-TBC54..336
Pchabaudi_PCHAS_0910900.1-p1_TBC-X2_RabGAP-TBC54..365
Pyoelii_PYYM_0934800.1-p1_TBC-X2_RabGAP-TBC54..365
Pfalciparum_PF3D7_1114200.1-p1_TBC-X2_RabGAP-TBC55..366
Preichenowi_PRCDC_1112700.1-p1_TBC-X2_RabGAP-TBC55..366
Pgallinaceum_PGAL8A_00341100.1-p1_TBC-X2_RabGAP-TBC53..363
Prelictum_PRELSG_0911300.1-p1_TBC-X2_RabGAP-TBC52..363
Pknowlesi_PKNH_0911800.1-p1_TBC-X2_RabGAP-TBC54..365
Povale_PocGH01_09020500.1-p1_TBC-X2_RabGAP-TBC79..392
Pvivax_PVX_091355.1-p1_TBC-X2_RabGAP-TBC54..365
Chominis_ChTU502y2012_406g0790-t32_1-p1_TBC-X2_RabGAP-TBC94..408
Cparvum_cgd6_2190-t26_1-p1_TBC-X2_RabGAP-TBC94..408
Cmuris_CMU_007680-t26_1-p1_TBC-X2_RabGAP-TBC114..431
Cvelia_Cvel_17475.t1-p1_TBC-X2_RabGAP-TBC152..462
Pmarinus_EER11537_TBC-X2_RabGAP-TBC1..232
Skawagutii_Skav201913_TBC-X2_RabGAP-TBC_RabGAP-TBC1..172
Sminutum_symbB.v1.2.026200.t1_TBC-X2_RabGAP-TBC55..360
Smicroadriaticum_Smic26821_TBC-X2_RabGAP-TBC355..660
Pmarinus_EER09896_TBC-X2_RabGAP-TBC118..426
Smicroadriaticum_Smic12901_TBC-X2_RabGAP-TBC1627..1880
Sminutum_symbB.v1.2.014537.t1_TBC-X2_RabGAP-TBC232..529
Vbrassicaformis_Vbra_15646.t1-p1_TBC-X2_RabGAP-TBC400..704
Sminutum_symbB.v1.2.013573.t2_TBC-X2_RabGAP-TBC194..517
Sminutum_symbB.v1.2.013573.t1_TBC-X2_RabGAP-TBC1..232
TBC-X2
0.4

Figure 4.24 Detailed phylogeny of TBC-Q-related sequences

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 224 taxon by 178 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.


TBC-K
TBC-Q
TBC-X1
TBC-X4
TBC-X3
TBC-X5
MrBayes/RAxML
1/95
0.9/75
0.8/50
98/18
1/100
1/94
1/79
1/
89
0.62/23
0.4

Initial homology searching identified 700 putative Arf/Arl homologues across the dataset (Online Appendix Table 4.11), of which 176 were inferred to be ARF homologues, 78 Sar1 homologues (not studied further here), and 255 homologues of pan-eukaryotic Arl proteins, as inferred by their presence in the distantly related *H. sapiens*. Additional phylogenetic analyses of the remaining sequences revealed that 88 fell into one of three well-supported LSP clades (detailed below), while the rest remained unclassified.

Although mammals possess up to six distinct ARFs, many eukaryotes possess fewer, often as little as one homologue[31,32]. As expected, therefore, every organism in the dataset encodes at least one ARF homologue (Figure 4.25); within the Apicomplexa, the only expansion is within *Cryptosporidium spp.*, which possess two ARFs (Online Appendix Table 4.11). Phylogenetic analysis of ARF homologues was attempted but has not yet sufficiently resolved the internal structure of this clade to provide a more elaborate classification.

Among the Arl proteins identified, Arl1, Arl2, and ARFRP1L (which hit ARFRP1 in BLASTp searches but did not easily resolve in phylogenetic analysis) were well-conserved across the dataset (Figure 4.25). Despite this, some notable absences, including that of Arl1 in piroplasmids and *Cryptosporidium spp.*, and of ARFRP1L in all Apicomplexa except coccidians and *G. niphandrodes*, were present in the Apicomplexa even for these otherwise conserved Arl paralogues. Conservation among the other Arl paralogues was sparser: Arls 3, 5, 6, 8, 13, and 16 were identified in chromerids (Arls 3, 5, and 6 only) and in a selection of outgroup taxa (Figure 4.25).

LSP sequences were combined with pan-eukaryotic Arl homologues in an initial large-scale phylogenetic analysis to confirm the initial classifications (Figure 4.26). This analysis also yielded insight into the relationships between some of the Arl paralogues. For example, Arl1 and Arl5 consistently branch together (82.1 SH-aLRT, 90 UF-bb), as do Arl2 and Arl3 (82.6 SH-aLRT, 95 UF-bb), and Arl6 and Arl8, albeit with weaker support (76 SH-aLRT, 97 UF-bb, Figure 4.26); regardless of the exact statistical support, all of these relationships were consistently observed across all phylogenetic reconstructions and so appear stable.

In addition to these stable pan-eukaryotic clades, three other monophyletic clades were found in the course of the analysis, referred to simply as ArlX1, ArlX2, and ArlX3 (Figure 4.26). ArlX1 is conserved across Apicomplexa, chromerids, and dinoflagellates, but homologues were also found in some ciliates as well as the rhizarian *B. natans*, suggesting a possible origin at the base of the SAR clade and then subsequent loss in stramenopiles. ArlX2, surprisingly, was conserved across Apicomplexa but otherwise absent from both chromerids and dinoflagellates, suggesting that it represents a recently derived paralogue, sometime

Figure 4.25 Arl conservation

This figure shows the conservation of select Arls in study taxa. Each column represents an Arl paralogue, as indicated at the top, while each row represents a different taxon. Filled circles represent the presence of a confidently inferred homologue while empty circles represent the failure to identify a homologue. Cladogram on the left represents the relationships between study taxa. Colour-coding is as follows: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids.

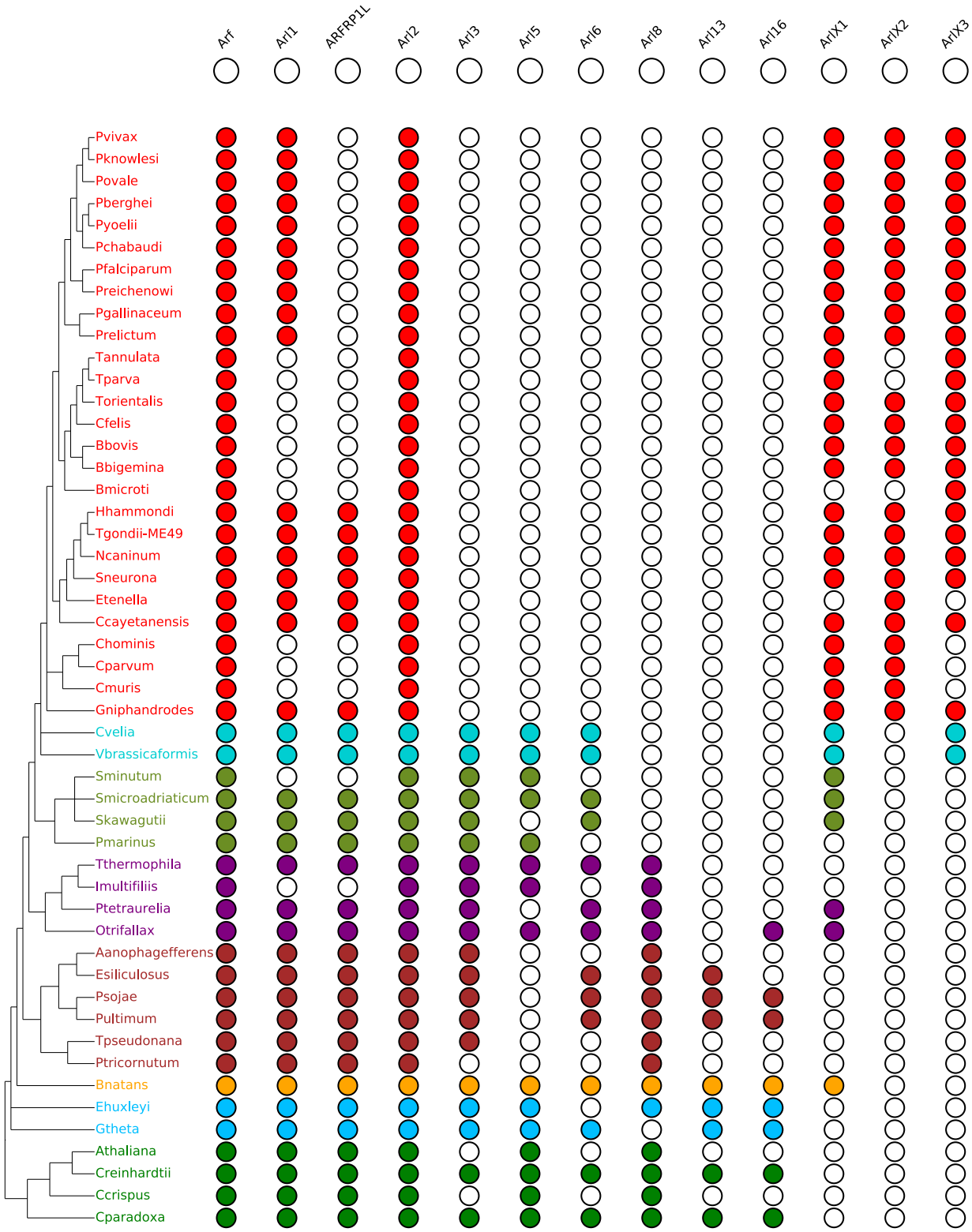
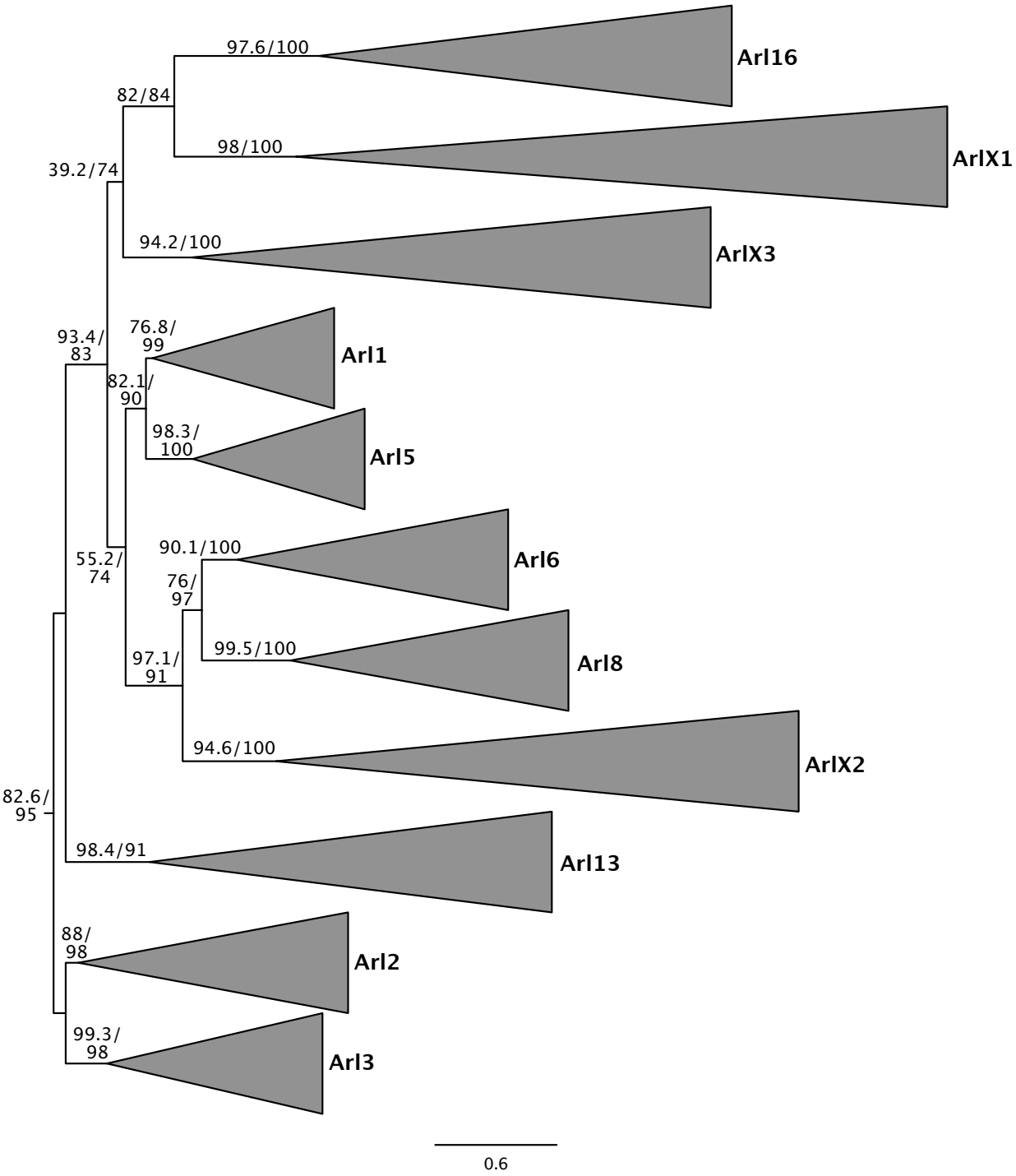

Arf
Arl1
ARFRP1L
Arl2
Arl3
Arl5
Arl6
Arl8
Arl13
Arl16
ArlX1
ArlX2
ArlX3
Pvivax
Pknowlesi
Povale
Pberghei
Pyoelii
Pchabaudi
Pfalciparum
Preichenowi
Pgallinaceum
Prelictum
Tannulata
Tparva
Torientalis
Cfelis
Bbovis
Bbigemina
Bmicroti
Hhammondi
Tgondii-ME49
Ncaninum
Sneurona
Etenella
Ccayetanensis
Chominis
Cparvum
Cmuris
Gniphandrodes
Cvelia
Vbrassicaformis
Sminutum
Smicroadriaticum
Skawagutii
Pmarinus
Tthermophila
Imultifiliis
Ptetraurelia
Otrifallax
Aanophagefferens
Esiliculosus
Psojae
Pultimum
Tpseudonana
Ptricornutum
Bnatans
Ehuxleyi
Gtheta
Athaliana
Creinhardtii
Ccrispus
Cparadoxa

Figure 4.26 Arl phylogeny

This figure shows an IQ-TREE phylogeny obtained from a 302 taxon by 161 position alignment of Arls. Clades are collapsed for simplicity and support values listed for all nodes in order of Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT)/ultra-fast bootstrap (UF-bb). Reliable clades are those with >= 80% SH-aLRT and >= 95% UF-bb support. Scale bar represents number of substitutions per site.


97.6/100
Arl16
82/84
98/100
ArlX1
39.2/74
94.2/100
ArlX3
93.4/
83
76.8/
99
Arl1
82.1/
90
98.3/
100
Arl5
55.2/
74
90.1/100
Arl6
76/
97
99.5/100
Arl8
97.1/
91
94.6/100
ArlX2
82.6/
95
98.4/91
Arl13
88/
98
Arl2
99.3/
98
Arl3
0.6

after the apicomplexan-chromerid split. Finally, ArlX3 is conserved across Apicomplexa, with the exception of *Cryptosporidium spp.*, and chromerids, but was not identified in any of the other myzozoan taxa included, suggesting an origin in the chromerid-apicomplexan ancestor (Figure 4.25).

Relationships of these LSPs to known Arl paralogues was not well-resolved in any analysis, except for the tendency of ArlX1 to consistently group with Arl16 (for example, 82 SH-aLRT and 84 UF-bb, Figure 4.26). ArlX2 also did not robustly group with any other clades in the majority of the analyses performed, although in the large-scale analysis did seem to group with the Arl6/8 group (97.1 SH-aLRT, 91 UF-bb, Figure 4.26). Finally, ArlX3 was the most enigmatic, failing to group with any other paralogues, although, it did branch most often as a basal outgroup to the combined ArlX1/Arl16 clade.

Given these initial relationships, an additional analysis was performed using more robust methods and excluding Arl2, Arl3, and Arl13, as these clades never grouped with any of the LSPs (Figure 4.27). Unfortunately, although the branching pattern observed using Bayesian reconstruction was consistent with likelihood-based approaches, support for the relationship of any Arl LSP to a pan-eukaryotic paralogue was lacking. Furthermore, additional analyses including only one LSP at a time failed to yield any significantly supported relationships. Notably, ArlX1 always branched as sister to Arl16; although this relationship is not well-supported in Figure 4.27 (0.66 pp, 26 bb), the ArlX1/Arl16 group achieved significant posterior probability (>= 0.8) in some analyses, although bootstrap support was always lacking.

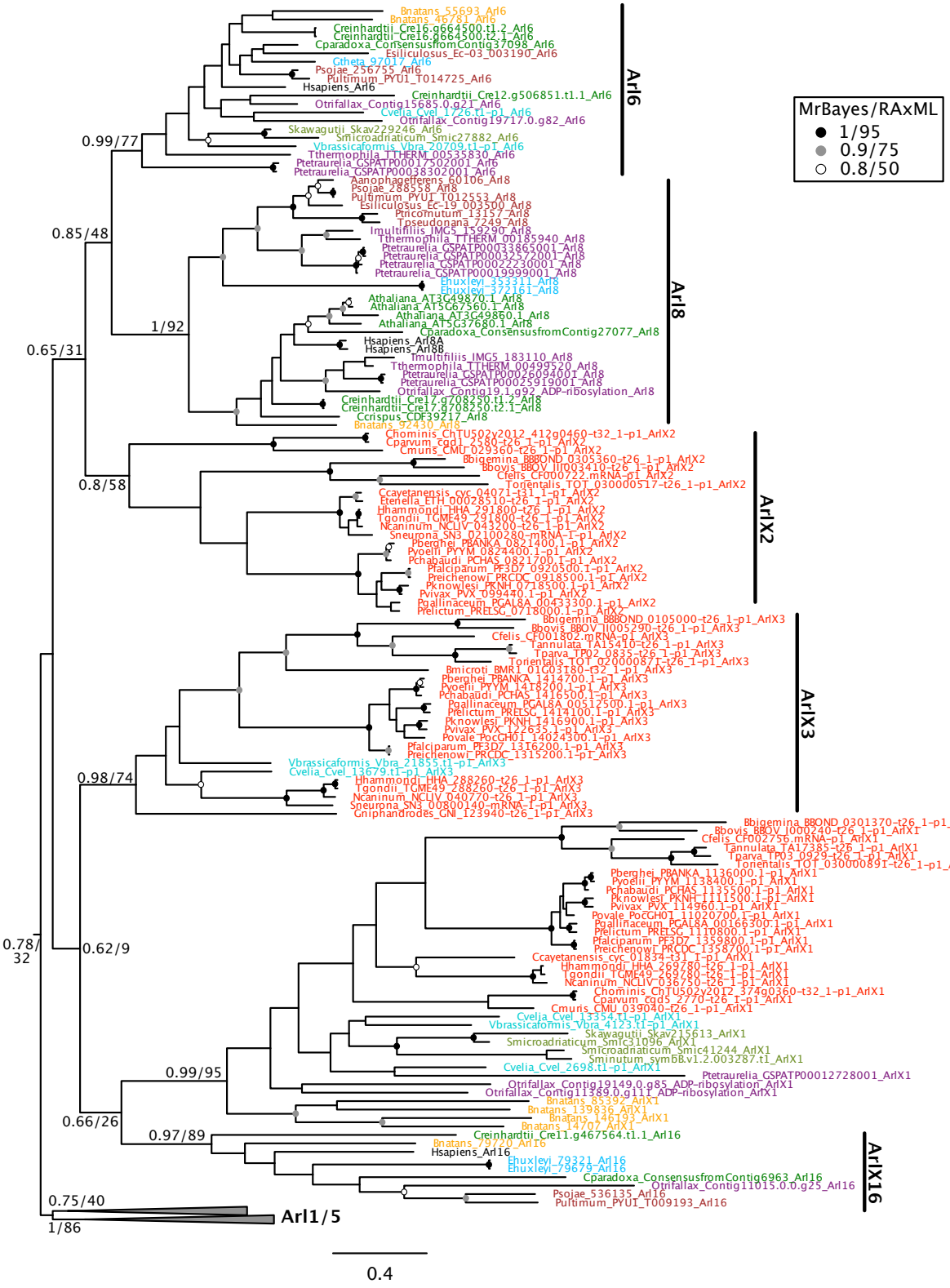
Hence, despite encoding a paucity of pan-eukaryotic Arl homologues, apicomplexans possess three Arl LSPs. It is likely that resolution of these LSPs to distinct pan-eukaryotic clades is complicated by their divergent nature, and future work is required to pin-point such relationships with satisfactory support.

4.5 Discussion

As discussed in section 4.1, a rational argument may be made for the need for Apicomplexa to encode additional MTS paralogues to mediate specific trafficking to the additional organelles found in this group. Three examples have been previously reported in the literature to represent additional paralogues in lineages including the Apicomplexa, namely the SAR (plus *G. theta*)-specific Rab1A[513] as well as the alveolate-specific Rab11B[390] and TBS proteins[36,686]. Hence, even before this work it was clear that Apicomplexa potentially encode additional MTS paralogues not found in other eukaryotes.

Figure 4.27 Detailed phylogeny of ArlX1/X2/X3

This figure shows results of Bayesian (MrBayes) and maximum-likelihood (RAxML) analysis of a 202 taxon by 155 position alignment. The best Bayesian topology is shown with posterior probability and associated bootstrap support for important clades listed on the figure (pp/bb); internal node support is denoted as per figure legend. Clades are labelled in bold; some clades are collapsed for clarity. Scale bar represents number of substitutions per site. Labels are colour-coded by taxon: red, Apicomplexa; teal, chromerids; olive, dinoflagellates; purple, ciliates; brown, stramenopiles; orange, *B. natans,* light blue, *E. huxleyi* and *G. theta*; green, archaeplastids; black, marker sequences.


MrBayes/RAxML
1/95
0.9/75
0.8/50
Arl6
Arl8
ArlX2
ArlX3
ArlX1
ArlX16
Arl1/5
0.99/77
0.85/48
1/92
0.65/31
0.8/58
0.98/74
0.78/32
0.62/9
0.99/95
0.66/26
0.97/89
0.75/40
1/86
0.4

This thesis describes the further presence of 15 other MTS LSPs in Apicomplexa: Stx12A, VAMP7B, Ykt6B, ArlX1, ArlX2, ArlX3, Rab1K, Rab5B, Rab5C, RabX1, TBC-X1, TBC-PI/X2, TBC-X3, TBC-X4, and TBC-X5 (Table 4.1). The number of identified LSPs is striking for several reasons. Firstly, the consideration that previous studies have implicated loss of MTS machinery in Apicomplexa, including the significant losses within the ESCRTs, APs, and MTCs as discussed in Chapter 2 (Figure 2.1). Secondly, because this pattern of loss was still associated with the MTS families described in this Chapter; for example, Apicomplexa encode only a minimal subset of the inferred pan-eukaryotic Rab GTPases (Figure 4.13) yet have undergone further paralogous expansion of some of these retained paralogues to yield additional complexity. Lastly, it is of note that this thesis does not represent an exhaustive survey of all MTS families, and that further LSPs may still be present which remain, to this point, undescribed.

As mentioned in section 4.1, at least two of these LSPs provide evidence that taxonomic distribution and function of MTS paralogues can be correlated, at least to some extent. Hence, it is interesting to consider, relying on the arguments of functional homology put forward in Chapter 3 and the notion of correlation between taxonomic distribution and function, what the function of some of these novel paralogues may be. Although additional paralogues of known MTS factors would theoretically be free to diverge and acquire new functions, the two examples touched upon previously, Rab5C and Rab11B, suggest that, at least in some cases, the function remains relatively close to that of the pan-eukaryotic paralogue; in the case of Rab5C, trafficking to the micronemes and rhoptries through the endosomal system[253], and for Rab11B, the transport and/or recycling of material to a structure near the cell surface[390].

Stx7 and Stx12 are both Stx12-like (StxE) proteins in mammalian cells. Stx7 is primarily responsible for homotypic late endosome and late endosome-lysosome fusion[688–691], which it accomplishes by forming complexes with the Q-SNAREs Vti1b and Stx8 and the R-SNAREs VAMP7 and VAMP8[688,691]. Consequently, Vti1, VAMP7, and VAMP8 have all been shown to be important for the fusion of structures, including autophagosomes, and late endosomes, with the lysosome[692–696]. Conversely, Stx12 (also referred to as Stx13), has been attributed diverse roles, including in recycling endosome to plasma membrane trafficking[697], autophagosome maturation[698], and homotypic early endosome fusion[699,700]; the latter two functions were also shown to involve Stx12-Vti1a interactions, but typically involve different Qc and R-SNAREs as those for Stx7. Overall though, it is clear that all of these components function at varying points throughout the endosomal system.

Table 4.1 Novel Arl Paralogues in *T. gondii*.

Overview of identified LSPs in *T. gondii*. Each LSP is listed along with the corresponding gene family (family), ToxoDB identifier (ToxoID), its presumed taxonomic spread (conservation), presumed relationship to a pan-eukaryotic paralogue (if known; presumed pan-euk origin), and the phenotypic score associated with the gene in a genome-wide CRISPR/CAS9-based screen[354]. Question marks represent uncertainty about the provenance of some LSPs, while NI (not identified) marks an inability to discern a confident relationship to any pan-eukaryotic paralogue.

Name	Family	ToxoID	Conservation	Presumed pan-euk origin	Phenotypic sccore
Stx12B	SNARE (Qa)	TGME49_204060	Myzozoa	Stx12	-0.97
VAMP7B	SNARE (R)	TGME49_248100	Myzozoa	VAMP7	-0.55
Ykt6B	SNARE (R)	TGME49_299180	Chromerids + Apicomplexa	Ykt6	-1.62
Rab1A	Rab	TGME49_258130	SAR + *G. theta*	Rab1	-3.58
Rab1K	Rab	TGME49_249170	Myzozoa	Rab1	-0.79
Rab5B	Rab	TGME49_207460	Alveolata	Rab5 (?)	-1.35
Rab5C	Rab	TGME49_219720	Myzozoa	Rab5	-4.24
Rab11B	Rab	TGME49_320480	Alveolata	Rab11	-5.06
RabX1	Rab	TGME49_277840	Myzozoa	Rab28 (?)	-1.35
TBC-X1	TBC	TGME49_289820	Alveolata	TBC-Q	-1.14
TBC-X2/PI	TBC	TGME49_261200	Myzozoa + Archaeplastid	NI	-1
TBC-X3	TBC	TGME49_203910	Alveolata	TBC-Q	0.53
TBC-X4	TBC	TGME49_213325	Alveolata	TBC-Q	-2.61
TBC-X5	TBC	TGME49_226850	Alveolata	TBC-Q	0.54
TBS1	ArfGEF/TBC	TGME49_266830	Alveolata	TBC-N	0.3
TBS2	ArfGEF/TBC	TGME49_312300	Alveolata	TBC-N	-1.19
ArlX1	Arl	TGME49_269780	SAR	NI	-1.28
ArlX2	Arl	TGME49_291800	Apicomplexa	NI	0.2
ArlX3	Arl	TGME49_288260	Chromerids + Apicomplexa	NI	-2.49

Therefore, it is interesting to note the presence of LSPs in Apicomplexa and related taxa for Stx7/12 and VAMP7 (Figures 4.4, 4.6, 4.10, and 4.11), as well as the additional homologues of Vti1 encoded (Online Appendix Table 4.6). Given the putative identity of micronemes and rhoptries as derived endolysosomes[236,431], and the remaining presence of other endosomal compartments in apicomplexan cells, including the ELC and VAC[245,247,248], one hypothesis is that at least some of these additional SNAREs localize to, and function in fusion at, micronemes and rhoptries. Further support for this theory is present in the correlation between the taxonomic distribution of these compartments and the identified Stx12B and VAMP7B LSPs: both are restricted to the Myzozoa (Figures 4.4 and 4.10). In addition to this prospect, and specifically regarding the presence of multiple Vti1 homologues in Apicomplexa (Online Appendix Table 4.6), it is interesting to note that Vti1a has been described also to function in dense core granule (DCG) biogenesis[701], similar to Stx6[702]. As discussed in section 3.3, dense granules in Apicomplexa superficially resemble DCGs, and disruption of TgStx6 impairs proper dense granule formation[409]. Therefore, at least one Vti1 homologue in Apicomplexa may also function in dense granule biogenesis.

Ykt6 is a R-SNARE associated with a number of diverse functions in cells including ER-Golgi transport[703], release of constitutive secretory vesicles[704], as well as transport of material through the late endosomal system, and both heterotypic and homotypic vacuole fusion, including with autophagosomes[705–707]. Given these diverse roles, it is unclear what role the additional Ykt6B LSP identified (Figures 4.10 and 4.12) might play. It is likely that a full understanding of Ykt6 function in Apicomplexa will require molecular characterization of both Ykt6A and Ykt6B paralogues.

Within the Arl proteins, it is curious to note that three Arl paralogues associated with cilia/flagella, Arl3, Arl6, and Arl13[708–714], are absent from Apicomplexa (Figure 4.25), which lack flagella in most of their lifecycle stages[270]. Interestingly, chromerids encode Arl3 and Arl6, but lack Arl13, and both *C. velia*[471] and *V. brassicaformis*[208] are similarly flagellated in only a portion of their lifecycle. Also of note is the apparent lack of Arl8, which has been shown to have key functions in lysosome positioning and control of lysosome fusion, the latter function likely mediated through interaction with the multimeric HOPS tethering complex[715–717]. Given that none of the myzozoan taxa studied to date encode the HOPS-specific subunit Vps39, and most Apicomplexa do not possess Vps41 either (Figure 2.1), it is tempting to speculate that VpsC complex function in Myzozoa is modified from that found in other eukaryotes.

In addition, three Arl LSPs were identified in Apicomplexa, one of which, ArlX2, represents the only LSP identified in this thesis to be restricted solely to this group (Figure 4.25). Although none of the Arl LSPs strongly grouped together with a pan-eukaryotic Arl paralogue (Figure 4.27), ArlX1 does appear to consistently group together with Arl16. Unfortunately, only one study in the literature investigating Arl16 function was identified; Arl16 was found to mediate antiviral activity by binding to the Retinoic acid-inducible gene I (RIG-I), a pattern recognition receptor of the mammalian immune system[718]. Hence, it is unclear what potential role any of the Arl LSPs might play based solely on functional homology arguments; these ArlX paralogues are the subject of Chapter 5.

Among the Rab proteins, Rab1A[513] and Rab11B[390] had been described previously as LSPs; the results presented here confirm and extend those of previous analyses, and suggest that Rab1A is indeed found across SAR, as well as in the cryptophyte *G. theta*, while Rab11B is an alveolate-specific Rab11 paralogue. As will be further discussed below, Rab5B and Rab5C had been described in the literature from a molecular perspective, but their identity as LSPs had not previously been recognized.

Rab1 is present at the ER but is also known to interact with Golgi-resident proteins such as GBF1, GM130, and Giantin[719,720]. Interaction of Rab1 with p115 is important for COP-II vesicle formation and resulting ER-Golgi trafficking[721,722], whereas interaction with GBF1 is important for COP-I-mediated ER retrieval from the Golgi[719]. In a previous study of Rab function in *T. gondii*, TgRab1B (the pan-eukaryotic Rab1 paralogue) was found to localize in the vicinity of the ER/Golgi, as expected, whereas TgRab1A was found more associated with endosomal compartments, as evident by co-localization with proM2AP and VP1[253]. The other Rab1 paralogue identified, Rab1K, is unique among the apicomplexan Rabs in that Rab1K homologues possess an N-terminal extension with variable numbers of Kelch repeats predicted to form β-propellers (Figure 4.16). Despite this unusual property, Rab1K groups with Rab1A sequences, consistent with a scenario in which Rab1 duplicated early in the evolution of the diaphoretickes (to yield Rab1A and Rab1B) and then again in the myzozoan common ancestor to yield Rab1K, which was subsequently lost in all Apicomplexa except for coccidians (Figure 4.13). Unfortunately, Rab1K was not included in the previous large-scale localization study in *T. gondii*[253] and it currently remains uncharacterized.

Among the Rabs, Rab5 is perhaps one of the best studied, as it has numerous roles in both tethering at, and maturation of, early endosomal compartments[124,150,371–373]. An early study into Rab conservation within the Apicomplexa suggested the presence of three Rab5 paralogues[369], which has been generally accepted within the literature. However, among these Rab5 paralogues, PfRab5B was previously shown to be N-terminally myristoylated[374], rather

than C-terminally prenylated like Rab5A and Rab5C. In addition, Rab5B could not be knocked out in *P. berghei*, leading to the conclusion that it is essential and its function cannot be compensated for by either of the other two Rab5 paralogues[374]. In *T. gondii*, only disruption of TgRab5A and TgRab5C function was found to alter localization of a subset of microneme and rhoptry proteins[253], again suggesting that Rab5B has a different function than Rab5A and Rab5C. The results presented here offer a potential explanation for these observations: either that Rab5B is a divergent Rab5 paralogue, or that it does not represent a *bona fide* Rab5 paralogue, but rather a separate Rab5-related protein that is conserved among alveolates (Figure 4.17). It is clear that Rab5C is a *bona fide* myzozoan-specific Rab5 paralogue, as it groups together was Rab5A with good support (Figure 4.17), and, at least in *T. gondii*, appears to possess similar function as well[253].

The role of TgRab5A and TgRab5C in trafficking to apical secretory organelles has been established, although this study relied on overexpression of both wild-type and dominant negative versions, which may produce artefactual results[253]. Future studies employing conditional disruption (knockdown and/or knockout) strategies may help to further understand the specific functions associated with each paralogue. Conversely, the function of Rab5B remains elusive; in *T. gondii* TgRab5B was shown to localize to the cell surface as well as to intracellular puncta, and overexpression was lethal[253]; TgRab5B has also been shown to interact with the retromer subunit TgVps26[375]. In *P. falciparum*, PfRab5B was similarly localized to both the plasma membrane to the food vacuole, a lysosome-like organelle in which ingested hemoglobin is trafficked for degradation[374]. More recently, PfRab5B was shown to partially co-localize with the AP-2μ[723], which would be consistent with a role for this protein either in uptake and trafficking to the digestive compartments, recycling from internal compartments back to the plasma membrane, or both.

Surprisingly, although all the other Rab LSPs could be confidently associated with a pan-eukaryotic Rab clade, the myzozon-specific LSP RabX1 did not group strongly with any of the Rab clades included in this study (Figure 4.19). Although lacking support, it does appear as though RabX1 is most similar to a subset of Rabs including Rab7, and among them, most similar to Rab28. Like Rab1K, this Rab was not included in the previous large-scale localization study of Rabs in *T. gondii*[253], and no characterization has been performed on it. Unfortunately, little is known about its closest paralogue, Rab28; dysfunction has been linked to cone-rod dystrophy, which itself may be linked to phagocytic uptake of cone outer segments[724]. It has also been suggested as a substrate for the GAP activity of the similar TBC1D1 and TBC1D4 proteins, and plays a role in trafficking of the GLUT4 glucose transporter[725]. In trypanosomes,

it has been suggested that Rab28 associates with endosomal machinery such as retromer and the ESCRTs to mediate both endosomal trafficking and degradation[726].

Intriguingly though, a recent study in *Caenorhabditis elegans* employing transcriptional profiling demonstrated that Rab28-GTP is capable of concentrating near the base of cilia and undergoing bi-directional intraflagellar transport (IFT). Although *rab28* null mutants did not show significant ciliary defects, overexpression of GDP- or GTP-bound variants perturbed normal ciliary function[727]. As the apical complex employs conserved flagellar machinery[230,231], it is intriguing to speculate that RabX1 might be involved in trafficking to the apical complex and/or apical organelles. This presumed function fits with RabX1 taxonomic distribution; it is conserved across Apicomplexa and found in at least one representative of both the chromerids and dinoflagellates as well (Figure 4.13). In addition, Rab28 has been shown, albeit without strong statistical support, to branch as a sister clade to another IFT-related GTPase, RabL4/IFT27, both in this study (Figure 4.19) and by previous detailed phylogenetic studies[110]. Alternatively, or perhaps, in addition, RabX1 may be involved in IFT/trafficking to the flagella in flagellated stages, for example, in the microgametes of *T. gondii*[270]. Despite these attractive hypotheses, further understanding of RabX1 localization and function will require detailed molecular studies.

Although TBC-PIA and TBC-PIB were previously reported to represent TBC paralogues restricted to a subset of the Archaeplastida[123], the results presented here (Figures 4.20 and 4.23) suggest that TBC-PI is also conserved in Myzozoa. This would be consistent with either gain of TBC-PI in the common ancestor of these lineages and subsequent loss in multiple SAR lineages, or with a horizontal transfer event between the two groups. Further studies investigating the conservation of TBC-PI across eukaryotes will be required to further understand the timing of its emergence and pattern of its retention.

Members of both the TBC-N (including TBC1D12 and TBC1D14) and TBC-Q (including TBC1D1, TBC1D4, EVI5, EVI5L, GYP5, and GYP5L) clades have been shown to bind Rab11, although none display Rab11 GAP activity *in vitro*[728,729]. The roles associated with these diverse paralogues are extensive: TBC1D1 and TBC1D4 have been suggested to function in trafficking and/or retention of the GLUT4 transporter[730], GYP5 and GYP5L function in polarised exocytosis[731–733], and TBC1D14 is involved in autophagy[734,735]. The common thread among all of these functions appears to be that TBC-Q and TBC-N homologues are primarily involved in endosomal trafficking events.

The presence of additional TBC-N TBC domains, in the form of TBS proteins, as well as the presence of multiple TBC-Q-like proteins (Figure 4.20) in Apicomplexa echo the results obtained for the SNARE proteins, and suggest that the emergence of micronemes and

rhoptries during apicomplexan evolution may have necessitated additional machinery for endosomal trafficking. Unlike what was observed for the SNARE proteins though, the majority of these additional TBC paralogues are also present in ciliates (Figure 4.20), suggesting a more broadly conserved function, such as trafficking to the alveoli/IMC. It is also possible that such a role may involve one or more of these additional paralogues interacting with the alveolate-specific Rab11B (Figure 4.18). However, these hypotheses remain to be tested.

This chapter has presented a detailed investigation into several large paralogous gene families in the Apicomplexa and associated outgroup taxa. It is clear that MTS evolution in Apicomplexa has involved not only loss of machinery, as outlined in Chapter 2 and throughout this Chapter, but also gain of additional machinery through the mechanism of gene duplication. In instances where duplication from a pan-eukaryotic paralogue can be clearly asserted, and the gene's function has been characterized, it appears that these LSPs mediate similar functions as their associated pan-eukaryotic paralogue. It is possible that this additional machinery primarily mediates trafficking of different cargoes, or to different subsets of a family of organelles, as shown for Rab5 proteins and the micronemes[253]. Regardless, the few examples present to date in the literature, combined with the large number of LSPs identified here, suggest that the gene duplication may represent a powerful mechanism to generate complexity within the MTS, as suggested by the OPH mechanism.

In cases where a relationship to a pan-eukaryotic paralogue could not be confidently assigned, including for the novel Arl paralogues, initial hypotheses based on the principle of functional homology cannot be generated, and it is unclear what roles these proteins may play. Therefore, Chapter 5 presents localization data for ArlX1, ArlX2, and ArlX3, as well as an in-depth characterization of ArlX3 function.

5. Characterization of three novel Arl proteins in *T. gondii*

5.1 Overview of novel paralogue characterization in *T. gondii*

Chapter 4 provided a detailed characterization of seven MTS families across the breadth of apicomplexan diversity and included multiple outgroups for comparison. In total, 18 lineage-specific paralogues (LSPs) were identified, including multiple members within the same gene family (for example, six LSPs from the Rab GTPase family). Based on the rationale proposed in Chapter 3, it is expected that at least some of these novel paralogues would be involved in trafficking to novel organelles. This has already been described in the literature for several of the Rab GTPases.

As discussed in section 1.5.1.1, the IMC is a network of flattened membranous sacs connected by proteinaceous sutures considered to be homologous to the alveoli of other alveolates[236], which are a defining feature of the group Alevolata[17,736]. The novelty of alveoli as a synapomorphic feature is supported by molecular data for alveolins and other IMC-resident proteins, which suggest that the majority of these proteins are restricted to alveolates[736,737], as expected. However, the exact relationship of alveoli with another cellular structure, which would convincingly establish homology, is difficult to pin down. Regardless of the exact nature of this structure, it is clear that trafficking to the plasma membrane and IMC of *T. gondii* is dependent on both the pan-eukaryotic Rab11A and alveolate-specific Rab11B[390,391].

Of the three Rab5 paralogues, Rab5A, Rab5B, and Rab5C, in Apicomplexa, Rab5A represents the pan-eukaryotic paralogue. The alveolate-specific Rab5B has been most extensively studied in *Plasmodium*, where it is suggested to be involved in the trafficking of ingested red blood cell cytosol to the digestive vacuole[374]. Additionally, Rab5B has been localized to the plasma membrane both in *T. gondii*[253] and *P. falciparum*[374], and in small cytosolic vesicles in *P. falciparum* by cryo-immunogold EM (iEM)[723]. In comparison, Rab5C, which is a clear duplication of Rab5A in Myzozoa (Figure 4.17), is involved in trafficking of a subset of microneme and rhoptry proteins in *T. gondii*[253]. TgRab5A and TgRab5C represent an interesting example of gene duplication and function. Both localize similarly to the Golgi/post-Golgi region in *T. gondii* and have similar detrimental effects with overexpression on rhoptry resident proteins, as well as a subset of microneme resident proteins (including, at least, MIC3, 8, and 11). Combined with stimulated emission depletion (STED) microscopy data that showed a lack of co-localization between some sets of microneme proteins, this was interpreted as the presence of independent microneme populations that use different trafficking pathways[253].

Considering the above examples, it does appear that the additional paralogues studied in the literature to date function in trafficking to novel organelles. However, the pattern is not

entirely clear, as pan-eukaryotic paralogues also perform similar functions, at least in the case of Rab11A/B and Rab5A/C. For TgRab11A, the differences in localization, and the key importance for TgRab11A, but not TgRab11B, for delivery of the plasma membrane protein TgSAG1[390,391], suggests that the functional divergence between the two paralogues is appreciable, and suggests that the IMC necessitated a second Rab11 copy to mediate trafficking that could not be accomplished through a single orthologue. For TgRab5A and TgRab5C, this difference is less clear based on available data; it would be interesting to determine what, if any, differences in function exist between these two paralogues. However, it is clear from these examples that the phylogenetic spread of each LSP mirrors the structure to which it directs trafficking; the alveolate-specific Rab11B mediates trafficking to the IMC (which is homologous to alveoli), and the myzozoan-specific Rab5C mediates trafficking to micronemes and rhoptries. As discussed in Chapter 3, this is exactly the pattern that would be expected under the OPH and assuming functional homology: namely, the restricted presence of a paralogue only in the lineage for which its function is required.

Of the gene families explored in Chapter 4, only the Rab GTPases have been studied to an appreciable extent in Apicomplexa. This stands to reason, as Rab GTPases are comparatively well-studied in model eukaryotes such as human, yeast, and plant cells, and are known to be key regulators of membrane trafficking[112]. Another class of G proteins involved in membrane trafficking are the ARF/Arl members of the Ras superfamily, first introduced in section 1.3.1.1. Although Arls have not been studied as extensively in model systems as Rabs, several have been shown to be important in facilitating organelle positioning/trafficking. This includes Arl and ARFRP1, which are involved in the recruitment of tethering factors at the *trans*-Golgi[28], Arl8, which functions in lysosome tethering/positioning[716,717], and Arls 3/6/13, which are involved in flagellar integrity/trafficking[708–712,714,738]. However, the vast majority of known Arl paralogues remain uncharacterized, or have poorly defined roles.

Given the observed role for some Rab LSPs (Rab5C/11B) in trafficking to novel organelles, the lack of characterization for most Arl paralogues, and the presence of three novel Arl paralogues (Figure 4.X) in Apicomplexa, it appears prudent to explore the function of Arl LSPs in Apicomplexa. Hence, this chapter explores the three Arl LSPs in *T. gondii*.

5.2 Materials and Methods

5.2.1 Parasite and host cell culture, and transfection

Human foreskin fibroblast (HFF; ATCC® designation SCRC-1041™) cells were grown in tissue culture grade plastics in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% foetal bovine serum (FBS), 2 mM L-glutamine, and 25 mg/mL gentamicin. *T. gondii* strain RH parasites were cultured on confluent HFF monolayers. All cells were maintained at 37ºC and 5% CO_2. In cases when fully egressed cultures were required but not available, late-stage (32 parasite) vacuole-containing cultures were mechanically released by scratching the host cell monolayer using a flat-based tool and then passing the culture medium through a 25-gauge needle three times. Transfection was carried out using an AMAXA 4D-Nucleofector™ (Lonza). Transient transfections used ~10 µg of purified DNA and ~$1x10^5$ freshly egressed parasites, whereas stable transfections used ~20-30 µg of purified DNA and ~$1x10^6$ freshly egressed parasites. In the case of stable transfection, integration was selected for by supplementing culture medium with 78µM mycophenolic acid and 230µM xanthine; selected pools were then cloned by limiting dilution in 96 well plates and individual clones picked and analyzed.

5.2.2 Genomic DNA Isolation, cloning, and PCR

To isolate genomic DNA from parasites, roughly $1x10^6$ fully egressed parasites were collected and then gDNA was isolated using Qiagen DNeasy Blood and Tissue Kit, as per the manufacturer's instructions. Amplification of DNA segments for cloning used Q5® high-fidelity DNA polymerase (NEB) whereas diagnostic PCR used standard *Taq* DNA polymerase (NEB). All restriction enzymes for cloning were purchased from NEB, using the high-fidelity (HF) versions, where available. Plasmid preps were made using Qiagen QIAprep Spin Miniprep Kit, as per the manufacturer's instructions. All primers used in this study for cloning and diagnostic PCR confirmation of stably transfected cell lines are provided in Supplementary Table S5.1 and were synthesized by Eurofins (UK). Information of all vector constructs used are provided in Supplementary Table S5.2.

