

**A systems-level investigation of the vesicle formation machinery in
*Giardia intestinalis***

by

Shweta Vinodrai Pipaliya

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

Doctor of Philosophy

Department of Medicine
University of Alberta

© Shweta Vinodrai Pipaliya, 2021

ABSTRACT

The endomembrane system is one of the hallmark features of all eukaryotes that distinguishes them from prokaryotes. Compared to bacteria and archaea, eukaryotic cells consist of biochemically and functionally distinct membranous compartments such as the endoplasmic reticulum, the Golgi apparatus, endosomes, lysosomes, and the plasma membrane. Cargo is transported to and from these organelles in distinct vesicle carriers whose formation and fusion at donor and recipient organellar membranes is facilitated by vesicle formation and fusion trafficking proteins. Together, these compartments with this molecular machinery form a complex interconnected network termed the membrane trafficking system that mediates extracellular export, intracellular import, and cargo-sorting between the cell and its environment. Proper functioning of this system is fundamental to the survival of all eukaryotes, and breakdown results in disease states or the inability to inhabit niche environments. Eukaryotic pathogens heavily rely on their membrane trafficking system for host-pathogen interactions, to transition between different lifecycle stages, and secrete virulence factors for immune avoidance and establishing disease. *Giardia intestinalis* is one such enteric microbial parasite of humans and animals that relies on its cargo secretory and endocytic processes to cause diarrheal infection within the animal gut and for environmental survival and propagation.

Microscopically, *Giardia*'s endomembrane organization is strikingly different from that present in model eukaryotes such as yeast, plants, animals, or even other eukaryotic parasites. It lacks conventional stacked Golgi and endo-lysosomal compartments, and instead, possesses numerous *Giardia*-specific organelles. These are the peripheral vacuoles, which are static-state vesicular compartments, a labyrinth tubulovesicular endoplasmic reticulum, and stage-specific encystation-specific vesicles. *Giardia* belongs to the sub-phylum Fornicata within the eukaryotic supergroup Metamonada. Fornicata is a lineage that consists of free-living heterotrophic flagellates such as *Carpodomonas membranifera* and *Carpodomonas*-like organisms, endobionts of ruminant animals such as the retortamonads, and the largely parasitic Diplomonadida, which comprises bi-nucleated parasites such as *Giardia*. The endomembrane organization in fornicates is variable in its complexity and one that is gradually reduced. Within the *Giardia* genus, different *Giardia intestinalis* assemblages have been proposed to be distinct species that cause disease in humans and animals. This thesis examined the molecular evolution of the membrane trafficking system, specifically the vesicle formation machinery, to understand how the minimal trafficking system

in *Giardia* arose and what evolutionary processes were at interplay. It also investigated the differences in the vesicle formation machinery between the different *Giardia intestinalis* assemblages, especially those implicated in causing human infections, through a population-level survey. Finally, upon identifying critical components from several important vesicle formation protein complexes, their molecular functions were assessed to study their roles in this parasite's divergent endomembrane system.

A systems approach using evolutionary bioinformatics (comparative genomics and phylogenetics), immunofluorescent microscopy, proteomics, and large-scale genome assembly determined how the vesicle formation machinery evolved in fornicates from an ancestral state, leading into parasitism and within the *Giardia* lineage, and what organelles they associate in the *Giardia* trophozoite cells. The culmination of work produced in this thesis determined that *Giardia*'s reduced trafficking system is a by-product of ancient losses, parasitism-associated streamlining, and *Giardia*-specific adaptations. A population-level survey of the vesicle formation proteins across isolates of the human-infecting *Giardia intestinalis* also revealed inter-assemblage differences in the molecular complement of these proteins. Finally, *in vitro* microscopy and proteomics investigations with key membrane trafficking proteins in the lab strain of *Giardia intestinalis* AWB (C6) revealed associations of these proteins primarily with the peripheral vacuoles, marking them as a singular yet a multi-dynamic destination for endo-lysosomal trafficking in this parasite. Unexpectedly, promiscuous roles of some machinery were also elucidated at the parasite mitosomes and the ER and have shed light on the plasticity of trafficking system proteins in eukaryotes in general. Overall, findings from this thesis unveiled new modes and tempo by which cargo transport processes evolved and take place in this parasite that is of substantial public health, clinical, and biomedical importance.