Endogenous tagging was carried out using the ligation-independent cloning (LIC) approach[355]. A C-terminal fragment of each gene to be tagged was amplified by PCR to contain a unique restriction site not present in the LIC vector backbone (for each gene, primers LIC fwd and LIC rev, Supplementary table S5.1). The LIC vector (pG514, Supplementary Table S5.2) was digested with *PacI* and then both backbone and insert were treated with T4 DNA polymerase (NEB) prior to ligation. For the vector, 6µl 10X NEB buffer 2, 3µl 100mM DTT, 2.4µl 100mM dGTP, 1.5µl T4 DNA polymerase, 0.6µl 100X BSA, and 1.2µg of digested vector prep were mixed on ice and the final volume adjusted to 60µl. For the PCR insert, 2µl NEB buffer 2, 1µl 100mM DTT, 0.8µl 100mM dCTP, 0.5µl T4 DNA polymerase, 0.2µl 100X BSA,

and 0.2pm PCR product were mixed on ice and the final volume adjusted to 20µl. Each separate prep was then incubated in a PCR thermocycler: 30 minutes at 22°C, 20 minutes at 75ºC, 4ºC hold; reactions were held on ice prior to annealing. To anneal, 1µl of treated vector and 2µl of treated insert were mixed and incubated for 10 minutes at room temperature before addition of 1µl 25mM EDTA and five minutes additional incubation. Annealed vectors were held on ice and used to transform competent bacteria. Each vector was linearized using the corresponding unique restriction enzyme prior to transfection.

CRISPR/CAS9 modification of parasites used a single vector encoding both CAS9-NLS-YFP enzyme and pTgU6-gRNA (pG474, Supplementary table S5.2). To generate vectors containing a specific gRNA, the gRNA was synthesized as complementary primers (for each gene, primers gRNA fwd and rev, Supplementary Table S5.1). Primers were suspended in annealing buffer (10mM Tris pH 7.5, 50mM NaCl, and 1mM EDTA), heated to 95ºC, and then allowed to cool to room temperature. The parental vector was digested with *BsaI* and then gRNA inserts were ligated into the digested vector using T4 DNA ligase (NEB).

Vector maps for all vectors constructed in this thesis, including the parental pG474 and pG514 vectors, are included in the Online Appendix and Supplementary Figure S5.1.

5.2.3 Induction of the ArlX3 inducible knockdown

The TATi-ArlX3 knockdown line was induced by supplementing the culture medium with 1 µg/mL anyhydrotetracycline (ATc) for the indicated amount of time in each assay.

5.2.4 Immunofluorescence analysis (IFA)

For IFAs, HFF cells were seeded onto glass coverslips and confluent monolayers infected with parasites. Parasites were fixed using 3% paraformaldehyde (PFA) at room temperature for 20 minutes before being washed three times with 1X phosphate-buffered saline (PBS). Subsequently, coverslips were permeabilized and blocked using blocking buffer (3% BSA and 0.2% Triton X-100 in 1X PBS (PBS-TX-100)) for one hour at room temperature. Primary antibodies were added to blocking buffer at the dilutions indicated in Supplementary Table S5.3, and cells stained for one hour at room temperature before being washed three times with PBS-TX-100. Similarly, Alexa Fluor 488, 594, or 647-conjugated secondary antibodies (Life Technologies) were added to blocking buffer at 1:3000 dilution and cells stained for one hour at room temperature in the dark. Samples were washed three more times with PBS-TX-100 and then coverslips were mounted using either mounting media alone or mounting media supplemented with 4’,6-diamidino-2-phenylindole (DAPI). Non-

permeabilizing IFAs were performed as above, but without the addition of TX-100 to blocking and wash buffers. The source of all antibodies is listed in Supplementary Table S5.3.

5.2.5 Western blotting and protein detection

Approximately $1x10^7$ freshly egressed parasites were harvested and pelleted by centrifugation at 5000 rpm followed by a single wash with 1X PBS. The parasite pellet was lysed on ice with NP-40 buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Nonident P-40, and 4mM EDTA) and incubated for five minutes on ice. Insoluble material was pelleted by centrifugation at 14,000 rpm at 4ºC. The supernatant was placed in a new tube together with 10X NuPage™ Sample Reducing Agent (Invitrogen™) and 4X loading buffer (125mM Tris-HCl pH 6.5, 50% v/v glycerol, 4% w/v SDS, 0.2% w/v orange G). Samples were boiled for 10 minutes at 95ºC, loaded onto 12% Mini-Protean® TGX™ Precast polyacrylamide gels (BioRad), and run at 130V. Samples were transferred to nitrocellulose using a Mini-Protean® transfer tank containing 1L of transfer buffer (48mM Tris, 39mM glycine, and 20% methanol) running at 400mA for one hour. Membranes were blocked using 5% skim milk powder in 1X PBS at room temperature for one hour. Primary antibodies were added at the appropriate concentration (Supplementary Table S5.3) in blocking buffer (5% skim milk powder in 1X PBS + 0.2% Tween-20 (PBS-TW-20) for one hour. Membranes were washed three times with PBS-TW-20, before addition of IRDye680RD and IRDye800RD secondary antibodies (Li-Cor, used at 1:15,000) in blocking buffer for a further hour. Membranes were washed three times in PBS-TW-20 followed by an additional wash in 1X PBS to remove Tween-20 prior to imaging. Detection of infrared signal was performed using a Li-Cor Odyssey with Image Studio 5.0 software (Li-Cor).

5.2.6 Structured illumination microscopy (SIM)

SIM imaging used an ELYRA PS.1 microscope (Zeiss) equipped with a Plan Apochromat 63x, 1.4 NA oil immersion lens and CoolSNAP HQ camera (Photometrics). SIM processing of captured images used ZEN Black software (Zeiss) and all subsequent processing used Fiji[739].

5.2.7 Quantitative fluorescence microscopy

Time-course quantification of TATi-ArlX3 knockdown protein levels was carried out as follows. Parental Δku80-TATi and TATi-ArlX3 parasite lines were induced for the relevant time periods as described above and processed for IFA using α-myc and α-GAP45 antibodies. Images were captured using the same excitation parameters on a Leica DiM8 widefield fluorescence microscope equipped with a HC PL APO 100x/1.44 oil immersion lens (Leica) and

C13440-20C CMOS camera (Hamamatsu). Image files were loaded into Fiji and z-stacks collapsed into 2D images by summation of individual slices. For each vacuole, a region of interest (ROI) was traced in Fiji, using the α-Gap45 signal to indicate the bounding region of parasites in each vacuole. These ROIs were subsequently used to measure area, integrated density, and mean grey value in the α-myc channel. For each ROI, similar measurements were also obtained for the local background in the α-myc channel where no vacuoles were present. Subsequently, corrected total cell fluorescence was calculated as integrated density – (vacuole area x mean background fluorescence), as described previously[292]. One hundred random vacuoles were quantified for each of three independent experiments.

5.2.8 Electron Microscopy

Samples were fixed in 2.5% glutaraldehyde and 4% paraformaldehyde in 0.1M cacodylate buffer, pH 7.2, washed in 0.1M cacodylate buffer, pH 7.2, and post-fixed in 1% OsO4 for 1 hour on ice. After several washes in the same buffer, the samples were stained *en bloc* with 0.5% uranyl acetate in water for 30 minutes. Afterwards, samples were washed with water, dehydrated in ascending acetone series and epoxy resin embedded. Ultrathin sections (50 nm thick) were sectioned in a Leica Ultramicrotome UC7 and collected on copper grids covered with formvar.

For cryo-immunolabeling, the samples were fixed in phosphate buffer, pH 7.2, containing 4% freshly prepared formaldehyde. After several washes in the same buffer, they were embedded in 10% gelatin at 37º C for 30 minutes. The material was spun down and the samples were left on ice for 30 minutes. After confirming the gelatin was solid, the pellet was removed from the tubes and infiltrated overnight in 2.1 M sucrose and rapidly frozen by immersion in liquid nitrogen. Cryo-sections (70 nm thick) of the frozen material were obtained at −120ºC using an Ultracut cryo-ultramicrotome (Leica Microsystems). The cryo-sections were collected on formvar-coated nickel grids, thawed, and put on a cushion of 2% gelatin. The grids were left for 20 minutes at 37ºC and then blocked in PBS containing 3% bovine serum albumin for 1 hour. After this time, they were incubated in the presence of primary antibody. Then they were washed several times in blocking buffer and incubated with 15 nm gold-conjugated Protein A (Aurion). The grids were washed several times in the blocking buffer, dried, and contrasted in a mixture of methylcellulose/uranyl acetate.

All images were captured on a Jeol 1200 transmission electron microscope (JEOL, Japan) operating at 80kV and analyzed/processed with Fiji software[739].

5.2.9 Five-day plaque assay

For plaque assays, $1x10^3$ freshly egressed parasites were added to a confluent HFF monolayer with or without addition of 1 µg/mL ATc. After five days, cultures were washed once with PBS and then fixed with ice-cold methanol for 20 minutes. Methanol was removed and cells stained with Giemsa, followed by three washes with PBS. All plaques in 10 random fields of view were measured using Fiji for three independent experiments.

5.2.10 Gliding assay

For gliding assays, $1x10^6$ freshly egressed parasites were suspended in gliding buffer (1mM EDTA and 100mM HEPES) and allowed to glide on glass coverslips coated with FBS for 30 minutes prior to fixation with 3% PFA. An IFA was performed using α-SAG1 primary antibody under non-permeabilising conditions to label deposited trails. One hundred random parasites were assessed for the presence/absence of trails in three independent experiments, and both the mean and SEM calculated.

5.2.11 Invasion assay

For invasion assays, $5x10^4$ freshly egressed parasites were allowed to invade confluent HFFs for one hour before four washes with PBS were performed to remove uninvaded parasites. 24 hours later, cells were fixed with 3% PFA and processed for IFA with α-GAP45 primary antibody. The number of vacuoles in 15 random 40X fields of view were counted for three independent experiments, and the mean and SEM calculated.

5.2.12 Egress assay

For egress assays, $5x10^4$ freshly egressed parasites were allowed to invade confluent HFFs as per invasion assays. 36 hours later, culture media was exchanged for pre-warmed DMEM with 2 µM calcium ionophore (A23187) to induce egress. Five minutes after media exchange, cells were fixed with 3% PFA and subsequently stained with α-SAG1 antibody under non-permeabilising conditions (together with DAPI to assess intracellular vacuoles). One hundred random vacuoles were assessed for egress ability in three independent experiments, and the mean and SEM calculated.

5.2.13 Marker mis-localization analysis

To assess the effect of ArlX3 knockdown on organelle markers, Δku80-TATi or TATi-ArlX3 parasites were grown with or without ATc for the relevant time period before being mechanically released and allowed to invade new confluent HFF monolayers, as described under invasion assay. 24 hours later, cells were fixed and processed for IFA using the

appropriate primary antibodies. One hundred random vacuoles were assessed for the phenotypes described in the main text for three independent experiments, and the mean and SEM calculated.

5.2.14 Statistical analysis

All statistical analysis was performed in R v3.6.1[740]. Comparisons among multiple means used one-way ANOVA followed by post-hoc Tukey's HSD test when assumptions of normality and equal variance were not significantly violated; in cases where violation did occur, Kruskal-Wallis followed by post-hoc Dunn's test was used instead. Comparison between plaque sizes in plaque assay used two-way ANOVA and means within each group were compared by a Wilcoxon signed rank test. Comparison of multiple populations within a single group (e.g. for phenotypic analysis) used Chi-square followed by post-hoc Fisher's exact test. All plots were made in R and the first instance of significant difference from controls indicated.

5.3 Results

5.3.1 Endogenous tagging and localization of Arl LSPs in *T. gondii*

To begin the characterization of all three Arl LSPs, localization studies were performed by vector-mediated endogenous gene tagging with a 3xHA epitope tag. A C-terminal fragment of each genomic locus was amplified with primers containing overhangs compatible with ligation-independent cloning, before annealing into a backbone containing an in-frame 3xHA tag and downstream HXGPRT selectable marker (Figure 5.1A, materials and methods)[355,741,742]. In all cases, endogenous tagging was carried out in a parasite line lacking the Ku80 gene product, but encoding dimerizable Cre recombinase to facilitate future genome editing and knockout strategies (parasite line RHΔku80::DiCre, hereafter referred to as DiCre). Clones isolated by serial dilution following drug selection were confirmed through integration PCR, using a gene-specific forward primer upstream of the intended integration site and a reverse primer in the 3xHA tag (P1 and P1' in Figure 5.1A). In all tagged lines, a clear band was observed at the expected size (ArlX1 – 1.7kb; ArlX2 – 2kb; ArlX3 – 0.47kb), but no amplification was observed in parental DiCre parasites (Figure 5.1B). To further confirm the accuracy and specificity of endogenous gene tagging in these lines, and to confirm that each Arl LSP is expressed in asexual tachyzoites, western blotting was also performed using anti-HA antibodies. Similar to the integration PCR, a clear band was observed at the expected size (ArlX1 – 29kDa; ArlX2 – 27kDa; ArlX3 – 63kDa) in each tagged line, but no signal was observed in parental DiCre parasites (Figure 5.1C).

To establish the localization of each Arl LSP, indirect immunofluorescence analysis (IFA) was carried out on fixed samples and imaging performed using super-resolution 3-dimensional structured illumination microscopy (3D-SIM, materials and methods). Intracellular parasites were typically imaged at the 4-parasite vacuole stage, as in larger vacuoles parasites can become densely clustered and obscure signal localization.

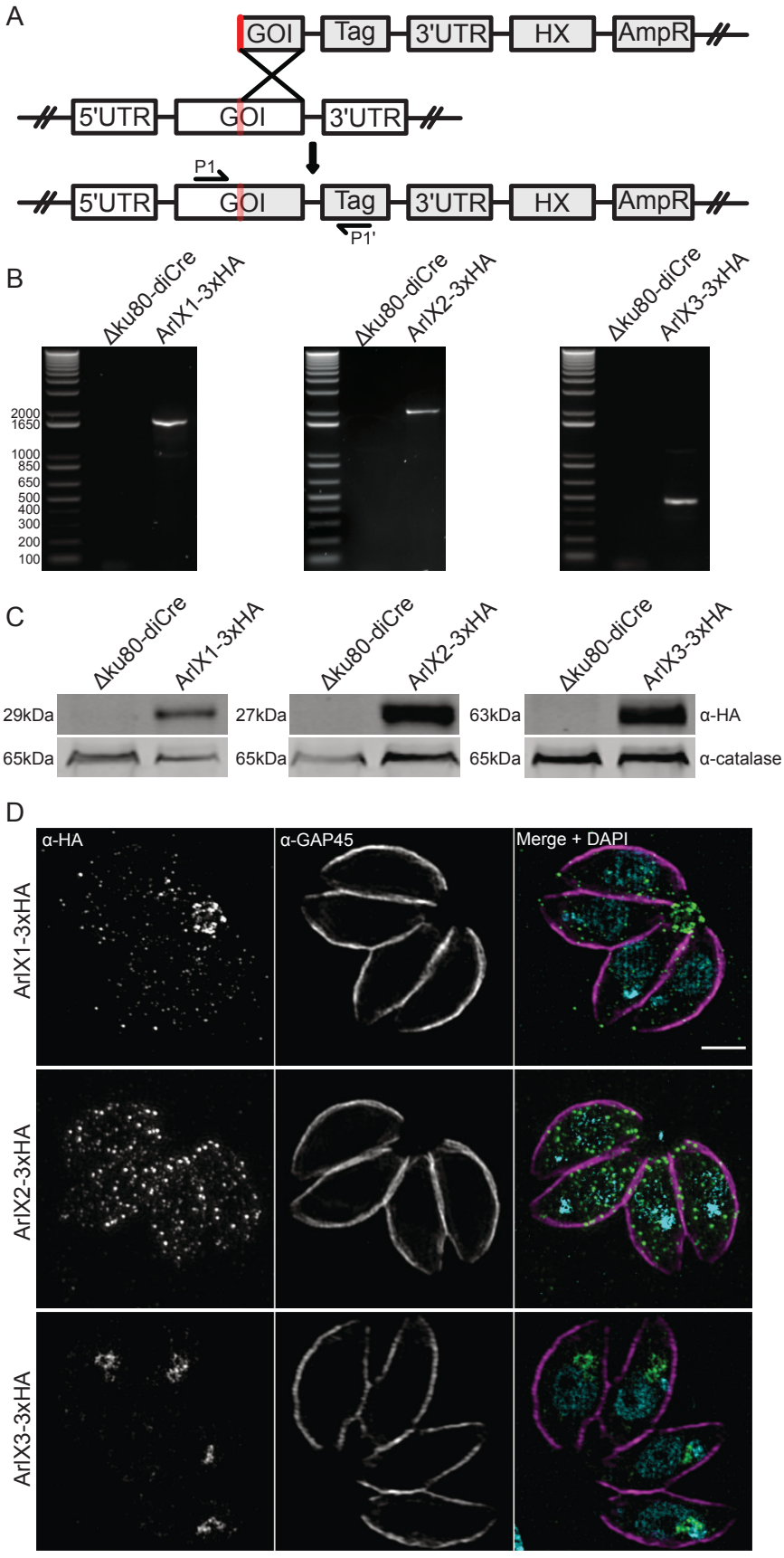
Anti-HA staining of intracellular DiCre::ArlX1-3xHA parasites, hereafter referred to just as ArlX1-3xHA, revealed a dynamic localization. Each individual zoite within a vacuole had a single bright concentration of signal at the apical end of the parasite, with additional puncta present throughout the cell body (Figure 5.1D, top row). In addition, some vacuoles had extensive signal in the space between parasites at the basal end, as demarcated by staining with an antibody to the glideosome protein GAP45, which bridges the parasite plasma membrane and IMC[288]. This signal, which is presumably present in the filamentous network connecting parasites, as well as in the basal body, was sometimes observed as one or more filament-like structures, but frequently adopted a ring-like structure (Figure 5.1D, top row). This is similar to the staining pattern recently described for filamentous (F)-actin[314].

In comparison, anti-HA signal in ArlX2-3xHA parasites was consistent, with multiple bright puncta in close proximity to the IMC, as assessed by anti-GAP45 staining. (Figure 5.1D, middle row). Interestingly, signal was frequently observed at the base of the IMC, which opens at the basal end of the parasite[743]. Finally, anti-HA staining in ArlX3-3xHA parasites was observed as a single conglomeration of staining just apical to the nucleus, with some signal occasionally observed further in the apical portion of the parasite (Figure 5.1D, bottom row). This localization pattern closely mirrors that of markers for the Golgi and/or TGN, including ERD[401], GRASP[243], and GalNAc[244,245,420].

The HA signal in the ArlX1-3xHA line was intriguing and prompted further investigation via co-staining and iEM. Overall, ArlX1 localization was dynamic, appearing to various extents throughout the filamentous network and within intracellular parasites between vacuoles in the same sample. Although apical staining was almost always observed, staining within the filamentous network was observed only in ~25-50% of vacuoles. To determine the relative position of the apical signal, ArlX1-3xHA parasites were transiently transfected with an expression construct for mCherry (mC)-α-tubulin, which is incorporated into both the conoid and sub-pellicular microtubules[226]. In these parasites, the anti-HA signal localizes apical to a large concentration of mC-α-tubulin, indicating that ArlX1 is present at the very apical tip of the parasite, above the conoid (Figure 5.2A, top row); apical localization was also confirmed through staining with an antibody to the IMC apical cap protein ISP1, which marks the IMC compartment closest to the conoid[421] (Figure 5.2A, middle row). Furthermore, signal was

Figure 5.1 Endogenous gene tagging of Arl LSPs in *T. gondii*

This figure demonstrates the strategy used, and results obtained, for C-terminal endogenous tagging of all three Arl LSPs. A) Schematic of endogenous gene tagging strategy; a tagging vector is created for each gene, which is linearized prior to transfection by a unique restriction enzyme present only within the target gene sequence (vertical red line). B) PCR confirmation of endogenous gene tagging; primer pair used as in panel A, predicted sizes: ArlX1 – 1.7kb; ArlX2 – 2kb; ArlX3 – 0.47kb. C) Western blot confirmation of 3xHA tagging; molecular weights are listed to the left and antibodies to the right. Predicted sizes: ArlX1 – 29kDa; ArlX2 – 27kDa; ArlX3 – 63kDa D) 3D-SIM images of intracellular ArlX1-3xHA (top row), ArlX2-3xHA (middle row), and ArlX3-3xHA (bottom row) tachyzoites. Abbreviations: GOI, gene of interest; UTR, untranslated region; HX, HXGPRT selectable marker. Scale bar is 2μm.


A
GOI
Tag
3'UTR
HX
AmpR
5'UTR
GOI
3'UTR
P1
5'UTR
GOI
Tag
3'UTR
HX
AmpR
P1'
B
Δku80-diCre
ArlX1-3xHA
Δku80-diCre
ArlX2-3xHA
Δku80-diCre
ArlX3-3xHA
2000
1650
1000
850
650
500
400
300
200
100
C
Δku80-diCre
ArlX1-3xHA
Δku80-diCre
ArlX2-3xHA
Δku80-diCre
ArlX3-3xHA
29kDa
27kDa
63kDa
α-HA
65kDa
65kDa
65kDa
α-catalase
D
α-HA
α-GAP45
Merge + DAPI
ArlX1-3xHA
ArlX2-3xHA
ArlX3-3xHA

observed in close proximity, and appeared to follow, the course of sub-pellicular microtubules (Figure 5.2A, top row), which was confirmed by investigation of individual z-stack sections as well (Supplementary Figure S5.2). Additional transient transfection of mC-MORN1, which is incorporated into the basal ring and localizes near the parasite nucleus during endodyogeny[744], confirmed that the signal between parasites is outside of tachyzoite cell bodies (Figure 5.2A, bottom row). To gain further insight into this latter observation, ArlX1-3xHA parasites were transiently transfected with a construct encoding a chromobody (an antibody fragment derived from camel) against F-actin fused to an Emerald fluorescent protein (CB-Em). As described previously, Cb-Em signal appeared diffusely throughout individual parasite's cytosol and in intense filamentous structures in the network between parasites (Figure 5.2B)[314]. Investigation of both maximum-intensity projections (MIP; Figure 5.2B, top row) and individual z-stack sections (Figure 5.2B, bottom rows), revealed anti-HA signal coincidental with Cb-Em, suggesting that ArlX1 is present in the network.

To confirm these unusual localization patterns, ArlX1-3xHA parasites were prepared for iEM and imaged (materials and methods). Gold labeling was observed at the extreme apical end of the parasite (Figure 5.3A), confirming the IFA localization data (Figure 5.2A, top row). In addition, more specific localizations than could be observed in IFA were determined. Gold labeling was frequently observed at the limiting membrane of vesicular structures in the apical end of the parasite and at the limiting membrane of rhoptries as well; occasional labeling of micronemes was also observed (Figure 5.3B). Additionally, labeling was also observed in the vacuolar space between parasites, in close proximity to membranous material and long tubules (Figure 5.3C). Overall, these results confirm and extend those based on IFA studies and suggest that ArlX1 is associated with multiple organelles in the apical end of the parasite, as well as in the network between parasites within a vacuole.

Although the observed localization pattern of ArlX2-3xHA is intriguing (Figure 5.1D), it was not pursued further, in part due to the predicted dispensability of ArlX2 in the asexual stage (see section 5.3.2) and limitations of time and resources.

Although the initial IFA staining pattern (Figure 5.1D) is consistent with Golgi localization, several organelles are present in close proximity in this region of the parasite. To further establish the localization of ArlX3, ArlX3-3xHA parasites were transiently transfected with the following markers of the early secretory pathway: P30-GFP-HDEL, which primarily marks the ER[240], ERD-GFP and GRASP-RFP, which both mark the *cis/medial*-Golgi[243,401], and GalNAc-YFP, which marks the TGN, and possibly further post-Golgi structures[244,245,420]. Anti-HA signal did not overlap appreciably with P30-GFP-HDEL (Figure 5.4, top row), but did appear

Figure 5.2 Detailed localization of ArlX1

This figure shows additional immunofluorescence localization data for intracellular ArlX1-3xHA parasites. A) ArlX1-3xHA parasites were transiently transfected with vectors encoding mCherry-α-tubulin (top row) or mCherry-MORN1 (bottom row), or stained with anti-ISP1 antibodies, and imaged by 3D-SIM. Arrows point to apical staining in all three rows; in the top row, arrowheads point to example regions of close apposition between ArlX1 staining and mCherry-α-tubulin staining. B) ArlX1-3xHA parasites were transiently transfected with a vector encoding Cb-Em targeting F-actin and imaged by 3D-SIM; top row shows maximum-intensity project (MIP), while bottom rows show individual z-stack sections, as indicated. All scale bars are 2μm.

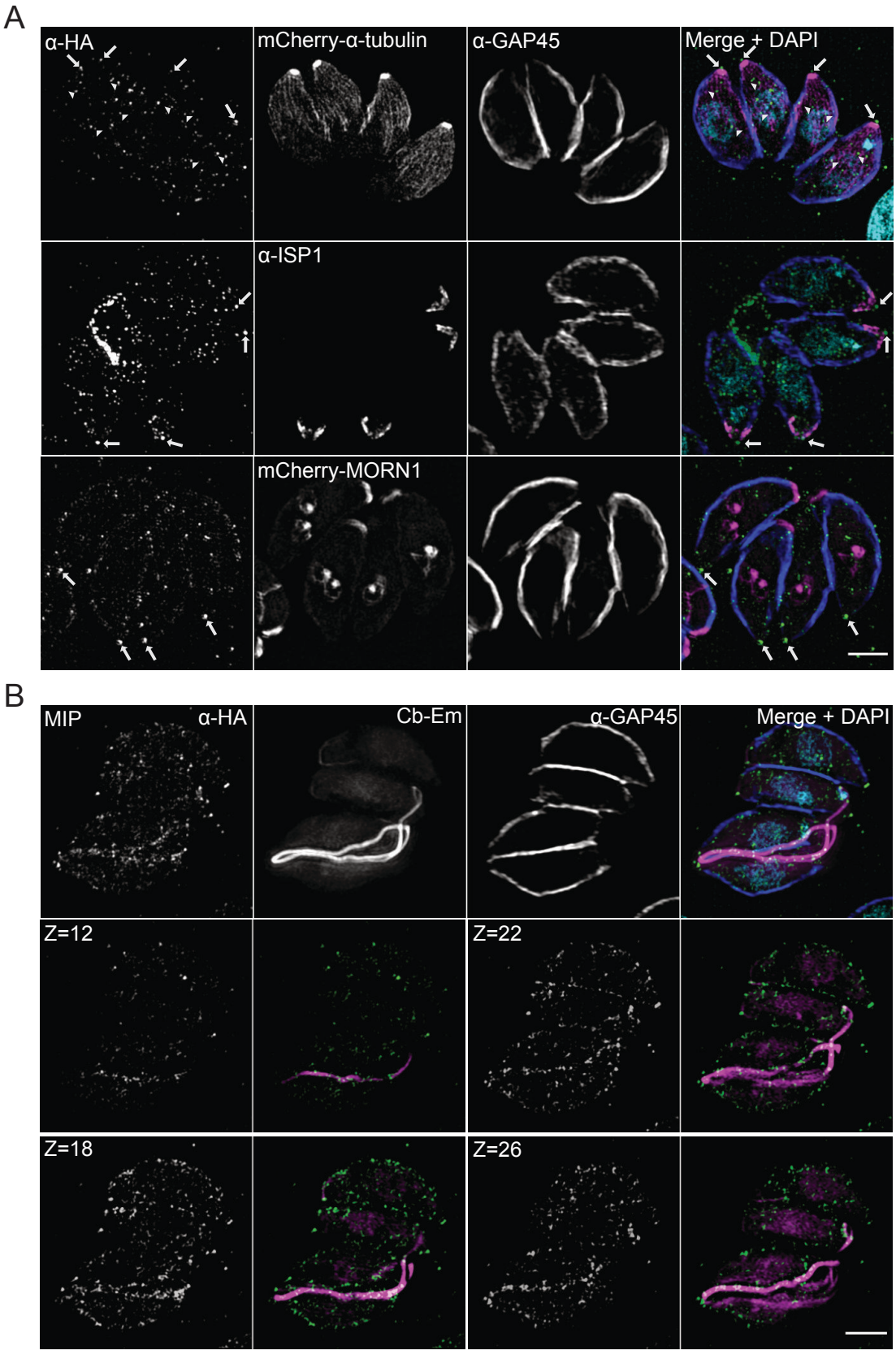
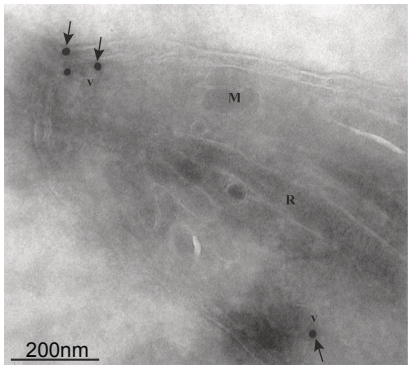
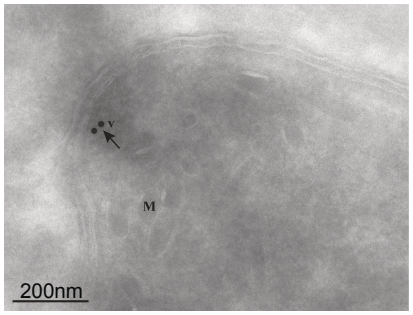

A
α-HA
mCherry-α-tubulin
α-GAP45
Merge + DAPI
α-ISP1
mCherry-MORN1
B
MIP
α-HA
Cb-Em
α-GAP45
Merge + DAPI
Z=12
Z=22
Z=18
Z=26

Figure 5.3 Cryo-immunogold electron microscopy of ArlX1

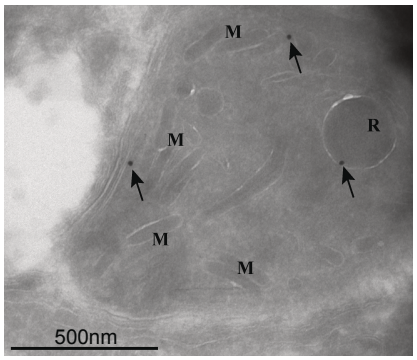
This figure shows the results obtained from performing iEM on intracellular ArlX1-3xHA tachyzoites. Arrows point to gold labeling in each section. Abbreviations: M, microneme; R, rhoptry; v, vesicle; Tg, *T. gondii* tachyzoites, RB, residual body. Scale bars are as indicated on each image. Sample preparation and imaging was performed by Dr. Leandro Lemgruber.

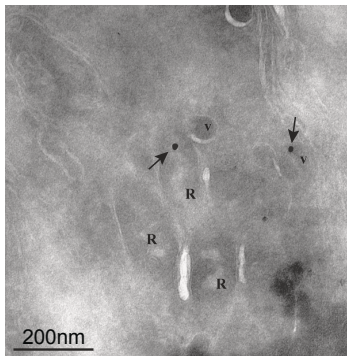
A

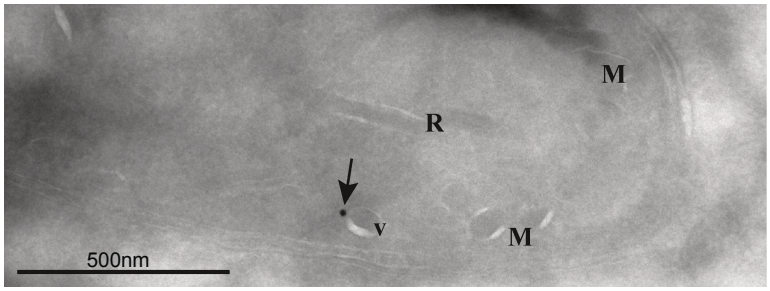

V
M
R
V
200nm


V
M
200nm

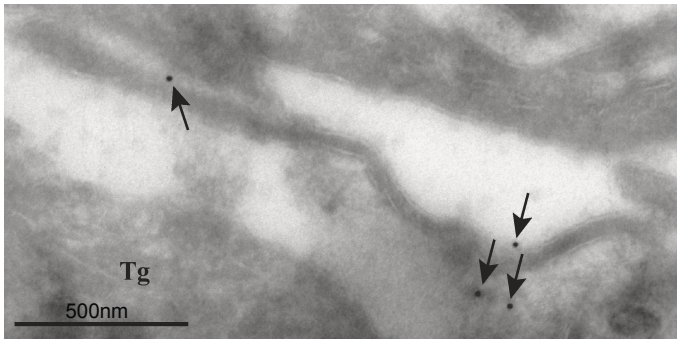
B

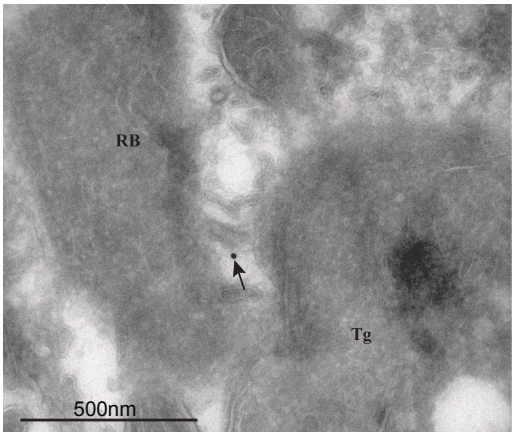

M
R
M
M
M
500nm


V
V
R
R
R
200nm


M
R
V
M
500nm

C


Tg
500nm


RB
Tg
500nm

in close proximity to the signal from the other three constructs (Figure 5.4, bottom rows), suggesting that ArlX3 is primarily present near the Golgi/TGN. To confirm these results, ArlX3-3xHA parasites were prepped for iEM (materials and methods). Gold labeling was observed at the Golgi (in 7 out of 12 section), as expected, but also occasionally at the limiting membrane of vesicular and tubular structures in the vicinity of the nucleus/Golgi; in some sections, extensive labeling of micronemes was also observed (Figure 5.5). Together, these results suggest that ArlX3 is present near the Golgi/TGN, but also on early post-Golgi carriers and micronemes.

5.3.2 Generation of an inducible TgArlX3 knockdown line

A recent study reported a large scale CRISPR/CAS9-based screen of all *T. gondii* genes during asexual growth, providing a "phenotypic score" for each gene based on the relative abundance of gRNA constructs in a pool before and after transfection[354]. In this scheme, scores at or below about -2.5 suggest a gene product is essential for asexual growth, scores between -2.5 and 0 suggest that a gene product contributes to fitness, and scores higher than 0 suggest that a gene product is dispensable for asexual growth[354]. The three Arl LSPs, surprisingly, fall into each of these categories: ArlX2 is predicted to be dispensable (score of 0.2), ArlX1 important, but not essential, for growth (score of -1.28), and ArlX3 essential for asexual growth (score of -2.49). ArlX3 also represents a unique phylogenetic pattern: it is present in Myzozoa, which suggests it might be involved in trafficking to organelles of the apical complex, but absent from both *Cryptosporidium spp.* and dinoflagellates, which have lost the apicoplast[457,458] and a true apical complex/apical organelles, respectively[232]; hence, ArlX3 was chosen for further analysis.

A conditional knockdown line was established using a combination of CRISPR/CAS9 and the tetracycline-transactivator (TATi) system[285,366]. A gRNA targeted to induce a double-strand break just upstream of the START codon of the endogenous ArlX3 gene was synthesized and ligated into an appropriate vector for transient expression (materials and methods). Concurrently, an ~2.5kb cassette was amplified from a vector backbone containing an HXGPRT selectable marker, seven *tetO* repeats, a minimal heterologous promoter derived from *sag1*, and a myc epitope tag, all flanked by 50bp homology arms and designed to place the myc tag at the N-terminus of ArlX3 in frame for proper expression (Figure 5.6A). Proper integration of this cassette is expected to abrogate any gene expression based on the endogenous promoter, as it would be unable to read through the 3' untranslated region of the selectable marker. Transfection into Δku80-TATi parasites resulted in Δku80-T7S1-myc-ArlX3 transfectants, hereafter referred to as TATi-ArlX3. Proper integration was confirmed for both

Figure 5.4 Detailed localization of ArlX3

This figure shows additional immunofluorescence localization data for intracellular ArlX3-3xHA parasites transiently transfected with vectors encoding P30-GFP-HDEL (top row), ERD-GDP (top middle row), GRASP-RFP (bottom middle row), and GalNAc-YFP (bottom row). All images captured using 3D-SIM and the scale bars is 2μm.

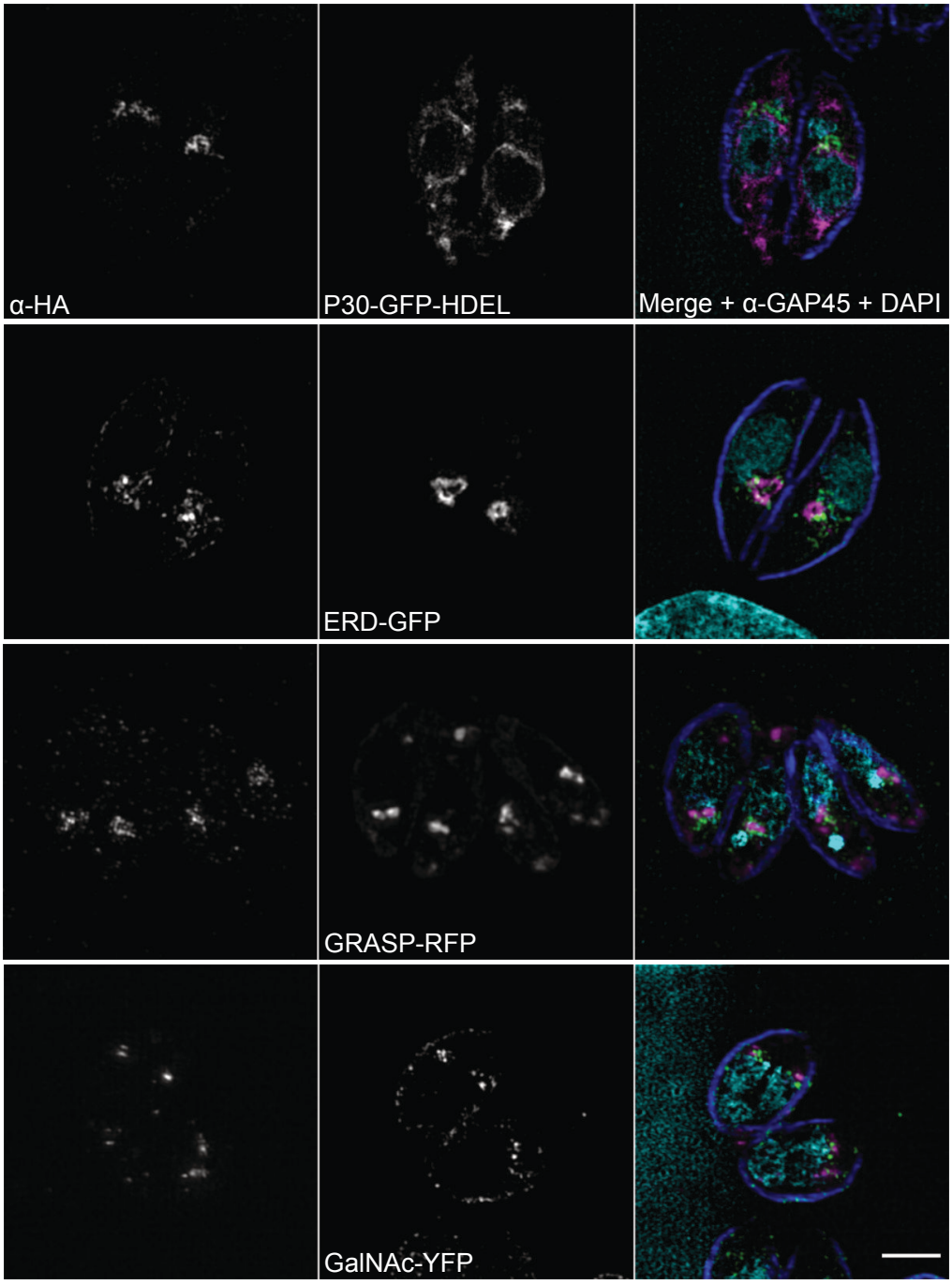
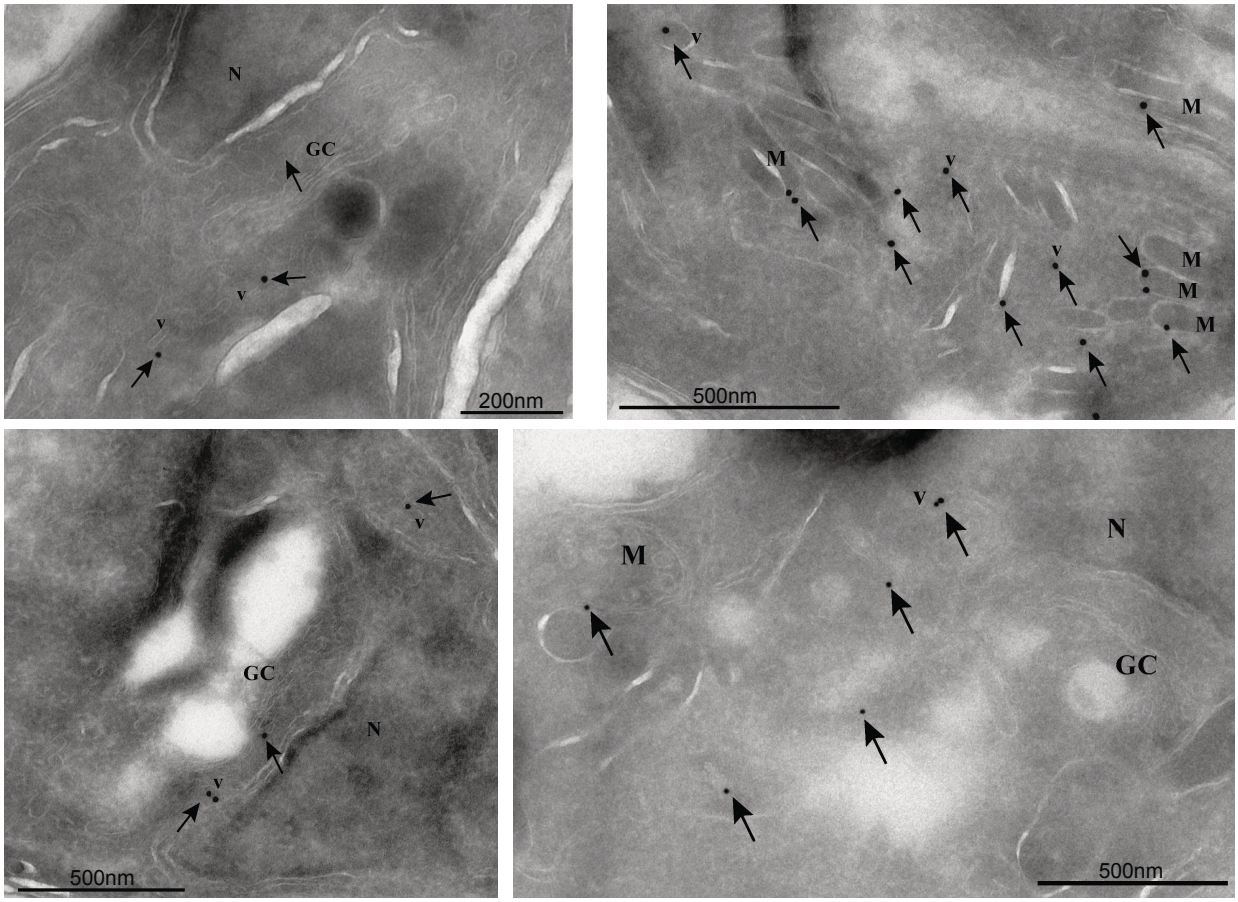

α-HA
P30-GFP-HDEL
Merge + α-GAP45 + DAPI
ERD-GFP
GRASP-RFP
GalNAc-YFP

Figure 5.5 Cryo-immunogold electron microscopy of ArlX3

This figure shows the results obtained from performing iEM on intracellular ArlX3-3xHA tachyzoites. Arrows point to gold labeling in each section. Abbreviations: GC, Golgi complex; M, microneme; v, vesicle; N, nucleus. Scale bars are as indicated on each image. Sample preparation and imaging was performed by Dr. Leandro Lemgruber.


N
GC
v
v
v
200nm
v
M
M
v
v
M
M
M
500nm
v
GC
N
v
500nm
M
v
N
GC
500nm

the 5’ and 3’ ends of the cassette by integration PCR (Figure 5.6B), and a single band of the expected size (60kDa) was observed by western blotting with an anti-myc antibody in the TATi-ArlX3, but not Δku80-TATi, line (Figure 5.6C). To confirm that replacement of the endogenous promoter did not alter ArlX3 localization, IFA was performed following transient transfection of the same panel of Golgi markers as for ArlX3-3xHA (Figure 5.4). Anti-myc antibody staining revealed the same pattern of localization as observed for ArlX3-3xHA (Figure 5.6D), suggesting that any change in gene expression timing and/or expression level did not significantly affect the localization of ArlX3. Additionally, co-staining with an antibody to the dynamin-related protein DrpB, which localizes to an undetermined structure in the vicinity of the Golgi[425], showed striking overlap (see white arrow in bottom row of Figure 5.6D), including in regions of the cell clearly separate from the main Golgi-localized signal (white arrow in Figure 5.6D, bottom row). These results confirm that ArlX3 localizes in the vicinity of the Golgi in *T. gondii* and suggest that promoter replacement does not appreciably alter ArlX3 localization.

In order to understand any phenotypes apparent with ArlX3 knockdown, knockdown kinetics were established using a combination of quantitative IFA (materials and methods) and western blotting. As expected, addition of anhydrotetracycline (ATc), which blocks transcription[285], resulted in a time-dependent decrease in anti-myc signal (Figure 5.7A). Quantification of corrected fluorescence confirmed that fluorescence dropped significantly after only six hours of ATc induction, and reached background level by 24 hours, which remained until 48 hours, the latest time point measured (Figure 5.7B). These results were confirmed by western blotting (Figure 5.7C). Hence, ArlX3 knockdown is rapid, with lack of detectable protein by 24 hours post-induction.

Figure 5.6 Generation of an inducible ArlX3 knockdown line

This figure demonstrates the strategy used to obtain an inducible ArlX3 knockdown line (TATi-ArlX3). A) Schematic of promoter displacement strategy. B) PCR confirmation of endogenous gene tagging; primer pairs used as in panel A, expected size: P1-P2’ (parental line) 666 bp, P1-P1’ 776 bp, P2-P2’ 408 bp, P3-P2’ 346 bp. C) Western blot confirmation of 3xHA tagging; molecular weights are listed to the left and antibodies to the right; TATi-ArlX3 predicted size is 60kDa. D) 3D-SIM images of intracellular TATi-ArlX3 tachyzoites transiently transfected with ERD-GFP (top row), GRASP-RFP (top middle row), and GalNAc-YFP (bottom middle row), or stained with anti-DrpB antibodies (bottom row). Scale bar is 2μm.

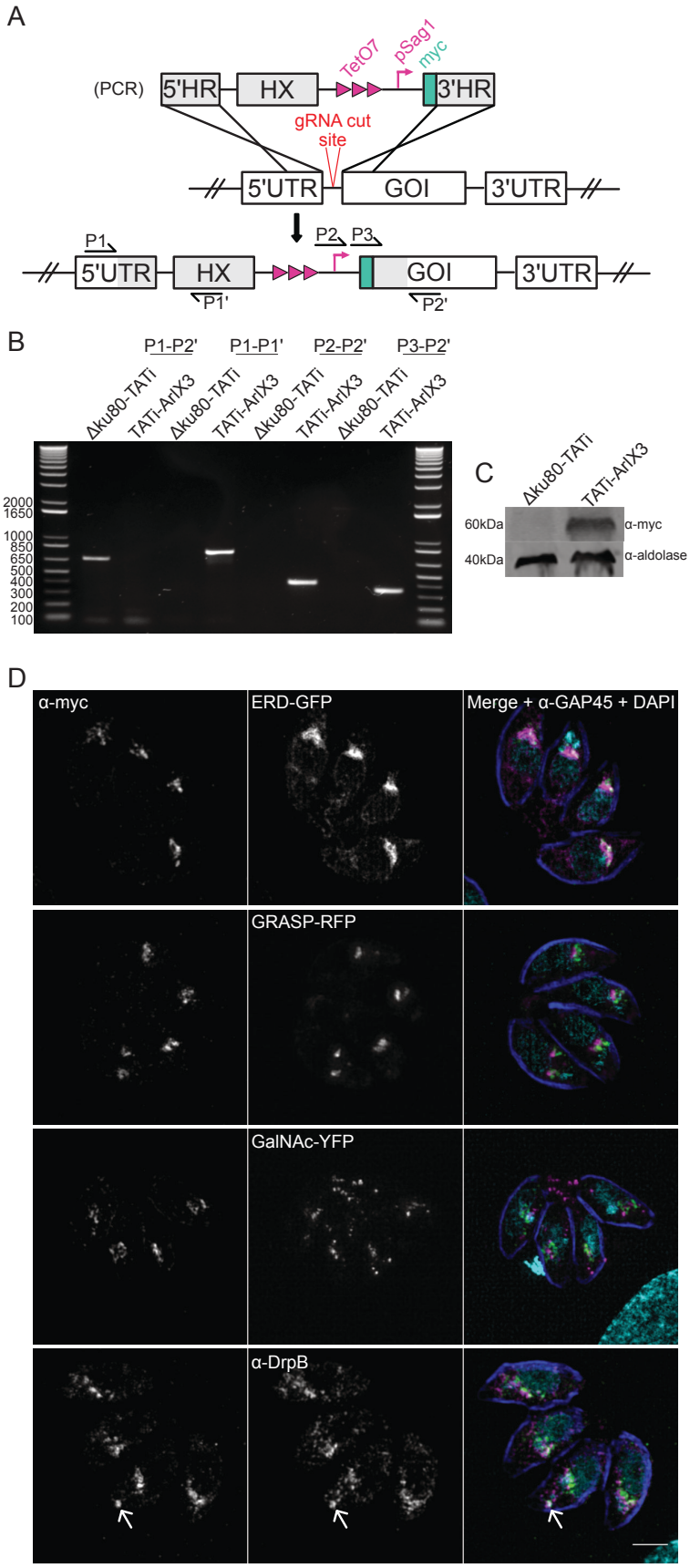
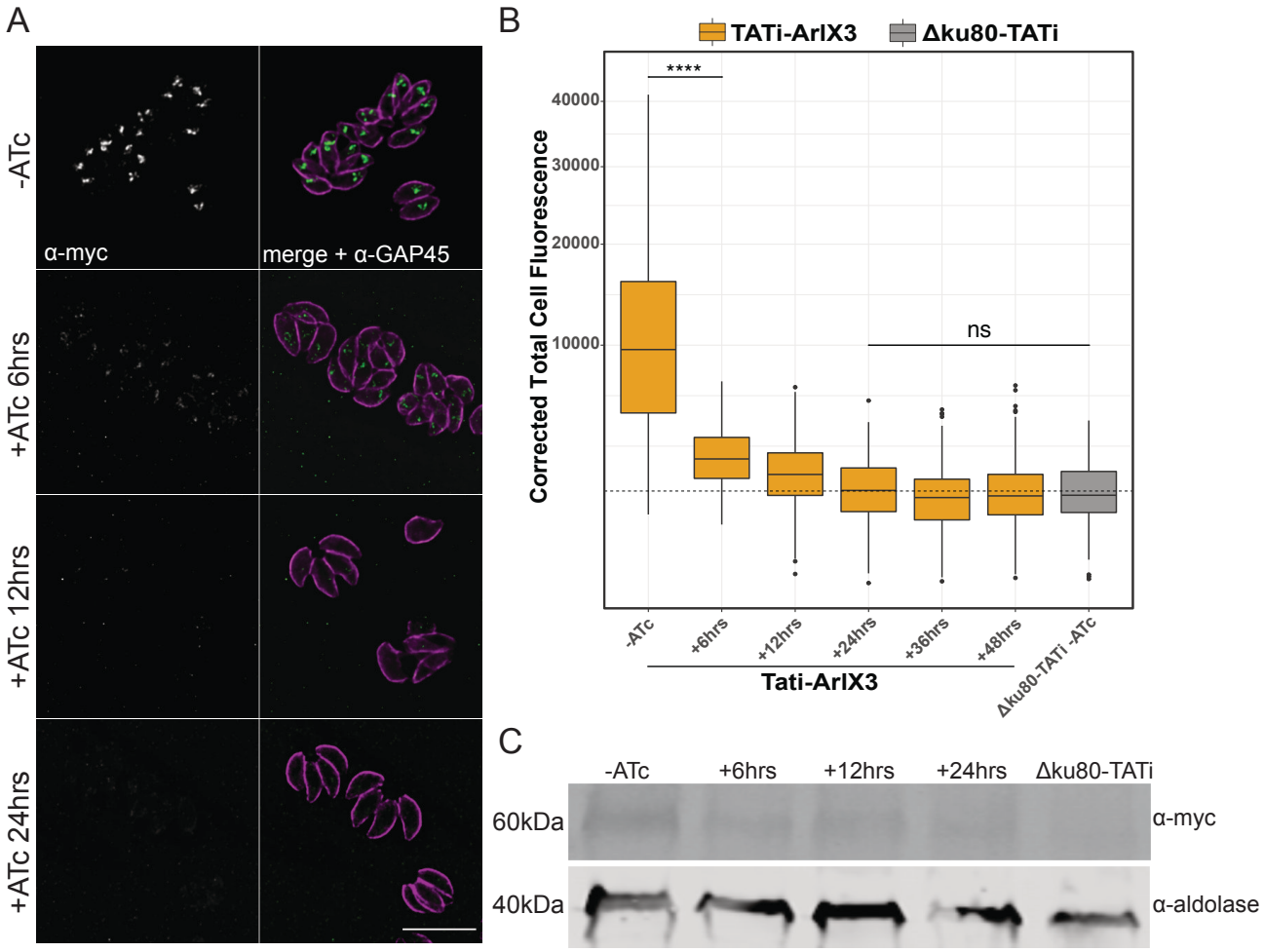

A
(PCR)
5'HR
HX
TetO7
pSag1
myc
3'HR
gRNA cut site
5'UTR
GOI
3'UTR
P1
P2
P3
P1'
P2'
B
P1-P2'
P1-P1'
P2-P2'
P3-P2'
ΔKu80-TATi
TATi-ArlX3
2000
1650
1000
850
650
500
400
300
200
100
C
60kDa
40kDa
α-myc
α-aldolase
D
α-myc
ERD-GFP
Merge + α-GAP45 + DAPI
GRASP-RFP
GalNAc-YFP
α-DrpB

Figure 5.7 ArlX3 knockdown kinetics

This figure shows evaluation of the rate at which ArlX3 protein levels decrease following knockdown induction via anhydrotetracycline (ATc). A) Representative deconvoluted widefield fluorescent images of TATi-ArlX3 parasites either untreated (top row) or induced with ATc for the indicated time periods (bottom row); scale bar is 5μm. B) Quantification of corrected total cell fluorescence, as in A; 100 random vacuoles were measured for each condition, and the experiment repeated three times. Significance was assessed by a Kruskal-Wallis test followed by post-hoc Dunn test with Bonferroni correction for multiple testing. Anti-myc signal showed a significant decrease starting from six hours after ATc addition and was not significantly different to the control by 24 hours after addition. Dashed horizontal line denotes the mean of control condition. **** denotes a p-value ≤ 0.0001, ns denotes not significant. C) Time-course western blot showing similar knockdown kinetics as in B.


A
-ATc
+ATc 6hrs
+ATc 12hrs
+ATc 24hrs
α-myc
merge + α-GAP45
B
TATi-ArlX3
Δku80-TATi
****
ns
Corrected Total Cell Fluorescence
40000
30000
20000
10000
-ATc
+6hrs
+12hrs
+24hrs
+36hrs
+48hrs
Δku80-TATi -ATc
Tati-ArlX3
C
-ATc
+6hrs
+12hrs
+24hrs
Δku80-TATi
60kDa
40kDa
α-myc
α-aldolase

5.3.3 TgArlX3 is crucial for asexual growth

ArlX3 is predicted to be essential for asexual growth[354]. To investigate this, five-day plaque assays were carried out to test the ability of parasites to form zones of clearing in host-cell monolayers (materials and methods). Briefly, parasite progression through the lytic cycle of egress, motility, and invasion (section 1.5.2) results in destruction of host cells within a confined area (a "plaque"); parasites impaired in one or more steps of the lytic cycle fail to form plaques. Parental Δku80-TATi and TATi-ArlX3 parasites were seeded onto host cell monolayers in the presence or absence of ATc and left undisturbed for five days. Whereas Δku80-TATi parasites formed plaques regardless of the presence/absence of ATc, TATi-ArlX3 parasite plaque formation was severely impaired in the presence of ATc (Figure 5.8A). Quantification of plaque sizes confirmed that there was no significant difference between Δku80-TATi plaque sizes ± ATc, demonstrating that ATc by itself has no impact on *T. gondii* asexual growth (Figure 5.8B). Additionally, plaque sizes were comparable between untreated Δku80-TATi and TATi-ArlX3 parasites, suggesting that replacement of the endogenous ArlX3 promoter did not appreciably affect parasite viability. In comparison, plaque sizes were significantly smaller in TATi-ArlX3 parasites treated with ATc, suggesting that ArlX3 knockdown results in a significant impairment in lytic cycle progression.

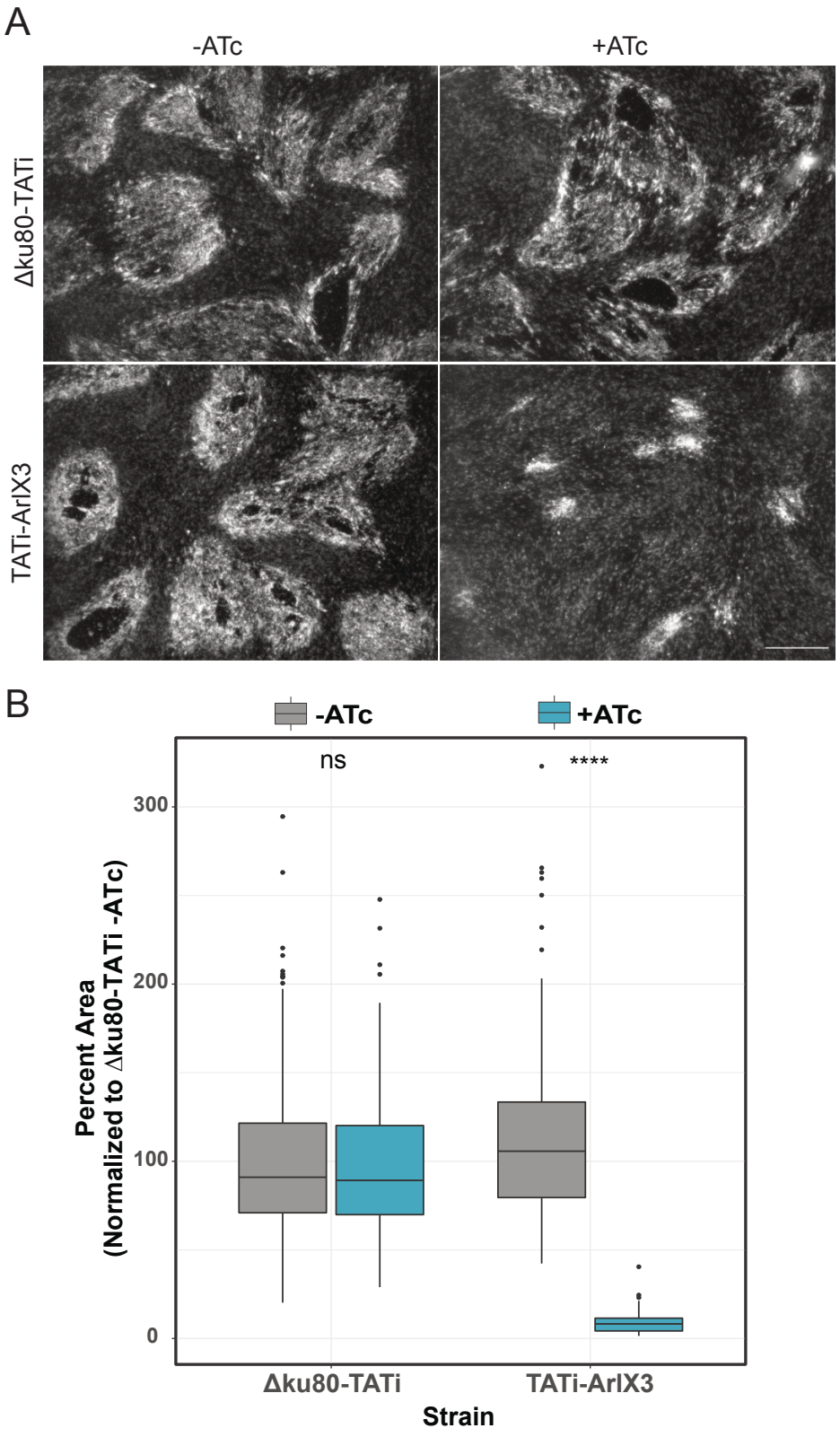
5.3.4 TgArlX3 knockdown impairs each step within the lytic cycle

Impaired plaque formation suggests that ArlX3 knockdown affects one or more steps of the lytic cycle. To determine which steps are impaired, additional assays were performed to determine the competence of parasites to egress, move by gliding motility, and invade host cells (materials and methods).

Egress can be artificially induced by modulating intracellular calcium levels, such as with addition of the calcium ionophore A23187[280]. To egress correctly, parasites must lyse first the parasitophorus vacuole (PV) membrane, followed by the host membrane, in part through the secretion of microneme-localized proteins such as TgPLP1[281], and then activate gliding motility to escape from the host cell. Staining parasites with anti-SAG1 antibody under non-permeabilizing conditions, as well as use of membrane-permeant DAPI nuclear stain,

Figure 5.8 ArlX3 is crucial for *T. gondii* asexual growth

A) Representative images of plaque formation following five-day incubation with or without ATc addition. Scale bar is 200μm. B) Quantification of relative plaque areas, as in A. For each condition, all plaques in 10 random fields of view were measured and the experiment performed three times. Significance was assessed by two-way ANOVA followed by post-hoc Tukey's HSD test, and comparison within strains performed using a Wilcoxon signed-rank test. **** denotes a p-value ≤ 0.0001, ns denotes not significant.


A
-ATc
+ATc
Δku80-TATi
TATi-ArlX3
B
-ATc
+ATc
ns
****
300
200
100
0
Percent Area
(Normalized to Δku80-TATi -ATc)
Δku80-TATi
TATi-ArlX3
Strain

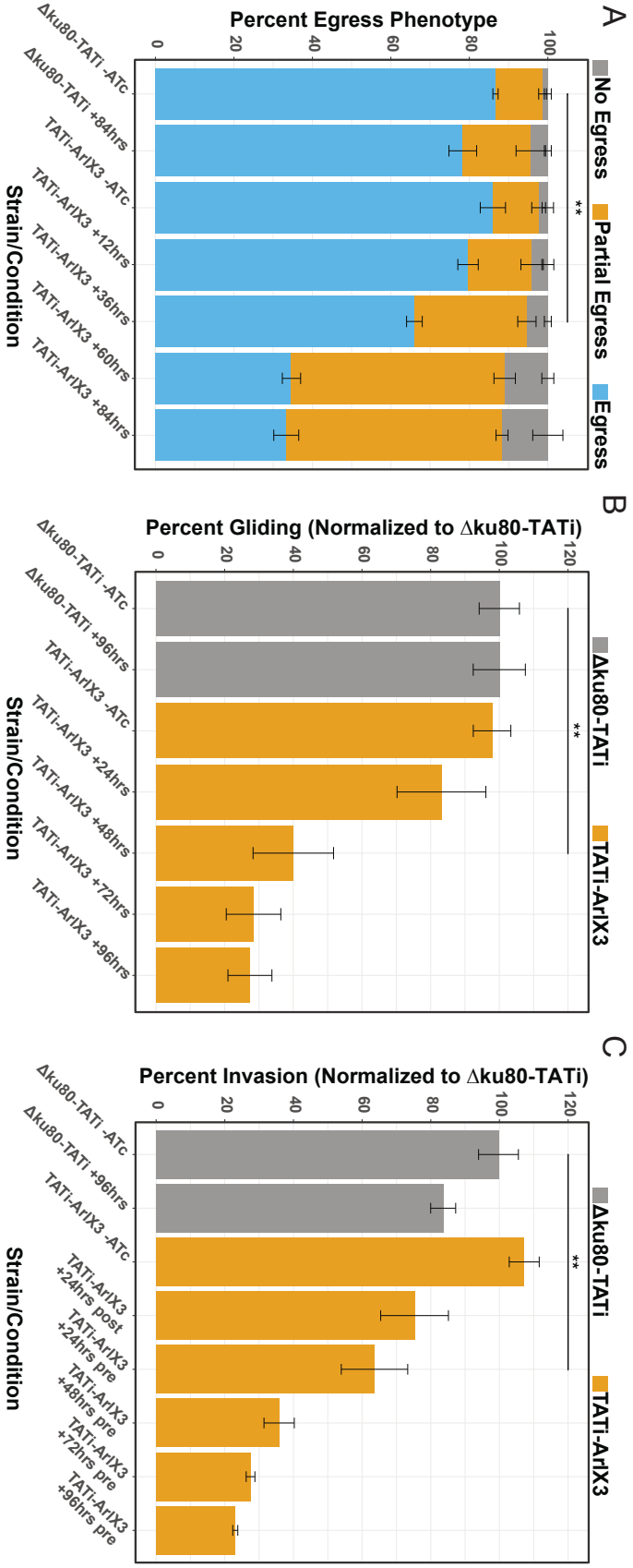
allows to discriminate between parasites that have not lysed membranes (stain with DAPI but not SAG1), those that have lysed membranes but not activated gliding motility (stain with both DAPI and SAG1, but remain together within the host cell), and those that egress completely (stain with both DAPI and SAG1, and are present outside the boundaries of the lysed host cell membrane. These three states are referred to here as "no egress", "partial egress", and "egress", respectively. Control Δku80-TATi parasites egress normally in roughly 80% of vacuoles, regardless of the presence/absence of ATc, although ATc treatment resulted in a slight (from ~2% to ~6%) increase in the population of parasites that did not initiate egress (Figure 5.9A). Untreated TATi-ArlX3 parasites were virtually indistinguishable from controls but showed time-dependent increase in the proportion of parasites showing partial (from ~12% to ~55%) or no (from ~3% to ~12%) egress with ATc treatment; the proportion of parasites failing to fully egress was significantly different from the control following 36hrs of ATc induction (Figure 5.9A). These results suggest that ArlX3 knockdown parasites remain largely able to secrete perforins and other components necessary to induce membrane lysis but are impaired in gliding motility activation.

Motility competence can be measured by a 2-D gliding assay; parasites are incubated on coverslips coated with bovine serum albumin (BSA) in conditions that promote gliding motility prior to fixation and staining with anti-SAG1 antibodies. Gliding parasites deposit membrane and membrane proteins in "trails", which appear visible with SAG1 staining. Quantification of trails provides a measure of the proportion of parasites able to initiate gliding motility; trail length is not considered in this fixed assay, as parasites initiating motility at different times will necessarily deposit trails of different length. Whereas pre-treatment with ATc for 96 hours had no effect on the ability of parental Δku80-TATi parasites to glide (Figure 5.9B), 48 hours of ATc pre-treatment significantly affected the ability of TATi-ArlX3 parasites to glide (Figure 5.9B). Further treatment did not appreciably lower gliding ability, which appeared to drop to roughly 30% of controls by the end of the time course. These results are consistent with those obtained for the egress assay and suggest that ArlX3 knockdown impairs the ability of *T. gondii* to initiate gliding motility.

Invasion is measured in a pulse-chase manner, whereby an equal number of parasites are allowed to invade host cells for one hour before uninvaded parasites are thoroughly washed off. Parasites are allowed to grow intracellularly for 24 hours, and then vacuoles within a prescribed number of fields of view (FOV) are counted. Parental Δku80-TATi parasites treated with ATc invaded slightly less, and untreated TATi-ArlX3 parasites slightly more, than untreated Δku80-TATi parasites, but these differences were not significant (Figure 5.9C). Similarly, ATc addition at the point of invasion ("+24hrs post"), lowered invasion efficiency

Figure 5.9 Effect of ArlX3 knockdown on the *T. gondii* lytic cycle

This figure demonstrates the effect of ArlX3 knockdown on egress (A), gliding motility (B), and invasion (C). A) Egress competency was divided into three categories: parasites that egress normally ("egress"), parasites that lyse host cell membranes but do not initiate gliding motility ("partial egress"), and parasites that fail to lyse host cell membranes ("no egress"). For each of three independent replicates, 100 random vacuoles were assessed for egress competency; distributions between conditions were compared via chi-square test followed by post-hoc Fisher's exact test with FDR correction for multiple testing. B) Gliding competency was assessed by allowing parasites to glide for 30 minutes on coated coverslips; the percentage of parasites depositing trails in 15 random fields of view was quantified for three independent experiments. C) Invasion competency was assessed by allowing parasites to invade host cell for one hour prior to washing; the number of vacuoles in 15 random fields of view was quantified for three independent experiments. For B and C, significance was assessed by one-way ANOVA followed by post-hoc Tukey's HSD test. For all panels, bar height represents the mean and error bars the SEM; bar colour is as per each panel legend. ** denotes a p-value ≤ 0.01.


A
Percent Egress Phenotype
No Egress
Partial Egress
Egress
**
Δku80-TATi -ATc
Δku80-TATi +84hrs
TATi-ArlX3 -ATc
TATi-ArlX3 +12hrs
TATi-ArlX3 +36hrs
TATi-ArlX3 +60hrs
TATi-ArlX3 +84hrs
Strain/Condition
B
Percent Gliding (Normalized to Δku80-TATi)
Δku80-TATi
TATi-ArlX3
**
Δku80-TATi -ATc
Δku80-TATi +96hrs
TATi-ArlX3 -ATc
TATi-ArlX3 +24hrs
TATi-ArlX3 +48hrs
TATi-ArlX3 +72hrs
TATi-ArlX3 +96hrs
Strain/Condition
C
Percent Invasion (Normalized to Δku80-TATi)
Δku80-TATi
TATi-ArlX3
**
Δku80-TATi -ATc
Δku80-TATi +96hrs
TATi-ArlX3 -ATc
TATi-ArlX3 +24hrs post
TATi-ArlX3 +24hrs pre
TATi-ArlX3 +48hrs pre
TATi-ArlX3 +72hrs pre
TATi-ArlX3 +96hrs pre
Strain/Condition

slightly, but this difference was not significant. However, starting with 24 hours pre-treatment (i.e. prior to t=0 where pulse invasion occurred), TATi-ArlX3 parasites showed a significant time-dependent decrease in invasion efficiency, reaching ~25% after the full 96 hours (Figure 5.9C). Hence, ArlX3 knockdown impairs *T. gondii* invasion ability.

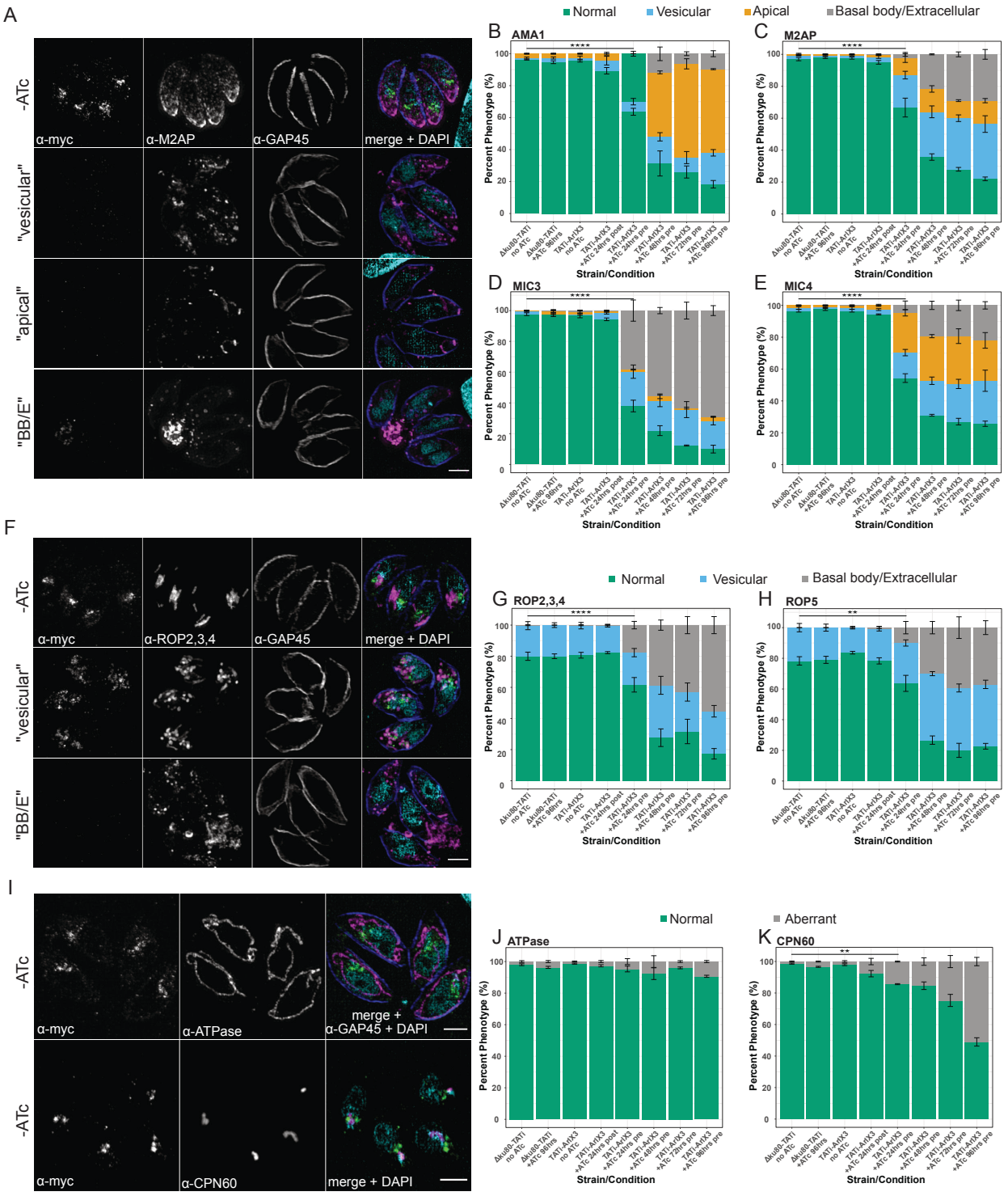
Taken together, these individual assays collectively suggest that ArlX3 knockdown has broad effects on the ability of *T. gondii* to egress, move by gliding, and invade host cells. Although the failure to form plaques (Figure 5.8) can be associated with a single block in the lytic cycle, ArlX3 knockdown parasites appear to be impaired, but not completely blocked, at each separate step. These combined defects are clearly detrimental to the parasite, and result in a significant impairment in the ability of the parasite to efficiently complete its asexual life cycle. Interestingly, secretion from both micronemes and rhoptries is known to be key to all three of these processes[281,294,493].

5.3.5 TgArlX3 knockdown causes mis-localization of resident microneme and rhoptry proteins

Given the defects observed in all steps of the lytic cycle (Figure 5.9), it is reasonable to investigate the integrity of the apical secretory organelles (micronemes and rhoptries). Previous studies on select trafficking machinery, including DrpB[425], clathrin[405], Rab5A and 5C[253], Vps10/sortilin[398], retromer[375], the VpsC complexes[407], and AP-1[258,396], showed variable defects in the localization of microneme and/or rhoptry proteins. In all cases where functions such as invasion, gliding, and egress were assayed in these mutants, these steps were invariably impaired; this is in keeping with the key role these organelles play in each step of the lytic cycle[281,294,493]. Hence, localization of a panel of microneme and rhoptry resident proteins was assessed over a time course following ATc addition (materials and methods). As for the invasion assay, an invasion pulse of one hour was followed by 24 hours intracellular growth prior to fixation, staining, and imaging.

A previous study focussing on the role of TgRab5A and TgRab5C to mediate trafficking to apical organelles revealed that only a subset of microneme proteins were affected following disruption of either paralogue[253]. Hence, staining was performed with proteins that were affected by TgRab5A/C disruption (MIC3), with proteins that were not affected (apical membrane antigen 1; AMA1, and MIC2-associated protein; M2AP), and with a protein that was not assessed in the previous study (MIC4). Representative images show the four staining patterns observed: normal apical microneme staining (Figure 5.10A, top row), “vesicular” staining throughout the cytosol (Figure 5.10A, second row), “apical” staining, where little

Figure 5.10 Effect of ArlX3 knockdown on localization of select organelle markers
This figure shows example images (A, F, and I), and subsequent quantification of phenotypes for select markers of: micronemes (AMA1, B; M2AP, C; MIC3, D; and MIC4, E), rhoptries (ROP2,3,4, G; ROP5, H), mitochondria (mitochondrial ATPase, J), and the apicoplast (CPN60, K). For each marker the phenotype of 100 random vacuoles was scored and the experiment repeated three times. For all graphs, bar height represents the mean and error bars the SEM. Bar colour is as per the legend defined above each group of panels. Significance was assessed by chi-square test followed by post-hoc Fisher's exact test with FDR correction for multiple testing. ** denotes a p-value ≤ 0.01, while **** denotes a p-value ≤ 0.0001. Scale bar for all images is 2μm.


A
-ATc
"vesicular"
"apical"
"BB/E"
α-myc
α-M2AP
α-GAP45
merge + DAPI
Normal
Vesicular
Apical
Basal body/Extracellular
B AMA1
C M2AP
D MIC3
E MIC4
Percent Phenotype (%)
Strain/Condition
F
α-ROP2,3,4
G ROP2,3,4
H ROP5
I
α-ATPase
α-CPN60
merge + α-GAP45 + DAPI
Aberrant
J ATPase
K CPN60

signal is observed throughout the rest of the parasite (Figure 5.10A, third row), and basal body/extracellular ("BB/E") staining, where signal is present either in the vacuolar network or in the vacuole itself (Figure 5.10A, bottom row). Occasional staining beyond the vacuole was also observed, including apparently in the host cell nucleus; this was classified as "BB/E".

Normal staining was almost exclusively observed in parental Δku80-TATi parasites regardless of presence/absence of ATc, and in TATi-ArlX3 parasites either without ATc, or where ATc was added at the point of invasion ("+ATc 24hrs post"; Figure 5.10B-E). However, starting with 24 hours pre-treatment and progressing to 96 hours pre-treatment, parasites displayed a significant time-dependent increase in the proportion of aberrant staining patterns for each microneme protein with a concomitant decrease in normal staining (down to between ~10 to 25%, Figure 5.10B-E). As well, the predominant staining patterns differed between different microneme proteins. MIC3 was the most extreme, where roughly 70% of parasites showed BB/E staining by the 96-hour time point, while AMA1 showed a preponderance for apical staining. M2AP and MIC4 staining with ArlX3 knockdown was split more evenly between the three aberrant staining patterns (Figure 5.10B-E). Hence, microneme proteins become mis-localized with ArlX3 knockdown, albeit with different consequences for each protein.

Rhoptry staining is typically observed as long (~1-2μm) foci ("normal", Figure 5.10F, top row), but can also be observed as punctate dots in the apical half of the parasite ("vesicular", Figure 5.10F, middle row). It is likely that this punctate staining coincides with forming "pro-rhoptries" during rhoptry maturation, and hence the staining pattern depends on the cell cycle state of the individual vacuoles[256,392]. Even with a pulse invasion, *T. gondii* cells do not remain synchronized in a homogeneous culture, unlike *Plasmodium*[745]. Finally, rhoptry staining could also be observed in a similar basal body/extracellular ("BB/E") pattern (Figure 5.10F, bottom row), similar to micronemes. As observed for microneme protein staining, rhoptry staining was virtually indistinguishable between Δku80-TATi lines with or without ATc and TATi-ArlX3 parasites either not treated, or for which ATc was added during the pulse invasion (Figure 5.10G,H). In these cells, roughly 20% of vacuoles showed vesicular staining, which suggests that under normal conditions roughly 20% of parasites in a given sample will be undergoing rhoptry biogenesis. A similar pattern as for micronemes was observed following pre-treatment with ATc for between 24 and 96 hours; significant time-dependent decrease in normal rhoptry staining to ~20%, and concomitant increase in vesicular and BB/E staining (Figure 5.10G,H); the proportions were similar between the two markers tested. Hence, rhoptry proteins become mis-localized with ArlX3 knockdown.

To test whether the observed defects in microneme and rhoptry protein mis-localization were specific or were due to a global impairment in organelle morphology and/or

integrity, two other organelles for which reliable antibodies are available – the mitochondrion and apicoplast – were assessed. The normal morphology for each organelle is shown in Figure 5.10I; abnormal staining was too variable to be classified confidently as for micronemes and rhoptries, and so any staining pattern not resembling the normal morphology was classified as "aberrant". Unlike the case with microneme/rhoptry proteins, no significant changes in mitochondrial staining was apparent in TATi-ArlX3 parasites (Figure 5.10J). In contrast, although ATc had no effect on the apicoplast marker CPN60 in parental Δku80-TATi parasites, a significant increase in aberrant staining was apparent starting from 24 hours of pre-treatment in the TATi-ArlX3 line, and this proportion reached ~50% by 96 hours (Figure 5.10K). Hence, although mitochondrial marker staining is unaffected by ArlX3 knockdown, there is a significant effect on the apicoplast CPN60 marker, although it is not as severely affected as microneme and rhoptry marker staining.

5.4 Discussion

This chapter focussed on initial characterization of three Arl LSPs in the model apicomplexan *T. gondii*. Despite increased interest in studying apicomplexan trafficking, some MTS families, including the Arls, have yet to be characterized in any apicomplexan.

The endogenous gene tagging studies confirmed publicly available data that all three Arl LSPs are expressed in *T. gondii* tachyzoites (Figure 5.1). Additionally, 3D-SIM microscopy revealed that all three localize to distinct parts of intracellular tachyzoites, suggesting that they may have non-overlapping functions within the parasite. ArlX2 localizes to multiple puncta throughout the parasite, but which are almost exclusively found in close proximity to the IMC (Figure 5.1D). However, this gene is predicted not to be essential (phenotypic score of 0.2)[354], which suggests that its function is dispensable. This seems at odds with the unique nature of its taxonomic distribution, as it is the only LSP identified that is restricted just to Apicomplexa, not being present even in closely related chromerid algae (Figure 4.25). It is possible that ArlX2 gain coincided with a function that is parasite-specific, yet dispensable or redundant with other cellular machinery, or that ArlX2 plays an important role in another lifecycle stage (for example, in bradyzoites or sporozoites), or is dispensable under the relatively permissive conditions of *in vitro* culture. This has been described for some other proteins; for example, the autophagy protein TgATG9 was dispensable *in vitro*, but disruption led to a significant decrease in *in vivo* virulence[746]. Future studies will be required to understand the role of ArlX2.

ArlX1 localization was unusual and apparently highly dynamic (Figure 5.1C). Although ArlX1 signal was always observed at the apical tip of intracellular tachyzoites, no other

consistent staining pattern was observed. Signal around the periphery of individual tachyzoites, in close proximity to, and following the curvature of, sub-pellicular microtubules was frequently observed, as was staining throughout the vacuolar network (Figure 5.2). Both the apical and network localization was confirmed by iEM (Figure 5.3), which additionally clarified that ArlX1 also labels micronemes, rhoptries, and vesicles in the apical end of the parasite. It has recently been proposed that microneme and rhoptry components may be recycled during endodyogeny[312], and live imaging of a MIC2-Cb fusion protein has allowed visualization of this protein moving through the vacuolar network in close proximity to F-actin (Javier Periz, personal communication). Hence, it is tempting to speculate that ArlX1 would be involved in these trafficking steps, and in the apical repositioning of recycled microneme and/or rhoptry components. In fact, I recently obtained an ArlX1 knockout line (Supplementary Figure S5.3), but initial characterization has not yielded a clear phenotype and future work is required to understand the role of ArlX1 in tachyzoites.

In comparison to both ArlX1 and ArlX2, which possessed unique and variable localization patterns not easily relatable to parasite organelles, ArlX3 localization is reminiscent of Golgi markers and Golgi/TGN proteins such as CHC[405] and Stx6[409]. Indeed, TgArlX3 signal overlaps with an array of Golgi-localized proteins, including ERD, GRASP, and GalNAc, as well as TgDrpB, which localizes near the Golgi in an uncharacterized structure[425] (Figures 5.1 and 5.6). This was also confirmed by iEM analysis, with gold labeling present at and near the Golgi complex (Figure 5.5).

Based on its taxonomic distribution (conserved in chromerid and most apicomplexans but absent from *Crytposporidium spp.* and dinoflagellates), its localization, and its phenotypic score (-2.49), ArlX3 was chosen for further characterization. The TATi system was chosen, as the DiCre system often results in heterogeneous populations following rapamycin-induced gene excision[258] and the rapid auxin system had not yet been successfully established when ArlX3 characterization was initiated[368]. A split-CAS9 system, similar to the DiCre system, was available for use, but the DSB induced in the absence of homologous repair DNA can lead to significant non-specific phenotypes within the first intracellular cycle following induction (Johannes Stortz, personal communication). A single step "promoter displacement" method readily yielded positive transfectants (Figure 5.6A), and the gene is amenable to knockdown as evident from the relatively rapid knockdown kinetics assessed (Figure 5.7). The fact that background expression level was attainable suggests that the endogenous promoter is not capable of driving ArlX3 expression in this line.

As suggested by the phenotypic score, plaque assays established that ArlX3 is indeed crucial during the asexual lytic cycle, as parasites in which ArlX3 knockdown is induced form

significantly smaller plaques (Figure 5.8). The size of the plaques may suggest that parasites are able to complete one or a small number of cycles. This is consistent both with the presence of detectable protein up to 24 hours post-induction with ATc (Figure 5.6), as well as the gradual decrease in parasite competency to egress, glide, and invade (Figure 5.9). This latter observation is itself consistent with the gradual decrease in correct localization of microneme and rhoptry resident proteins (Figure 5.10). As micronemes and rhoptries are important players in all steps of the lytic cycle[281,294,493], parasites possessing comparatively normal organelles under ArlX3 knockdown would be expected to be able to egress, glide, and invade as normal. Further support for this notion comes from the observation that the minimum levels of gliding and invasion observed (~20% compared to controls) correlate well with the proportion of parasites maintaining normal staining for microneme and rhoptry markers (again, ~20-25%; Figures 5.9, 5.10). It is also interesting to note that parasites in which the key glideosome component, TgACT1, is conditionally knocked out show similar levels of gliding and invasion[292], suggesting that compensatory mechanisms exist. For example, some residual invasion in *myoA* knockout lines has been suggested to involve contributions from the host cell[307]. It is also possible that some background expression exists in the population, and this residual expression is responsible for some or all of the parasites not showing a strong phenotype; future studies employing a knockout-based strategy may be able to confirm or refute this possibility.

Based on the observed defects in the lytic cycle, localization of representative microneme and rhoptry resident proteins was investigated. By and large, prior studies investigating localization of apical organelle proteins have invoked a purely binary split between correctly localized and "mis-localized" signal (for example, Breinich et al. (2009)[425]); even in studies acknowledging a diversity of observed phenotypes, these are rarely, if ever, quantified (for example, Morlon-Guyot et al. (2015)[407]). Therefore, distinct staining patterns were explicitly quantified to provide insight into any underlying complexity within trafficking pathways. This is especially important given evidence for multiple trafficking pathways to the micronemes in *T. gondii*, whereby only a subset of microneme proteins required TgRab5A and TgRab5C for their localization[253]. In comparison, ArlX3 knockdown resulted in mis-localization of a diverse array of microneme proteins (Figure 5.10B-E), suggesting that ArlX3 may function in multiple pathways, or in a single pathway upstream of TgRab5A/5C. Similarly, although only two proteins were assessed, ArlX3 knockdown appears to impair correct localization of rhoptry proteins (Figure 5.10G,H).

By stratifying mis-localized signal into several categories, it is apparent that, at least for microneme proteins, the same knockdown can result in very different signal distributions

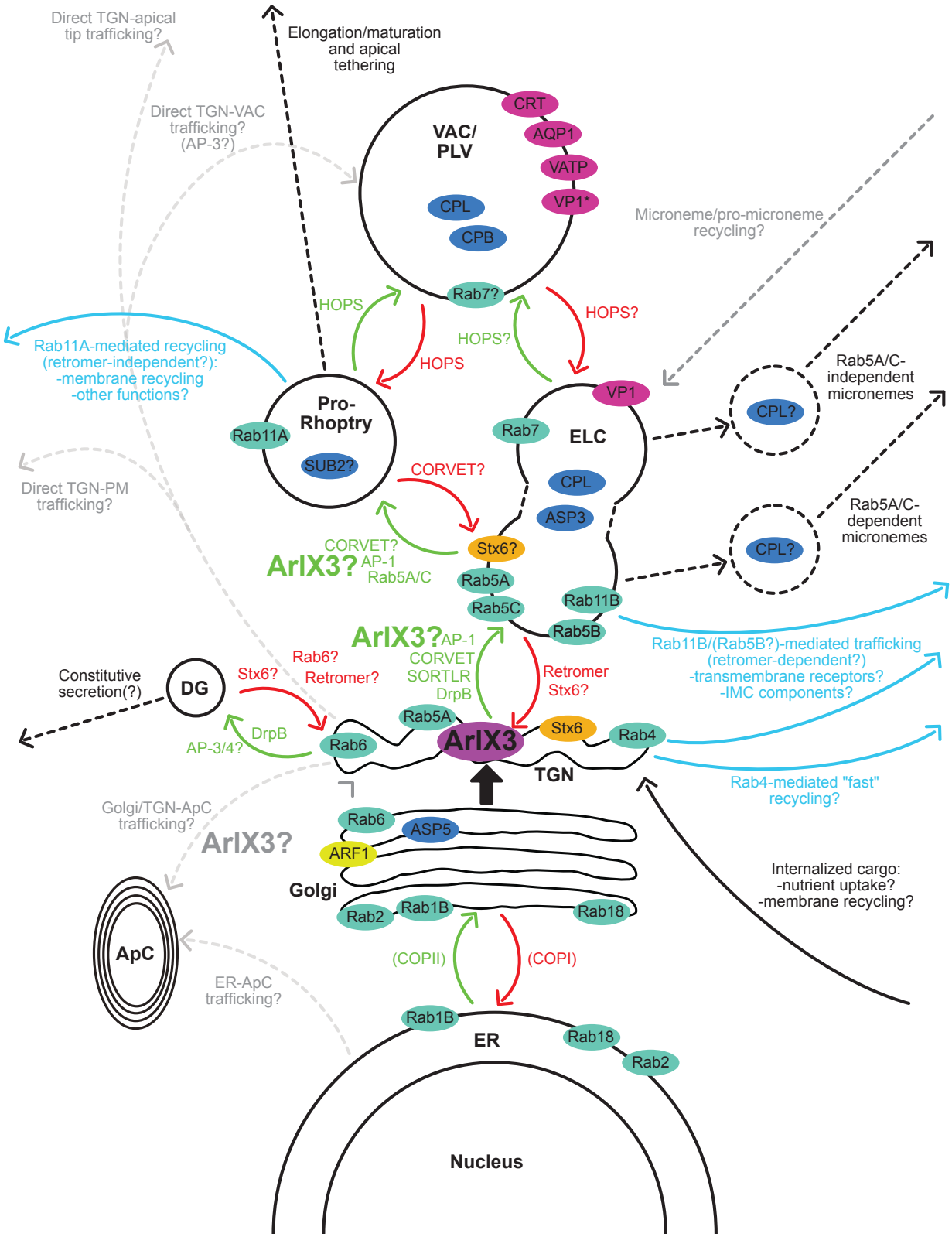
(Figure 5.8B-E). Of these, vesicular staining is relatively difficult to interpret; it could result from failure to traffic structures containing newly derived microneme proteins, a failure during microneme protein recycling following endodyogeny, or some other defect. Similarly, the signal within the basal body may be a result of mis-localization and/or a failure during recycling. Extracellular staining is typically associated with aberrant incorporation of proteins into nascent dense granules formed from the TGN, as these are thought to represent the "default" constitutive secretory pathway of the parasite[423,424]. The apical signal is enigmatic; it was previously proposed to represent a novel trafficking pathway to a sub-population of apical micronemes[407,747]. However, this is inconsistent with the notion that disruption of the Vps10-like TgSORTLR, which is responsible for forward translocation of diverse microneme and rhoptry proteins from the TGN, also results in apical staining in some of the representative images provided in Sloves et al. (2012)[398]. Considering the rhoptry staining, it is clear that, although some vesicular staining is always present and probably corresponds to nascent forming rhoptries, which emerge from Rab5A-positive compartments[258], it is increased in ArlX3 knockdown cells. This may indicate that pro-rhoptries are impaired in trafficking or fusion, or that rhoptry protein-containing compartments become increasingly fragmented in the absence of ArlX3. As with microneme proteins, the increase in BB/E staining could similarly be envisioned to be due to a multitude of factors.