PREFACE

(Mandatory due to collaborative work)

Bioinformatic and molecular functional work presented in this thesis are a product of several previous and on-going collaborations.

The beginning portions of Chapter 1 are currently under preparation for a first-author review article conceptualized between Dr. Joel B. Dacks, Dr. Kristina Záhonová, Dr. Lynora Saxinger, and myself. I performed all literature surveys under the guidance of JBD and LS. Figure 1.1 from this chapter was reproduced and modified from the following published dispatch where I was a co-author: Rueckert, S., Pipaliya, S. V., & Dacks, J. B. (2019). Evolution: Parallel Paths to Parasitism in the Apicomplexa. *Current Biology*, 29(17), R836–R839. <https://doi.org/10.1016/j.cub.2019.07.047>. Figure 1.2A from this chapter was reproduced from the following primary research article where I was also a co-author: Karnkowska, A., Treitli, S. C., Brzoň, O., Novák, L., Vacek, V., Soukal, P., Barlow, L. D., Herman, E. K., Pipaliya, S. V., Pánek, T., Žihala, D., Petrželková, R., Butenko, A., Eme, L., Stairs, C. W., Roger, A. J., Eliáš, M., Dacks, J. B., & Hampl, V. (2019). The Oxymonad Genome Displays Canonical Eukaryotic Complexity in the Absence of a Mitochondrion. *Molecular Biology and Evolution*, 36(10), 2292–2312. <https://doi.org/10.1093/molbev/msz147>. Current Biology is an Elsevier Journal that grants permission for article content reuse by its original publication authors. Molecular Biology and Evolution is an open-access Oxford University Press journal that allows the reproduction of its content under the Creative Commons CC-BY-NC licence. Therefore, no additional permissions from either publisher were necessary for the inclusion of these figures in this thesis.

Bioinformatic investigations presented in Chapter 2 were performed as part of the *Carpediemonas membranifera* genome project conceived by Dr. Dayana Salas-Leiva and Dr. Andrew J. Roger at Dalhousie University (Nova Scotia, Canada). Dr. Joel B. Dacks and I designed specific comparative genomics and phylogenetic analyses of the vesicle coat proteins. I performed all bioinformatic investigations and interpreted the results. Both JBD and I composed the written sections to include these results into a larger manuscript currently in preparation. The limited inclusion of the fornicate genomes and transcriptomes in this chapter compared to later parts of the thesis was due to the following reasons: 1) to match the sampling determined for the genome project, and 2) their lack of availability during the initial stages of my Ph.D. Several of the analyzed fornicate genomes were also not yet publicly available in 2017 when these bioinformatic investigations were first pursued. Dr. Goro Tanifuji provided early versions of the *Kipferlia bialata* genome, and Dr. Feifei Xu and Dr. Staffan Svärd shared the *Giardia muris* genome for analyses.

Chapter 3 (exclusive of the preface and afterword) has been published as the following first-author research publication: Pipaliya, S. V., Thompson, L. A., & Dacks, J. B. (2021). The reduced ARF regulatory system in