The observed differences in the distribution of aberrant phenotypes among microneme proteins in ArlX3 knockdown parasites (Figure 5.10) suggests that these proteins are either differentially capable of unassisted forward targeting or of entering alternate trafficking pathways, or both. As discussed in section 1.6, many microneme proteins contain a propeptide that is cleaved prior to trafficking to the micronemes, possibly in the ELC. Curiously, removal of propeptides results in arrest of microneme proteins in different cellular compartments: TgM2APΔpro arrests in the ELC[245], TgMIC3Δpro ends up in the basal body/PV[437], and TgMIC5Δpro arrests in the early secretory pathway, mostly in the ER[436]. Although MIC5 was not included in the marker panel investigated here, the phenotypes observed for M2AP (Figure 5.10C) and MIC3 (Figure 5.10D) are similar to those described for removal of the propeptide. This is consistent with the notion that blocking forward translocation of microneme proteins prior to removal of their propeptides in a post-Golgi compartment results in differential mis-localization, presumably by inclusion of the accumulating protein into non-physiological trafficking pathways. It further suggests that ArlX3 is indeed involved in the forward trafficking of diverse microneme proteins.

Revisiting the model of *T. gondii* endosomal trafficking presented in section 1.6 (Figure 5.11), it is proposed that ArlX3 localizes in the vicinity of the TGN and mediates early TGN-

Figure 5.11 Proposed model of *T. gondii* endosomal trafficking including ArlX3

Organelles are depicted with associated molecular markers. Potential trafficking routes between organelles are also shown, along with known or hypothesized machinery for each step; for simplicity, not all organelles and markers are shown. Dashed lines represent uncertainty in organelle identity or trafficking step. Solid green arrows represent anterograde, and red arrows retrograde, trafficking steps while black arrows represent internalization and light blue arrows putative recycling steps. The thick black arrow between the Golgi and TGN represents cisternal maturation. Dashed grey lines represent trafficking steps that are not described, but that may exist on the basis of other evidence. Bold text denotes organelle labels while all other text denotes pathways or trafficking machinery. Markers are shown as filled ovals, with the colour corresponding to the type of marker: magenta, various trans-membrane proteins; deep blue, proteases/maturases; teal, Rabs; orange, SNAREs; yellow, ARFs. Proposed ArlX3 localization is shown by purple oval, and proposed involvement of ArlX3 in trafficking steps is shown as for other factors; for visibility, all ArlX3-related labels are in large bold font. Abbreviations: CRT, chloroquine resistance transporter; AQP, aquaporin; VATP, vacuolar ATPase; VP, vacuolar-H+-pyrophosphatase; CPL/B, cathepsin protease L/B; SUB, subtilisin-like protease; ASP, aspartyl protease; VAC/PLV, plant-like vacuolar compartment; ELC, endosome-like compartment; ApC, apicoplast; DG, dense granule; TGN, *trans*-Golgi network; ER, endoplasmic reticulum.


Direct TGN-apical tip trafficking?
Elongation/maturation and apical tethering
Direct TGN-VAC trafficking? (AP-3?)
VAC/ PLV
CRT
AQP1
VATP
VP1*
CPL
CPB
Microneme/pro-microneme recycling?
Rab7?
HOPS
HOPS
HOPS?
HOPS?
Rab11A-mediated recycling (retromer-independent?): -membrane recycling -other functions?
Pro-Rhoptry
Rab11A
SUB2?
VP1
Rab7
ELC
CPL?
Rab5A/C-independent micronemes
CORVET?
CPL
ASP3
Direct TGN-PM trafficking?
CORVET?
ArlX3?
AP-1
Rab5A/C
Stx6?
Rab5A
Rab5C
Rab11B
Rab5B
CPL?
Rab5A/C-dependent micronemes
ArlX3?
AP-1
CORVET
SORTLR
DrpB
Rab6?
Retromer?
Retromer
Stx6?
Rab11B/(Rab5B?)-mediated trafficking (retromer-dependent?) -transmembrane receptors? -IMC components?
Constitutive secretion(?)
DG
Stx6?
Rab5A
Stx6
Rab4
DrpB
AP-3/4?
Rab6
ArlX3
TGN
Rab4-mediated "fast" recycling?
Golgi/TGN-ApC trafficking?
ArlX3?
Rab6
ASP5
ARF1
Golgi
Internalized cargo: -nutrient uptake? -membrane recycling?
Rab2
Rab1B
Rab18
ApC
(COPII)
(COPI)
ER-ApC trafficking?
Rab1B
ER
Rab18
Rab2
Nucleus

ELC trafficking, and possibly also a later step involving ELC-pro-rhoptry trafficking as well. A direct role for ArlX3 in microneme and rhoptry integrity is less likely, given that addition of ATc at the point of pulse invasion, which would lead to non-detectable levels of ArlX3 by the time of fixation and staining, had little effect (Figure 5.10). Instead, a minimum of 24 hours pre-treatment was necessary to observe clear phenotypes, which suggests that ArlX3 absence prior to intracellular replication (and organelle biogenesis/recycling) is associated with the clearest phenotypic consequence, which would be consistent with a direct role for ArlX3 in trafficking. In this interpretation, ArlX3 knockdown results in an accumulation of microneme and rhoptry components in the Golgi/TGN, and possibly earlier in the secretory pathway as well. At the TGN, some of these proteins can enter alternative pathways; the apical signal may be due to aberrant inclusion into a trafficking pathway for delivery of components to the apical complex, while the extracellular signal is likely due to inclusion of proteins into nascent dense granules. The vesicular microneme staining, and the increase in vesicular rhoptry staining, is difficult to fully explain. In some EM sections, the Golgi appears disrupted and fragmented in ArlX3 knockdown parasites (Supplementary Figure S5.4), similar to that observed on overexpression of the TGN-resident SNARE TgStx6[409]. Although this requires further investigation, it is possible that disrupted post-Golgi trafficking leads to Golgi fragmentation, which would result in numerous vesicular structures that could potentially get trapped in the residual body during endodyogeny, explaining both the vesicular and basal body staining.

The proposed location of ArlX3 at the TGN and proposed function in post-Golgi trafficking has interesting ramifications for apicoplast trafficking. Although the kinetics were not as fast as for microneme and rhoptry proteins, localization of CPN60, a chaperone protein located in the apicoplast lumen[410], was affected by ArlX3 knockdown (Figure 5.10K). As discussed in section 1.6.2.2, the trafficking route of proteins to the apicoplast is not fully elucidated, with studies suggesting trafficking via both the ER[412,413] and Golgi[414]. The results presented here are consistent with trafficking of CPN60 via the Golgi, including the lack of ArlX3 in aplastidic cryptosporidians[457,458] (Figure 4.25). However, it might also be the case that ArlX3 knockdown presents pleiotropic defects earlier in the secretory pathway, or eventually causes non-specific defects in the apicoplast. It is also of note then that that ArlX3 is absent in dinoflagellates (Figure 4.25), which have plastid-bearing lineages and wherein the ancestral plastid is still present outside the ER but bound by three membranes[672]. As dinoflagellates have secondarily lost a true apical complex, including microneme and rhoptry-like organelles[232], it is possible that the presence of multiple additional post-Golgi pathways,

rather than a specific pathway itself, necessitated an additional Arl paralogue, similar to current understanding of TgRab5A/5C function[253].

Regardless, the data presented here suggest a possible role for ArlX3 in apicoplast trafficking. That CPN60 staining was not perturbed to the same extent as that of other organelle resident proteins might be due to redundant pathways, or due to a difference in protein stability and/or organelle inheritance during endodyogeny. It is also possible that the observed effect is indeed due to non-specific abrogation of trafficking, which would be consistent with its slower kinetics. Further study is required to understand the role of ArlX3, if any, on the delivery of nuclear-encoded apicoplast proteins.

Overall, the effect of ArlX3 knockdown on microneme and rhoptry proteins is consistent with the proposed hypothesis regarding paralogous expansion and organellogenesis in extant eukaryotes. Although ArlX3 was not found in any of the three *Symbiodinium* species or *P. marinus*, it is still possible that ArlX3 arose in the common myzozoan ancestor and was subsequently lost in some or all of these taxa. The apical complex, together with micronemes and rhoptries, is known to be present in a variety of forms across Myzozoa, including in a highly modified form in both chomerids and dinoflagellates[232]. Although an apicomplexan-like apical complex, complete with micronemes and rhoptries, is present in *Perkinsus*[232], the failure to identify ArlX3 in *P. marinus* could be due to technical, rather than biological, factors. Genomic data from additional basal taxa on both sides of the apicomplexan/dinoflagellate myzozoan split will be instrumental in confirming the timing of origin of ArlX3. Although the data presented here are insufficient to unambiguously tie a specific organelle (from among micronemes, rhoptries, and the apicoplast) to the advent of ArlX3, either timing for the advent of ArlX3 (either at the base of Myzozoa or in the common ancestor of Apicomplexa/chromerids) would be consistent with the proposed hypothesis relating paralogue distribution and gene function.

6. Final Discussion

6.1. Overview

The primary purpose of this thesis has been to discuss the relationships between gene conservation, gene function, and the generation of new organelles. To this end, both a large pan-eukaryotic literature analysis (Chapter 3) as well as more focussed analyses on Apicomplexa as a model group (Chapters 2, 4, and 5) were discussed. This chapter will synthesize these disparate analyses and provide argued conclusions regarding the utility of simultaneously considering the taxonomic distribution of factors (specifically, genes and organelles) and the function of those factors in order to inform not only gene and organism function, but also to expand on previously proposed evolutionary mechanisms.

6.2. The organelle paralogy hypothesis

Central to the organelle paralogy hypothesis (OPH) is the notion that no single gene encodes specificity among trafficking steps; rather, it is the combinatorial interaction between genes from different families that provides trafficking specificity. Organelles differ in many ways, including in morphology, number, lipid composition, and the sum total of proteins in their membranes and/or luminal space. However, organelle identity is maintained only through specific trafficking. Take for example the key process of membrane fusion. If no specificity and/or regulation were provided for the membrane fusion step of trafficking, then organelles would fuse at random, likely in order of their proximity, and eventually no distinct compartments would remain. As organelle identity is intimately tied to the specificity of trafficking between organelles, it stands to reason that new organelles can only emerge concurrently with additional trafficking pathways.

The OPH dictates that paralogous expansion of MTS family members (e.g. Rabs, SNAREs, etc.), and subsequent divergence of additional paralogues, allowed for an increase in the number of distinct trafficking pathways, and therefore, in the number of distinct autogenous eukaryotic organelles[453]. This is generally invoked to explain the diversity of organelles inferred as present in the LECA, and is supported to some extent by the paucity of eukaryotic endomembrane factors in the Asgardarchaeota, the closest known archaeal relatives of eukaryotes[446,447,748]. That is, that the paucity of such factors in eukaryotic relatives suggests the advent of a large number of factors during the FECA-LECA transition, which is consistent with the concomitant increase in the complexity of internal membrane bound compartments in eukaryotes compared to prokaryotes.

However, although the OPH mechanism was likely important during early eukaryotic evolution, there is nothing to suggest that this mechanism has stopped operating in eukaryotic lineages since the LECA. For example, subsequent expansions of ancestral eukaryotic

proteins, including endosomal syntaxins, Rab5, and the lone beta subunits shared between AP-1 and AP-2 complexes, have been previously noted and suggested to represent modern examples of further elaboration on the LECA complement[394]. Further studies, although not explicitly invoking the OPH in discussion of the data, have demonstrated further expansions within specific lineages for a number of MTS factors, most notably large expansions of Rabs in, for example, the Metazoa[110]. Hence, it is likely that continued sculpting of the MTS, both in terms of expansions and losses, has occurred across the eukaryotic tree.

This raises questions regarding the extent to which the function of any given paralogue can be considered conserved. If additional paralogues can arise and acquire new functions, it is possible that either they, or the paralogue from which they originally arose, may become redundant; loss of function in the latter would critically undermine efforts to map function across eukaryotes. In addition, loss of MTS machinery may be taken to mean that the function and/or pathway mediated by that machinery no longer exists. In extreme cases, it could suggest that entire organelles are absent compared to the LECA complement. The ability of investigators to form initial hypotheses for organisms in which only genomic data are available therefore critically rely on the validity of assuming functional homology.

6.3. Functional homology

Functional homology, the hypothesis that homologues, especially orthologues, should retain the same or similar function across organisms, was discussed at length in Chapter 3. Although functional homology is an implicit assumption in all comparative genomic analyses, it has generally been regarded as safe despite the fact that no previous systematic analysis of its assumptions had been performed.

The existence of functional homology is related to the OPH; under the OPH, paralogous expansion of MTS factors is intimately tied to organellogenesis, which suggests that a paralogue arising concurrently with an organelle would be constrained to function in trafficking to, or at, that organelle. The expectation therefore is that, in an organism possessing either that organelle, or an organelle that can be confidently inferred as homologous to it, the relevant paralogue should be present and be constrained to function in trafficking to/at that organelle. Furthermore, it suggests that, as long as ancestral organelles are present, existing machinery will be constrained and further novel organelles can only be generated through additional expansion of MTS machinery.

In a large, pan-eukaryotic, analysis of the literature it was found that, although eukaryotes differ substantially in the organization of their endomembrane systems, homologues largely possess similar function when considering homologous organelles and

trafficking pathways between them (Chapter 3). Although this conclusion is not surprising given the above arguments regarding paralogue presence, it does suggest that, in cases where additional complexity is warranted, further expansion of paralogues, rather than repurposing of existing machinery, is the predominant mechanism in eukaryotes.

6.4. The OPH in the FECA-LECA transition and extant eukaryotes

Although the OPH mechanism is suggested to have played a key role in the early evolution of eukaryotes[446], and such a role is supported by simulation studies[454], it is difficult, or even impossible, to envision an experiment that could explicitly test this. However, as discussed above, the OPH has likely contributed to eukaryotic evolution post-LECA as well, with the additional benefit that testing OPH-related mechanisms in extant eukaryotes is feasible. In this thesis, apicomplexan parasites are used as a model system in which to test the hypothesis that the advent of novel organelles in extant eukaryotes requires the same OPH mechanism, i.e. paralogous expansion and divergence, as envisioned to explain the advent of organelle complexity in the LECA.

6.5. Apicomplexa as a model to study the OPH in extant eukaryotes

As discussed in Section 1.4, Apicomplexa represent an extremely successful group of obligate intracellular parasites affecting human health and well-being on a global scale. Therefore, there is significant impetus to study these parasites to reduce the morbidity and mortality they cause. However, they also represent a fascinating system for cell biological study. This latter point is due, at least partly, to the unique evolutionary history and organelle complement of Apicomplexa. The discovery first of the apicoplast[261,262], and then later of chromerid algae[207,208], confirmed that Apicomplexa are descended from a photosynthetic, likely free-living or symbiotic, relative. In addition to the apicoplast, a number of other organelles, discussed more in depth in Section 4.1 and below, including the IMC, dense granules, micronemes, and rhoptries, all represent additional intracellular locations to which specific material must be trafficked. It should be noted that, although the OPH is usually discussed in terms of autogenous organelles, the endosymbiotic origin of the apicoplast does not change the fact that it represents an additional organelle outside of the ER[260] and hence likely receives at least some material via trafficking steps.

These additional organelles represent attractive candidates for the involvement of the OPH. Any additional paralogues can be identified through informatic methods with relatively high throughput and low cost, as discussed in section 4.3. Furthermore, the presence of fully sequenced genomes for groups outside of the Myzozoa, in which the majority of these

organelles are restricted, including the ciliates, stramenopiles, rhizarians, archaeplastids, and other related algae, provide the necessary depth with which to pinpoint the timing of emergence of additional paralogues. In addition, as discussed in section 1.4.2, several apicomplexans, notably *P. falciparum* and *T. gondii*, have been developed as model systems with a large number of resources including genetic tools, antibodies, and specific phenotypic assays. Hence, the function of any novel paralogues identified can be investigated.

By combining information of the taxonomic distribution and function of each paralogue, their ability to support or refute the OPH as a mechanism for organellogenesis can be determined. Under the OPH hypothesis, and considering functional homology, it is expected that the taxonomic distribution of a factor can provide insight into its function; if paralogues arise concurrently with a function, which is subsequently maintained through vertical descent, it stands to reason that those paralogues should only be present in taxa that require their function. As mentioned, multiple times throughout this thesis, this has already been supported in the apicomplexan literature for Rab11B and Rab5C[253,390]. Due to all of these arguments, Apicomplexa were selected as a model system and investigated for additional MTS paralogues not found in other eukaryotes.

6.6 Identification of novel MTS paralogues in Apicomplexa

Both in Chapter 2 as well as Chapter 4, some assumptions were made regarding the assessment of homology and absence of homologues across genomes, which, although widely used, must still be kept in mind. The first is the use of sequence similarity as a criterion for homology. By definition, homologues are genes that are related through common ancestry regardless of the exact nature of the relationship. Hence, sequence similarity is used as a proxy for inferring homology, as it is expected that homologues will be significantly more similar to each other than expected under a statistical distribution generated based on the properties of the sequences under comparison[464]. Therefore, sequences that are significantly similar to each other are expected to be homologous, but the absence of significant sequence similarity may indicate either a lack of homology or the divergence of homologues to the point where they are no longer detectable by such methods. Therefore, in this thesis, putative absence of homologues based on similarity searching algorithms is described as a failure to identify a homologue rather than the true absence of a homologue.

Chapter 4 described data for seven paralogous MTS families in Apicomplexa and select outgroup taxa. As outlined in section 4.3 and Figure 4.1, a phylogenetic workflow was conceived and implemented in order to discern LSPs from pan-eukaryotic paralogues. This was based mainly on running separate phylogenies with sequences that could not be classified

as pan-eukaryotic. In studies aiming to examine the conservation of factors across eukaryotes, these sequences are typically labeled as "unclassified" or "singletons", and not investigated further. However, if the interest is in conservation of additional factors within a lineage, these factors represent the only viable candidates for LSPs and therefore require further analysis to determine their identity and taxonomic spread.

Overall, 19 LSPs were confidently identified (Table 4.1), spread across five gene families (TBS encompasses both ArfGEF and RabGAP families, and is considered here as being an LSP for both). Furthermore, with few exceptions, the identity of the pan-eukaryotic clade from which each LSP is derived could be determined with confidence. Although LSPs are theoretically free to diverge and acquire new function, as discussed in section 4.1, previously described examples in Apicomplexa appear to retain similar function to their pan-eukaryotic paralogues, at least maintaining function within the same system. Hence, clear identification of the most closely related pan-eukaryotic paralogue for each LSP can inform an initial hypothesis into its function. This can be further combined with an examination of the taxonomic distribution of the LSP, as, under the OPH model, additional paralogues are expected to be restricted to lineages in which their function is required. If novel cellular features, specifically organelles, are present within the same set of lineages, the LSP could conceivably mediate trafficking to that organelle, or be involved in a trafficking step either up- or down-stream of the organelle itself.

Having taken an unbiased approach, it is interesting to note that, as discussed in section 4.5, a large number of the identified LSPs are predicted to function in the endosomal system. For example, additional paralogues of Stx12 (Figures 4.4, 4.6), VAMP7 (Figures 4.10, 4.11), and Vti1 (Online Appendix Table 4.6) may be involved in fusion of late endosomal structures[688–691], as well as the TBS proteins and one or more of the TBC-Q-related paralogues (Figure 4.20)[728–730,734,735]. As discussed in section 1.6.2.5, numerous lines of evidence suggest micronemes and rhoptries are derived from endosomal compartments and reside within the endosomal system[236,431]. Hence, it is tempting to speculate that at least some of these additional paralogues might function in trafficking to, and/or fusion of carriers at, micronemes and rhoptries. It is clear from multiple studies of endosomal trafficking machinery, including Rab5[253], DrpB[425], retromer[375], Vps10/sortilin[398], AP-1[258,396,399], and the VpsC complexes[407], that microneme and rhoptry trafficking involves at least some of the same machinery and likely traverses the apicomplexan MTS up to a post-Golgi branching point. Specificity of trafficking between these compartments has been suggested to involve transcriptional timing[310]; although this is a reasonable explanation, it is likely that specific machinery must at least be involved in the later steps of trafficking following the branch point. This is supported

at least in part by the lack of effect on only a subset of microneme proteins with disruption of TgRab5A/C[253], which suggests that trafficking of different subsets of microneme proteins involves distinct mechanisms.

For Stx12B (Figure 4.4) and VAMP7B (Figure 4.10) at least, the pattern of conservation (across Myzozoa but absent from ciliates) is consistent with an OPH-like duplication concurrent with the advent of these specialized secretory organelles[232,749]. Other machinery, including Ykt6B (conserved in Apicomplexa and chromerids only, Figure 4.10) and TBS/TBC-X1/X3/X4/X5 (conserved in alveolates, Figure 4.20) do not show the same clear pattern. Ykt6B may still be a myzozoan innovation that was subsequently lost in the few taxa included under study; distinct cellular features to which trafficking could be mediated only in chromerids and apicomplexans are not inherently obvious, although additional complexity within microneme populations[253] may represent a possible hypothesis for function. The most obvious candidate for alveolate-specific factors are the alveoli/IMC compartment(s), which are a synapomorphic feature for this group[17]. Combined with the known interactions between TBC-Q and TBC-N proteins with Rab11[728,729], and the presence of the alveolate-specific Rab11B, which mediates trafficking to the IMC in *T. gondii*[390], support this hypothesis. Under the assumption of this function, the presence of a single additional Rab paralogue (Rab11B) with multiple (in *T. gondii,* six) additional TBC paralogues, might suggest that regulation of activated Rab11 in alveoli/IMC trafficking is crucially important. However, not all identified TBC paralogues need function in this pathway; others could provide similar regulation of other machinery within the expanded endosomal system of alveolates[236,245,247,248,533,600,750].

The presence of up to three Rab1 paralogues in Apicomplexa (Figure 4.13) is also interesting. A previous study demonstrated that Rab1A and Rab1B exhibit different localization patterns, with the pan-eukaryotic Rab1B localizing primarily to the ER, as expected, and Rab1A localizing in early endosomal structures[253]. Based on this difference alone, it is expected that these two paralogues likely perform different functions in *T. gondii*. Rab1K is unusual for several reasons; it likely arose in the myzozoan ancestor yet has been lost from numerous extant apicomplexan lineages (Figure 4.13); furthermore, Rab1K homologues possess extended N-termini that are predicted to adopt a β-propeller tertiary structure (Figure 4.16). B-propellers are found in numerous classes of proteins including coat complexes[751–754], nuclear porins[755,756], and GEF proteins[757], precluding any concrete hypothesis regarding the function of the Rab1K N-terminus. However, these structures typically mediate protein-protein interaction[687]; therefore, identification of putative Rab1K binding partners will likely be important in understanding its function, including why it has been lost at least twice independently in Apicomplexa (Figure 4.13).

In addition to those paralogues whose function can be predicted based on their closest pan-eukaryotic paralogue, four paralogues, ArlX1, ArlX2, ArlX3, and RabX1, were identified whose closest pan-eukaryotic homologue could not be confidently identified (Figures 4.25, 4.27). Arls will be discussed in section 6.7. In the case of RabX1, a consistent and somewhat supported association with Rab28 was observed (Figure 4.19). A recent report suggesting that Rab28 functions in IFT in *C. elegans*[727], combined with the association of known flagellar proteins such as SAS-6 and SFA with the apical complex[230,231,233], leads to an intriguing hypothesis regarding possible RabX1 function. Despite having flagellated microgametes[270], *T. gondii* lacks homologues of known ciliary/flagellar machinery including Arl3[709,711], Arl6[714,758], Arl13[710,712], and IFT27[713] (Figures 4.13, 4.25). In addition, the flagellar ancestry of the apical complex, which itself is absent in flagellated microgametes but present in other life-cycle stages[270], suggests that machinery for trafficking to, and within, the flagella may still be important in this organism. Although *T. gondii* does encode a divergent homologue of Rab8, which has been implicated in ciliogenesis and protein import into cilia[388,759], as well as Rab23, which has been implicated in trafficking cargo to cilia[760], other Apicomplexa appear to have lost these Rabs (Figure 4.13). Hence, it is unclear how flagellar and apical complex biogenesis and trafficking function in Apicomplexa; RabX1, which is universally conserved in Apicomplexa (Figure 4.13), offers an attractive candidate for such function.

In addition to these possible broader functional implications of novel paralogues across apicomplexan lineages, it is also clear that not all these LSPs are conserved across extant members of the phylum. It is possible that some of these changes may reflect a role for the corresponding LSP in a function that has become unnecessary, or is compensated for in some other way, in lineages that have secondarily lost it. Further expansions in some lineages, for example the presence of an additional Stx2-like paralogue in piroplasmids and *Plasmodium spp.* (Figure 4.6) inject further complexity to this observation. Overall, it will be important to continue to functionally characterize each LSP to better understand the patterns of retention and loss across Apicomplexa.

Hence, the work presented in this thesis confirms and extends previous studies suggesting the presence of additional paralogues in Apicomplexa that are not conserved across eukaryotic diversity. These LSPs are expected to represent promising candidates for downstream analysis, and, in many cases, initial hypotheses regarding their function based on homology can be explicitly tested.

6.7 Novel Arl paralogues: an explicit test of the OPH?

Despite that Apicomplexa possess three Arl LSPs (Figure 4.25), none of them could be confidently established as having duplicated from a specific ancestral Arl protein (Figure 4.27), although ArlX1 appears similar to Arl16. Of these, ArlX1 is present across SAR, ArlX2 only in Apicomplexa, and ArlX3 restricted to chromerids and apicomplexans. To understand better the possible role of these Arl LSPs, Chapter 5 presented localization data for all three, together with detailed characterization of ArlX3.

ArlX1 presented the most variable localization pattern, being present at the apical tip and associated apical secretory organelles, as well as being found to varying degrees throughout the cell periphery and filamentous network among vacuoles within each sample (Figures 5.1-3). ArlX2 was found in a punctate distribution throughout the cell body, including at the periphery in close apposition to the IMC (Figure 5.1). ArlX3 localized to the Golgi, as well as to diverse post-Golgi carriers and compartments, including micronemes (Figures 5.1, 5.4, and 5.5). Hence, despite their enigmatic phylogenetic affiliation, all three Arls are clearly expressed and localize to distinct regions within *T. gondii* tachyzoites.

Although I recently obtained an ArlX1 knockout (Supplementary Figure S5.3), there is as yet no apparent associated phenotype. Similarly, no functional data yet exist for ArlX2. As discussed in section 5.4, an intriguing hypothesis is that ArlX1 is involved in the transport of material, potentially including MIC2-containing vesicles, between parasites via the filamentous network. It has been established that connections between parasites, with their associated F-actin network, is important for regulating synchronicity among developing parasites[314,315], and may have other functions as well. ArlX1 localization is consistent with a possible role in one or more of these processes: it is present throughout the network in intracellular vacuoles, and throughout the periphery of tachyzoites, mirroring the localization of microtubules (Figure 5.2). Furthermore, iEM analysis supports the presence of ArlX1 at the limiting membranes of micronemes, rhoptries, and apically localized vesicles of unknown providence (Figure 5.3). In depth characterization of the ArlX1 knockout line will be required to test the validity of these hypotheses.

The apparent dispensability of ArlX2 (phenotypic score of 0.2)[354] is at odds with its conservation solely in Apicomplexa (Figure 4.25), suggesting an advent following the chromerid-apicomplexan split. As discussed in section 5.4, this may be due to a number of factors, including the permissive conditions of *in vitro* culturing, and/or a dispensable role in the asexual stage, yet essential role in one or more other lifecycle stages. Although ArlX2 does not necessarily require an important or essential role in Apicomplexa, it seems unlikely that a parasitic lineage which has frequently lost MTS components (see for example Chapter 2) would have recently acquired an additional paralogue without some important function.

ArlX3 was investigated, both for its predicted importance during asexual growth (phenotypic score of -2.49)[354], and for its presence in chromerids and apicomplexans, yet absence from included dinoflagellate taxa, as well as from cryptosporidians (Figure 4.25). ArlX3 knockdown was found to significantly impair lytic cycle progression (Figure 5.8), which could be traced to defects in egress, initiation of gliding motility, and invasion (Figure 5.9). As all three of these processes require secretion of micronemes and/or rhoptries, localizations of resident proteins for these organelles were investigated (Figure 5.10). It was found that a significant fraction of vacuoles displayed mis-localization of diverse microneme and rhoptry markers (Figure 5.10A-H), and that this was associated with the loss of detectable levels of ArlX3 prior to host cell invasion. This latter result suggests that, rather than playing a direct role in the integrity and/or secretion of apical organelles, ArlX3 is instead involved in trafficking of organelle components during endodyogeny. In addition, investigation of endosymbiotic organelles, the mitochondrion and apicoplast, revealed no significant effect on mitochondria (Figure 5.10J), yet, as for microneme and rhoptry proteins, a significant mis-localization of the apicoplast marker CPN60[410] in ArlX3 knockdown parasites (Figure 5.10K).

Although the results presented here are inconclusive as to the exact role played by ArlX3 in trafficking to the apical organelles and/or apicoplast, they are consistent with a role for ArlX3 in at least some post-Golgi trafficking (Figure 5.11). In addition, mis-localization of markers found both to be affected by TgRab5A/C disruption (MIC3) and not affected (AMA1, M2AP), suggests that ArlX3 may function upstream of the branch point in trafficking to the putative subpopulations of micronemes[253]. Hence, like TgRab5C, ArlX3 is a LSP restricted to (a subset of) Myzozoa with a proposed role in trafficking to micronemes and rhoptries. This is consistent with the hypothesis put forward in this thesis, that the advent of additional organelles, here micronemes and rhoptries, in the presence of ancestral compartments, here the ELC and VAC, is likely to have occurred concurrently with gene duplication and diversification.

However, the absence of ArlX3 in *Cryptosporidium spp.*, which still retain both micronemes and rhoptries[761], suggests that this simple explanation may belie additional complexity. One possibility, which is at least partially supported by the data presented in this thesis (Figure 5.10K), is that ArlX3 is also involved in trafficking to the apicoplast. Given the localization of TgArlX3 primarily near the Golgi/TGN (Figures 5.4-5.6), this would suggest the possibility of post-Golgi apicoplast trafficking. As discussed in section 1.6.2.2, the identity of trafficking pathways to the apicoplast remains unclear, with various studies suggesting either direct trafficking from the ER[412,413], or via the Golgi[414]. The data here appear to suggest the latter, and additionally suggest that the number of putative post-Golgi destinations may have

necessitated an additional LSP for trafficking. This might explain why both dinoflagellates, which lack a true apical complex and associated organelles[232], and cryptosporidians, which lack an apicoplast[457,458], do not encode ArlX3 homologues. Just as both Rab11B and Rab5C mediate similar trafficking pathways as their pan-eukaryotic homologues in apicomplexans[253,390], so too might the presence of both the apical complex and apicoplast necessitate additional complexity in post-Golgi trafficking.

The results presented here argue, at least, for future work in these lineages to clarify both the detailed taxonomic distribution and function of the identified LSPs. ArlX3 loss having significant effects on three organelles not present outside of the Myzozoa (micronemes, rhoptries, and the apicoplast), are wholly consistent with the hypothesis put forward here regarding organellogenesis in extant eukaryotes; namely, that such events would likely occur concurrently with MTS gene duplication.

6.8 An overall picture of apicomplexan MTS evolution

This thesis provides the first comprehensive picture of the evolutionary pattern of MTS evolution in Apicomplexa during the transition from a free-living or symbiotic algal ancestor to intracellular parasitism. As discussed in Chapter 2, and subsequently observed in Chapter 4, the predominant pattern is that of selective, lineage-specific loss. This is consistent with the traditional school of thought that dictates parasites lose non-essential machinery as they become increasingly dependent on their hosts[460], and also reflects the differences in host and host cell preference, morphology, lifecycle, and physiology of extant apicomplexan lineages. However, it is clear from the results presented in Chapters 4 and 5 that Apicomplexa have also acquired additional paralogues of select MTS machinery, at least one of which now plays an essential role during asexual growth in *T. gondii*. Hence, it can be appreciated that the picture of MTS evolution in Apicomplexa, and indeed their close relatives, is likely more complex than has been previously envisioned.

6.9 A philosophical plea for continued study of organellogenesis

This thesis outlines a clear hypothesis for the evolution of novel organelles in extant eukaryotic lineages. Just as the duplication and divergence of MTS paralogues during the FECA-LECA transition is hypothesized to have been involved in the expansion of a primordial set of autogenously generated compartments[446,453], so too is it possible that continued expansion of such factors could be associated with further expansion of these compartments in modern eukaryotes. This is especially true in cases where homologues of canonical (i.e. ancestral) eukaryotic organelles can be confidently inferred in the lineage(s) in question; as

discussed extensively in Chapter 3, functional homology then dictates that retained ancestral paralogues would be constrained in function and unable to mediate additional trafficking pathways required for organellogenesis.

Although the contents of this thesis may give the impression that such mechanisms only warrant study in enigmatic eukaryotic lineages, such duplications have been reported in model systems such as humans, yeast, and *A. thaliana*[394]. Furthermore, the presence of additional organelles may represent a more subjective matter than immediately apparent. For example, despite being a continuous structure, plasma membrane subdomains in polarized cell types such as epithelial and neural cells, receive different cargoes at both the apical and basolateral faces[762]. Many factors are involved including the AP-1 complex, in which an additional mu subunit paralogue, μ1B, is expressed in polarized cells to help mediate basolateral cargo sorting[763]. Furthermore, lysosome-related organelles (LROs) are present in numerous cell types and represent enticing candidates for study. For example, Rab38, which represents an additional paralogue beyond the ancestral Rab complement[110], and Stx13(12), which has paralogously expanded in some lineages (e.g. Stx7 in humans), have both been implicated in melanosome biogenesis (a form of LRO)[764,765]. Hence, the relationship between additional paralogues and cellular complexity probably exists across eukaryotes.

It is likely that such a general mechanism as envisioned here, duplication and divergence through relaxed constraints, is capable of generating diversity at multiple levels throughout the trafficking system. A full understanding of this mechanism may help to explain unique features within specialized cell types in metazoans and multicellular plants, novel organelles in diverse eukaryotic lineages, and may also inform knowledge regarding the origin of eukaryotes. Therefore, it is crucial that future studies investigate the correlation between machinery and function; in addition to expanding the view of the mechanistic aspects of this model, it is likely that initial hypotheses for function based on homology will prove an invaluable tool for molecular biology.

6.10 Proposals for future work

Given the arguments outlined in section 6.9, it is proposed that characterization of all identified LSPs in Apicomplexa be undertaken in future work. Given the hypotheses put forward regarding their potential involvement in novel organelles of Apicomplexa, and the known importance of these organelles to apicomplexan pathogenesis, they represent promising candidates for study. In addition, only by characterizing each LSP using in depth molecular methods can a full picture of apicomplexan MTS evolution be achieved and the

validity of the OPH model to explain the advent of micronemes, rhoptries, the IMC, the apicoplast, and perhaps even the dense granules, evaluated.

In addition, although extensively studied, the families presented in Chapters 2 and 4 do not represent all known MTS families (for example, COPI/II coat complexes, DENN and ELMOD GTPase regulators, etc). Hence, expanding the informatic characterization to include these additional families will be required to achieve a truly universal picture of trafficking evolution. In addition, despite continued sequencing efforts, the vast majority of complete genome sequences within the Myzozoa are restricted to the Apicomplexa, which, although useful, drastically limits the possibility to time character evolution within the Myzozoa as a whole. Although a large number of transcriptomes for dinoflagellates are available, for example through MMETSP[766], the nature of transcriptomic analysis precludes any inference based on absence of machinery in such datasets. Therefore, it is expected that additional genomes and genomic data for both dinoflagellates, as well as basal taxa on both the apicomplexan and dinoflagellate branches of the myzozoan tree, in the future will allow reinvestigation of the taxonomic distribution of LSPs with unprecedented accuracy.

Regarding the results presented in Chapter 5, characterization of the ArlX1 knockout, and disruption and characterization of ArlX2, will be required to understand their roles in *T. gondii*. Although the characterization of ArlX3 is extensive, future work could greatly improve understanding of its function. Application of fluorescent proteins or appropriate gene tags for live cell imaging will allow better understanding of the relationships between ArlX3 and resident proteins for both apical organelles and the apicoplast, specifically as it pertains to their trafficking. Additional apicoplast resident proteins, such as the recently described ATrx thioredoxin proteins[767], should be investigated to determine whether mis-localization of CPN60 represents an isolated phenotype or a more broad effect on apicoplast proteins. Some microneme and rhoptry proteins undergo proteolytic maturation which, based on localization and BFA-sensitivity data, is presumed to occur in a post-Golgi compartment[245,434,435]. Similarly, some apicoplast proteins are processed during translocation through one or more of the four membranes in the organelle[410]. Although not investigated here, the processing state of these proteins in ArlX3 knockdown cells is expected to further understanding of the nature of the mis-localized signal by placing bounds on the timing of that mis-localization. Finally, secretion assays for both micronemes[497] and rhoptries[768] have been described and can be used to support the inference that protein mis-localization coincides with organelle absence, or at least, the impaired formation of normal organelles in the absence of ArlX3.

6.11 Conclusions

This thesis has presented a rational argument for the ability of the OPH evolutionary mechanism to generate additional complexity, including in the advent of novel organelles, in extant eukaryotes. This relies in part on the presence of novel paralogues, but also on the presence/absence of ancestral organelles and associated machinery, as functional homology dictates that function is largely conserved among orthologues, even in distantly related eukaryotes. Adopting apicomplexan parasites as a model system, it is clear that novel paralogues are present and, at least in one case, function in a manner consistent with the proposed hypothesis. The results presented here are important for understanding the evolution of some of the world's deadliest parasites but are also generally applicable to all eukaryotes and deepen understanding of the mechanistic forces underlying the advent and retention of eukaryotic organelles.

References

1. Sogin, M. L. Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* **1**, 457–463 (1991).
2. Vossbrinck, C., Maddox, J., Friedman, S., Debrunner-Vossbrinck, B. & Woese, C. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* **326**, 411–414 (1987).
3. Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. & Peattie, D. A. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from Giardia lamblia. *Science.* **243**, 75–77 (1989).
4. Vossbrinck, C. R. & Woese, C. R. Eukaryotic ribosomes that lack a 5.8S RNA. *Nature* **320**, 287–288 (1986).
5. Cavalier-Smith, T. Archamoebae: the ancestral eukaryotes? *BioSystems* **25**, 25–38 (1991).
6. Cavalier-Smith, T. Eukaryotes with no mitochondria. *Nature* **326**, 332–333 (1987).
7. Cavalier-Smith, T. Molecular phylogeny. Archaebacteria and Archezoa. *Nature* **339**, l00–l01 (1989).
8. Stiller, J. W. & Hall, B. D. Long-Branch Attraction and the rDNA Model of Early Eukaryotic Evolution. *Mol. Biol. Evol.* **16**, 1270–1279 (1999).
9. Felsenstein, J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25**, 471–92 (1973).
10. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376 (1981).
11. Rannala, B. & Yang, Z. Probability Distribution of Molecular Evolutionary Trees: A New Method of Phylogenetic Inference. *J. Mol. Evol.* **43**, 304–311 (1996).
12. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
13. Delsuc, F., Brinkmann, H. & Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375 (2005).
14. Cavalier-Smith, T. A revised six-kingdom system of life. *Biol. Rev.* 203–266 (1998).
15. Van Der Giezen, M. Hydrogenosomes and mitosomes: Conservation and evolution of functions. *J. Eukaryot. Microbiol.* **56**, 221–231 (2009).
16. Karnkowska, A. *et al.* A eukaryote without a mitochondrial organelle. *Curr. Biol.* **26**, 1274–1284 (2016).
17. Adl, S. M. *et al.* Revisions to the Classification, Nomenclature, and Diversity of

Eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119 (2019).

18. Adl, S. M. *et al.* The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429–93 (2012).
19. Brown, M. W. *et al.* Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc. R. Soc. B* **280**, 20131755 (2013).
20. Burki, F., Okamoto, N., Pombert, J.-F. & Keeling, P. J. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc. R. Soc. B* **279**, 2246–54 (2012).
21. Bonifacino, J. S. & Glick, B. S. The mechanisms of vesicle budding and fusion. *Cell* **116**, 153–66 (2004).
22. Pasqualato, S., Renault, L. & Cherfils, J. Arf, Arl, Arp and Sar proteins: A family of GTP-binding proteins with a structural device for 'front-back' communication. *EMBO Rep.* **3**, 1035–1041 (2002).
23. Goldberg, J. Structural Basis for Activation of ARF GTPase: Mechanisms of Guanine Nucleotide Exchange and GTP-Myristoyl Switching. *Cell* **95**, 237–248 (1998).
24. Wennerberg, K., Rossman, K. L. & Der, C. J. The Ras superfamily at a glance. *J. Cell Sci.* **118**, 843–846 (2005).
25. Rojas, A. M., Fuentes, G., Rausell, A. & Valencia, A. The Ras protein superfamily: Evolutionary tree and role of conserved amino acids. *J. Cell Biol.* **196**, 189–201 (2012).
26. Gillingham, A. K. & Munro, S. The small G proteins of the Arf family and their regulators. *Annu. Rev. Cell Dev. Biol.* **23**, 579–611 (2007).
27. Donaldson, J. G. & Jackson, C. L. ARF family G proteins and their regulators: roles in membrane transport, development and disease. *Nat. Rev. Mol. Cell Biol.* **12**, 362–75 (2011).
28. Behnia, R., Panic, B., Whyte, J. R. C. & Munro, S. Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. *Nat. Cell Biol.* **6**, 405–413 (2004).
29. Setty, S. R. G., Strochlic, T. I., Tong, A. H. Y., Boone, C. & Burd, C. G. Golgi targeting of Arf-like GTPase Arl3p requires its Nα-acetylation and the integral membrane protein Sys1p. *Nat. Cell Biol.* **6**, 414–419 (2004).
30. Hofmann, I. & Munro, S. An N-terminally acetylated Arf-like GTPase is localised to lysosomes and affects their motility. *J. Cell Sci.* **119**, 1494–1503 (2006).
31. Li, Y. *et al.* Functional genomic analysis of the ADP-ribosylation factor family of GTPases: phylogeny among diverse eukaryotes and function in C. elegans. *FASEB J.* **18**, 1834–1850 (2004).

32. Schlacht, A., Mowbrey, K., Elias, M., Kahn, R. A. & Dacks, J. B. Ancient complexity, opisthokont plasticity, and discovery of the 11th subfamily of Arf GAP proteins. *Traffic* **14**, 636–49 (2013).
33. Schlacht, A. & Dacks, J. B. Unexpected ancient paralogues and an evolutionary model for the COPII coat complex. *Genome Biol. Evol.* **7**, 1098–1109 (2015).
34. Chardin, P. *et al.* A human exchange factor for ARF contains Sec7- and pleckstrin-homology domains. *Nature* **384**, 481–484 (1996).
35. Peyroche, A., Paris, S. & Jackson, C. L. Nucleotide exchange on ARF mediated by yeast Gea1 protein. *Nature* **384**, 479–481 (1996).
36. Pipaliya, S. V., Schlacht, A., Klinger, C. M., Kahn, R. A. & Dacks, J. Ancient complement and lineage-specific evolution of the Sec7 ARF GEF proteins in eukaryotes. *Mol. Biol. Cell* **30**, 1846–1863 (2019).
37. Cukierman, E., Huber, I., Rotman, M. & Cassel, D. The ARF1 GTPase-Activating Protein: Zinc Finger Motif and Golgi Complex Localization. *Science.* **270**, 1999–2002 (1995).
38. Barlowe, C. & Schekman, R. SEC12 encodes a guanine-nucleotide-exchange factor essential for transport vesicle budding from the ER. *Nature* **365**, 347–349 (1993).
39. Sato, K. & Nakano, A. Dissection of COPII subunit-cargo assembly and disassembly kinetics during Sar1p-GTP hydrolysis. *Nat. Struct. Mol. Biol.* **12**, 167–174 (2005).
40. Liu, Y.-W., Huang, C.-F., Huang, K.-B. & Lee, F.-J. S. Role for Gcs1p in Regulation of Arl1p at Trans-Golgi Compartments. *Mol. Biol. Cell* **16**, 4024–33 (2005).
41. Chen, K.-Y. *et al.* Syt1p promotes activation of Arl1p at the late Golgi to recruit Imh1p. *J. Cell Sci.* **123**, 3478–3489 (2010).
42. Bowzard, J. B., Cheng, D., Peng, J. & Kahn, R. A. ELMOD2 is an Arl2 GTPase-activating protein that also acts on Arfs. *J. Biol. Chem.* **282**, 17568–17580 (2007).
43. East, M. P., Bowzard, J. B., Dacks, J. B. & Kahn, R. A. ELMO domains, evolutionary and functional characterization of a novel GTPase-activating protein (GAP) domain for Arf protein family GTPases. *J. Biol. Chem.* **287**, 39538–39553 (2012).
44. Veltel, S., Gasper, R., Eisenacher, E. & Wittinghofer, A. The retinitis pigmentosa 2 gene product is a GTPase-activating protein for Arf-like 3. *Nat. Struct. Mol. Biol.* **15**, 373–380 (2008).
45. Keen, J. H. Clathrin assembly proteins: Affinity purification and a model for coat assembly. *J. Cell Biol.* **105**, 1989–1998 (1987).
46. Ohno, H. *et al.* Interaction of tyrosine-based sorting signals with clathrin-associated proteins. *Science.* **269**, 1872–1875 (1995).
47. Janvier, K. *et al.* Recognition of dileucine-based sorting signals from HIV-1 Nef and

LIMP-II by the AP-1 γ-σ1 and AP-3 δ-σ3 hemicomplexes. *J. Cell Biol.* **163**, 1281–1290 (2003).

48. Hirst, J. *et al.* The fifth adaptor protein complex. *PLoS Biol.* **9**, e1001170 (2011).
49. Hirst, J. *et al.* Characterization of TSET, an ancient and widespread membrane trafficking complex. *Elife* **3**, e02866 (2014).
50. Dell'Angelica, E. C., Klumperman, J., Stoorvogel, W. & Bonifacino, J. S. Association of the AP-3 Adaptor Complex with Clathrin. *Science.* **280**, 431–434 (1998).
51. Peden, A. A., Rudge, R. E., Lui, W. W. Y. & Robinson, M. S. Assembly and function of AP-3 complexes in cells expressing mutant subunits. *J. Cell Biol.* **156**, 327–336 (2002).
52. Hirst, J. *et al.* Interaction between AP-5 and the hereditary spastic paraplegia proteins SPG11 and SPG15. *Mol. Biol. Cell* **24**, 2558–69 (2013).
53. Barlowe, C. *et al.* COPII: A membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* **77**, 895–907 (1994).
54. D'Enfert, C., Wuestehube, L. J., Lila, T. & Schekman, R. Sec12p-dependent membrane binding of the small GTP-binding protein Sar1p promotes formation of transport vesicles from the ER. *J. Cell Biol.* **114**, 663–670 (1991).
55. D'Enfert, C., Barlowe, C., Nishikawa, S., Nakano, A. & Schekman, R. Structural and functional dissection of a membrane glycoprotein required for vesicle budding from the endoplasmic reticulum. *Mol. Cell. Biol.* **11**, 5727–5734 (1991).
56. Bi, X., Corpina, R. A. & Goldberg, J. Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. *Nature* **419**, 271–277 (2002).
57. Yoshihisa, T., Barlowe, C. & Schekman, R. Requirement for a GTPase-activating protein in vesicle budding from the endoplasmic reticulum. *Science.* **259**, 1466–1468 (1993).
58. Kuehn, M. J., Herrmann, J. M. & Schekman, R. COPII-cargo interactions direct protein sorting into ER-derived transport vesicles. *Nature* **391**, 187–190 (1998).
59. Bi, X., Mancias, J. D. & Goldberg, J. Insights into COPII Coat Nucleation from the Structure of Sec23•Sar1 Complexed with the Active Fragment of Sec31. *Dev. Cell* **13**, 635–645 (2007).
60. Kung, L. F. *et al.* Sec24p and Sec16p cooperate to regulate the GTP cycle of the COPII coat. *EMBO J.* **31**, 1014–1027 (2012).
61. Antonny, B., Madden, D., Hamamoto, S., Orci, L. & Schekman, R. Dynamics of the COPII coat with GTP and stable analogues. *Nat. Cell Biol.* **3**, 531–537 (2001).
62. Cai, H. *et al.* TRAPPI tethers COPII vesicles by binding the coat subunit Sec23. *Nature* **445**, 941–4 (2007).
63. Lord, C. *et al.* Sequential interactions with Sec23 control the direction of vesicle traffic.

Nature **473**, 181–6 (2011).

64. Bhandari, D. *et al.* Sit4p/PP6 regulates ER-to-Golgi traffic by controlling the dephosphorylation of COPII coat subunits. *Mol. Biol. Cell* **24**, 2727–2738 (2013).
65. Duden, R., Griffiths, G., Frank, R., Argos, P. & Kreis, T. E. β-COP, a 110 kd protein associated with non-clathrin-coated vesicles and the golgi complex, shows homology to β-adaptin. *Cell* **64**, 649–665 (1991).
66. Serafini, T. *et al.* A coat subunit of Golgi-derived non-clathrin-coated vesicles with homology to the clathrin-coated vesicle coat protein β-adaptin. *Nature* **349**, 215–220 (1991).
67. Waters, M. G., Serafini, T. & Rothman, J. E. 'Coatomer': A cytosolic protein complex containing subunits of non-clathrin-coated Golgi transport vesicles. *Nature* **349**, 248–251 (1991).
68. Bremser, M. *et al.* Coupling of coat assembly and vesicle budding to packaging of putative cargo receptors. *Cell* **96**, 495–506 (1999).
69. Spang, A., Matsuoka, K., Hamamoto, S., Schekman, R. & Orci, L. Coatomer, Arf1p, and nucleotide are required to bud coat protein complex I-coated vesicles from large synthetic liposomes. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11199–11204 (1998).
70. García-Mata, R., Szul, T., Alvarez, C. & Sztul, E. ADP-Ribosylation Factor/COPI-dependent Events at the Endoplasmic Reticulum-Golgi Interface Are Regulated by the Guanine Nucleotide Exchange Factor GBF1. *Mol. Biol. Cell* **14**, 2250–2261 (2003).
71. Hara-Kuge, S. *et al.* En bloc incorporation of coatomer subunits during the assembly of COP- coated vesicles. *J. Cell Biol.* **124**, 883–892 (1994).
72. Eugster, A., Frigerio, G., Dale, M. & Duden, R. The α- and β'-COP WD40 Domains Mediate Cargo-selective Interactions with Distinct Di-lysine Motifs. *Mol. Biol. Cell* **15**, 1011–1023 (2004).
73. Jackson, L. P. *et al.* Molecular Basis for Recognition of Dilysine Trafficking Motifs by COPI. *Dev. Cell* **23**, 1255–1262 (2012).
74. Tanigawa, G. *et al.* Hydrolysis of bound GTP by ARF protein triggers uncoating of golgi-derived COP-coated vesicles. *J. Cell Biol.* **123**, 1365–1371 (1993).
75. Poon, P. P. *et al.* Retrograde transport from the yeast Golgi is mediated by two ARF GAP proteins with overlapping function. *EMBO J.* **18**, 555–564 (1999).
76. Frigerio, G., Grimsey, N., Dale, M., Majoul, I. & Duden, R. Two human ARFGAPs associated with COP-I-coated vesicles. *Traffic* **8**, 1644–1655 (2007).
77. Pearse, B. M. F. Coated vesicles from pig brain: Purification and biochemical characterization. *J. Mol. Biol.* **97**, 93–98 (1975).

78. Keen, J. H., Willingham, M. C. & Pastan, I. H. Clathrin-coated vesicles: Isolation, dissociation and factor-dependent reassociation of clathrin baskets. *Cell* **16**, 303–312 (1979).
79. Zaremba, S. & Keen, J. H. Assembly polypeptides from coated vesicles mediate reassembly of unique clathrin coats. *J. Cell Biol.* **97**, 1339–1347 (1983).
80. Ungewickell, E. & Branton, D. Assembly units of clathrin coats. *Nature* **289**, 420–422 (1981).
81. Payne, G. S., Baker, D., Van Tuinen, E. & Schekman, R. Protein transport to the vacuole and receptor-mediated endocytosis by clathrin heavy chain-deficient yeast. *J. Cell Biol.* **106**, 1453–1461 (1988).
82. Payne, G. S. & Schekman, R. Clathrin: A role in the intracellular retention of a Golgi membrane protein. *Science.* **245**, 1358–1365 (1989).
83. Seeger, M. & Payne, G. S. A role for clathrin in the sorting of vacuolar proteins in the Golgi complex of yeast. *EMBO J.* **11**, 2811–2818 (1992).
84. Morris, S. M. & Cooper, J. A. Disabled-2 colocalizes with the LDLR in clathrin-coated pits and interacts with AP-2. *Traffic* **2**, 111–123 (2001).
85. He, G. *et al.* ARH is a modular adaptor protein that interacts with the LDL receptor, clathrin, and AP-2. *J. Biol. Chem.* **277**, 44044–44049 (2002).
86. Hirst, J. *et al.* A family of proteins with γ-adaptin and VHS domains that facilitate trafficking between the trans-golgi network and the vacuole/lysosome. *J. Cell Biol.* **149**, 67–80 (2000).
87. Dell'Angelica, E. C. *et al.* GGAs: A family of ADP ribosylation factor-binding proteins related to adaptors and associated with the Golgi complex. *J. Cell Biol.* **149**, 81–94 (2000).
88. Field, M. C., Gabernet-Castello, C. & Dacks, J. B. Reconstructing the evolution of the endocytic system: insights from genomics and molecular cell biology. *Adv. Exp. Med. Biol.* **607**, 84–96 (2007).
89. Manna, P. T., Gadelha, C., Puttick, A. E. & Field, M. C. ENTH and ANTH domain proteins participate in AP2-independent clathrin-mediated endocytosis. *J. Cell Sci.* **128**, 2130–2142 (2015).
90. Martina, J. A., Bonangelino, C. J., Aguilar, R. C. & Bonifacino, J. S. Stonin 2: An adaptor-like protein that interacts with components of the endocytic machinery. *J. Cell Biol.* **153**, 1111–1120 (2001).
91. Ungewickell, E. *et al.* Role of auxilin in uncoating clathrin-coated vesicles. *Nature* **378**, 632–635 (1995).

92. Taylor, M. J., Perrais, D. & Merrifield, C. J. A high precision survey of the molecular dynamics of mammalian clathrin-mediated endocytosis. *PLoS Biol.* **9**, e1000604 (2011).
93. Seaman, M. N. J., McCaffery, J. M. & Emr, S. D. A Membrane Coat Complex Essential for Endosome-to-Golgi Retrograde Transport in Yeast. *J. Cell Biol.* **142**, 665–681 (1998).
94. Horazdovsky, B. F. *et al.* A sorting nexin-1 homologue, Vps5p, forms a complex with Vps17p and is required for recycling the vacuolar protein-sorting receptor. *Mol. Biol. Cell* **8**, 1529–1541 (1997).
95. Norwood, S. J. *et al.* Assembly and solution structure of the core retromer protein complex. *Traffic* **12**, 56–71 (2011).
96. Koumandou, V. L. *et al.* Evolutionary reconstruction of the retromer complex and its function in Trypanosoma brucei. *J. Cell Sci.* **124**, 1496–509 (2011).
97. Seaman, M. N. J. Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer. *J. Cell Biol.* **165**, 111–122 (2004).
98. Fjorback, A. W. *et al.* Retromer Binds the FANSHY Sorting Motif in SorLA to Regulate Amyloid Precursor Protein Sorting and Processing. *J. Neurosci.* **32**, 1467–1480 (2012).
99. Nielsen, M. S. *et al.* Sorting by the Cytoplasmic Domain of the Amyloid Precursor Protein Binding Receptor SorLA. *Mol. Cell. Biol.* **27**, 6842–6851 (2007).
100. Seaman, M. N. J., Marcusson, E. G., Cereghino, J. L. & Emr, S. D. Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. *J. Cell Biol.* **137**, 79–92 (1997).
101. Arighi, C. N., Harmell, L. M., Aguilar, R. C., Haft, C. R. & Bonifacino, J. S. Role of the mammalian retromer in sorting of the cation-independent mannose 6-phosphate receptor. *J. Cell Biol.* **165**, 123–133 (2004).
102. Tabuchi, M., Yanatori, I., Kawai, Y. & Kishi, F. Retromer-mediated direct sorting is required for proper endosomal recycling of the mammalian iron transporter DMT1. *J. Cell Sci.* **123**, 756–766 (2010).
103. Temkin, P. *et al.* SNX27 mediates retromer tubule entry and endosome-to-plasma membrane trafficking of signalling receptors. *Nat. Cell Biol.* **13**, 715–723 (2011).
104. Lucas, M. *et al.* Structural Mechanism for Cargo Recognition by the Retromer Complex. *Cell* **167**, 1623-1635.e14 (2016).
105. Strochlic, T. I., Setty, T. G., Sitaram, A. & Burd, C. G. Grd19/Snx3p functions as a cargo-specific adapter for retromer-dependent endocytic recycling. *J. Cell Biol.* **177**, 115–125 (2007).

106. Rojas, R. *et al.* Regulation of retromer recruitment to endosomes by sequential action of Rab5 and Rab7. *J. Cell Biol.* **183**, 513–526 (2008).
107. Seaman, M. N. J., Harbour, M. E., Tattersall, D., Read, E. & Bright, N. Membrane recruitment of the cargo-selective retromer subcomplex is catalysed by the small GTPase Rab7 and inhibited by the Rab-GAP TBC1D5. *J. Cell Sci.* **122**, 2371–2382 (2009).
108. Jia, D. *et al.* Structural and mechanistic insights into regulation of the retromer coat by TBC1d5. *Nat. Commun.* **7**, 13305 (2016).
109. Alexandrov, K. *et al.* Characterization of the ternary complex between Rab7, REP-1 and Rab geranylgeranyl transferase. *Eur. J. Biochem.* **265**, 160–170 (1999).
110. Elias, M., Brighouse, A., Gabernet-Castello, C., Field, M. C. & Dacks, J. B. Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. *J. Cell Sci.* **125**, 2500–8 (2012).
111. Diekmann, Y. *et al.* Thousands of rab GTPases for the cell biologist. *PLoS Comput. Biol.* **7**, e1002217 (2011).
112. Stenmark, H. Rab GTPases as coordinators of vesicle traffic. *Nat. Rev. Mol. Cell Biol.* **10**, 513–25 (2009).
113. Gibieža, P. & Prekeris, R. Rab GTPases and cell division. *Small GTPases* **9**, 107–115 (2018).
114. Carney, D. S., Davies, B. A. & Horazdovsky, B. F. Vps9 domain-containing proteins: Activators of Rab5 GTPases from yeast to neurons. *Trends Cell Biol.* **16**, 27–35 (2006).
115. Horiuchi, H. *et al.* A Novel Rab5 GDP/GTP Exchange Factor Complexed to Rabaptin-5 Links Nucleotide Exchange to Effector Recruitment and Function. *Cell* **90**, 1149–1159 (1997).
116. Herman, E. K., Ali, M., Field, M. C. & Dacks, J. B. Regulation of early endosomes across eukaryotes: Evolution and functional homology of Vps9 proteins. *Traffic* **19**, 546–563 (2018).
117. Marat, A. L., Dokainish, H. & McPherson, P. S. DENN domain proteins: Regulators of Rab GTPases. *J. Biol. Chem.* **286**, 13791–13800 (2011).
118. Jones, S., Newman, C., Liu, F. & Segev, N. The TRAPP Complex Is a Nucleotide Exchanger for Ypt1 and Ypt31/32. *Mol. Biol. Cell* **11**, 4403–4411 (2000).
119. Nordmann, M. *et al.* The Mon1-Ccz1 complex is the GEF of the late endosomal Rab7 homolog Ypt7. *Curr. Biol.* **20**, 1654–1659 (2010).
120. Pusapati, G. V., Luchetti, G. & Pfeffer, S. R. Ric1-Rgp1 complex is a guanine nucleotide exchange factor for the late Golgi Rab6A GTPase and an effector of the medial Golgi

Rab33B GTPase. *J. Biol. Chem.* **287**, 42129–42137 (2012).

121. Nagano, F. *et al.* Molecular cloning and characterization of the noncatalytic subunit of the Rab3 subfamily-specific GTPase-activating protein. *J. Biol. Chem.* **273**, 24781–24785 (1998).
122. Pan, X., Eathiraj, S., Munson, M. & Lambright, D. G. TBC-domain GAPs for Rab GTPases accelerate GTP hydrolysis by a dual-finger mechanism. *Nature* **442**, 303–306 (2006).
123. Gabernet-Castello, C., O'Reilly, A. J., Dacks, J. B. & Field, M. C. Evolution of Tre-2/Bub2/Cdc16 (TBC) Rab GTPase-activating proteins. *Mol. Biol. Cell* **24**, 1574–83 (2013).
124. Christoforidis, S., McBride, H. M., Burgoyne, R. D. & Zerial, M. The rab5 effector EEA1 is a core component of endosome docking. *Nature* **397**, 621–625 (1999).
125. Murray, D. H. *et al.* An endosomal tether undergoes an entropic collapse to bring vesicles together. *Nature* **537**, 107–111 (2016).
126. Koumandou, V. L., Dacks, J. B., Coulson, R. M. R. & Field, M. C. Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. *BMC Evol. Biol.* **23**, 29 (2007).
127. Klinger, C. M., Klute, M. J. & Dacks, J. B. Comparative genomic analysis of multi-subunit tethering complexes demonstrates an ancient pan-eukaryotic complement and sculpting in Apicomplexa. *PLoS One* **8**, e76278 (2013).
128. Yu, I.-M. & Hughson, F. M. Tethering Factors as Organizers of Intracellular Vesicular Traffic. *Annu. Rev. Cell Dev. Biol.* **26**, 137–156 (2010).
129. Chou, H.-T., Dukovski, D., Chambers, M. G., Reinisch, K. M. & Walz, T. CATCHR, HOPS and CORVET tethering complexes share a similar architecture. *Nat. Struct. Mol. Biol.* **23**, 761–763 (2016).
130. Kraynack, B. A. *et al.* 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. *Mol. Biol. Cell* **16**, 3963–77 (2005).
131. Kamena, F. & Spang, A. Tip20p Prohibits Back-Fusion of COPII Vesicles with the Endoplasmic Reticulum. *Science.* **304**, 286–289 (2004).
132. Frigerio, G. The Saccharomyces cerevisiae early secretion mutant tip20 is synthetic lethal with mutants in yeast coatomer and the SNARE proteins Sec22p and Ufe1p. *Yeast* **14**, 633–646 (1998).
133. Andag, U. & Schmitt, H. D. Dsl1p, an essential component of the Golgi-endoplasmic reticulum retrieval system in yeast, uses the same sequence motif to interact with different subunits of the COPI vesicle coat. *J. Biol. Chem.* **278**, 51722–34 (2003).

134. Conibear, E., Cleck, J. N. & Stevens, T. H. Vps51p Mediates the Association of the GARP (Vps52/53/54) Complex with the Late Golgi t-SNARE Tlg1p. *Mol. Biol. Cell* **14**, 1610–1623 (2003).
135. Reggiori, F., Wang, C. W., Stromhaug, P. E., Shintani, T. & Klionsky, D. J. Vps51 is part of the yeast Vps fifty-three tethering complex essential for retrograde traffic from the early endosome and Cvt vesicle completion. *J. Biol. Chem.* **278**, 5009–5020 (2003).
136. Pérez-Victoria, F. J. & Bonifacino, J. S. Dual roles of the mammalian GARP complex in tethering and SNARE complex assembly at the trans-golgi network. *Mol. Cell. Biol.* **29**, 5251–63 (2009).
137. Pérez-Victoria, F. J. *et al.* Ang2/fat-free is a conserved subunit of the Golgi-associated retrograde protein complex. *Mol. Biol. Cell* **21**, 3386–3395 (2010).
138. Ungar, D. *et al.* Characterization of a mammalian Golgi-localized protein complex, COG, that is required for normal Golgi morphology and function. *J. Cell Biol.* **157**, 405–415 (2002).
139. Shestakova, A., Suvorova, E., Pavliv, O., Khaidakova, G. & Lupashin, V. Interaction of the conserved oligomeric Golgi complex with t-SNARE Syntaxin5a/Sed5 enhances intra-Golgi SNARE complex stability. *J. Cell Biol.* **179**, 1179–1192 (2007).
140. Zolov, S. N. & Lupashin, V. V. Cog3p depletion blocks vesicle-mediated Golgi retrograde trafficking in HeLa cells. *J. Cell Biol.* **168**, 747–759 (2005).
141. Heider, M. R. *et al.* Subunit connectivity, assembly determinants and architecture of the yeast exocyst complex. *Nat. Struct. Mol. Biol.* **23**, 59–66 (2016).
142. Mei, K. *et al.* Cryo-EM structure of the exocyst complex. *Nat. Struct. Mol. Biol.* **25**, 139–146 (2018).
143. TerBrush, D. R., Maurice, T., Roth, D. & Novick, P. The Exocyst is a multiprotein complex required for exocytosis in Saccharomyces cerevisiae. *EMBO J.* **15**, 6483–94 (1996).
144. Guo, W., Roth, D., Walch-Solimena, C. & Novick, P. The exocyst is an effector for Sec4p, targeting secretory vesicles to sites of exocytosis. *EMBO J.* **18**, 1071–1080 (1999).
145. Gromley, A. *et al.* Centriolin anchoring of exocyst and SNARE complexes at the midbody is required for secretory-vesicle-mediated abscission. *Cell* **123**, 75–87 (2005).
146. Bodemann, B. O. *et al.* RalB and the exocyst mediate the cellular starvation response by direct activation of autophagosome assembly. *Cell* **144**, 253–267 (2011).
147. Zuo, X., Guo, W. & Lipschutz, J. H. The Exocyst Protein Sec10 Is Necessary for Primary Ciliogenesis and Cystogenesis In Vitro. *Mol. Biol. Cell* **20**, 2522–2529 (2009).
148. Seals, D. F., Eitzen, G., Margolis, N., Wickner, W. T. & Price, A. A Ypt/Rab effector

complex containing the Sec1 homolog Vps33p is required for homotypic vacuole fusion. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9402–7 (2000).

149. Sato, T. K., Rehling, P., Peterson, M. R. & Emr, S. D. Class C Vps protein complex regulates vacuolar SNARE pairing and is required for vesicle docking/fusion. *Mol. Cell* **6**, 661–671 (2000).
150. Peplowska, K., Markgraf, D. F., Ostrowicz, C. W., Bange, G. & Ungermann, C. The CORVET tethering complex interacts with the yeast Rab5 homolog Vps21 and is involved in endo-lysosomal biogenesis. *Dev. Cell* **12**, 739–50 (2007).
151. Wurmser, A. E., Sato, T. K. & Emr, S. D. New component of the vacuolar class C-Vps complex couples nucleotide exchange on the Ypt7 GTPase to SNARE-dependent docking and fusion. *J. Cell Biol.* **151**, 551–562 (2000).
152. Brett, C. L. *et al.* Efficient termination of vacuolar Rab GTPase signaling requires coordinated action by a GAP and a protein kinase. *J. Cell Biol.* **182**, 1141–1151 (2008).
153. Sacher, M. *et al.* TRAPP, a highly conserved novel complex on the cis-Golgi that mediates vesicle docking and fusion. *EMBO J.* **17**, 2494–503 (1998).
154. Sacher, M., Barrowman, J., Schieltz, D., Yates, J. R. & Ferro-Novick, S. Identification and characterisation of five new subunits of TRAPP. *Eur. J. Cell Biol.* **79**, 71–80 (2000).
155. Sacher, M. *et al.* TRAPP I implicated in the specificity of tethering in ER-to-Golgi transport. *Mol. Cell* **7**, 433–42 (2001).
156. Yamasaki, A. *et al.* mTrs130 is a component of a mammalian TRAPPII complex, a Rab1 GEF that binds to COPI-coated vesicles. *Mol. Biol. Cell* **20**, 4205–4215 (2009).
157. Liang, Y., Morozova, N., Tokarev, A. A., Mulholland, J. W. & Segev, N. The Role of Trs65 in the Ypt/Rab Guanine Nucleotide Exchange Factor Function of the TRAPP II Complex. *Mol. Biol. Cell* **18**, 2533–2541 (2007).
158. Choi, C. *et al.* Organization and assembly of the TRAPPII complex. *Traffic* **12**, 715–25 (2011).
159. Scrivens, P. J. *et al.* TRAPPC2L is a novel, highly conserved TRAPP-interacting protein. *Traffic* **10**, 724–36 (2009).
160. Kim, J. J., Lipatova, Z. & Segev, N. TRAPP Complexes in Secretion and Autophagy. *Front. Cell Dev. Biol.* **4**, 20 (2016).
161. Meiling-Wesse, K. *et al.* Trs85 (Gsg1), a component of the TRAPP complexes, is required for the organization of the preautophagosomal structure during selective autophagy via the Cvt pathway. *J. Biol. Chem.* **280**, 33669–78 (2005).
162. Nazarko, T. Y., Huang, J., Nicaud, J.-M., Klionsky, D. J. & Sibirny, A. A. Trs85 is required for macroautophagy, pexophagy and cytoplasm to vacuole targeting in Yarrowia

lipolytica and Saccharomyces cerevisiae. *Autophagy* **1**, 37–45 (2005).

163. Reese, C., Heise, F. & Mayer, A. Trans-SNARE pairing can precede a hemifusion intermediate in intracellular membrane fusion. *Nature* **436**, 410–414 (2005).
164. Lu, X., Zhang, Y. & Shin, Y. K. Supramolecular SNARE assembly precedes hemifusion in SNARE-mediated membrane fusion. *Nat. Struct. Mol. Biol.* **15**, 700–706 (2008).
165. McDargh, Z. A., Polley, A. & O'Shaughnessy, B. SNARE-mediated membrane fusion is a two-stage process driven by entropic forces. *FEBS Lett.* **592**, 3504–3515 (2018).
166. Söllner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H. & Rothman, J. E. A protein assembly-disassembly pathway in vitro that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. *Cell* **75**, 409–418 (1993).
167. Yoshizawa, A. C. *et al.* Extracting sequence motifs and the phylogenetic features of SNARE-dependent membrane traffic. *Traffic* **7**, 1104–1118 (2006).
168. Kloepper, T. H., Kienle, C. N. & Fasshauer, D. An elaborate classification of SNARE proteins sheds light on the conservation of the eukaryotic endomembrane system. *Mol. Biol. Cell* **18**, 3463–3471 (2007).
169. Hata, Y., Slaughter, C. A. & Südhof, T. C. Synaptic vesicle fusion complex contains unc-18 homologue bound to syntaxin. *Nature* **366**, 347–351 (1993).
170. Dulubova, I. *et al.* A conformational switch in syntaxin during exocytosis: Role of munc18. *EMBO J.* **18**, 4372–4382 (1999).
171. Yu, H. *et al.* SNARE zippering requires activation by SNARE-like peptides in Sec1/Munc18 proteins. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8421–E8429 (2018).
172. Asao, H. *et al.* Hrs Is Associated with STAM, a Signal-transducing Adaptor Molecule. Its suppressive effect on cytokine-induced cell growth. *J. Biol. Chem.* **272**, 32785–32791 (1997).
173. Katzmann, D. J., Babst, M. & Emr, S. D. Ubiquitin-Dependent Sorting into the Multivesicular Body Pathway Requires the Function of a Conserved Endosomal Protein Sorting Complex, ESCRT-I. *Cell* **106**, 145–155 (2001).
174. Chu, T., Sun, J., Saksena, S. & Emr, S. D. New component of ESCRT-I regulates endosomal sorting complex assembly. *J. Cell Biol.* **175**, 815–823 (2006).
175. Curtiss, M., Jones, C. & Babst, M. Efficient Cargo Sorting by ESCRT-I and the Subsequent Release of ESCRT-I from Multivesicular Bodies Requires the Subunit Mvb12. *Mol. Biol. Cell* **18**, 636–645 (2007).
176. Babst, M., Katzmann, D. J., Snyder, W. B., Wendland, B. & Emr, S. D. Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body. *Dev. Cell* **3**, 283–289 (2002).

177. Babst, M., Katzmann, D. J., Estepa-Sabal, E. J., Meerloo, T. & Emr, S. D. ESCRT-III: An endosome-associated heterooligomeric protein complex required for MVB sorting. *Dev. Cell* **3**, 271–282 (2002).
178. Babst, M., Sato, T. K., Banta, L. M. & Emr, S. D. Endosomal transport function in yeast requires a novel AAA-type ATPase, Vps4p. *EMBO J.* **16**, 1820–1831 (1997).
179. Odorizzi, G., Katzmann, D. J., Babst, M., Audhya, A. & Emr, S. D. Bro1 is an endosome-associated protein that functions in the MVB pathway in Saccharomyces cerevisiae. *J. Cell Sci.* **116**, 1893–1903 (2003).
180. Luhtala, N. & Odorizzi, G. Bro1 coordinates deubiquitination in the multivesicular body pathway by recruiting Doa4 to endosomes. *J. Cell Biol.* **166**, 717–729 (2004).
181. Lottridge, J. M., Flannery, A. R., Vincelli, J. L. & Stevens, T. H. Vta1p and Vps46p regulate the membrane association and ATPase activity of Vps4p at the yeast multivesicular body. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6202–6207 (2006).
182. Leung, K. F., Dacks, J. B. & Field, M. C. Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. *Traffic* **9**, 1698–716 (2008).
183. Blanc, C. *et al.* Dictyostelium Tom1 participates to an ancestral ESCRT-0 complex. *Traffic* **10**, 161–71 (2009).
184. Filimonenko, M. *et al.* Functional multivesicular bodies are required for autophagic clearance of protein aggregates associated with neurodegenerative disease. *J. Cell Biol.* **179**, 485–500 (2007).
185. Lee, J. A., Beigneux, A., Ahmad, S. T., Young, S. G. & Gao, F. B. ESCRT-III Dysfunction Causes Autophagosome Accumulation and Neurodegeneration. *Curr. Biol.* **17**, 1561–1567 (2007).
186. Rusten, T. E. *et al.* ESCRTs and Fab1 Regulate Distinct Steps of Autophagy. *Curr. Biol.* **17**, 1817–1825 (2007).
187. Carlton, J. G., Agromayor, M. & Martin-Serrano, J. Differential requirements for Alix and ESCRT-III in cytokinesis and HIV-1 release. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10541–10546 (2008).
188. WHO. *World Malaria Report 2018*. (2018). doi:ISBN 978 92 4 1564403
189. Navin, T. R. & Hardy, A. M. Cryptosporidiosis in Patients with AIDS. *J. Infect. Dis.* **155**, 150 (1987).
190. DuPont, H. L. *et al.* The Infectivity of Cryptosporidium parvum in Healthy Volunteers. *N. Engl. J. Med.* **332**, 855–9 (1995).
191. Wang, R., Li, J., Chen, Y., Zhang, L. & Xiao, L. Widespread occurrence of Cryptosporidium infections in patients with HIV/AIDS: Epidemiology, clinical feature,

diagnosis, and therapy. *Acta Trop.* **187**, 257–263 (2018).

192. Chapman, H. D. Milestones in avian coccidiosis research : A review. *Poult. Sci.* **93**, 501–11 (2014).
193. Morrison, W. I. The aetiology, pathogenesis and control of theileriosis in domestic animals. *Rev. Sci. Tech.* **34**, 599–611 (2015).
194. McAllister, M. M. Diagnosis and Control of Bovine Neosporosis. *Vet. Clin. North Am. Food Anim. Pract.* **32**, 443–463 (2016).
195. Janouškovec, J. *et al.* Apicomplexan-like parasites are polyphyletic and widely but selectively dependent on cryptic plastid organelles. *Elife* **8**, e49662 (2019).
196. Thompson, G. A. & Nozawa, Y. Tetrahymena: A system for studying dynamic membrane alterations within the eukaryotic cell. *Biochim. Biophys. Acta* **472**, 55–92 (1977).
197. Prescott, D. M. The DNA of Ciliated Protozoa. *Microbiol. Rev.* **58**, 233–267 (1994).
198. Cavalier-Smith, T. & Chao, E. E. Protalveolate phylogeny and systematics and the origins of Sporozoa and dinoflagellates (phylum Myzozoa nom. nov.). *Eur. J. Protistol.* **40**, 185–212 (2004).
199. Hackett, J. D., Anderson, D. M., Erdner, D. L. & Bhattacharya, D. Dinoflagellates: a remarkable evolutionary experiment. *Am. J. Bot.* **91**, 1523–1534 (2004).
200. Kopp, C. *et al.* Subcellular investigation of photosynthesis-driven carbon assimilation and utilization in the symbiotic reef coral Pocillopora damicornis. *MBio* **6**, e02299-14 (2015).
201. Dorrell, R. G. & Howe, C. J. Integration of plastids with their hosts: Lessons learned from dinoflagellates. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10247–54 (2015).
202. Janouškovec, J., Horák, A., Oborník, M., Lukes, J. & Keeling, P. J. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10949–54 (2010).
203. Ševčíková, T. *et al.* Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci. Rep.* **5**, 10134 (2015).
204. Waller, R. F., Gornik, S. G., Koreny, L. & Pain, A. Metabolic pathway redundancy within the apicomplexan-dinoflagellate radiation argues against an ancient chromalveolate plastid. *Commun. Integr. Biol.* **9**, e1116653 (2016).
205. Janouškovec, J., Horák, A., Barott, K. L., Rohwer, F. L. & Keeling, P. J. Environmental distribution of coral-associated relatives of apicomplexan parasites. *ISME J.* **7**, 444–7 (2013).

206. Janouškovec, J., Horák, A., Barott, K. L., Rohwer, F. L. & Keeling, P. J. Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Curr. Biol.* **22**, R518-9 (2012).
207. Moore, R. B. *et al.* A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959–63 (2008).
208. Oborník, M. *et al.* Morphology, ultrastructure and life cycle of Vitrella brassicaformis n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist* **163**, 306–23 (2012).
209. De Niz, M. & Heussler, V. T. Rodent malaria models: insights into human disease and parasite biology. *Curr. Opin. Microbiol.* **46**, 93–101 (2018).
210. Rug, M. & Maier, A. G. Transfection of Plasmodium falciparum. *Methods Mol. Biol.* **923**, 75–98 (2013).
211. Collins, C. R. *et al.* Robust inducible Cre recombinase activity in the human malaria parasite Plasmodium falciparum enables efficient gene deletion within a single asexual erythrocytic growth cycle. *Mol. Microbiol.* **88**, 687–701 (2013).
212. Ghorbal, M. *et al.* Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nat. Biotechnol.* **32**, 819–821 (2014).
213. Wagner, J. C., Platt, R. J., Goldfless, S. J., Zhang, F. & Niles, J. C. Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum. *Nat. Methods* **11**, 915–918 (2014).
214. Bones, A. J. *et al.* Past and future trends of Cryptosporidium in vitro research. *Exp. Parasitol.* **196**, 28–37 (2019).
215. Morada, M. *et al.* Continuous culture of Cryptosporidium parvum using hollow fiber technology. *Int. J. Parasitol.* **46**, 21–29 (2016).
216. Miller, C. N. *et al.* A cell culture platform for Cryptosporidium that enables long-term cultivation and new tools for the systematic investigation of its biology. *Int. J. Parasitol.* **48**, 197–201 (2018).
217. Vinayak, S. *et al.* Genetic modification of the diarrhoeal pathogen Cryptosporidium parvum. *Nature* **523**, 477–480 (2015).
218. Beyer, T. V., Svezhova, N. V., Sidorenko, N. V. & Khokhlov, S. E. Cryptosporidium parvum (coccidia, apicomplexa): Some new ultrastructural observations on its endogenous development. *Eur. J. Protistol.* **36**, 151–159 (2000).
219. Cavalier-Smith, T. Gregarine site-heterogeneous 18S rDNA trees, revision of gregarine higher classification, and the evolutionary diversification of Sporozoa. *Eur. J. Protistol.* **50**, 472–495 (2014).

220. Suarez, C. E., Bishop, R. P., Alzan, H. F., Poole, W. A. & Cooke, B. M. Advances in the application of genetic manipulation methods to apicomplexan parasites. *Int. J. Parasitol.* **47**, 701–710 (2017).
221. Arranz-Solís, D., Regidor-Cerrillo, J., Lourido, S., Ortega-Mora, L. M. & Saeij, J. P. J. Toxoplasma CRISPR/Cas9 constructs are functional for gene disruption in Neospora caninum. *Int. J. Parasitol.* **48**, 597–600 (2018).
222. Langreth, S. G., Jensen, J. B., Reese, R. T. & Trager, W. Fine Structure of Human Malaria In Vitro. *J. Protozool.* **25**, 443–452 (1978).
223. Ferguson, D. J. P. & Dubremetz, J.-F. *The Ultrastructure of Toxoplasma gondii*. *Toxoplasma Gondii: The Model Apicomplexan - Perspectives and Methods: Second Edition* (Elsevier, 2014). doi:10.1016/B978-0-12-396481-6.00002-7
224. Morrissette, N. S. & Sibley, L. D. Cytoskeleton of apicomplexan parasites. *Microbiol. Mol. Biol. Rev.* **66**, 21–38; table of contents (2002).
225. Nichols, B. A. & Chiappino, M. L. Cytoskeleton of Toxoplasma gondii. *J. Protozool.* **34**, 217–226 (1987).
226. Hu, K., Roos, D. S. & Murray, J. M. A novel polymer of tubulin forms the conoid of Toxoplasma gondii. *J. Cell Biol.* **156**, 1039–1050 (2002).
227. Hu, K. *et al.* Cytoskeletal components of an invasion machine--the apical complex of Toxoplasma gondii. *PLoS Pathog.* **2**, e13 (2006).
228. Russell, D. G. & Burns, R. G. The polar ring of coccidian sporozoites: a unique microtubule-organizing centre. *J. Cell Sci.* **65**, 193–207 (1984).
229. Leung, J. M. *et al.* Stability and function of a putative microtubule-organizing center in the human parasite *Toxoplasma gondii*. *Mol. Biol. Cell* **28**, 1361–1378 (2017).
230. Francia, M. E. *et al.* Cell Division in Apicomplexan Parasites Is Organized by a Homolog of the Striated Rootlet Fiber of Algal Flagella. *PLoS Biol.* **10**, e1001444 (2012).
231. de Leon, J. C. *et al.* A SAS-6-Like Protein Suggests that the Toxoplasma Conoid Complex Evolved from Flagellar Components. *Eukaryot. Cell* **12**, 1009–1019 (2013).
232. Okamoto, N. & Keeling, P. J. The 3D structure of the apical complex and association with the flagellar apparatus revealed by serial TEM tomography in Psammosa pacifica, a distant relative of the Apicomplexa. *PLoS One* **9**, e84653 (2014).
233. Portman, N., Foster, C., Walker, G. & Šlapeta, J. Evidence of intraflagellar transport and apical complex formation in a free-living relative of the apicomplexa. *Eukaryot. Cell* **13**, 10–20 (2014).
234. Chen, A. L. *et al.* Novel Components of the Toxoplasma Inner Membrane Complex Revealed by BioID. *MBio* **6**, e02357-14 (2015).

235. Chen, A. L. *et al.* Novel insights into the composition and function of the Toxoplasma IMC sutures. *Cell. Microbiol.* **19**, 1–16 (2017).
236. Klinger, C. M., Nisbet, R. E., Ouologuem, D. T., Roos, D. S. & Dacks, J. B. Cryptic organelle homology in apicomplexan parasites: insights from evolutionary cell biology. *Curr. Opin. Microbiol.* **16**, 424–31 (2013).
237. Ouologuem, D. T. & Roos, D. S. Dynamics of the Toxoplasma gondii inner membrane complex. *J. Cell Sci.* **127**, 3320–30 (2014).
238. Harding, C. R. *et al.* Alveolar proteins stabilize cortical microtubules in Toxoplasma gondii. *Nat. Commun.* **10**, 401 (2019).
239. Nichols, B. A., Chiappino, M. L. & Pavesio, C. E. N. Endocytosis at the micropore of Toxoplasma gondii. *Parasitol. Res.* **80**, 91–98 (1994).
240. Hager, K. M., Striepen, B., Tilney, L. G. & Roos, D. S. The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite Toxoplasma gondii. *J. Cell Sci.* **112**, 2631–2638 (1999).
241. Van Wye, J. *et al.* Identification and localization of rab6, separation of rab6 from ERD2 and implications for an 'unstacked' Golgi, in Plasmodium falciparum. *Mol. Biochem. Parasitol.* **83**, 107–120 (1996).
242. Struck, N. S. *et al.* Re-defining the Golgi complex in Plasmodium falciparum using the novel Golgi marker PfGRASP. *J. Cell Sci.* **118**, 5603–5613 (2005).
243. Pelletier, L. *et al.* Golgi biogenesis in Toxoplasma gondii. *Nature* **418**, 548–52 (2002).
244. Wojczyk, B. S. *et al.* cDNA cloning and expression of UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase T1 from Toxoplasma gondii. *Mol. Biochem. Parasitol.* **131**, 93–107 (2003).
245. Harper, J. M. *et al.* A cleavable propeptide influences Toxoplasma infection by facilitating the trafficking and secretion of the TgMIC2–M2AP invasion complex. *Mol. Biol. Cell* **17**, 4551–4563 (2006).
246. Robibaro, B. *et al.* Toxoplasma gondii Rab5 enhances cholesterol acquisition from host cells. *Cell. Microbiol.* **4**, 139–152 (2002).
247. Miranda, K. *et al.* Characterization of a novel organelle in Toxoplasma gondii with similar composition and function to the plant vacuole. *Mol. Microbiol.* **76**, 1358–75 (2010).
248. Parussini, F., Coppens, I., Shah, P. P., Diamond, S. L. & Carruthers, V. B. Cathepsin L occupies a vacuolar compartment and is a protein maturase within the endo/exocytic system of Toxoplasma gondii. *Mol. Microbiol.* **76**, 1340–57 (2010).
249. Garnham, P. C. C., Baker, J. R. & Bird, R. G. Fine structure of cystic form of toxoplasma gondii. *Br. Med. J.* **1**, 83–84 (1962).

250. Dubois, D. J. & Soldati-Favre, D. Biogenesis and secretion of micronemes in *Toxoplasma gondii*. *Cell. Microbiol.* **21**, e13018 (2019).
251. Hammoudi, P. M., Maco, B., Dogga, S. K., Frénal, K. & Soldati-Favre, D. Toxoplasma gondii TFP1 is an essential transporter family protein critical for microneme maturation and exocytosis. *Mol. Microbiol.* **109**, 225–244 (2018).
252. Bannister, L. H. *et al.* Plasmodium falciparum apical membrane antigen 1 (PfAMA-1) is translocated within micronemes along subpellicular microtubules during merozoite development. *J. Cell Sci.* **116**, 3825–34 (2003).
253. Kremer, K. *et al.* An overexpression screen of Toxoplasma gondii Rab-GTPases reveals distinct transport routes to the micronemes. *PLoS Pathog.* **9**, e1003213 (2013).
254. Bradley, P. J. *et al.* Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in Toxoplasma gondii. *J. Biol. Chem.* **280**, 34245–58 (2005).
255. Lemgruber, L., Lupetti, P., De Souza, W. & Vommaro, R. C. New details on the fine structure of the rhoptry of Toxoplasma gondii. *Microsc. Res. Tech.* **74**, 812–8 (2011).
256. Nishi, M., Hu, K., Murray, J. M. & Roos, D. S. Organellar dynamics during the cell cycle of Toxoplasma gondii. *J. Cell Sci.* **121**, 1559–68 (2008).
257. Shaw, M. K., Roos, D. S. & Tilney, L. G. Acidic compartments and rhoptry formation in Toxoplasma gondii. *Parasitology* **117**, 435–43 (1998).
258. Venugopal, K. *et al.* Dual role of the Toxoplasma gondii clathrin adaptor AP1 in the sorting of rhoptry and microneme proteins and in parasite division. *PLoS Pathog.* **13**, e1006331 (2017).
259. Mercier, C. & Cesbron-Delauw, M. F. Toxoplasma secretory granules: One population or more? *Trends Parasitol.* **31**, 604 (2015).
260. Lemgruber, L. *et al.* Cryo-electron tomography reveals four-membrane architecture of the Plasmodium apicoplast. *Malar. J.* **12**, 25 (2013).
261. Köhler, S. *et al.* A Plastid of Probable Green Algal Origin in Apicomplexan Parasites. *Science.* **275**, 1485–1489 (1997).
262. McFadden, G. I., Reith, M. E., Munholland, J. & Lang-Unnasch, N. Plastid in human parasites. *Nature* **381**, 482 (1996).
263. Ovciarikova, J., Lemgruber, L., Stilger, K. L. & Sullivan, W. J. Mitochondrial behaviour throughout the lytic cycle of Toxoplasma gondii. *Sci. Rep.* **7**, 42746 (2017).
264. Frenkel, J. K., Dubey, J. P. & Miller, N. L. Toxoplasma gondii in Cats: Fecal Stages Identified as Coccidian Oocysts. *Science.* **167**, 893–896 (1970).
265. Dlugońska, H. Toxoplasma gondii and the host cells. *Ann. Parasitol.* **60**, 83–88 (2014).

266. Montoya, J. G. & Liesenfeld, O. Toxoplasmosis. *Lancet* **363**, 1965–76 (2004).
267. Dubey, J. P. & Frenkel, J. K. Cyst-Induced Toxoplasmosis in Cats. *J. Protozool.* **19**, 155–177 (1972).
268. Dubey, J. P. & Frenkel, J. K. Feline toxoplasmosis from acutely infected mice and the development of Toxoplasma cysts. *J. Protozool.* **23**, 537–546 (1976).
269. Speer, C. A. & Dubey, J. P. Ultrastructural differentiation of Toxoplasma gondii schizonts (types B to E) and gamonts in the intestines of cats fed bradyzoites. *Int. J. Parasitol.* **35**, 193–206 (2005).
270. Ferguson, D. J., Hutchison, W. M., Dunachie, J. F. & Siim, J. C. Ultrastructural study of early stages of asexual multiplication and microgametogony of Toxoplasma gondii in the small intestine of the cat. *Acta Pathol. Microbiol. Scand. B. Microbiol. Immunol.* **82**, 167–81 (1974).
271. Ferguson, D. J., Hutchison, W. M. & Siim, J. C. The ultrastructural development of the macrogamete and formation of the oocyst wall of Toxoplasma gondii. *Acta Pathol Microbiol Scand B* **83**, 491–505 (1975).
272. Ferguson, D. J., Birch-Andersen, A., Siim, J. C. & Hutchison, W. M. Ultrastructural studies on the sporulation of oocysts of Toxoplasma gondii. I. Development of the Zygote and Formation of the Sporoblasts. *Acta Pathol Microbiol Scand B* **87**, 171–181 (1979).
273. Ferguson, D. J., Birch-Andersen, A., Siim, J. C. & Hutchison, W. M. Ultrastructural studies on the sporulation of oocysts of Toxoplasma gondii. II. Formation of the Sporocyst and Structure of the Sporocyst Wall. *Acta Pathol Microbiol Scand B* **87**, 183–190 (1979).
274. Ferguson, D. J., Birch-Andersen, A., Siim, J. C. & Hutchison, W. M. Ultrastructural studies on the sporulation of oocysts of Toxoplasma gondii. III. Formation of the Sporozoites within the Sporocysts. *Acta Pathol Microbiol Scand B* **87**, 253–260 (1979).
275. Speer, C. A., Clark, S. & Dubey, J. P. Ultrastructure of the oocysts, sporocysts, and sporozoites of Toxoplasma gondii. *J. Parasitol.* **84**, 505–12 (1998).
276. Wainwright, K. E. *et al.* Chemical Inactivation of Toxoplasma gondii Oocysts in Water. *J. Parasitol.* **93**, 925–931 (2007).
277. Wainwright, K. E. *et al.* Physical Inactivation of Toxoplasma gondii Oocysts in Water. *Appl. Environ. Microbiol.* **73**, 5663–5666 (2007).
278. Dumètre, A. *et al.* Effects of ozone and ultraviolet radiation treatments on the infectivity of Toxoplasma gondii oocysts. *Vet. Parasitol.* **153**, 209–213 (2008).
279. Blader, I. J., Coleman, B. I., Chen, C.-T. & Gubbels, M.-J. Lytic Cycle of Toxoplasma

gondii: 15 Years Later. *Annu. Rev. Microbiol.* **69**, 463–85 (2015).

280. Endo, T., Sethi, K. K. & Piekarski, G. Toxoplasma gondii: Calcium Ionophore A23187-mediated exit of trophozoites from infected murine macrophages. *Exp. Parasitol.* **53**, 179–188 (1982).
281. Kafsack, B. F. C. *et al.* Rapid Membrane Disruption by a Perforin-Like Protein Facilitates Parasite Exit from Host Cells. *Science.* **323**, 530–533 (2009).
282. Garrison, E. *et al.* A Forward Genetic Screen Reveals that Calcium-dependent Protein Kinase 3 Regulates Egress in Toxoplasma. *PLoS Pathog.* **8**, e1003049 (2012).
283. Opitz, C. & Soldati, D. 'The glideosome': A dynamic complex powering gliding motion and host cell invasion by Toxoplasma gondii. *Mol. Microbiol.* **45**, 597–604 (2002).
284. Ryning, F. W. & Remington, J. S. Effect of cytochalasin D on Toxoplasma gondii Cell Entry. *Infect. Immun.* **20**, 739–43 (1978).
285. Meissner, M., Schlüter, D. & Soldati, D. Role of Toxoplasma gondii myosin A in powering parasite gliding and host cell invasion. *Science.* **298**, 837–40 (2002).
286. Herm-Götz, A. *et al.* Toxoplasma gondii myosin A and its light chain: A fast, single-headed, plus-end-directed motor. *EMBO J.* **21**, 2149–2158 (2002).
287. Gaskins, E. *et al.* Identification of the membrane receptor of a class XIV myosin in Toxoplasma gondii. *J. Cell Biol.* **165**, 383–393 (2004).
288. Frénal, K. *et al.* Functional dissection of the apicomplexan glideosome molecular architecture. *Cell Host Microbe* **8**, 343–357 (2010).
289. Nebl, T. *et al.* Quantitative in vivo analyses reveal calcium-dependent phosphorylation sites and identifies a novel component of the toxoplasma invasion motor complex. *PLoS Pathog.* **7**, e1002222 (2011).
290. Egarter, S. *et al.* The toxoplasma acto-myoA motor complex is important but not essential for gliding motility and host cell invasion. *PLoS One* **9**, e91819 (2014).
291. Andenmatten, N. *et al.* Conditional genome engineering in Toxoplasma gondii uncovers alternative invasion mechanisms. *Nat. Methods* **10**, 125–127 (2013).
292. Whitelaw, J. A. *et al.* Surface attachment, promoted by the actomyosin system of Toxoplasma gondii is important for efficient gliding motility and invasion. *BMC Biol.* **15**, 1 (2017).
293. Münter, S. *et al.* Plasmodium Sporozoite Motility Is Modulated by the Turnover of Discrete Adhesion Sites. *Cell Host Microbe* **6**, 551–562 (2009).
294. Gras, S. *et al.* An endocytic-secretory cycle participates in Toxoplasma gondii in motility. *PLoS Biol.* **17**, e3000060 (2019).
295. Frénal, K., Dubremetz, J.-F., Lebrun, M. & Soldati-Favre, D. Gliding motility powers

invasion and egress in Apicomplexa. *Nat. Rev. Microbiol.* **15**, 645–660 (2017).

296. Carruthers, V. & Boothroyd, J. C. Pulling together: an integrated model of Toxoplasma cell invasion. *Curr. Opin. Microbiol.* **10**, 83–9 (2007).
297. Lebrun, M. *et al.* The rhoptry neck protein RON4 re-localizes at the moving junction during Toxoplasma gondii invasion. *Cell. Microbiol.* **7**, 1823–33 (2005).
298. Alexander, D. L., Mital, J., Ward, G. E., Bradley, P. & Boothroyd, J. C. Identification of the moving junction complex of Toxoplasma gondii: A collaboration between distinct secretory organelles. *PLoS Pathog.* **1**, e17 (2005).
299. Mital, J., Meissner, M., Soldati, D. & Ward, G. E. Conditional Expression of Toxoplasma gondii Apical Membrane Antigen-1 (TgAMA1) Demonstrates That TgAMA1 Plays a Critical Role in Host Cell Invasion. *Mol. Biol. Cell* **16**, 4341–49 (2005).
300. Besteiro, S., Michelin, A., Poncet, J., Dubremetz, J.-F. & Lebrun, M. Export of a Toxoplasma gondii rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. *PLoS Pathog.* **5**, e1000309 (2009).
301. Tyler, J. S. & Boothroyd, J. C. The C-terminus of Toxoplasma RON2 provides the crucial link between AMA1 and the host-associated invasion complex. *PLoS Pathog.* **7**, e1001282 (2011).
302. Lamarque, M. *et al.* The RON2-AMA1 interaction is a critical step in moving junction-dependent invasion by apicomplexan parasites. *PLoS Pathog.* **7**, e1001276 (2011).
303. Bichet, M. *et al.* The toxoplasma-host cell junction is anchored to the cell cortex to sustain parasite invasive force. *BMC Biol.* **12**, 773 (2014).
304. Gonzalez, V. *et al.* Host Cell Entry by Apicomplexa Parasites Requires Actin Polymerization in the Host Cell. *Cell Host Microbe* **5**, 259–272 (2009).
305. Delorme-Walker, V. *et al.* Toxofilin upregulates the host cortical actin cytoskeleton dynamics, facilitating Toxoplasma invasion. *J. Cell Sci.* **125**, 4333–4342 (2012).
306. Gaji, R. Y., Huynh, M. H. & Carruthers, V. B. A Novel High Throughput Invasion Screen Identifies Host Actin Regulators Required for Efficient Cell Entry by Toxoplasma gondii. *PLoS One* **8**, e64693 (2013).
307. Bichet, M. *et al.* Genetic impairment of parasite myosin motors uncovers the contribution of host cell membrane dynamics to Toxoplasma invasion forces. *BMC Biol.* **14**, 97 (2016).
308. Sibley, L. D., Niesman, I. R., Parmley, S. F. & Cesbron-Delauw, M. F. Regulated secretion of multi-lamellar vesicles leads to formation of a tubulo-vesicular network in host-cell vacuoles occupied by Toxoplasma gondii. *J. Cell Sci.* **108**, 1669–77 (1995).
309. Mercier, C. *et al.* Biogenesis of Nanotubular Network in Toxoplasma Parasitophorous

Vacuole Induced by Parasite Proteins. *Mol. Biol. Cell* **13**, 2397–2409 (2002).

310. Behnke, M. S. *et al.* Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of Toxoplasma gondii. *PLoS One* **5**, e12354 (2010).
311. Goldman, M., Carver, R. K. & Sulzer, A. J. Reproduction of Toxoplasma gondii by Internal Budding. *J. Parasitol.* **44**, 161–171 (1958).
312. Jimenez-Ruiz, E., Morlon-Guyot, J., Daher, W. & Meissner, M. Vacuolar protein sorting mechanisms in apicomplexan parasites. *Mol. Biochem. Parasitol.* **209**, 18–25 (2016).
313. Muñiz-Hernández, S. *et al.* Contribution of the residual body in the spatial organization of toxoplasma gondii tachyzoites within the parasitophorous vacuole. *J. Biomed. Biotechnol.* **2011**, 473983 (2011).
314. Periz, J. *et al.* Toxoplasma gondii F-actin forms an extensive filamentous network required for material exchange and parasite maturation. *Elife* **6**, e24119 (2017).
315. Frénal, K. *et al.* Myosin-dependent cell-cell communication controls synchronicity of division in acute and chronic stages of Toxoplasma gondii. *Nat. Commun.* **8**, 15710 (2017).
316. Tosetti, N., Pacheco, N. D. S., Soldati-Favre, D. & Jacot, D. Three F-actin assembly centers regulate organelle inheritance , cell-cell communication and motility in Toxoplasma gondii. *Elife* **8**, e42669 (2019).
317. Jeffers, V., Tampaki, Z., Kim, K. & Sullivan, W. J. J. A latent ability to persist: differentiation in Toxoplasma gondii. *Cell. Mol. Life Sci.* **75**, 2355–2373 (2018).
318. Dubey, J. P., Lindsay, D. S. & Speer, C. A. Structures of Toxoplasma gondii tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. *Clin. Microbiol. Rev.* **11**, 267–299 (1998).
319. Ferguson, D. J. P. & Hutchison, W. M. An ultrastructural study of the early development and tissue cyst formation of Toxoplasma gondii in the brains of mice. *Parasitol. Res.* **73**, 483–491 (1987).
320. Ferguson, D. J. P., Hutchison, W. M. & Pettersen, E. Tissue cyst rupture in mice chronically infected with Toxoplasma gondii - An immunocytochemical and ultrastructural study. *Parasitol. Res.* **75**, 599–603 (1989).
321. Rougier, S., Montoya, J. G. & Peyron, F. Lifelong Persistence of Toxoplasma Cysts: A Questionable Dogma? *Trends Parasitol.* **33**, 93–101 (2017).
322. Dunay, I. R., Gajurel, K., Dhakal, R., Liesenfeld, O. & Montoya, J. G. Treatment of Toxoplasmosis: Historical Perspective , Animal Models, and Current Clinical Practice. *Clin. Microbiol. Rev.* **31**, e00057-17 (2018).
323. Tenter, A. M., Heckeroth, A. R. & Weiss, L. M. Toxoplasma gondii: from animals to

humans. *Int. J. Parasitol.* **30**, 1217–58 (2000).

324. Jones, J. L. *et al.* Toxoplasma gondii infection in the United States, 2011-2014. *Am. J. Trop. Med. Hyg.* **98**, 551–557 (2018).
325. Schurer, J. M. *et al.* Toxoplasmosis and toxocariasis: An assessment of human immunodeficiency virus comorbidity and health-care costs in Canada. *Am. J. Trop. Med. Hyg.* **95**, 168–174 (2016).
326. Thomas, M. K. *et al.* Estimates of the burden of foodborne illness in Canada for 30 specified pathogens and unspecified agents, Circa 2006. *Foodborne Pathog. Dis.* **10**, 639–648 (2013).
327. Messier, V. *et al.* Seroprevalence of Toxoplasma gondii among Nunavik Inuit (Canada). *Zoonoses Public Health* **56**, 188–97 (2009).
328. Jenkins, E. J. *et al. Tradition and Transition: Parasitic Zoonoses of People and Animals in Alaska, Northern Canada, and Greenland*. *Advances in Parasitology* **82**, (Elsevier, 2013).
329. Porter, S. B. & Sande, M. A. Toxoplasmosis of the Central Nervous System in the Acquired Immunodeficiency Syndrome. *N. Engl. J. Med.* **327**, 1643–8 (1992).
330. Bosch-Driessen, L. E. H., Berendschot, T. T. J. M., Ongkosuwito, J. V & Rothova, A. Ocular toxoplasmosis: Clinical features and prognosis of 154 patients. *Ophthalmology* **109**, 869–878 (2002).
331. McAuley, J. B. Congenital Toxoplasmosis. *J. Pediatric Infect. Dis. Soc.* **3**, S30–S35 (2014).
332. Sibley, L. D. & Boothroyd, J. C. Virulent strains of Toxoplasma gondii comprise a single clonal lineage. *Nature* **359**, 82–85 (1992).
333. Howe, D. K. & Sibley, L. D. Toxoplasma gondii Comprises Three Clonal Lineages: Correlation of Parasite Genotype with Human Disease. *J. Infect. Dis.* **172**, 1561–1566 (1995).
334. Khan, A. *et al.* Genetic analyses of atypical Toxoplasma gondii strains reveal a fourth clonal lineage in North America. *Int. J. Parasitol.* **41**, 645–655 (2011).
335. Su, C. *et al.* Globally diverse Toxoplasma gondii isolates comprise six major clades originating from a small number of distinct ancestral lineages. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5844–5849 (2012).
336. Behnke, M. S., Dubey, J. P. & Sibley, L. D. Genetic Mapping of Pathogenesis Determinants in Toxoplasma gondii. *Annu. Rev. Microbiol.* **70**, 63–81 (2016).
337. Khan, A. *et al.* Genetic divergence of Toxoplasma gondii strains associated with ocular toxoplasmosis, Brazil. *Emerg. Infect. Dis.* **12**, 942–949 (2006).

338. Demar, M. *et al.* Fatal Outbreak of Human Toxoplasmosis along the Maroni River: Epidemiological, Clinical, and Parasitological Aspects. *Clin. Infect. Dis.* **45**, e88–e95 (2007).
339. Pfefferkorn, E. R. & Pfefferkorn, L. C. Toxoplasma gondii: Isolation and preliminary characterization of temperature-sensitive mutants. *Exp. Parasitol.* **39**, 365–376 (1976).
340. Pfefferkorn, E. R. & Kasper, L. H. Toxoplasma gondii: Genetic crosses reveal phenotypic suppression of hydroxyurea resistance by fluorodeoxyuridine resistance. *Exp. Parasitol.* **55**, 207–218 (1983).
341. Pfefferkorn, L. C. & Pfefferkorn, E. R. Toxoplasma gondii: Genetic recombination between drug resistant mutants. *Exp. Parasitol.* **50**, 305–316 (1980).
342. Soldati, D. & Boothroyd, J. C. Transient Transfection and Expression in the Obligate Intracellular Parasite Toxoplasma gondii. *Science.* **260**, 349–352 (1993).
343. Donald, R. G. & Roos, D. S. Stable molecular transformation of Toxoplasma gondii: a selectable dihydrofolate reductase-thymidylate synthase marker based on drug-resistance mutations in malaria. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11703–7 (1993).
344. Kim, K., Soldati, D. & Boothroyd, J. C. Gene replacement in Toxoplasma gondii with chloramphenicol acetyltransferase as selectable marker. *Science.* **262**, 911–914 (1993).
345. Pfefferkorn, E. R. Interferon gamma blocks the growth of Toxoplasma gondii in human fibroblasts by inducing the host cells to degrade tryptophan. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 908–12 (1984).
346. Sibley, L. D., Messina, M. & Niesman, I. R. Stable DNA transformation in the obligate intracellular parasite Toxoplasma gondii by complementation of tryptophan auxotrophy. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5508–5512 (1994).
347. Donald, R. G. K., Carter, D., Ullman, B. & Roos, D. S. Insertional Tagging, Cloning, and Expression of the Toxoplasma gondii Hypoxanthine-Xanthine-Guanine Phosphoribosyltransferase Gene. Use as a selectable marker for stable transformation. *J. Biol. Chem.* **271**, 14010–14019 (1996).
348. Pfefferkorn, E. R. & Borotz, S. E. Toxoplasma gondii: Characterization of a Mutant Resistant to 6-Thioxanthine. *Exp. Parasitol.* **79**, 374–82 (1994).
349. Pfefferkorn, E. R. & Pfefferkorn, L. C. Toxoplasma gondii: Characterization of a mutant resistant to 5-fluorodeoxyuridine. *Exp. Parasitol.* **42**, 44–55 (1977).
350. Bohne, W. & Roos, D. S. Stage-specific expression of a selectable marker in Toxoplasma gondii permits selective inhibition of either tachyzoites or bradyzoites. *Mol. Biochem.*

Parasitol. **88**, 115–126 (1997).

351. Donald, R. G. & Roos, D. S. Insertional mutagenesis and marker rescue in a protozoan parasite: cloning of the uracil phosphoribosyltransferase locus from Toxoplasma gondii. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5749–53 (1995).
352. Matrajt, M., Donald, R. G. K., Singh, U. & Roos, D. S. Identification and characterization of differentiation mutants in the protozoan parasite Toxoplasma gondii. *Mol. Microbiol.* **44**, 735–747 (2002).
353. Gubbels, M.-J. *et al.* Forward genetic analysis of the apicomplexan cell division cycle in Toxoplasma gondii. *PLoS Pathog.* **4**, e36 (2008).
354. Sidik, S. M. *et al.* A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell* **166**, 1423-1435.e12 (2016).
355. Huynh, M. H. & Carruthers, V. B. Tagging of endogenous genes in a Toxoplasma gondii strain lacking Ku80. *Eukaryot. Cell* **8**, 530–539 (2009).
356. Fox, B. A. *et al.* Type II Toxoplasma gondii KU80 knockout strains enable functional analysis of genes required for Cyst development and latent infection. *Eukaryot. Cell* **10**, 1193–1206 (2011).
357. Meissner, M., Brecht, S., Bujard, H. & Soldati, D. Modulation of myosin A expression by a newly established tetracycline repressor-based inducible system in Toxoplasma gondii. *Nucleic Acids Res.* **29**, E115 (2001).
358. Sheiner, L. *et al.* A systematic screen to discover and analyze apicoplast proteins identifies a conserved and essential protein import factor. *PLoS Pathog.* **7**, e1002392 (2011).
359. Banaszynski, L. A., Chen, L., Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. *Cell* **126**, 995–1004 (2006).
360. Armstrong, C. M. & Goldberg, D. E. An FKBP destabilization domain modulates protein levels in Plasmodium falciparum. *Nat. Methods* **4**, 1007–1009 (2007).
361. Herm-Götz, A. *et al.* Rapid control of protein level in the apicomplexan Toxoplasma gondii. *Nat. Methods* **4**, 1003–1005 (2007).
362. Jiménez-Ruiz, E., Wong, E. H., Pall, G. S. & Meissner, M. Advantages and disadvantages of conditional systems for characterization of essential genes in Toxoplasma gondii. *Parasitology* **141**, 1390–1398 (2014).
363. Jacot, D., Meissner, M., Sheiner, L., Soldati-Favre, D. & Striepen, B. Genetic Manipulation of Toxoplasma gondii. in *Toxoplasma Gondii: The Model Apicomplexan - Perspectives and Methods: Second Edition* 577–611 (2014). doi:10.1016/B978-0-12-

396481-6.00017-9

364. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science.* **315**, 1709–1712 (2007).
365. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337**, 816–821 (2012).
366. Shen, B., Brown, K. M., Lee, T. D. & Sibley, L. D. Efficient Gene Disruption in Diverse Strains of Toxoplasma gondii Using CRISPR/CAS9. *MBio* **5**, e01114-14 (2014).
367. Di Cristina, M. & Carruthers, V. B. New and emerging uses of CRISPR/Cas9 to genetically manipulate apicomplexan parasites. *Parasitology* **145**, 1119–1126 (2018).
368. Brown, K. M., Long, S. & Sibley, L. D. Plasma membrane association by N-acylation governs PKG function in Toxoplasma gondii. *MBio* **8**, e00375-17 (2017).
369. Langsley, G. *et al.* Comparative genomics of the Rab protein family in Apicomplexan parasites. *Microbes Infect.* **10**, 462–470 (2008).
370. Langemeyer, L., Fröhlich, F. & Ungermann, C. Rab GTPase Function in Endosome and Lysosome Biogenesis. *Trends Cell Biol.* **28**, 957–970 (2018).
371. Shin, H.-W. *et al.* An enzymatic cascade of Rab5 effectors regulates phosphoinositide turnover in the endocytic pathway. *J. Cell Biol.* **170**, 607–618 (2005).
372. Lachmann, J., Glaubke, E., Moore, P. S. & Ungermann, C. The Vps39-like TRAP1 is an effector of Rab5 and likely the missing Vps3 subunit of human CORVET. *Cell. Logist.* **4**, e970840 (2014).
373. Poteryaev, D., Datta, S., Ackema, K., Zerial, M. & Spang, A. Identification of the switch in early-to-late endosome transition. *Cell* **141**, 497–508 (2010).
374. Ezougou, C. N. *et al.* Plasmodium falciparum Rab5B is an N-terminally myristoylated Rab GTPase that is targeted to the parasite's plasma and food vacuole membranes. *PLoS One* **9**, e87695 (2014).
375. Sangaré, L. O. *et al.* Unconventional endosome-like compartment and retromer complex in Toxoplasma gondii govern parasite integrity and host infection. *Nat. Commun.* **7**, 11191 (2016).
376. Matanis, T. *et al.* Bicaudal-D regulates COPI-independent Golgi-ER transport by recruiting the dynein-dynactin motor complex. *Nat. Cell Biol.* **4**, 986–992 (2002).
377. Burguete, A. S., Fenn, T. D., Brunger, A. T. & Pfeffer, S. R. Rab and Arl GTPase Family Members Cooperate in the Localization of the Golgin GCC185. *Cell* **132**, 286–298 (2008).
378. Hill, E., Clarke, M. & Barr, F. A. The Rab6-binding kinesin, Rab6-KIFL, is required for cytokinesis. *EMBO J.* **19**, 5711–5719 (2000).

379. Ayala, C. I., Kim, J. & Neufeld, T. P. Rab6 promotes insulin receptor and cathepsin trafficking to regulate autophagy induction and activity in Drosophila. *J. Cell Sci.* **131**, jcs216127 (2018).
380. Fujiwara, T., Oda, K., Yokota, S., Takatsuki, A. & Ikehara, Y. Brefeldin A causes disassembly of the Golgi complex and accumulation of secretory proteins in the endoplasmic reticulum. *J. Biol. Chem.* **263**, 18545–18552 (1988).
381. Stedman, T. T., Sussmann, A. R. & Joiner, K. A. Toxoplasma gondii Rab6 mediates a retrograde pathway for sorting of constitutively secreted proteins to the Golgi complex. *J. Biol. Chem.* **278**, 5433–5443 (2003).
382. McEwan, D. G. *et al.* PLEKHM1 regulates autophagosome-lysosome fusion through HOPS complex and LC3/GABARAP proteins. *Mol. Cell* **57**, 39–54 (2015).
383. Hegedűs, K. *et al.* The Ccz1-Mon1-Rab7 module and Rab5 control distinct steps of autophagy. *Mol. Biol. Cell* **27**, 3132–3142 (2016).
384. Plemel, R. L. *et al.* Subunit organization and Rab interactions of Vps-C protein complexes that control endolysosomal membrane traffic. *Mol. Biol. Cell* **22**, 1353–63 (2011).
385. Krai, P., Dalal, S. & Klemba, M. Evidence for a Golgi-to-endosome protein sorting pathway in Plasmodium falciparum. *PLoS One* **9**, e89771 (2014).
386. Ullrich, O., Reinsch, S., Urbé, S., Zerial, M. & Parton, R. G. Rab11 Regulates Recycling through the Pericentriolar Recycling Endosome. *J. Cell Biol.* **135**, 913–924 (1996).
387. Nakayama, K. Regulation of cytokinesis by membrane trafficking involving small GTPases and the ESCRT machinery. *Crit. Rev. Biochem. Mol. Biol.* **51**, 1–6 (2016).
388. Knödler, A. *et al.* Coordination of Rab8 and Rab11 in primary ciliogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6346–6351 (2010).
389. Szatmári, Z. *et al.* Rab11 facilitates cross-talk between autophagy and endosomal pathway through regulation of Hook localization. *Mol. Biol. Cell* **25**, 522–531 (2014).
390. Agop-Nersesian, C. *et al.* Biogenesis of the inner membrane complex is dependent on vesicular transport by the alveolate specific GTPase Rab11B. *PLoS Pathog.* **6**, e1001029 (2010).
391. Agop-Nersesian, C. *et al.* Rab11A-controlled assembly of the inner membrane complex is required for completion of apicomplexan cytokinesis. *PLoS Pathog.* **5**, e1000270 (2009).
392. Venugopal, K. & Marion, S. Secretory organelle trafficking in Toxoplasma gondii: A long story for a short travel. *Int. J. Med. Microbiol.* **308**, 751–760 (2018).
393. Liendo, A. *et al.* Toxoplasma gondii ADP-ribosylation Factor 1 Mediates Enhanced

Release of Constitutively Secreted Dense Granule Proteins. *J. Biol. Chem.* **276**, 18272–18281 (2001).

394. Dacks, J. B., Poon, P. P. & Field, M. C. Phylogeny of endocytic components yields insight into the process of nonendosymbiotic organelle evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 588–93 (2008).
395. Woo, Y. H. *et al.* Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *Elife* **4**, e06974 (2015).
396. Ngô, H. M. *et al.* AP-1 in Toxoplasma gondii mediates biogenesis of the rhoptry secretory organelle from a post-Golgi compartment. *J. Biol. Chem.* **278**, 5343–52 (2003).
397. Labesse, G. *et al.* ROP2 from Toxoplasma gondii: A Virulence Factor with a Protein-Kinase Fold and No Enzymatic Activity. *Structure* **17**, 139–146 (2009).
398. Sloves, P.-J. *et al.* Toxoplasma sortilin-like receptor regulates protein transport and is essential for apical secretory organelle biogenesis and host infection. *Cell Host Microbe* **11**, 515–27 (2012).
399. Kibria, K. M. K. *et al.* A role for adaptor protein complex 1 in protein targeting to rhoptry organelles in Plasmodium falciparum. *Biochim. Biophys. Acta - Mol. Cell Res.* **1853**, 699–710 (2015).
400. Fomovska, A. *et al.* Novel N-benzoyl-2-hydroxybenzamide disrupts unique parasite secretory pathway. *Antimicrob. Agents Chemother.* **56**, 2666–82 (2012).
401. Pfluger, S. L. *et al.* Receptor for Retrograde Transport in the Apicomplexan Parasite Toxoplasma gondii. *Eukaryot. Cell* **4**, 432–42 (2005).
402. Adisa, A. *et al.* Re-assessing the locations of components of the classical vesicle-mediated trafficking machinery in transfected Plasmodium falciparum. *Int. J. Parasitol.* **37**, 1127–1141 (2007).
403. Struck, N. S. *et al.* Spatial dissection of the cis- and trans-Golgi compartments in the malaria parasite Plasmodium falciparum. *Mol. Microbiol.* **67**, 1320–30 (2008).
404. Lee, M. C. S., Moura, P. A., Miller, E. A. & Fidock, D. A. Plasmodium falciparum Sec24 marks transitional ER that exports a model cargo via a diacidic motif. *Mol. Microbiol.* **68**, 1535–46 (2008).
405. Pieperhoff, M. S., Schmitt, M., Ferguson, D. J. P. & Meissner, M. The Role of Clathrin in Post-Golgi Trafficking in Toxoplasma gondii. *PLoS One* **8**, e77620 (2013).
406. kleine Balderhaar, H. J. & Ungermann, C. CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion. *J. Cell Sci.* **126**, 1307–16 (2013).
407. Morlon-Guyot, J., Pastore, S., Berry, L., Lebrun, M. & Daher, W. Toxoplasma gondii

Vps11, a subunit of HOPS and CORVET tethering complexes, is essential for the biogenesis of secretory organelles. *Cell. Microbiol.* **17**, 1157–78 (2015).

408. Morlon-Guyot, J. *et al.* A proteomic analysis unravels novel CORVET and HOPS proteins involved in Toxoplasma gondii secretory organelles biogenesis. *Cell. Microbiol.* **20**, e12870 (2018).
409. Jackson, A. J., Clucas, C., Mamczur, N. J., Ferguson, D. J. & Meissner, M. Toxoplasma gondii Syntaxin 6 is required for vesicular transport between endosomal-like compartments and the Golgi complex. *Traffic* **14**, 1166–81 (2013).
410. Agrawal, S., van Dooren, G. G., Beatty, W. L. & Striepen, B. Genetic evidence that an endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins. *J. Biol. Chem.* **284**, 33683–33691 (2009).
411. Striepen, B. *et al.* The Plastid of Toxoplasma gondii Is Divided by Association with the Centrosomes. *J. Cell Biol.* **151**, 1423–1434 (2000).
412. DeRocher, A., Gilbert, B., Feagin, J. E. & Parsons, M. Dissection of brefeldin A-sensitive and -insensitive steps in apicoplast protein targeting. *J. Cell Sci.* **118**, 565–574 (2005).
413. Tonkin, C. J., Struck, N. S., Mullin, K. A., Stimmler, L. M. & McFadden, G. I. Evidence for Golgi-independent transport from the early secretory pathway to the plastid in malaria parasites. *Mol. Microbiol.* **61**, 614–630 (2006).
414. Heiny, S. R., Pautz, S., Recker, M. & Przyborski, J. M. Protein Traffic to the Plasmodium falciparum Apicoplast: Evidence for a Sorting Branch Point at the Golgi. *Traffic* **15**, 1290–304 (2014).
415. Dooren, G. G. Van, Yeoh, L. M., Striepen, B. & Mcfadden, G. I. The Import of Proteins into the Mitochondrion of Toxoplasma. *J. Biol. Chem.* **291**, 19335–19350 (2016).
416. Brydges, S. D. & Carruthers, V. B. Mutation of an unusual mitochondrial targeting sequence of SODB2 produces multiple targeting fates in Toxoplasma gondii. *J. Cell Sci.* **116**, 4675–4685 (2003).
417. Pino, P. *et al.* Dual Targeting of Antioxidant and Metabolic Enzymes to the Mitochondrion and the Apicoplast of Toxoplasma gondii. *PLoS Pathog.* **3**, e115 (2007).
418. Chaudhari, R., Narayan, A. & Patankar, S. A novel trafficking pathway in Plasmodium falciparum for the organellar localization of glutathione peroxidase-like thioredoxin peroxidase. *FEBS J.* **279**, 3872–88 (2012).
419. Hoppe, H. C., Ngô, H. M., Yang, M. & Joiner, K. A. Targeting to rhoptry organelles of Toxoplasma gondii involves evolutionarily conserved mechanisms. *Nat. Cell Biol.* **2**, 449–56 (2000).

420. McGovern, O. L., Rivera-Cuevas, Y., Kannan, G., Narwold, A. J. & Carruthers, V. B. Intersection of endocytic and exocytic systems in Toxoplasma gondii. *Traffic* **19**, 336–353 (2018).
421. Beck, J. R. *et al.* A novel family of Toxoplasma IMC proteins displays a hierarchical organization and functions in coordinating parasite division. *PLoS Pathog.* **6**, e1001094 (2010).
422. Heaslip, A. T., Nelson, S. R. & Warshaw, D. M. Dense granule trafficking in Toxoplasma gondii requires a unique class 27 myosin and actin filaments. *Mol. Biol. Cell* **27**, 2080–2089 (2016).
423. Karsten, V. *et al.* The Protozoan Parasite Toxoplasma gondii Targets Proteins to Dense Granules and the Vacuolar Space Using Both Conserved and Unusual Mechanisms. *J. Cell Biol.* **141**, 1323–1333 (1998).
424. Chaturvedi, S. *et al.* Constitutive calcium-independent release of Toxoplasma gondii dense granules occurs through the NSF/SNAP/SNARE/Rab machinery. *J. Biol. Chem.* **274**, 2424–2431 (1999).
425. Breinich, M. S. *et al.* A dynamin is required for the biogenesis of secretory organelles in Toxoplasma gondii. *Curr. Biol.* **19**, 277–86 (2009).
426. Stewart, R. J. *et al.* Phosphorylation of αSNAP is required for secretory organelle biogenesis in Toxoplasma gondii. *Traffic* **17**, 102–16 (2015).
427. Hammoudi, P.-M. *et al.* Fundamental Roles of the Golgi-Associated Toxoplasma Aspartyl Protease, ASP5, at the Host-Parasite Interface. *PLoS Pathog.* **11**, e1005211 (2015).
428. Coffey, M. J. *et al.* Aspartyl Protease 5 Matures Dense Granule Proteins That Reside at the Host-Parasite Interface in Toxoplasma gondii. *MBio* **9**, e01796-18 (2018).
429. Francia, M. E. *et al.* A Toxoplasma gondii protein with homology to intracellular type Na+/H+ exchangers is important for osmoregulation and invasion. *Exp. Cell Res.* **317**, 1382–1396 (2011).
430. Que, X. *et al.* The cathepsin B of Toxoplasma gondii, toxopain-1, is critical for parasite invasion and rhoptry protein processing. *J. Biol. Chem.* **277**, 25791–25797 (2002).
431. Ngô, H. M., Yang, M. & Joiner, K. A. Are rhoptries in Apicomplexan parasites secretory granules or secretory lysosomal granules? *Mol. Microbiol.* **52**, 1531–41 (2004).
432. Tomavo, S. Evolutionary repurposing of endosomal systems for apical organelle biogenesis in Toxoplasma gondii. *Int. J. Parasitol.* **44**, 133–8 (2014).
433. Dou, Z., McGovern, O. L., Di Cristina, M. & Carruthers, V. B. Toxoplasma gondii ingests and digests host cytosolic proteins. *MBio* **5**, e01188-14 (2014).
434. Soldati, D., Lassen, A., Dubremetz, J.-F. & Boothroyd, J. C. Processing of Toxoplasma

ROP1 protein in nascent rhoptries. *Mol. Biochem. Parasitol.* **96**, 37–48 (1998).

435. Rabenau, K. E. *et al.* TgM2AP participates in Toxoplasma gondii invasion of host cells and is tightly associated with the adhesive protein TgMIC2. *Mol. Microbiol.* **41**, 537–547 (2001).
436. Brydges, S. D., Harper, J. M., Parussini, F., Coppens, I. & Carruthers, V. B. A transient forward-targeting element for microneme-regulated secretion in Toxoplasma gondii. *Biol. Cell* **100**, 253–64 (2008).
437. El Hajj, H. *et al.* Molecular signals in the trafficking of Toxoplasma gondii protein MIC3 to the micronemes. *Eukaryot. Cell* **7**, 1019–1028 (2008).
438. Gaji, R. Y., Flammer, H. P. & Carruthers, V. B. Forward targeting of Toxoplasma gondii proproteins to the micronemes involves conserved aliphatic amino acids. *Traffic* **12**, 840–53 (2011).
439. Mueller, C. *et al.* The Toxoplasma protein ARO mediates the apical positioning of rhoptry organelles, a prerequisite for host cell invasion. *Cell Host Microbe* **13**, 289–301 (2013).
440. Dettmer, J., Hong-Hermesdorf, A., Stierhof, Y.-D. & Schumacher, K. Vacuolar H+-ATPase Activity Is Required for Endocytic and Secretory Trafficking in Arabidopsis. *Plant Cell* **18**, 715–730 (2006).
441. Viotti, C. *et al.* Endocytic and secretory traffic in Arabidopsis merge in the trans-Golgi network/early endosome, an independent and highly dynamic organelle. *Plant Cell* **22**, 1344–1357 (2010).
442. Day, K. J., Casler, J. C. & Glick, B. S. Budding Yeast Has a Minimal Endomembrane System. *Dev. Cell* **44**, 56-72.e4 (2018).
443. Huotari, J. & Helenius, A. Endosome maturation. *EMBO J.* **30**, 3481–500 (2011).
444. van der Sluijs, P. *et al.* The small GTP-binding protein rab4 controls an early sorting event on the endocytic pathway. *Cell* **70**, 729–740 (1992).
445. Lindsay, M. R., Webb, R. I. & Fuerst, J. A. Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus Pirellula. *Microbiology* **143**, 739–48 (1997).
446. Klinger, C. M., Spang, A., Dacks, J. B. & Ettema, T. J. G. Tracing the archaeal origins of eukaryotic membrane-trafficking system building blocks. *Mol. Biol. Evol.* **33**, 1528–1541 (2016).
447. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
448. Schlacht, A., Herman, E. K., Klute, M. J., Field, M. C. & Dacks, J. B. Missing Pieces of an Ancient Puzzle: Evolution of the Eukaryotic Membrane-Trafficking System. *Cold*

Spring Harb. Perspect. Biol. **6**, a016048 (2014).
449. Ohno, S. *Evolution by Gene Duplication*. *Springer-Verlag* (1970).
450. Lynch, M. & Force, A. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473 (2000).
451. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science.* **290**, 1151–1155 (2000).
452. He, X. & Zhang, J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157–1164 (2005).
453. Dacks, J. B. & Field, M. C. Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J. Cell Sci.* **120**, 2977–85 (2007).
454. Ramadas, R. & Thattai, M. New organelles by gene duplication in a biophysical model of eukaryote endomembrane evolution. *Biophys. J.* **104**, 2553–2563 (2013).
455. Brugerolle, G. Colpodella vorax: ultrastructure, predation, life-cycle, mitosis, and phylogenetic relationships. *Eur. J. Protistol.* **38**, 113–125 (2002).
456. Gubbels, M.-J. & Duraisingh, M. T. Evolution of apicomplexan secretory organelles. *Int. J. Parasitol.* **42**, 1071–81 (2012).
457. Zhu, G., Marchewka, M. J. & Keithly, J. S. *Cryptosporidium parvum* appears to lack a plastid genome. *Microbiology* **146**, 315–321 (2000).
458. Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, Cryptosporidium parvum. *Science* **304**, 441–5 (2004).
459. Leander, B. S., Kuvardina, O. N., Aleshin, V. V, Mylnikov, A. P. & Keeling, P. J. Molecular Phylogeny and Surface Morphology of Colpodella edax (Alveolata): Insights into the Phagotrophic Ancestry of Apicomplexans. *J. Eukaryot. Microbiol.* **50**, 334–340 (2003).
460. Klinger, C. M., Karnkowska, A., Herman, E. K., Hampl, V. & Dacks, J. B. Phylogeny and evolution. in *Molecular Parasitology: Protozoan Parasites and their Molecules* (2016). doi:10.1007/978-3-7091-1416-2_12
461. Shoguchi, E. *et al.* Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Curr. Biol.* **23**, 1399–408 (2013).
462. Eisen, J. A. *et al.* Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. *PLoS Biol.* **4**, e286 (2006).
463. Aury, J.-M. *et al.* Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. *Nature* **444**, 171–8 (2006).
464. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).

465. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
466. Field, H. I., Coulson, R. M. R. & Field, M. C. An automated graphics tool for comparative genomics: the Coulson plot generator. *BMC Bioinformatics* **14**, 141 (2013).
467. Baum, J., Gilberger, T.-W., Frischknecht, F. & Meissner, M. Host-cell invasion by malaria parasites: insights from Plasmodium and Toxoplasma. *Trends Parasitol.* **24**, 557–63 (2008).
468. Nevin, W. D. & Dacks, J. B. Repeated secondary loss of adaptin complex genes in the Apicomplexa. *Parasitol. Int.* **58**, 86–94 (2009).
469. Hirst, J., Itzhak, D. N., Antrobus, R., Borner, G. H. H. & Robinson, M. S. Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. *PLoS Biol.* **16**, e2004411 (2018).
470. Liu, J. & Guo, W. The exocyst complex in exocytosis and cell migration. *Protoplasma* **249**, 587–97 (2012).
471. Oborník, M. *et al.* Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of apicomplexa, Chromera velia. *Protist* **162**, 115–30 (2011).
472. Ren, X. & Hurley, J. H. VHS domains of ESCRT-0 cooperate in high-avidity binding to polyubiquitinated cargo. *EMBO J.* **29**, 1045–1054 (2010).
473. Blanc, C. *et al.* Dictyostelium Tom1 participates to an ancestral ESCRT-0 complex. *Traffic* **10**, 161–171 (2009).
474. Shields, S. B. *et al.* ESCRT ubiquitin-binding domains function cooperatively during MVB cargo sorting. *J. Cell Biol.* **185**, 213–224 (2009).
475. Slagsvold, T. *et al.* Eap45 in mammalian ESCRT-II binds ubiquitin via a phosphoinositide- interacting GLUE domain. *J. Biol. Chem.* **280**, 19600–19606 (2005).
476. Katzmann, D. J., Stefan, C. J., Babst, M. & Emr, S. D. Vps27 recruits ESCRT machinery to endosomes during MVB sorting. *J. Cell Biol.* **162**, 413–423 (2003).
477. Kostelansky, M. S. *et al.* Structural and Functional Organization of the ESCRT-I Trafficking Complex. *Cell* **125**, 113–126 (2006).
478. Gill, D. J. *et al.* Structural insight into the ESCRT-I/-II link and its role in MVB trafficking. *EMBO J.* **26**, 600–612 (2007).
479. Teo, H., Perisic, O., González, B. & Williams, R. L. ESCRT-II, an Endosome-Associated Complex Required for Protein Sorting: Crystal Structure and Interactions with ESCRT-III and Membranes. *Dev. Cell* **7**, 559–569 (2004).
480. Teo, H. *et al.* ESCRT-I Core and ESCRT-II GLUE Domain Structures Reveal Role for GLUE in Linking to ESCRT-I and Membranes. *Cell* **125**, 99–111 (2006).

481. Hanson, P. I., Roth, R., Lin, Y. & Heuser, J. E. Plasma membrane deformation by circular arrays of ESCRT-III protein filaments. *J. Cell Biol.* **180**, 389–402 (2008).
482. Saksena, S., Wahlman, J., Teis, D., Johnson, A. E. & Emr, S. D. Functional Reconstitution of ESCRT-III Assembly and Disassembly. *Cell* **136**, 97–109 (2009).
483. Wollert, T., Wunder, C., Lippincott-Schwartz, J. & Hurley, J. H. Membrane scission by the ESCRT-III complex. *Nature* **458**, 172–177 (2009).
484. Wollert, T. & Hurley, J. H. Molecular mechanism of multivesicular body biogenesis by ESCRT complexes. *Nature* **464**, 864–869 (2010).
485. Yang, M. *et al.* The Plasmodium falciparum Vps4 homolog mediates multivesicular body formation. *J. Cell Sci.* **117**, 3831–8 (2004).
486. Zhou, F. *et al.* Rab5-dependent autophagosome closure by ESCRT. *J. Cell Biol.* **218**, 1908–1927 (2019).
487. Karasmanis, E. P. *et al.* A Septin Double Ring Controls the Spatiotemporal Organization of the ESCRT Machinery in Cytokinetic Abscission. *Curr. Biol.* **29**, 2174-2182.e7 (2019).
488. Schlüter, A. *et al.* The evolutionary origin of peroxisomes: an ER-peroxisome connection. *Mol. Biol. Evol.* **23**, 838–45 (2006).
489. Gabaldón, T. Peroxisome diversity and evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365**, 765–73 (2010).
490. Kumar, H. *et al.* The Exocyst Complex and Rab5 are required for Abscission by Localizing ESCRT III Subunits to the Cytokinetic Bridge. *J. Cell Sci.* jcs226001 (2019).
491. Singh, S. *et al.* Exocyst Subcomplex Functions in Autophagosome Biogenesis by Regulating Atg9 Trafficking. *J. Mol. Biol.* **431**, 2821–2834 (2019).
492. Morgera, F. *et al.* Regulation of exocytosis by the exocyst subunit Sec6 and the SM protein Sec1. *Mol. Biol. Cell* **23**, 337–46 (2012).
493. Carruthers, V. B. & Sibley, L. D. Sequential protein secretion from three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts. *Eur. J. Cell Biol.* **73**, 114–123 (1997).
494. Brown, K. M. & Sibley, L. D. Essential cGMP Signaling in Toxoplasma Is Initiated by a Hybrid P-Type ATPase-Guanylate Cyclase. *Cell Host Microbe* **24**, 804-816.e6 (2018).
495. Lourido, S. *et al.* Calcium-dependent protein kinase 1 is an essential regulator of exocytosis in Toxoplasma. *Nature* **465**, 359–62 (2010).
496. Farrell, A. *et al.* A DOC2 protein identified by mutational profiling is essential for apicomplexan parasite exocytosis. *Science.* **335**, 218–21 (2012).
497. Bullen, H. E. *et al.* Phosphatidic Acid-Mediated Signaling Regulates Microneme Secretion in Toxoplasma. *Cell Host Microbe* **19**, 349–360 (2016).

498. Coleman, B. I. *et al.* A Member of the Ferlin Calcium Sensor Family Is Essential for Toxoplasma gondii Rhoptry Secretion. *MBio* **9**, e01510-18 (2018).
499. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
500. Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–60 (2005).
501. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
502. Choi, Y. H. & Kim, J. K. Dissecting Cellular Heterogeneity Using Single-Cell RNA Sequencing. *Mol. Cells* **42**, 189–199 (2019).
503. Koonin, E. V. Orthologs, Paralogs, and Evolutionary Genomics. *Annu. Rev. Genet.* **39**, 309–338 (2005).
504. Klinger, C. M. *et al.* Resolving the homology-function relationship through comparative genomics of membrane-trafficking machinery and parasite cell biology. *Mol. Biochem. Parasitol.* **209**, 88–103 (2016).
505. Koumandou, V. L. *et al.* Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.* **48**, 373–96 (2013).
506. Fitch, W. M. Distinguishing homologous from analogous proteins. *Syst. Zool.* **19**, 99–113 (1970).
507. Barlowe, C. K. & Miller, E. A. Secretory protein biogenesis and traffic in the early secretory pathway. *Genetics* **193**, 383–410 (2013).
508. Cai, H., Reinisch, K. & Ferro-Novick, S. Coats, tethers, Rabs, and SNAREs work together to mediate the intracellular destination of a transport vesicle. *Dev. Cell* **12**, 671–82 (2007).
509. Perez-Linero, A. M. & Muñiz, M. Membrane trafficking: Returning to the fold(ER). *Curr. Biol.* **25**, R288–R290 (2015).
510. Robinson, M. S. Forty Years of Clathrin-coated Vesicles. *Traffic* **16**, 1210–38 (2015).
511. Manna, P. T., Boehm, C., Leung, K. F., Natesan, S. K. & Field, M. C. Life and times: synthesis, trafficking, and evolution of VSG. *Trends Parasitol.* **30**, 251–8 (2014).
512. Field, M. C., Sali, A. & Rout, M. P. Evolution: On a bender-BARs, ESCRTs, COPs, and finally getting your coat. *J. Cell Biol.* **193**, 963–972 (2011).
513. Elias, M., Patron, N. J. & Keeling, P. J. The RAB family GTPase Rab1A from Plasmodium falciparum defines a unique paralog shared by chromalveolates and rhizaria. *J. Eukaryot. Microbiol.* **56**, 348–56 (2009).
514. Scheuring, D. *et al.* Multivesicular Bodies Mature from the Trans-Golgi Network/Early Endosome in Arabidopsis. *Plant Cell* **23**, 3463–81 (2011).
515. Briguglio, J. S. & Turkewitz, A. P. Tetrahymena thermophila: A divergent perspective

on membrane traffic. *J. Exp. Zool. B. Mol. Dev. Evol.* **322**, 500–16 (2014).

516. Faso, C. & Hehl, A. B. Membrane trafficking and organelle biogenesis in Giardia lamblia: use it or lose it. *Int. J. Parasitol.* **41**, 471–80 (2011).
517. Jackson, A. P. *et al.* Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Curr. Biol.* **26**, 161–172 (2016).
518. Field, M. C. & Carrington, M. The trypanosome flagellar pocket. *Nat. Rev. Microbiol.* **7**, 775–86 (2009).
519. Teixeira, J. E. & Huston, C. D. Evidence of a continuous endoplasmic reticulum in the protozoan parasite Entamoeba histolytica. *Eukaryot. Cell* **7**, 1222–1226 (2008).
520. Ralston, K. S. Taking a bite: Amoebic trogocytosis in Entamoeba histolytica and beyond. *Curr Opin Microbiol* **28**, 26–35 (2015).
521. Saito-Nakano, Y., Yasuda, T., Nakada-Tsukui, K., Leippe, M. & Nozaki, T. Rab5-associated vacuoles play a unique role in phagocytosis of the enteric protozoan parasite Entamoeba histolytica. *J. Biol. Chem.* **279**, 49497–49507 (2004).
522. Tarnita, C. E., Washburne, A., Martinez-Garcia, R., Sgro, A. E. & Levin, S. A. Fitness tradeoffs between spores and nonaggregating cells can explain the coexistence of diverse genotypes in cellular slime molds. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2776–81 (2015).
523. Neuhaus, E. M., Almers, W. & Soldati, T. Morphology and Dynamics of the Endocytic Pathway in Dictyostelium discoideum. *Mol. Biol. Cell* **13**, 1390–1407 (2002).
524. Müller-Taubenberger, A., Kortholt, A. & Eichinger, L. Simple system - substantial share: The use of Dictyostelium in cell biology and molecular medicine. *Eur. J. Cell Biol.* **92**, 45–53 (2013).
525. Robinson, M. S. Adaptable adaptors for coated vesicles. *Trends Cell Biol.* **14**, 167–174 (2004).
526. Bonifacino, J. S. Adaptor proteins involved in polarized sorting. *J. Cell Biol.* **204**, 7–17 (2014).
527. Hirst, J. *et al.* Distinct and overlapping roles for AP-1 and GGAs revealed by the 'knocksideways' system. *Curr. Biol.* **22**, 1711–1716 (2012).
528. Park, M. *et al.* Arabidopsis μ-adaptin subunit AP1M of adaptor protein complex 1 mediates late secretory and vacuolar traffic and is required for growth. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10318–23 (2013).
529. Teh, O. K. *et al.* The AP-1 μ adaptin is required for KNOLLE localization at the cell plate to mediate cytokinesis in Arabidopsis. *Plant Cell Physiol.* **54**, 838–847 (2013).
530. Wang, J.-G. *et al.* HAPLESS13, the Arabidopsis μ1 adaptin, is essential for protein

sorting at the trans-Golgi network/early endosome. *Plant Physiol.* **162**, 1897–910 (2013).

531. Song, J., Lee, M. H., Lee, G.-J., Yoo, C. M. & Hwang, I. Arabidopsis EPSIN1 Plays an Important Role in Vacuolar Trafficking of Soluble Cargo Proteins in Plant Cells via Interactions with Clathrin, AP-1, VTI11, and VSR1. *Plant Cell* **18**, 2258–2274 (2006).
532. Nishimura, K. *et al.* The tyrosine-sorting motif of the vacuolar sorting receptor VSR4 from Arabidopsis thaliana, which is involved in the interaction between VSR4 and AP1M2, μ1-adaptin type 2 of clathrin adaptor complex 1 subunits, participates in the post-Golgi sorting of VS. *Biosci. Biotechnol. Biochem.* **80**, 694–705 (2016).
533. Elde, N. C., Morgan, G., Winey, M., Sperling, L. & Turkewitz, A. P. Elucidation of clathrin-mediated endocytosis in tetrahymena reveals an evolutionarily convergent recruitment of dynamin. *PLoS Genet.* **1**, e52 (2005).
534. Tomavo, S., Slomianny, C., Meissner, M. & Carruthers, V. B. Protein trafficking through the endosomal system prepares intracellular parasites for a home invasion. *PLoS Pathog.* **9**, e1003629 (2013).
535. Touz, M. C., Kulakova, L. & Nash, T. E. Adaptor protein complex 1 mediates the transport of lysosomal proteins from a Golgi-like organelle to peripheral vacuoles in the primitive eukaryote Giardia lamblia. *Mol. Biol. Cell* **15**, 3053–3060 (2004).
536. Rivero, M. R. *et al.* Vacuolar protein sorting receptor in Giardia lamblia. *PLoS One* **7**, e43712 (2012).
537. Zoltner, M., Leung, K. F., Alsford, S., Horn, D. & Field, M. C. Modulation of the Surface Proteome through Multiple Ubiquitylation Pathways in African Trypanosomes. *PLoS Pathog.* **11**, e1005236 (2015).
538. Tazeh, N. N. *et al.* Role of AP-1 in developmentally regulated lysosomal trafficking in Trypanosoma brucei. *Eukaryot. Cell* **8**, 1352–1361 (2009).
539. Marion, S., Laurent, C. & Guillén, N. Signalization and cytoskeleton activity through myosin IB during the early steps of phagocytosis in Entamoeba histolytica: A proteomic approach. *Cell. Microbiol.* **7**, 1504–1518 (2005).
540. Lefkir, Y. *et al.* The AP-1 Clathrin-adaptor Is Required for Lysosomal Enzymes Sorting and Biogenesis of the Contractile Vacuole Complex in Dictyostelium Cells. *Mol. Biol. Cell* **14**, 1835–1851 (2003).
541. Lefkir, Y. *et al.* Involvement of the AP-1 Adaptor Complex in Early Steps of Phagocytosis and Macropinocytosis. *Mol. Biol. Cell* **15**, 861–869 (2004).
542. Mercanti, V., Blanc, C., Lefkir, Y., Cosson, P. & Letourneur, F. Acidic clusters target transmembrane proteins to the contractile vacuole in Dictyostelium cells. *J. Cell Sci.*

119, 837–845 (2006).

543. Reider, A. & Wendland, B. Endocytic adaptors - social networking at the plasma membrane. *J. Cell Sci.* **124**, 1613–22 (2011).
544. Bashline, L., Li, S., Anderson, C. T., Lei, L. & Gu, Y. The endocytosis of cellulose synthase in Arabidopsis is dependent on μ2, a clathrin-mediated endocytosis adaptin. *Plant Physiol.* **163**, 150–160 (2013).
545. Di Rubbo, S. *et al.* The Clathrin Adaptor Complex AP-2 Mediates Endocytosis of BRASSINOSTEROID INSENSITIVE1 in Arabidopsis. *Plant Cell* **25**, 2986–2997 (2013).
546. Fan, L. *et al.* Dynamic analysis of Arabidopsis AP2 σ subunit reveals a key role in clathrin-mediated endocytosis and plant development. *Development* **140**, 3826–37 (2013).
547. Kim, S. Y. *et al.* Adaptor protein complex 2-mediated endocytosis is crucial for male reproductive organ development in Arabidopsis. *Plant Cell* **25**, 2970–85 (2013).
548. Yamaoka, S. *et al.* Identification and Dynamics of Arabidopsis Adaptor Protein-2 Complex and Its Involvement in Floral Organ Development. *Plant Cell* **25**, 2958–2969 (2013).
549. Wang, C. *et al.* Differential Regulation of Clathrin and Its Adaptor Proteins during Membrane Recruitment for Endocytosis. *Plant Physiol.* **171**, 215–229 (2016).
550. Barth, M. & Holstein, S. E. H. Identification and functional characterization of Arabidopsis AP180, a binding partner of plant alphaC-adaptin. *J. Cell Sci.* **117**, 2051–2062 (2004).
551. Gadeyne, A. *et al.* The TPLATE adaptor complex drives clathrin-mediated endocytosis in plants. *Cell* **156**, 691–704 (2014).
552. Macro, L., Jaiswal, J. K. & Simon, S. M. Dynamics of clathrin-mediated endocytosis and its requirement for organelle biogenesis in Dictyostelium. *J. Cell Sci.* **125**, 5721–5732 (2012).
553. Sosa, R. T., Weber, M. M., Wen, Y. & O'Halloran, T. J. A Single β Adaptin Contributes to AP1 and AP2 Complexes and Clathrin Function in Dictyostelium. *Traffic* **13**, 305–316 (2012).
554. Bennett, N., Letourneur, F., Ragno, M. & Louwagie, M. Sorting of the v-SNARE VAMP7 in Dictyostelium discoideum: a role for more than one Adaptor Protein (AP) complex. *Exp. Cell Res.* **314**, 2822–33 (2008).
555. Gotthardt, D. *et al.* High-Resolution Dissection of Phagosome Maturation Reveals Distinct Membrane Trafficking Phases. *Mol. Biol. Cell* **13**, 3508–3520 (2002).
556. Wen, Y. *et al.* AP180-Mediated Trafficking of Vamp7B Limits Homotypic Fusion of

Dictyostelium Contractile Vacuoles. *Mol. Biol. Cell* **20**, 4278–4288 (2009).

557. Charette, S. J., Mercanti, V., Letourneur, F., Bennett, N. & Cosson, P. A role for adaptor protein-3 complex in the organization of the endocytic pathway in Dictyostelium. *Traffic* **7**, 1528–1538 (2006).
558. Rivero, M. R. *et al.* Adaptor protein 2 regulates receptor-mediated endocytosis and cyst formation in Giardia lamblia. *Biochem. J.* **428**, 33–45 (2010).
559. Manna, P. T., Kelly, S. & Field, M. C. Adaptin evolution in kinetoplastids and emergence of the variant surface glycoprotein coat in African trypanosomatids. *Mol. Phylogenet. Evol.* **67**, 123–8 (2013).
560. Braun, V. *et al.* AP-1 and ARF1 Control Endosomal Dynamics at Sites of FcR-mediated Phagocytosis. *Mol. Biol. Cell* **18**, 4921–4931 (2007).
561. Henne, W. M., Buchkovich, N. J. & Emr, S. D. The ESCRT pathway. *Dev. Cell* **21**, 77–91 (2011).
562. Manil-Segalén, M., Lefebvre, C., Culetto, E. & Legouis, R. Need an ESCRT for autophagosomal maturation? *Commun. Integr. Biol.* **5**, 566–571 (2012).
563. Herman, E. K., Walker, G., van der Giezen, M. & Dacks, J. B. Multivesicular bodies in the enigmatic amoeboflagellate Breviata anathema and the evolution of ESCRT 0. *J. Cell Sci.* **124**, 613–21 (2011).
564. Spitzer, C. *et al.* The Arabidopsis elch mutant reveals functions of an ESCRT component in cytokinesis. *Development* **133**, 4679–4689 (2006).
565. Winter, V. & Hauser, M. T. Exploring the ESCRTing machinery in eukaryotes. *Trends Plant Sci.* **11**, 115–123 (2006).
566. Cai, Y., Zhuang, X., Gao, C., Wang, X. & Jiang, L. The Arabidopsis Endosomal Sorting Complex Required for Transport III Regulates Internal Vesicle Formation of the Prevacuolar Compartment and Is Required for Plant Development. *Plant Physiol.* **165**, 1328–1343 (2014).
567. Katsiarimpa, A. *et al.* The deubiquitinating enzyme AMSH1 and the ESCRT-III subunit VPS2.1 are required for autophagic degradation in Arabidopsis. *Plant Cell* **25**, 2236–52 (2013).
568. Haas, T. J. *et al.* The Arabidopsis AAA ATPase SKD1 is involved in multivesicular endosome function and interacts with its positive regulator LYST-INTERACTING PROTEIN5. *Plant Cell* **19**, 1295–1312 (2007).
569. Shahriari, M. *et al.* The AAA-type ATPase AtSKD1 contributes to vacuolar maintenance of Arabidopsis thaliana. *Plant J.* **64**, 71–85 (2010).
570. Spitzer, C. *et al.* The ESCRT-related CHMP1A and B proteins mediate multivesicular

body sorting of auxin carriers in Arabidopsis and are required for plant development. *Plant Cell* **21**, 749–766 (2009).

571. Spitzer, C. *et al.* The Endosomal Protein CHARGED MULTIVESICULAR BODY PROTEIN1 Regulates the Autophagic Turnover of Plastids in Arabidopsis. *Plant Cell* **27**, 391–402 (2015).
572. Cardona-López, X. *et al.* ESCRT-III-Associated Protein ALIX Mediates High-Affinity Phosphate Transporter Trafficking to Maintain Phosphate Homeostasis in Arabidopsis. *Plant Cell* **27**, 2560–81 (2015).
573. Isono, E. *et al.* The deubiquitinating enzyme AMSH3 is required for intracellular trafficking and vacuole biogenesis in Arabidopsis thaliana. *Plant Cell* **22**, 1826–37 (2010).
574. Gao, C. *et al.* A Unique plant ESCRT component, FREE1, regulates multivesicular body protein sorting and plant growth. *Curr. Biol.* **24**, 2556–2563 (2014).
575. Gao, C. *et al.* Dual roles of an Arabidopsis ESCRT component FREE1 in regulating vacuolar protein transport and autophagic degradation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1886–1891 (2015).
576. Kolb, C. *et al.* FYVE1 is essential for vacuole biogenesis and intracellular trafficking in Arabidopsis. *Plant Physiol.* **167**, 1361–1373 (2015).
577. Reyes, F. C. *et al.* A novel endosomal sorting complex required for transport (ESCRT) component in Arabidopsis thaliana controls cell expansion and development. *J. Biol. Chem.* **289**, 4980–4988 (2014).
578. Wampfler, P. B., Tosevski, V., Nanni, P., Spycher, C. & Hehl, A. B. Proteomics of secretory and endocytic organelles in Giardia lamblia. *PLoS One* **9**, e94089 (2014).
579. Dutta, S., Saha, N., Ray, A. & Sarkar, S. Significantly Diverged Did2 / Vps46 Orthologues from the Protozoan Parasite Giardia lamblia. *Curr. Microbiol.* **71**, 333–340 (2015).
580. Chung, W.-L., Leung, K. F., Carrington, M. & Field, M. C. Ubiquitylation is required for degradation of transmembrane surface proteins in trypanosomes. *Traffic* **9**, 1681–97 (2008).
581. Silverman, J. S., Muratore, K. A. & Bangs, J. D. Characterization of the late endosomal ESCRT machinery in Trypanosoma brucei. *Traffic* **14**, 1078–90 (2013).
582. López-Reyes, I. *et al.* Detection of the endosomal sorting complex required for transport in Entamoeba histolytica and characterization of the EhVps4 protein. *J. Biomed. Biotechnol.* **2010**, 890674 (2010).
583. Bañuelos, C. *et al.* EhADH112 Is a Bro1 Domain-Containing Protein Involved in the

Entamoeba histolytica Multivesicular Bodies Pathway. *J. Biomed. Biotechnol.* **2012**, 657942 (2012).

584. Mattei, S., Klein, G., Satre, M. & Aubry, L. Trafficking and developmental signaling: Alix at the crossroads. *Eur. J. Cell Biol.* **85**, 925–936 (2006).
585. Seaman, M. N. J. The retromer complex - endosomal protein recycling and beyond. *J. Cell Sci.* **125**, 4693–702 (2012).
586. Oliviusson, P. *et al.* Plant retromer, localized to the prevacuolar compartment and microvesicles in Arabidopsis, may interact with vacuolar sorting receptors. *Plant Cell* **18**, 1239–1252 (2006).
587. Kleine-Vehn, J. *et al.* Differential degradation of PIN2 auxin efflux carrier by retromer-dependent vacuolar targeting. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17812–17817 (2008).
588. Yamazaki, M. *et al.* Arabidopsis VPS35, a retromer component, is required for vacuolar protein sorting and involved in plant growth and leaf senescence. *Plant Cell Physiol.* **49**, 142–156 (2008).
589. Zelazny, E. *et al.* Mechanisms governing the endosomal membrane recruitment of the core retromer in Arabidopsis. *J. Biol. Chem.* **288**, 8815–8825 (2013).
590. Munch, D. *et al.* Retromer Contributes to Immunity-Associated Cell Death in Arabidopsis. *Plant Cell* **27**, 463–79 (2015).
591. Jaillais, Y. *et al.* The Retromer Protein VPS29 Links Cell Polarity and Organ Initiation in Plants. *Cell* **130**, 1057–1070 (2007).
592. Pourcher, M. *et al.* Analyses of sorting nexins reveal distinct retromer-subcomplex functions in development and protein sorting in Arabidopsis thaliana. *Plant Cell* **22**, 3980–3991 (2010).
593. Niemes, S. *et al.* Retromer recycles vacuolar sorting receptors from the trans-Golgi network. *Plant J.* **61**, 107–121 (2009).
594. Zelazny, E., Santambrogio, M. & Gaude, T. Retromer association with membranes: plants have their own rules! *Plant Signal. Behav.* **8**, e25312 (2013).
595. Shimada, T. *et al.* AtVPS29, a putative component of a retromer complex, is required for the efficient sorting of seed storage proteins. *Plant Cell Physiol.* **47**, 1187–1194 (2006).
596. Hashiguchi, Y. *et al.* Loss-of-function mutations of retromer large subunit genes suppress the phenotype of an Arabidopsis zig mutant that lacks Qb-SNARE VTI11. *Plant Cell* **22**, 159–72 (2010).
597. Nodzynski, T. *et al.* Retromer subunits VPS35A and VPS29 mediate prevacuolar

compartment (PVC) function in Arabidopsis. *Mol. Plant* **6**, 1849–1862 (2013).

598. Fuji, K. *et al.* Arabidopsis vacuolar sorting mutants (green fluorescent seed) can be identified efficiently by secretion of vacuole-targeted green fluorescent protein in their seeds. *Plant Cell* **19**, 597–609 (2007).
599. Briguglio, J. S., Kumar, S. & Turkewitz, A. P. Lysosomal sorting receptors are essential for secretory granule biogenesis in Tetrahymena. *J. Cell Biol.* **203**, 537–50 (2013).
600. Kumar, S., Briguglio, J. S. & Turkewitz, A. P. An aspartyl cathepsin, CTH3, is essential for proprotein processing during secretory granule maturation in Tetrahymena thermophila. *Mol. Biol. Cell* **25**, 2444–60 (2014).
601. Touz, M. C., Rivero, M. R., Miras, S. L. & Bonifacino, J. S. Lysosomal protein trafficking in Giardia lamblia: common and distinct features. *Front. Biosci.* **4**, 1898–1909 (2012).
602. Miras, S. L., Merino, M. C., Gottig, N., Rópolo, A. S. & Touz, M. C. The giardial VPS35 retromer subunit is necessary for multimeric complex assembly and interaction with the vacuolar protein sorting receptor. *Biochim. Biophys. Acta* **1833**, 2628–38 (2013).
603. Nakada-Tsukui, K., Saito-Nakano, Y., Ali, V. & Nozaki, T. A retromerlike complex is a novel Rab7 effector that is involved in the transport of the virulence factor cysteine protease in the enteric protozoan parasite Entamoeba histolytica. *Mol. Biol. Cell* **16**, 5294–5303 (2005).
604. Gillingham, A. K., Sinka, R., Torres, I. L., Lilley, K. S. & Munro, S. Toward a Comprehensive Map of the Effectors of Rab GTPases. *Dev. Cell* **31**, 358–373 (2014).
605. Rink, J., Ghigo, E., Kalaidzidis, Y. & Zerial, M. Rab conversion as a mechanism of progression from early to late endosomes. *Cell* **122**, 735–749 (2005).
606. Ueda, T., Yamaguchi, M., Uchimiya, H. & Nakano, A. Ara6, a plant-unique novel type Rab GTPase, functions in the endocytic pathway of Arabidopsis thaliana. *EMBO J.* **20**, 4730–4741 (2001).
607. Lee, G.-J., Sohn, E. J., Lee, M. H. & Hwang, I. The Arabidopsis rab5 homologs rha1 and ara7 localize to the prevacuolar compartment. *Plant Cell Physiol.* **45**, 1211–1220 (2004).
608. Ebine, K. *et al.* A membrane trafficking pathway regulated by the plant-specific RAB GTPase ARA6. *Nat. Cell Biol.* **13**, 853–859 (2011).
609. Beck, M., Zhou, J., Faulkner, C., MacLean, D. & Robatzek, S. Spatio-Temporal Cellular Dynamics of the Arabidopsis Flagellin Receptor Reveal Activation Status-Dependent Endosomal Sorting. *Plant Cell* **24**, 4205–19 (2012).
610. Singh, M. K. *et al.* Protein delivery to vacuole requires SAND protein-dependent Rab GTPase conversion for MVB-vacuole fusion. *Curr. Biol.* **24**, 1383–1389 (2014).

611. Ebine, K. *et al.* Plant vacuolar trafficking occurs through distinctly regulated pathways. *Curr. Biol.* **24**, 1375–1382 (2014).
612. Ueda, T., Uemura, T., Sato, M. H. & Nakano, A. Functional differentiation of endosomes in Arabidopsis cells. *Plant J.* **40**, 783–789 (2004).
613. Sohn, E. J. *et al.* Rha1, an Arabidopsis Rab5 homolog, plays a critical role in the vacuolar trafficking of soluble cargo proteins. *Plant Cell* **15**, 1057–1070 (2003).
614. Elliott, D. A. *et al.* Four distinct pathways of hemoglobin uptake in the malaria parasite Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2463–8 (2008).
615. Pal, A., Hall, B. S., Jeffries, T. R. & Field, M. C. Rab5 and Rab11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in Trypanosoma brucei. *Biochem. J.* **374**, 443–51 (2003).
616. Hall, B., Allen, C. L., Goulding, D. & Field, M. C. Both of the Rab5 subfamily small GTPases of Trypanosoma brucei are essential and required for endocytosis. *Mol. Biochem. Parasitol.* **138**, 67–77 (2004).
617. Pal, A., Hall, B. S., Nesbeth, D. N., Field, H. I. & Field, M. C. Differential endocytic functions of Trypanosoma brucei Rab5 isoforms reveal a glycosylphosphatidylinositol-specific endosomal pathway. *J. Biol. Chem.* **277**, 9529–9539 (2002).
618. Okada, M. *et al.* Kinetics and strain variation of phagosome proteins of Entamoeba histolytica by proteomic analysis. *Mol Biochem Parasitol* **145**, 171–183 (2006).
619. Okada, M. & Nozaki, T. New insights into molecular mechanisms of phagocytosis in Entamoeba histolytica by proteomic analysis. *Arch. Med. Res.* **37**, 244–252 (2006).
620. Chavrier, P., Parton, R. G., Hauri, H. P., Simons, K. & Zerial, M. Localization of Low-Molecular-Weight Gtp Binding-Proteins to Exocytic and Endocytic Compartments. *Cell* **62**, 317–329 (1990).
621. Price, A., Seals, D., Wickner, W. & Ungermann, C. The docking stage of yeast vacuole fusion requires the transfer of proteins from a cis-SNARE complex to a Rab/Ypt protein. *J. Cell Biol.* **148**, 1231–1238 (2000).
622. Cui, Y. *et al.* Activation of the Rab7 GTPase by the MON1-CCZ1 Complex Is Essential for PVC-to-Vacuole Trafficking and Plant Growth in Arabidopsis. *Plant Cell* **26**, 2080–2097 (2014).
623. Kwon, S. Il *et al.* The Rab GTPase RabG3b functions in autophagy and contributes to tracheary element differentiation in Arabidopsis. *Plant J.* **64**, 151–164 (2010).
624. Kwon, S. Il, Cho, H. J., Kim, S. R. & Park, O. K. The Rab GTPase RabG3b Positively Regulates Autophagy and Immunity-Associated Hypersensitive Cell Death in Arabidopsis. *Plant Physiol.* **161**, 1722–1736 (2013).

625. Jacobs, M. E. *et al.* The Tetrahymena thermophila phagosome proteome. *Eukaryot. Cell* **5**, 1990–2000 (2006).
626. Bright, L. J., Kambesis, N., Nelson, S. B., Jeong, B. & Turkewitz, A. P. Comprehensive analysis reveals dynamic and evolutionary plasticity of Rab GTPases and membrane traffic in Tetrahymena thermophila. *PLoS Genet.* **6**, e1001155 (2010).
627. Silverman, J. S., Schwartz, K. J., Hajduk, S. L. & Bangs, J. D. Late endosomal Rab7 regulates lysosomal trafficking of endocytic but not biosynthetic cargo in Trypanosoma brucei. *Mol. Microbiol.* **82**, 664–678 (2011).
628. Biller, L. *et al.* The cell surface proteome of Entamoeba histolytica. *Mol. Cell. Proteomics* **13**, 132–44 (2014).
629. Srinivasan, S. *et al.* Proteomic analysis of a developmentally regulated secretory vesicle. *Proteomics* **1**, 1119–1127 (2001).
630. Gotthardt, D. *et al.* Proteomics fingerprinting of phagosome maturation and evidence for the role of a Galpha during uptake. *Mol. Cell. Proteomics* **5**, 2228–2243 (2006).
631. Buczynski, G., Bush, J., Zhang, L., Rodriguez-Paris, J. & Cardelli, J. Evidence for a recycling role for Rab7 in regulating a late step in endocytosis and in retention of lysosomal enzymes in Dictyostelium discoideum. *Mol. Biol. Cell* **8**, 1343–1360 (1997).
632. Rupper, A., Grove, B. & Cardelli, J. Rab7 regulates phagosome maturation in Dictyostelium. *J. Cell Sci.* **114**, 2449–2460 (2001).
633. Rupper, A., Lee, K., Knecht, D. & Cardelli, J. Sequential activities of phosphoinositide 3-kinase, PKB/Aakt, and Rab7 during macropinosome formation in Dictyostelium. *Mol. Biol. Cell* **12**, 2813–2824 (2001).
634. Welz, T., Wellbourne-Wood, J. & Kerkhoff, E. Orchestration of cell surface proteins by Rab11. *Trends Cell Biol.* **24**, 407–415 (2014).
635. Feraru, E. *et al.* BEX5/RabA1b Regulates trans-Golgi Network-to-Plasma Membrane Protein Trafficking in Arabidopsis. *Plant Cell* **24**, 3074–3086 (2012).
636. Choi, S.-W. *et al.* RABA members act in distinct steps of subcellular trafficking of the FLAGELLIN SENSING2 receptor. *Plant Cell* **25**, 1174–87 (2013).
637. Asaoka, R. *et al.* Arabidopsis RABA1 GTPases are involved in transport between the trans-Golgi network and the plasma membrane, and are required for salinity stress tolerance. *Plant J.* **73**, 240–249 (2013).
638. Asaoka, R. *et al.* New insights into the role of Arabidopsis RABA1 GTPases in salinity stress tolerance. *Plant Signal. Behav.* **8**, e25377 (2013).
639. Chow, C.-M., Neto, H., Foucart, C. & Moore, I. Rab-A2 and Rab-A3 GTPases define a trans-golgi endosomal membrane domain in Arabidopsis that contributes substantially

to the cell plate. *Plant Cell* **20**, 101–23 (2008).

640. Lunn, D., Gaddipati, S. R., Tucker, G. A. & Lycett, G. W. Null Mutants of Individual RABA Genes Impact the Proportion of Different Cell Wall Components in Stem Tissue of Arabidopsis thaliana. *PLoS One* **8**, e75724 (2013).
641. Preuss, M. L., Serna, J., Falbel, T. G., Bednarek, S. Y. & Nielsen, E. The Arabidopsis Rab GTPase RabA4b localizes to the tips of growing root hair cells. *Plant Cell* **16**, 1589–1603 (2004).
642. Thole, J. M., Vermeer, J. E. M., Zhang, Y., Gadella Jr, T. W. J. & Nielsen, E. ROOT HAIR DEFECTIVE4 Encodes a Phosphatidylinositol-4-Phosphate Phosphatase Required for Proper Root Hair Development in Arabidopsis thaliana. *Plant Cell* **20**, 381–395 (2008).
643. Kang, B. H., Nielsen, E., Preuss, M. L., Mastronarde, D. & Staehelin, L. A. Electron Tomography of RabA4b- and PI-4Kβ1-Labeled Trans Golgi Network Compartments in Arabidopsis. *Traffic* **12**, 313–329 (2011).
644. Szumlanski, A. L. & Nielsen, E. The Rab GTPase RabA4d regulates pollen tube tip growth in Arabidopsis thaliana. *Plant Cell* **21**, 526–544 (2009).
645. Preuss, M. L. *et al.* A role for the RabA4b effector protein PI-4Kβ1 in polarized expansion of root hair cells in Arabidopsis thaliana. *J. Cell Biol.* **172**, 991–998 (2006).
646. Castillo-Romero, A. *et al.* Rab11 and actin cytoskeleton participate in Giardia lamblia encystation, guiding the specific vesicles to the cyst wall. *PLoS Negl. Trop. Dis.* **4**, e697 (2010).
647. Gabernet-Castello, C., DuBois, K. N., Nimmo, C. & Field, M. C. Rab11 function in trypanosoma brucei: Identification of conserved and novel interaction partners. *Eukaryot. Cell* **10**, 1082–1094 (2011).
648. Niyogi, S., Mucci, J., Campetella, O. & Docampo, R. Rab11 Regulates Trafficking of Trans-sialidase to the Plasma Membrane through the Contractile Vacuole Complex of Trypanosoma cruzi. *PLoS Pathog.* **10**, e1004224 (2014).
649. Temesvari, L. A., Harris, E. N., Stanley Jr., S. L. & Cardelli, J. A. Early and late endosomal compartments of Entamoeba histolytica are enriched in cysteine proteases, acid phosphatase and several Ras-related Rab GTPases. *Mol. Biochem. Parasitol.* **103**, 225–241 (1999).
650. McGugan Jr., G. C. & Temesvari, L. A. Characterization of a Rab11-like GTPase, EhRab11, of Entamoeba histolytica. *Mol. Biochem. Parasitol.* **129**, 137–146 (2003).
651. Mitra, B. N., Saito-Nakano, Y., Nakada-Tsukui, K., Sato, D. & Nozaki, T. Rab11B small GTPase regulates secretion of cysteine proteases in the enteric protozoan parasite Entamoeba histolytica. *Cell. Microbiol.* **9**, 2112–25 (2007).

652. Parkinson, K. *et al.* Calcium-dependent regulation of Rab activation and vesicle fusion by an intracellular P2X ion channel. *Nat. Cell Biol.* **16**, 87–98 (2014).
653. Harris, E., Yoshida, K., Cardelli, J. & Bush, J. Rab11-like GTPase associates with and regulates the structure and function of the contractile vacuole system in dictyostelium. *J. Cell Sci.* **114**, 3035–45 (2001).
654. Dieckmann, R. *et al.* The Balance in the Delivery of ER Components and the Vacuolar Proton Pump to the Phagosome Depends on Myosin IK in Dictyostelium. *Mol. Cell. Proteomics* **11**, 886–900 (2012).
655. Francia, M. E. & Striepen, B. Cell division in apicomplexan parasites. *Nat. Rev. Microbiol.* **12**, 125–36 (2014).
656. Harding, C. R. *et al.* Gliding Associated Proteins Play Essential Roles during the Formation of the Inner Membrane Complex of Toxoplasma gondii. *PLOS Pathog.* **12**, e1005403 (2016).
657. Giansanti, M. G. *et al.* Exocyst-Dependent Membrane Addition Is Required for Anaphase Cell Elongation and Cytokinesis in Drosophila. *PLoS Genet.* **11**, e1005632 (2015).
658. Essid, M., Gopaldass, N., Yoshida, K., Merrifield, C. & Soldati, T. Rab8a regulates the exocyst-mediated kiss-and-run discharge of the Dictyostelium contractile vacuole. *Mol. Biol. Cell* **23**, 1267–1282 (2012).
659. Komsic-Buchmann, K., Stephan, L. M. & Becker, B. The SEC6 protein is required for contractile vacuole function in Chlamydomonas reinhardtii. *J. Cell Sci.* **125**, 2885–2895 (2012).
660. Komsic-Buchmann, K., Wöstehoff, L. & Becker, B. The contractile vacuole as a key regulator of cellular water flow in Chlamydomonas reinhardtii. *Eukaryot. Cell* **13**, 1421–1430 (2014).
661. Ulrich, P. N. *et al.* Identification of contractile vacuole proteins in Trypanosoma cruzi. *PLoS One* **6**, e18013 (2011).
662. Rauch, L., Hennings, K. & Aepfelbacher, M. A role for exocyst in maturation and bactericidal function of staphylococci-containing endothelial cell phagosomes. *Traffic* **15**, 1083–1098 (2014).
663. Chen, D. *et al.* Clathrin and AP2 Are Required for Phagocytic Receptor-Mediated Apoptotic Cell Clearance in Caenorhabditis elegans. *PLoS Genet.* **9**, e1003517 (2013).
664. Chen, D. *et al.* Retromer is required for apoptotic cell clearance by phagocytic receptor recycling. *Science* **327**, 1261–1264 (2010).
665. Morrison, H. G. *et al.* Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. *Science* **317**, 1921–1926 (2007).

666. Papanikou, E., Day, K. J., Austin II, J. & Glick, B. S. COPI selectively drives maturation of the early Golgi. *Elife* **4**, e13232 (2015).
667. Kim, T., Gondré-Lewis, M. C., Arnaoutova, I. & Loh, Y. P. Dense-Core Secretory Granule Biogenesis. *Physiology* **21**, 124–133 (2006).
668. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498–511 (2002).
669. Kissinger, J. C., Gajria, B., Li, L., Paulsen, I. T. & Roos, D. S. ToxoDB: accessing the Toxoplasma gondii genome. *Nucleic Acids Res.* **31**, 234–236 (2003).
670. Adung'a, V. O., Gadelha, C. & Field, M. C. Proteomic analysis of clathrin interactions in trypanosomes reveals dynamic evolution of endocytosis. *Traffic* **14**, 440–57 (2013).
671. Lebrun, M., Carruthers, V. B. & Cesbron-Delauw, M.-F. Toxoplasma Secretory Proteins and Their Roles in Cell Invasion and Intracellular Survival. in *Toxoplasma Gondii: The Model Apicomplexan - Perspectives and Methods: Second Edition* 389–453 (Elsevier, 2014). doi:10.1016/B978-0-12-396481-6.00012-X
672. Cavalier-Smith, T. Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* **5**, 174–182 (2000).
673. Yeoman, J. A. *et al.* Tracking Glideosome-Associated Protein 50 Reveals the Development and Organization of the Inner Membrane Complex of Plasmodium falciparum. *Eukaryot. Cell* **10**, 556–564 (2011).
674. Hu, K. *et al.* Daughter Cell Assembly in the Protozoan Parasite Toxoplasma gondii. *Mol. Biol. Cell* **13**, 593–606 (2002).
675. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **40**, D290–D301 (2012).
676. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
677. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* **54**, 5.6.1-5.6.37 (2016).
678. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
679. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
680. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

681. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
682. Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E. & Stamatakis, A. How many bootstrap replicates are necessary? *J. Comput. Biol.* **17**, 337–354 (2010).
683. Ronquist, F. *et al.* Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
684. Kück, P., Mayer, C., Wägele, J.-W. & Misof, B. Long branch effects distort maximum likelihood phylogenies in simulations despite selection of the correct model. *PLoS One* **7**, e36593 (2012).
685. Lin, S. *et al.* The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science.* **350**, 691–694 (2015).
686. Mouratou, B. *et al.* The domain architecture of large guanine nucleotide exchange factors for the small GTP-binding protein Arf. *BMC Genomics* **6**, 20 (2005).
687. Hudson, A. M. & Cooley, L. Phylogenetic, structural and functional relationships between WD- and Kelch-repeat proteins. *Subcell. Biochem.* **48**, 6–19 (2008).
688. Antonin, W. *et al.* A SNARE complex mediating fusion of late endosomes defines conserved properties of SNARE structure and function. *EMBO J.* **19**, 6453–6464 (2000).
689. Mullock, B. M. *et al.* Syntaxin 7 Is Localized to Late Endosome Compartments, Associates with Vamp 8, and Is Required for Late Endosome-Lysosome Fusion. *Mol. Biol. Cell* **11**, 3137–3153 (2000).
690. Ward, D. M., Pevsner, J., Scullion, M. A., Vaughn, M. & Kaplan, J. Syntaxin 7 and VAMP-7 are Soluble N-Ethylmaleimide-sensitive Factor Attachment Protein Receptors Required for Late Endosome-Lysosome and Homotypic Lysosome Fusion in Alveolar Macrophages. *Mol. Biol. Cell* **11**, 2327–2333 (2000).
691. Pryor, P. R. *et al.* Combinatorial SNARE complexes with VAMP7 or VAMP8 define different late endocytic fusion events. *EMBO Rep.* **5**, 590–595 (2004).
692. Nozawa, T., Minowa-Nozawa, A., Aikawa, C. & Nakagawa, I. The STX6-VTI1B-VAMP3 complex facilitates xenophagy by regulating the fusion between recycling endosomes and autophagosomes. *Autophagy* **13**, 57–69 (2017).
693. Furuta, N., Fujita, N., Noda, T., Yoshimori, T. & Amano, A. Combinatorial Soluble N-Ethylmaleimide-sensitive Factor Attachment Protein Receptor Proteins VAMP8 and Vti1b Mediate Fusion of Antimicrobial and Canonical Autophagosomes with Lysosomes. *Mol. Biol. Cell* **21**, 1001–1010 (2010).
694. Chaineau, M., Danglot, L. & Galli, T. Multiple roles of the vesicular-SNARE TI-VAMP in post-Golgi and endosomal trafficking. *FEBS Lett.* **583**, 3817–26 (2009).

695. Fader, C. M., Sánchez, D. G., Mestre, M. B. & Colombo, M. I. TI-VAMP/VAMP7 and VAMP3/cellubrevin: two v-SNARE proteins involved in specific steps of the autophagy/multivesicular body pathways. *Biochim. Biophys. Acta - Mol. Cell Res.* **1793**, 1901–1916 (2009).
696. Diao, J. *et al.* ATG14 promotes membrane tethering and fusion of autophagosomes to endolysosomes. *Nature* **520**, 563–566 (2015).
697. Prekeris, R., Klumperman, J., Chen, Y. A. & Scheller, R. H. Syntaxin 13 mediates cycling of plasma membrane proteins via tubulovesicular recycling endosomes. *J. Cell Biol.* **143**, 957–971 (1998).
698. Lu, Y., Zhang, Z., Sun, D., Sweeney, S. T. & Gao, F. B. Syntaxin 13, a genetic modifier of mutant CHMP2B in frontotemporal dementia, is required for autophagosome maturation. *Mol. Cell* **52**, 264–271 (2013).
699. Mcbride, H. M. *et al.* Oligomeric Complexes Link Rab5 Effectors with NSF and Drive Membrane Fusion via Interactions between EEA1 and Syntaxin 13. *Cell* **98**, 377–386 (1999).
700. Brandhorst, D. *et al.* Homotypic fusion of early endosomes: SNAREs do not determine fusion specificity. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2701–2706 (2006).
701. Walter, A. M. *et al.* The SNARE protein vti1a functions in dense-core vesicle biogenesis. *EMBO J.* **33**, 1681–1697 (2014).
702. Wendler, F., Page, L., Urbé, S. & Tooze, S. A. Homotypic Fusion of Immature Secretory Granules During Maturation Requires Syntaxin 6. *Mol. Biol. Cell* **12**, 1699–1709 (2001).
703. Zhang, T. & Hong, W. Ykt6 Forms a SNARE Complex with Syntaxin 5, GS28, and Bet1 and Participates in a Late Stage in Endoplasmic Reticulum-Golgi Transport. *J. Biol. Chem.* **276**, 27480–27487 (2001).
704. Gordon, D. E. *et al.* VAMP3/Syb and YKT6 are required for the fusion of constitutive secretory carriers with the plasma membrane. *PLoS Genet.* **13**, e1006698 (2017).
705. Takáts, S. *et al.* Non-canonical role of the SNARE protein Ykt6 in autophagosome-lysosome fusion. *PLoS Genet.* **14**, e1007359 (2018).
706. Ungermann, C. *et al.* Three v-SNAREs and Two t-SNAREs, Present in a Pentameric cis-SNARE Complex on Isolated Vacuoles, Are Essential for Homotypic Fusion. *J. Cell Biol.* **145**, 1435–1442 (1999).
707. Kweon, Y., Rothe, A., Conibear, E. & Stevens, T. H. Ykt6p Is a Multifunctional Yeast R-SNARE That Is Required for Multiple Membrane Transport Pathways to the Vacuole. *Mol. Biol. Cell* **14**, 1868–1881 (2003).
708. Zhang, Y. *et al.* The unusual flagellar-targeting mechanism and functions of the

trypanosome ortholog of the ciliary GTPase Arl13b. *J. Cell Sci.* **131**, jcs219071 (2018).
709. Alkanderi, S. *et al.* ARL3 Mutations Cause Joubert Syndrome by Disrupting Ciliary Protein Composition. *Am. J. Hum. Genet.* **103**, 612–620 (2018).
710. Rafiullah, R. *et al.* A novel homozygous ARL13B variant in patients with Joubert syndrome impairs its guanine nucleotide-exchange factor activity. *Eur. J. Hum. Genet.* **25**, 1324–1334 (2017).
711. Hanke-Gogokhia, C. *et al.* Arf-like protein 3 (ARL3) regulates protein trafficking and ciliogenesis in mouse photoreceptors. *J. Biol. Chem.* **291**, 7142–7155 (2016).
712. Hanke-Gogokhia, C. *et al.* The guanine nucleotide exchange factor Arf-like protein 13b is essential for assembly of the mouse photoreceptor transition zone and outer segment. *J. Biol. Chem.* **292**, 21442–21456 (2017).
713. Liew, G. M. *et al.* The intraflagellar transport protein IFT27 promotes BBSome exit from cilia through the GTPase ARL6/BBS3. *Dev. Cell* **31**, 265–78 (2014).
714. Mouraõ, A., Nager, A. R., Nachury, M. V. & Lorentzen, E. Structural basis for membrane targeting of the BBSome by ARL6. *Nat. Struct. Mol. Biol.* **21**, 1035–1041 (2014).
715. Sasaki, A. *et al.* Arl8/ARL-8 functions in apoptotic cell removal by mediating phagolysosome formation in Caenorhabditis elegans. *Mol. Biol. Cell* **24**, 1584–1592 (2013).
716. Khatter, D. *et al.* The small GTPase Arl8b regulates assembly of the mammalian HOPS complex on lysosomes. *J. Cell Sci.* **128**, 1746–1761 (2015).
717. Marwaha, R. *et al.* The Rab7 effector PLE KHM1 binds Arl8b to promote cargo traffic to lysosomes. *J. Cell Biol.* **216**, 1051–1070 (2017).
718. Yang, Y. K. *et al.* ARF-like protein 16 (ARL16) inhibits RIG-I by binding with its C-terminal domain in a GTP-dependent manner. *J. Biol. Chem.* **286**, 10568–10580 (2011).
719. Monetta, P., Slavin, I., Romero, N. & Alvarez, C. Rab1b Interacts with GBF1 and Modulates both ARF1 Dynamics and COPI Association. *Mol. Biol. Cell* **18**, 2400–2410 (2007).
720. Beard, M., Satoh, A., Shorter, J. & Warren, G. A cryptic Rab1-binding site in the p115 tethering protein. *J. Biol. Chem.* **280**, 25840–25848 (2005).
721. Allan, B. B., Moyer, B. D. & Balch, W. E. Rab1 recruitment of p115 into a cis-SNARE complex: Programming budding COPII vesicles for fusion. *Science.* **289**, 444–448 (2000).
722. Moyer, B. D., Allan, B. B. & Balch, W. E. Rab1 interaction with a GM130 effector complex regulates COPII vesicle cis-Golgi tethering. *Traffic* **2**, 268–276 (2001).

723. Henrici, R. C. *et al.* Modification of an atypical clathrin-independent AP-2 adaptin complex of Plasmodium falciparum reduces susceptibility to artemisinin. *bioRxiv* 1–46 (2019). doi:https://doi.org/10.1101/621078
724. Ying, G. *et al.* The small GTPase RAB28 is required for phagocytosis of cone outer segments by the murine retinal pigmented epithelium. *J. Biol. Chem.* **293**, 17546–17558 (2018).
725. Zhou, Z. *et al.* Rab28 is a TBC1D1/TBC1D4 substrate involved in GLUT4 trafficking. *FEBS Lett.* **591**, 88–96 (2017).
726. Lumb, J. H., Leung, K. F., Dubois, K. N. & Field, M. C. Rab28 function in trypanosomes: interactions with retromer and ESCRT pathways. *J. Cell Sci.* **124**, 3771–83 (2011).
727. Jensen, V. L. *et al.* Whole-Organism Developmental Expression Profiling Identifies RAB-28 as a Novel Ciliary GTPase Associated with the BBSome and Intraflagellar Transport. *PLoS Genet.* **12**, e1006469 (2016).
728. Westlake, C. J. *et al.* Identification of Rab11 as a small GTPase binding protein for the Evi5 oncogene. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1236–1241 (2007).
729. Oguchi, M. E., Noguchi, K. & Fukuda, M. TBC1D12 is a novel Rab11-binding protein that modulates neurite outgrowth of PC12 cells. *PLoS One* **12**, e0174883 (2017).
730. Chadt, A. *et al.* Deletion of both rab-GTPase-activating proteins TBC1D1 and TBC1D4 in mice eliminates insulin- and AICAR-stimulated glucose transport. *Diabetes* **64**, 746–759 (2015).
731. Chesneau, L. *et al.* Gyp5p and Gyl1p are involved in the control of polarized exocytosis in budding yeast. *J. Cell Sci.* **117**, 4757–4767 (2004).
732. Chesneau, L. *et al.* Interdependence of the Ypt/RabGAP Gyp5p and Gyl1p for recruitment to the sites of polarized growth. *Traffic* **9**, 608–622 (2008).
733. Prigent, M. *et al.* The RabGAP proteins Gyp5p and Gyl1p recruit the BAR domain protein Rvs167p for polarized exocytosis. *Traffic* **12**, 1084–1097 (2011).
734. Longatti, A. *et al.* TBC1D14 regulates autophagosome formation via Rab11- and ULK1-positive recycling endosomes. *J. Cell Biol.* **197**, 659–675 (2012).
735. Lamb, C. A. *et al.* TBC1D14 regulates autophagy via the TRAPP complex and ATG9 traffic. *EMBO J.* **35**, 281–301 (2016).
736. Gould, S. B., Tham, W. H., Cowman, A. F., McFadden, G. I. & Waller, R. F. Alveolins, a new family of cortical proteins that define the protist infrakingdom Alveolata. *Mol. Biol. Evol.* **25**, 1219–1230 (2008).
737. Kono, M. *et al.* Evolution and architecture of the inner membrane complex in asexual and sexual stages of the malaria parasite. *Mol. Biol. Evol.* **29**, 2113–32 (2012).

738. Liew, G. M. *et al.* The intraflagellar transport protein ift27 promotes bbsome exit from cilia through the gtpase ARL6/BBS3. *Dev. Cell* **31**, 265–278 (2014).
739. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Meth* **9**, 676–682 (2012).
740. Team, R. C. R: A Language and Environment for Statistical Computing. (2015). Available at: https://www.r-project.org/. (Accessed: 9th March 2015)
741. Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069–74 (1990).
742. Stols, L. *et al.* A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif.* **25**, 8–15 (2002).
743. Harding, C. R. & Meissner, M. The inner membrane complex through development of Toxoplasma gondii and Plasmodium. *Cell. Microbiol.* **16**, 632–641 (2014).
744. Gubbels, M.-J., Vaishnava, S., Boot, N., Dubremetz, J.-F. & Striepen, B. A MORN-repeat protein is a dynamic component of the Toxoplasma gondii cell division apparatus. *J. Cell Sci.* **119**, 2236–2245 (2006).
745. Lambros, C. & Vanderberg, J. P. Synchronization of Plasmodium falciparum Erythrocytic Stages in Culture. *J. Parasitol.* **65**, 418–420 (1979).
746. Nguyen, H. M. *et al.* Toxoplasma gondii autophagy-related protein ATG9 is crucial for the survival of parasites in their host. *Cell. Microbiol.* **19**, e12712 (2017).
747. Matthiesen, S. H., Shenoy, S. M., Kim, K., Singer, R. H. & Satir, B. H. A parafusin-related Toxoplasma protein in Ca2+-regulated secretory organelles. *Eur. J. Cell Biol.* **80**, 775–783 (2001).
748. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
749. Okamoto, N., Horák, A. & Keeling, P. J. Description of two species of early branching dinoflagellates, Psammosa pacifica n. g., n. sp. and P. atlantica n. sp. *PLoS One* **7**, e34900 (2012).
750. Turkewitz, A. P. & Bright, L. J. A Rab-based view of membrane traffic in the ciliate Tetrahymena thermophila. *Small GTPases* **2**, 222–226 (2011).
751. Ter Haar, E., Musacchio, A., Harrison, S. C. & Kirchhausen, T. Atomic structure of clathrin: A β propeller terminal domain joins an α zigzag linker. *Cell* **95**, 563–573 (1998).
752. Saxena, K. *et al.* Analysis of the physical properties and molecular modeling of Sec13: A WD repeat protein involved in vesicular traffic. *Biochemistry* **35**, 15215–15221 (1996).

753. Stagg, S. M. *et al.* Structure of the Sec13/31 COPII coat cage. *Nature* **439**, 234–238 (2006).
754. Fath, S., Mancias, J. D., Bi, X. & Goldberg, J. Structure and Organization of Coat Proteins in the COPII Cage. *Cell* **129**, 1325–1336 (2007).
755. Devos, D. *et al.* Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* **2**, e380 (2004).
756. Devos, D. *et al.* Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2172–2177 (2006).
757. Renault, L. *et al.* The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller. *Nature* **392**, 97–101 (1998).
758. Wiens, C. J. *et al.* Bardet-biedl syndrome-associated small GTPase ARL6 (BBS3) functions at or near the ciliary gate and modulates Wnt signaling. *J. Biol. Chem.* **285**, 16218–16230 (2010).
759. Madugula, V. & Lu, L. A ternary complex comprising transportin1, Rab8 and the ciliary targeting signal directs proteins to ciliary membranes. *J. Cell Sci.* **129**, 3922–3934 (2016).
760. Lim, Y. S. & Tang, B. L. A role for Rab23 in the trafficking of Kif17 to the primary cilium. *J. Cell Sci.* **128**, 2996–3008 (2015).
761. Petry, F. Structural Analysis of Cryptosporidium parvum. *Microsc. Microanal.* 586–601 (2004).
762. Stoops, E. H. & Caplan, M. J. Trafficking to the Apical and Basolateral Membranes in Polarized Epithelial Cells. *J. Am. Soc. Nephrol.* **25**, 1375–1386 (2014).
763. Sugimoto, H. *et al.* Differential Recognition of Tyrosine-based Basolateral Signals by AP-1B Subunit μ1B in Polarized Epithelial Cells. *Mol. Biol. Cell* **13**, 2374–2382 (2002).
764. Jani, R. A., Purushothaman, L. K., Rani, S., Bergam, P. & Setty, S. R. G. STX13 regulates cargo delivery from recycling endosomes during melanosome biogenesis. *J. Cell Sci.* **128**, 3263–3276 (2015).
765. Bultema, J. J., Ambrosio, A. L., Burek, C. L. & Di Pietro, S. M. BLOC-2, AP-3, and AP-1 proteins function in concert with Rab38 and Rab32 proteins to mediate protein trafficking to lysosome-related organelles. *J. Biol. Chem.* **287**, 19550–63 (2012).
766. Keeling, P. J. *et al.* The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol.* **12**, e1001889 (2014).
767. Biddau, M. *et al.* Two essential Thioredoxins mediate apicoplast biogenesis, protein import, and gene expression in Toxoplasma gondii. *PLoS Pathog.* **14**, e1006836 (2018).

768. Håkansson, S., Charron, A. J. & Sibley, L. D. Toxoplasma evacuoles: A two-step process of secretion and fusion forms the parasitophorous vacuole. *EMBO J.* **20**, 3132–3144 (2001).

Appendix 1: Chapter 4 supplementary material

Supplementary Table S4.1 Genomic datasets

This table provides an overview of all predicted proteomes used in Chapter 4. Each organism is listed along with information for strain, proteome version, source, download website, date downloaded, and citation, if available.

Name	Strain	Version	Source	Website	Date	Citation
Aureococcus anophagefferens	clone 1984	1	JGI	http://genome.jgi.doe.gov/Auran1/Auran1.download.html	2017-09-08	Proc Natl Acad Sci U S A. 2011 Mar 15;108(11):4352-7.
Arabidopsis thaliana		TAIR10	JGI-Phytozome	http://genome.jgi.doe.gov/Phytozome/Phytozome.download.html	2017-09-07	Nature. 2000 Dec 14;408(6814):796-815.
Babesia bigemina	BOND	33	PiroplasmaDB	http://piroplasmadb.org/common/downloads/	2017-09-08	Nucl Acid Res. 2014 Jul 1; 42(11):7113-7131.
Babesia bovis	T2Bo	33	PiroplasmaDB	http://piroplasmadb.org/common/downloads/	2017-09-08	PLoS Pathog. 2007 Oct 19; 3(10):1401-13.
Babesia microti	RI	33	PiroplasmaDB	http://piroplasmadb.org/common/downloads/	2017-09-08	Nucl Acid Res. 2012 Oct; 40(18):9102-14.
Bigelowiella natans	CCMP2755	1	JGI	http://genome.jgi.doe.gov/Bigna1/Bigna1.download.html	2017-09-08	Nature. 2012 Dec 6;492(7427):59-65.
Cyclospora cayetanensis	CHN_HEN01	33	ToxoDB	http://toxodb.org/common/downloads/	2017-09-08	Genome Announc. 2015 Nov-Dec; 3(6):e01324-15.
Chondrus crispus		ASM35022v2	Ensembl	http://plants.ensembl.org/Chondrus_crispus/Info/Index	2017-09-07	Proc Natl Acad Sci U S A. 2013 Mar 26;110(13):5247-52.
Cytauxzoon felis	Winnie	33	PiroplasmaDB	http://piroplasmadb.org/common/downloads/	2017-09-08	Tarigo, Jaime Lynn. Doctoral Thesis. NC State, 2013-02-26
Cryptosporidium hominis	TU502	33	CryptoDB	http://cryptodb.org/common/downloads/	2017-09-08	Nature. 2004 Oct 28;431(7012):1107-12.
Cryptosporidium muris	RN66	33	CryptoDB	http://cryptodb.org/common/downloads/	2017-09-08	(CryptoDB)
Cyanophora paradoxa				http://cyanophora.rutgers.edu/cyanophora/blast.php	2017-09-07	Science. 2012 Feb 17;335(6070):843-7.
Cryptosporidium parvum	Iowall	33	CryptoDB	http://cryptodb.org/common/downloads/	2017-09-08	Science. 2004 Apr 16;304(5669):441-5.
Chlamydomonas reinhardtii	281	5.5	JGI-Phytozome	http://genome.jgi.doe.gov/Phytozome/Phytozome.download.html	2017-09-07	Trends Plant Sci. 2014 Oct;19(10):672-80
Chromera velia	CCMP2878	33	CryptoDB	http://cryptodb.org/common/downloads/	2017-09-08	Elife. 2015 Jul 15;4:e06974.
Emiliania huxleyi		1	JGI	http://genome.jgi.doe.gov/Emihu1/Emihu1.download.html	2017-09-08	Nature. 2013 Jul 11;499(7457):209-13.
Ectocarpus siliculosus				https://bioinformatics.psb.ugent.be/gdb/ectocarpus/	2017-09-08	Nature. 2010 Jun 3;465(7298):617-21.
Eimeria tenella	Houghton	33	ToxoDB	http://toxodb.org/common/downloads/	2017-09-08	(ToxoDB)
Gregarina niphandrodes		33	CryptoDB	http://cryptodb.org/common/downloads/	2017-09-08	(CryptoDB)
Guillardia theta	CCMP2712		JGI	http://genome.jgi.doe.gov/Guith1/Guith1.download.html	2017-09-08	Nature. 2012 Dec 6;492(7427):59-65.
Hammondia hammondi	HH34	33	ToxoDB	http://toxodb.org/common/downloads/	2017-09-08	Proc Natl Acad Sci U S A. 2013 Apr 30;110(18):7446-51.
Ichthyopthirius multifiliis			IchDB	http://ich.ciliate.org/index.php/home/downloads	2017-09-08	Genome Biol. 2011 Oct 17;12(10):R100.
Neospora caninum	LIV	33	ToxoDB	http://toxodb.org/common/downloads/	2017-09-08	PLoS Pathog. 2012;8(3):e1002567.
Oxytricha trifallax			OxyDB	http://oxy.ciliate.org/index.php/home/downloads	2017-09-08	PLoS Biol. 2013;11(1):e1001473.
Plasmodium berghei	ANKA	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Science. 2005 Jan 7;307(5706):82-6.
Plasmodium chabaudi	chabaudi	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Science. 2005 Jan 7;307(5706):82-6.
Plasmodium falciparum	3D7	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Nature. 2002 Oct 3;419(6906):498-511.

Plasmodium gallinaceum	8A	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Genome Res. 2018 Apr;28(4):547-560.
Plasmodium knowlesi	H	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Nature. 2008 Oct 9;455(7214):799-803.
Perkinsus marinus	ATCC50983	JCVI_PMG_1_0	Ensembl	http://protists.ensembl.org/Perkinsus_marinus_atcc_50983/Info/Index	2017-09-08	(TIGR)
Plasmodium ovale	GH01	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Nature. 2017 Feb 2;542(7639):101-104.
Plasmodium reichenowi	CDC	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Nat Commun. 2014 Sep 9;5:4754.
Plasmodium relictum	SGS1	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Genome Res. 2018 Apr;28(4):547-560.
Phythophthora sojae		3	JGI	http://genome.jgi.doe.gov/Physo3/Physo3.download.html	2017-09-08	Science. 2006 Sep 1;313(5791):1261-6.
Paramecium tetraurelia			ParameciumDB	http://paramecium.cgm.cnrs-gif.fr/download/fasta/	2017-09-08	Nature. 2006 Nov 9;444(7116):171-8.
Phaeodactylum tricornutum	CCAP1055/1	2	JGI	http://genome.jgi.doe.gov/Phatr2/Phatr2.download.html	2017-09-08	Nature. 2008 Nov 13;456(7219):239-44.
Pythium ultimum		pug	Ensembl	http://protists.ensembl.org/Pythium_ultimum/Info/Index?db=core	2017-09-08	Genome Biol. 2010;11(7):R73.
Plasmodium vivax	Sal1	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Wellcome Open Res. 2016 Nov 15;1:4.
Plasmodium yoelii	yoelii YM	33	PlasmoDB	http://plasmodb.org/common/downloads/	2017-09-08	Nature. 2002 Oct 3;419(6906):512-9.
Symbiodinium kawagutii		1		http://web.malab.cn/symka_new/download.jsp	2017-09-08	Science. 2015 Nov 6;350(6261):691-4.
Symbiodinium microadriaticum		1		http://smic.reefgenomics.org/download/	2017-09-08	Sci Reports. 2016 Dec 22;6:39734.
Symbiodinium minutum		1	OIST	http://marinegenomics.oist.jp/symb/viewer/download?project_id=21	2017-09-08	Curr Biol. 2013 Aug 5;23(15):1399-408.
Sarcocystis neurona	SN3	33	ToxoDB	http://toxodb.org/common/downloads/	2017-09-08	MBio. 2015 Feb 10;6(1).
Theileria annulata	Ankara	33	PiroplasmaDB	http://piroplasmadb.org/common/downloads/	2017-09-08	Science. 2005 Jul 1;309(5731):131-3.
Toxoplasma gondii	ME49	33	ToxoDB	http://toxodb.org/common/downloads/	2017-09-08	Nucl Acid Res. 2003 Jan 1;31(1):234-6.
Theileria orientalis	Shintoku	33	PiroplasmaDB	http://piroplasmadb.org/common/downloads/	2017-09-08	BMC Genomics. 2018; 19: 298.
Theileria parva	Muguga	33	PiroplasmaDB	http://piroplasmadb.org/common/downloads/	2017-09-08	Sciencce. 2005 Jul 1;309(5731):134-7.
Thalassiosira pseudonana	CCMP 1335	3	JGI	http://genome.jgi-psf.org/Thaps3/Thaps3.download.html	2017-09-08	Armbrust EV, *et al*, Science 2004
Tetrahymena thermophila			TGD	http://ciliate.org/index.php/home/downloads	2017-09-08	PLoS Biol. 2006 Sep;4(9):e286.
Vitrella brassicaformis	CCMP3155	33	CryptoDB	http://cryptodb.org/common/downloads/	2017-09-08	Elife. 2015 Jul 15;4:e06974.

Appendix 2: Chapter 5 supplementary material

Supplementary Table S5.1 Primers

This table provides a list of primers used in Chapter 5. Each primer is listed with a descriptive name (referred to in some cases in the main text), sequence (5’-3’), associated figure and label within that figure (if applicable), and a description of what the primer was used for.

Primer Name	Sequence	Figure	Fig. label	Description
ArlX1 LIC fwd	TACTTCCAATCCAATTTAATGCAGACGTGAGGGTCATCTCTG	Fig 5.1	NA	LIC cloning of ArlX1-3xHA endogenous tagging vector
ArlX1 LIC rev	TCCTCCACTTCCAATTTTAGCAGATTTCCCTTTTTGCTTTTTTTTC	Fig 5.1	NA	LIC cloning of ArlX1-3xHA endogenous tagging vector
ArlX2 LIC fwd	TACTTCCAATCCAATTTAATGCTCGGAAGGCAAGAACCATG	Fig 5.1	NA	LIC cloning of ArlX2-3xHA endogenous tagging vector
ArlX2 LIC rev	TCCTCCACTTCCAATTTTAGCCTGGACTTTACAGCATGCATAG	Fig 5.1	NA	LIC cloning of ArlX2-3xHA endogenous tagging vector
ArlX3 LIC fwd	TACTTCCAATCCAATTTAATGCGGCTCTGCAGGATCTTCG	Fig 5.1	NA	LIC cloning of ArlX3-3xHA endogenous tagging vector
ArlX3 LIC rev	TCCTCCACTTCCAATTTTAGCCCAGTGAAACTCCACAGAC	Fig 5.1	NA	LIC cloning of ArlX3-3xHA endogenous tagging vector
ArlX1 LIC int fwd	GCAAGTCAAAAGCAAGCGGC	Fig 5.1	P1	Integration PCR for ArlX1-3xHA
ArlX2 LIC int fwd	GTCCTGGCAGGTCCAGTAGG	Fig 5.1	P1	Integration PCR for ArlX2-3xHA
ArlX3 LIC int fwd	CGCGTCCTCCGTTTCTAAGC	Fig 5.1	P1	Integration PCR for ArlX3-3xHA
HA tag rev	GGATAGCCAGCGTAGTCCGGG	Fig 5.1	P1’	Integration PCR for endogenous HA tagging
ArlX3 gRNA fwd	AAGTTGACTCGAGGCAGACAGACTTG	Fig 5.6	NA	ArlX3 N-terminal gRNA cloning
ArlX3 gRNA rev	AAAACAAGTCTGTCTGCCTCGAGTCA	Fig 5.6	NA	ArlX3 N-terminal gRNA cloning
ArlX3 TATi fwd	cctcccgcgtgcttactcttggctttttgcgctggaatttctcgccgaag GATTTTGCCGATTTCGGCCTATTGG	Fig 5.6	NA	T7S1-myc-ArlX3 cassette PCR
ArlX3 TATi rev	actgacatgtggctgcgcgactccgagcgcatccacccagaggctgccat CAGGTCCTCCTCGGAGATGAGCTTCTGCTCCATTTTG	Fig 5.6	NA	T7S1-myc-ArlX3 cassette PCR
ArlX3 5’int fwd	TTGGCGAATCTCCAGGGTGTCCCG	Fig 5.6	P1	TATi-ArlX3 5’ integration
HX 5’UTR rev	CGCACGGCAGTCAGATAACAGGTGTA	Fig 5.6	P1’	TATi-ArlX3 5’ integration
Sag1 fwd	GCTGCACCACTTCATTATTTCTTCTGG	Fig 5.6	P2	TATi-ArlX3 3’ integration
Myc tag fwd	GGAGCAGAAGCTCATCTCCGAG	Fig 5.6	P3	TATi-ArlX3 3’ integration
ArlX3 1st exon rev	TTCGTCAAAGAGTCCGTCCG	Fig 5.6	P2’	TATi-ArlX3 3’ integration
ArlX1 1st exon gRNA fwd	AAGTTGGACCCACAAACATGATATGG	Fig 5.S3	NA	ArlX1 1st exon gRNA cloning
ArlX1 1st exon gRNA rev	AAAACCATATCATGTTTGTGGGTCCA	Fig 5.S3	NA	ArlX1 1st exon gRNA cloning
ArlX1 1st exon STOP	cttcgattcaggcgcttccttgtataaatgtgcgcctcatgcTAAgcTAAgcTAAgcTAAgcg atcatgtttgtgggtcctcctggagtgggaaagacgacac	Fig 5.S3	NA	Synthetic oligo to introduce STOP codons into ArlX1 1st exon together with gRNA
ArlX1 1st exon seq fwd	GCAAGTCAAAAGCAAGCGGC	Fig 5.S3	P1	PCR/sequencing across ArlX1 STOP insertion
ArlX1 1st exon seq rev	GAAAGGCACAAGCACAGACG	Fig 5.S3	P2	PCR/sequencing across ArlX1 STOP insertion

Supplementary Table S5.2 Vectors

This table provides an overview of all vectors used in Chapter 5, including those that were used for transient transfection and those that were used for/created by cloning. A plasmid identifier (pG number) is provided, along with a description of each plasmid, its purpose, source, and citation (if applicable). Plasmids used for cloning, as well as those created in this study have associated vector maps in the Online Appendix.

Plasmid ID	Description	Purpose	Source	Citation
pG40	GRASP-RFP	Transient transfection; marks Golgi	Kristin Hager	Nature. 2002 Aug 1;418(6897):548-52.
pG41	ptubmCherryFP-TgTubA1-CAT	Transient transfection; marks microtubules	John Murray	J Cell Biol. 2002 Mar 18;156(6):1039-50.
pG42	GalNacYFP	Transient transfection; marks TGN	Manami Nishi	(Unpublished)
pG47	TgERD-GFP	Transient transfection; marks Golgi	Kristin Hager	Eukaryot Cell. 2005 Feb;4(2):432-42.
pG145	P30-GFP-HDEL/sagCAT	Transient transfection; marks ER	Boris Striepen	J Cell Sci. 1999 Aug;112 (Pt 16):2631-8.
pG256	pmorn1cherryMORN1/sagCAT	Transient transfection	Marc-Jan Gubbels	J Cell Sci. 2006 Jun 1;119(Pt 11):2236-45.
pG474	Tub-Cas9-YFP-pU6-ccdB-tracrRNA	gRNA cloning for CRISPR/CAS9	Meissner	NA
pG514	LIC 3xHA HX	LIC cloning of endogenous tagging vectors	Meissner	Eukaryot Cell. 2009 Apr;8(4):530-9.
pG661	pT8TATi1-HX-tetO751myc	PCR amplification for ArlX3 promoter displacement	Dominique Soldati	EMBO J. 2013 Jun 12;32(12):1702-16.
pG738	LIC ArlX1 C-ter-3xHA (pG514)	Endogenous tagging of ArlX1	This work	NA
pG734	LIC ArlX2 C-ter-3xHA (pG514)	Endogenous tagging of ArlX2	This work	NA
pG736	LIC ArlX3 C-ter-3xHA (pG514)	Endogenous tagging of ArlX3	This work	NA
pG828	ArlX3 5'UTR gRNA (pG474)	CRISPR-CAS9 DSB just prior to ArlX3 START codon	This work	NA
pG838	ArlX1 1st exon gRNA in (pG474)	CRISPR-CAS9 DSB within ArlX1 1st exon	This work	NA

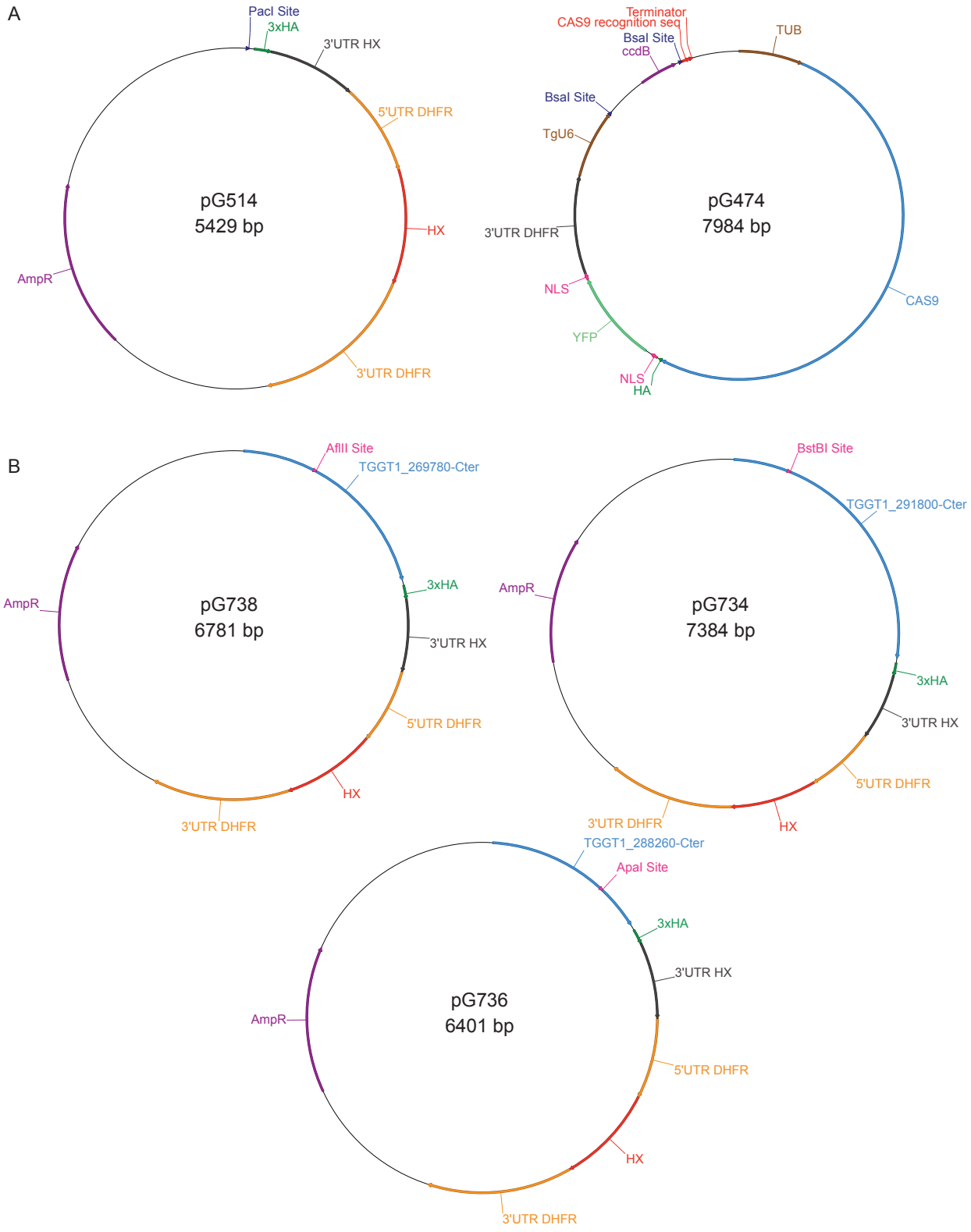
Supplementary Table S5.3 Antibodies

This table provides a list of all antibodies used in Chapter 5. The name (target) of each antibody is listed, along with its identity as either primary (P) or secondary (S), dilution used for both IFA and western blotting (WB), and its source. Hyphens indicate that the antibody was not used for the purpose indicated (either IFA or WB).

Antibody	Primary/Secondary	Dilution IFA	Dilution WB	Source
c-myc tag	P	1:1000	1:1000	Sigma, cat # M-4439
HA tag	P	1:1000	1:1000	Roche, cat # 1187431001
Aldolase	P	-	1:5000	David Sibley
AMA1	P	1:500	-	Gary Ward
ATPase	P	1:3000	-	Peter Bradley
CPN60	P	1:1000	-	Lilach Sheiner/Swati Agrawal
Catalase	P	-	1:3000	Meissner
DrpB	P	1:500	-	Peter Bradley
GAP45	P	1:3000	-	Dominique Soldati
ISP1	P	1:1000	-	Peter Bradley
M2AP	P	1:1000	-	Vern Carruthers
MIC3	P	1:300	-	Meissner
MIC4	P	1:3000	-	Dominique Soldati
ROP2,3,4	P	1:500	-	Meissner
ROP5	P	1:1000	-	Maryse Lebrun
AlexaFluor488	S	1:3000	-	Life Technologies
AlexaFluor594	S	1:3000	-	Life Technologies
AlexaFluor647	S	1:3000	-	Life Technologies
IRDye680	S	-	1:15000	Li-Cor
IRDye800	S	-	1:15000	Li-Cor

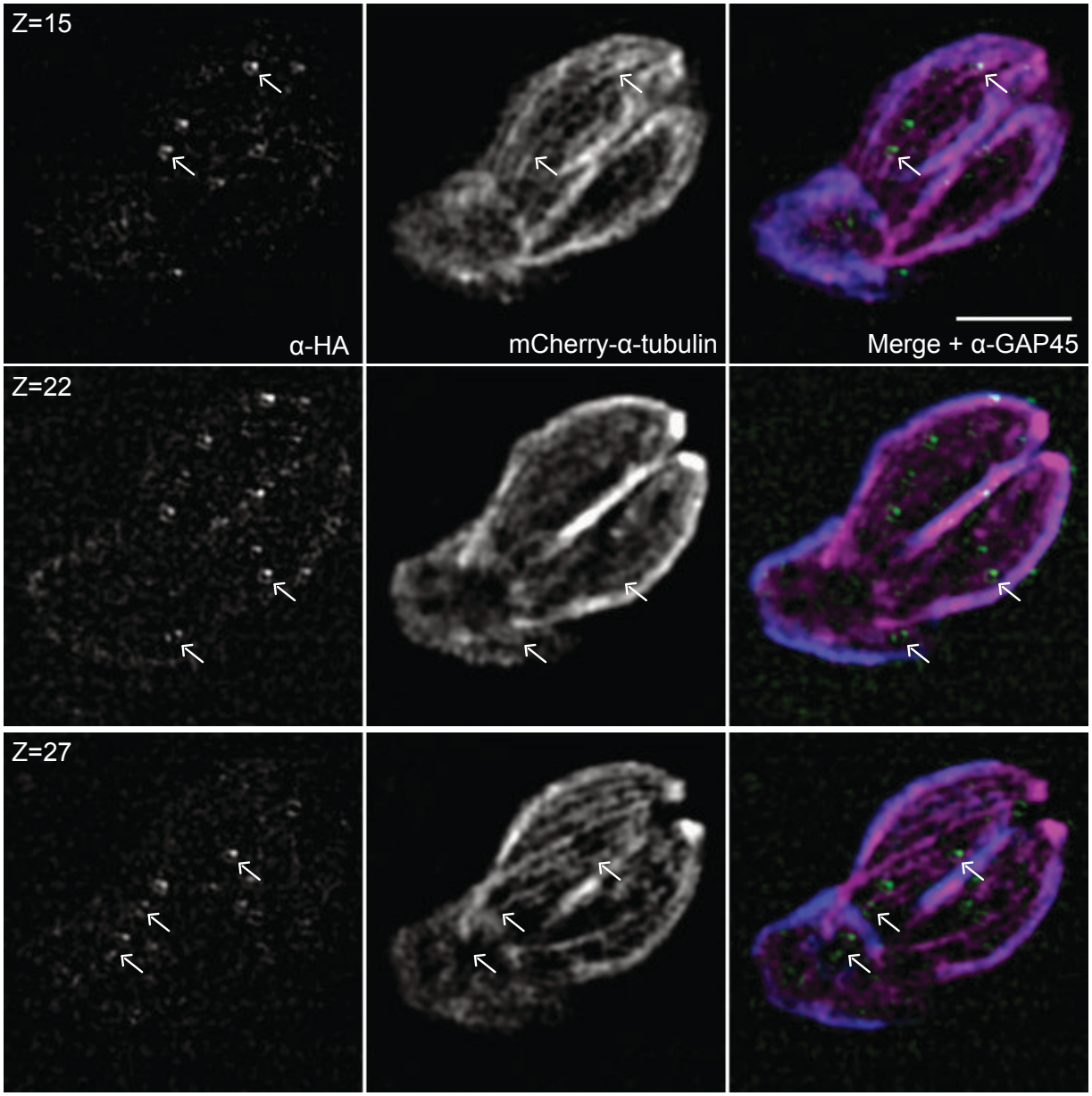
Supplementary Figure S5.1 Main vectors used

This figure provides an overview of all vectors used for genetic manipulation (not counting those for transient expression of marker proteins). A) Parental vectors used to generate vectors specific to a given locus. pG514 (left) contains a PacI site for ligation-independent cloning of PCR products followed by an in-frame 3xHA epitope tag, 3'UTR, and HXGPRT selectable marker for stable transfection. pG474 (right) contains two BsaI sites to clone specific gRNAs for transient expression. B) Three endogenous tagging vectors derived from pG514 for ArlX1 (pG738), ArlX2 (pG734), and ArlX3 (pG736). A C-terminal fragment of varying size was amplified and annealed following PacI digestion and T4 polymerase treatment. The corresponding unique restriction site for each gene is shown in magenta; vectors were digested using this site to linearize them prior to transfection and selection. Note that no derived pG474 vectors are shown for simplicity; each one differs from pG474 through removal of the ccdB coding region and addition of a unique 20 bp gRNA sequence in its place. Abbreviations: AmpR, ampicillin resistance cassette; HX, HXGPRT resistance cassette; UTR, untranslated region; TgU6, *T. gondii* U6 promoter; TUB, tubulin promoter; NLS, nuclear localization signal; Cter, C-terminal fragment.


A
PacI Site
3xHA
3'UTR HX
5'UTR DHFR
HX
3'UTR DHFR
AmpR
pG514
5429 bp
Terminator
CAS9 recognition seq
BsaI Site
ccdB
TUB
BsaI Site
TgU6
3'UTR DHFR
NLS
YFP
NLS
HA
CAS9
pG474
7984 bp
B
AflII Site
TGGT1_269780-Cter
3xHA
3'UTR HX
5'UTR DHFR
HX
3'UTR DHFR
AmpR
pG738
6781 bp
BstBI Site
TGGT1_291800-Cter
3xHA
3'UTR HX
5'UTR DHFR
HX
3'UTR DHFR
AmpR
pG734
7384 bp
TGGT1_288260-Cter
ApaI Site
3xHA
3'UTR HX
5'UTR DHFR
HX
3'UTR DHFR
AmpR
pG736
6401 bp

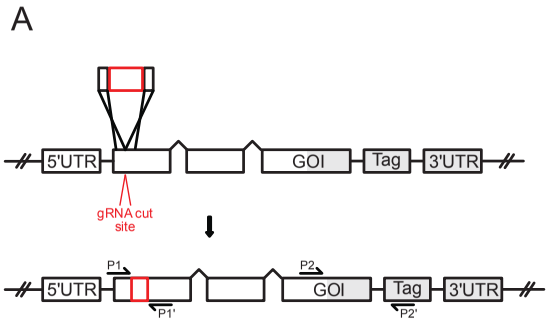
Supplementary Figure S5.2 ArlX1 signal follows sub-pellicular microtubules

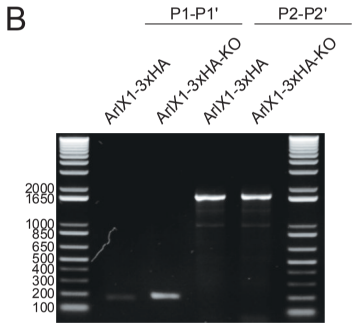
This figure provides another example of how α-HA signal in ArlX1-3xHA parasites is observed in close proximity to sub-pellicular microtubules. Three independent z-sections from the same z-stack are shown, of intracellular tachyzoites transfected with an expression vector for mCherry-α-tubulin. Arrows point either to close association of ArlX1 puncta with microtubules or the presence of ArlX1 signal in the basal end of the dividing mother cell. Scale bar is 2μm.

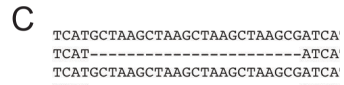

Z=15
α-HA
mCherry-α-tubulin
Merge + α-GAP45
Z=22
Z=27

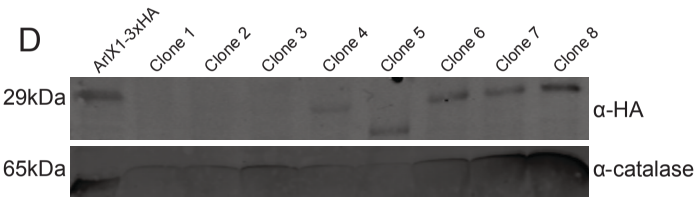
Supplementary Figure S5.3 Generation of an ArlX1 KO line

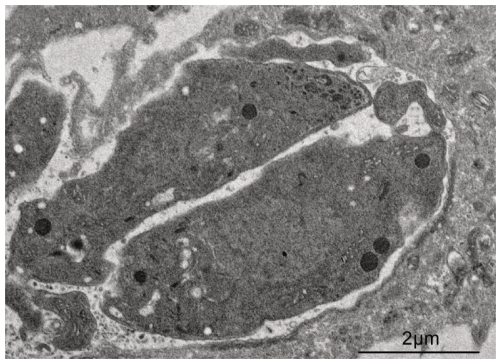
This figure provides an overview of the strategy used to generate a knockout of ArlX1 and the initial confirmation and EM results. A) Schematic of strategy used. A gRNA was designed to create a DSB within the 1st exon of the endogenous ArlX1 locus and co-transfected with an oligo containing STOP codons in all three coding frames, flanked by overlapping regions of homology 5' and 3' of the cut site. B) PCR confirmation of one selected ArlX1-KO clone, showing a slight shift in the P1-P1' product without altering the presence of the C-terminal HA tag. C) Sanger sequencing of the P1-P1' product from B, showing the presence of the exogenous sequence with STOP codons. D) WB confirmation of multiple clones isolated from a single KO transfection; note that C1-C3 no longer express an HA-tagged protein, despite the presence of the C-terminal HA tag (B), while C5 appears to have undergone a small deletion in response to the DSB and C4, C6-8 retain the original product at the same size. E,F) Transmission electron micrographs of intracellular ArlX1-3xHA (E) versus ArlX1-KO (F) tachyzoites. Note the conspicuous decrease in the number of apical micronemes, and the appearance of unusual multi-membranous structures in the parasitophorus vacuole in the KO parasites.

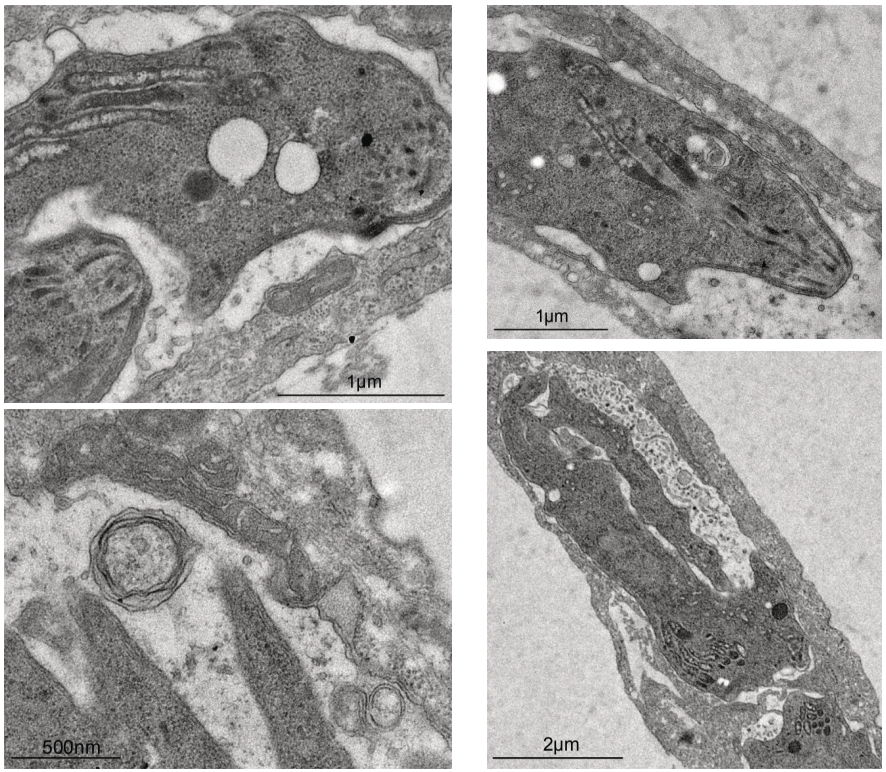

A
5'UTR
GOI
Tag
3'UTR
gRNA cut site
P1
P2
P1'
P2'


B
P1-P1'
P2-P2'
ArlX1-3xHA
ArlX1-3xHA-KO
ArlX1-3xHA
ArlX1-3xHA-KO
2000
1650
1000
850
650
500
400
300
200
100


C
TCATGCTAAGCTAAGCTAAGCTAAGCGATCA
TCAT----------ATCA
TCATGCTAAGCTAAGCTAAGCTAAGCGATCA
**** ****

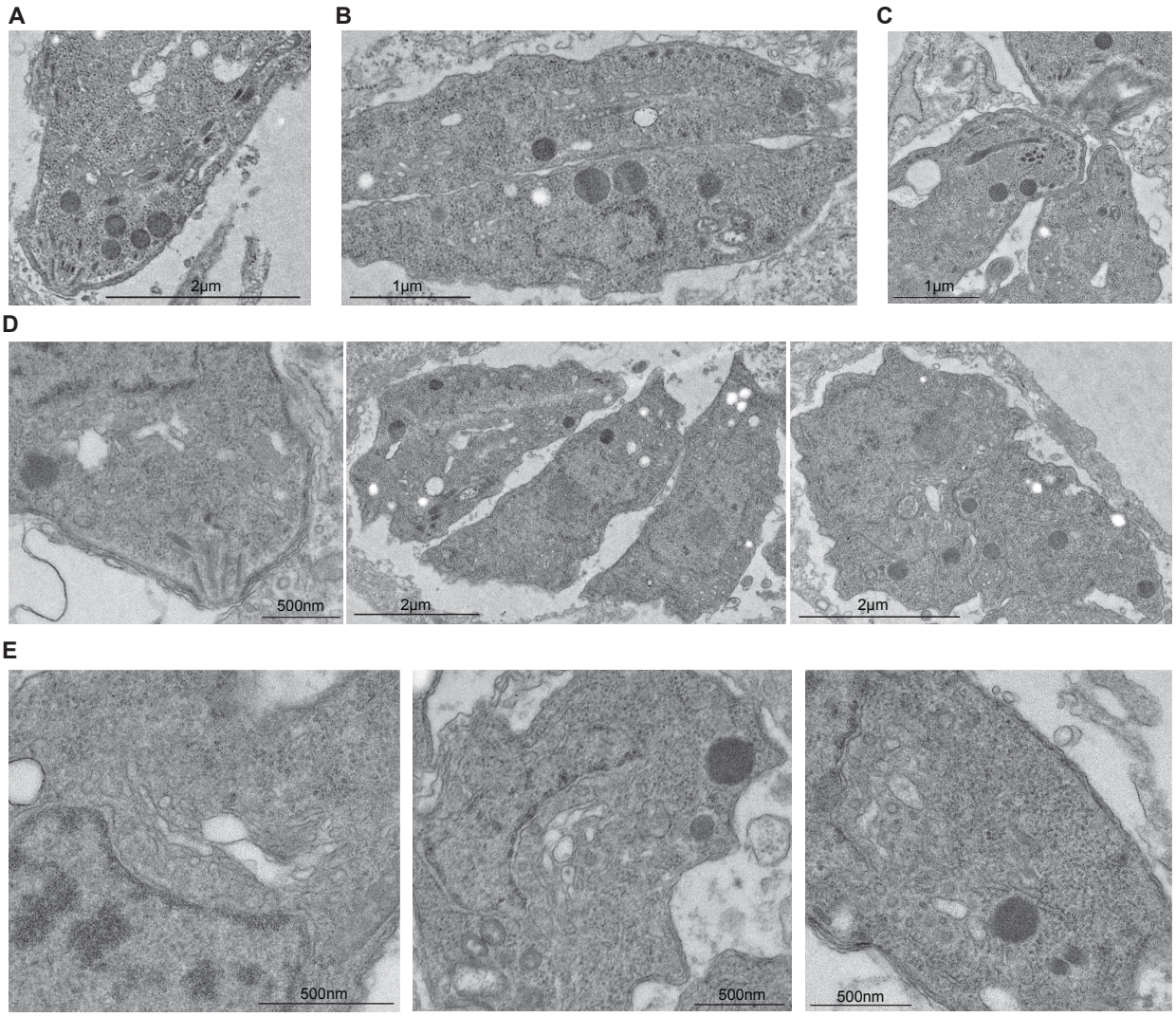

D
ArlX1-3xHA
Clone 1
Clone 2
Clone 3
Clone 4
Clone 5
Clone 6
Clone 7
Clone 8
29kDa
α-HA
65kDa
α-catalase


E
2μm
1μm


F
1μm
500nm
1μm
2μm

Supplementary Figure S5.4 ArlX3 knockdown causes Golgi fragmentation.

This figure shows the effect of ArlX3 loss on the ultrastructure of intracellular *T. gondii* tachyzoites, as assessed by transmission electron microscopy. A-C), untreated parental (A) and TATi-ArlX3 (C) parasites, as well as ATc-induced parental (B) parasites, showing no obvious defects in morphology or organelle complement. D) TATi-ArlX3 parasites induced for 48 hours with ATc showing almost complete lack of micronemes, and fewer conspicuous rhoptries, but with no obvious effects on dense granules. E) Close examination of the Golgi region in TATi-ArlX3 parasites induced for 48 hours with ATc showing various degrees of Golgi fragmentation/vesiculation.


A
2μm
B
1μm
C
1μm
D
500nm
2μm
2μm
E
500nm
500nm
500nm

Online Appendix

More materials can be found in the Online Appendix.