Giardia intestinalis pre-dates the transition to parasitism in the lineage Fornicata. *International Journal for Parasitology*, 51(10), 825-839. <https://doi.org/10.1016/j.ijpara.2021.02.004>. *International Journal for Parasitology* is an Elsevier Journal that allows its authors to reproduce and re-use articles for dissertations, and therefore, no additional licences were required. Text within this chapter is unmodified from the published article. Joel Dacks and I conceived the project and the design of the bioinformatic studies. Under my guidance, Alexa Thompson performed all comparative genomic analyses pertaining to the evolution and distribution of ADAP across eukaryotes. I performed all remaining comparative genomic and phylogenetic analyses of the ARF regulatory system proteins in fornicates. I also composed all of the figures and initial drafts of the manuscript, which all authors edited. This chapter and publication were conceived as a follow-up to the investigation that determined the evolution of the ARF GEF proteins across eukaryotes where I was a first author: Pipaliya, S. V., Schlacht, A., Klinger, C. M., Kahn, R. A., & Dacks, J. (2019). Ancient complement and lineage-specific evolution of the Sec7 ARF GEF proteins in eukaryotes. *Molecular biology of the cell*, 30(15), 1846–1863. <https://doi.org/10.1091/mbc.E19-01-0073>. Although findings from that study were not explicitly included in this thesis, they nonetheless contributed to our understanding of the ancient ARF GEF complement and provided numerous starting hypotheses for testing regarding the distribution of these proteins in fornicates.

Chapter 4 (exclusive of the preface and afterword) is published original research article: Pipaliya, S.V., Santos, R.R., Salas-Leiva, D., Balmer, E.A., Wirdnam, C.D., Roger, A.J., Hehl, A.B., Faso, C., and Dacks, J.B. (2021). Unexpected organellar locations of ESCRT machinery in *Giardia intestinalis* and complex evolutionary dynamics spanning the transition to parasitism in the lineage Fornicata. *BMC Biology*, 19(1), 167. *BMC Biology* is a Springer Nature open access journal under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>). No additional permissions were necessary to reproduce this work for this thesis. Minor typographical and syntax edits have been made to the original text to improve clarity and flow with the rest of the thesis. Dr. Joel Dacks, Rui Santos, Dr. Adrian Hehl, Dr. Carmen Faso, and I conceived the project and design of the informatics and molecular functional experiments. I performed all bioinformatic analyses and some immunofluorescence assays and microscopy during my three-month research visit to the Institute of Parasitology (University of Zurich, Switzerland) in the laboratory of Dr. Adrian Hehl in 2019. I also performed the subcellular fractionation experiments with HA-CHMP7 during my eight-month research visit to the University of Bern from Fall 2020 until Spring 2021, under the guidance of Dr. Corina Wirdnam and Dr. Carmen Faso. All remaining experimental work was performed by RRS, EAB, CDW, and CF. Personnel at the Functional Genomics Centre Zurich (ETH Zurich and the University of Zurich, Switzerland) performed the mass-spectrometry experiments. Imaging was performed with the equipment provided by the Center of Microscopy and Image Analysis (ZMB) at the University of Zurich and the Microscopy Imaging Center at the University of Bern. RS, JBD, CF, ABH, and I interpreted all results, and RS and I composed early versions of the manuscript, which was edited and approved by all authors.

The data presented in Chapter 5 are in preparation for eventual submission of a first-author publication. Work from this chapter was produced in collaboration with Dr. Carmen Faso (University of Bern, Switzerland), Dr. Adrian Hehl (University of Zurich, Switzerland), myself, and Dr. Joel Dacks. Dr. Carmen Faso and Dr. Corina Wirdnam generated the episomal constructs. Dr. Manfred Heller and Dr. Sophie Braga at the Proteomics Mass Spectrometry Core Facility (University of Bern) performed the mass-spectrometry experiments. I performed all remaining molecular functional experiments (*i.e.*, parasite culture and propagation, parasite transfections, immunofluorescence assays, widefield and confocal microscopy, and co-immunoprecipitation and western blotting) during my two research visits to the University of Zurich and the University of Bern (described above), with guidance from Rui Santos, Dr. Corina Wirdnam, Dr. Adrian Hehl, and Dr. Carmen Faso. Technical guidance with confocal image deconvolution and FIJI/ImageJ microscopy image analyses was provided by Dr. Yury Belyaev at the Theodor Kocher Institute (University of Bern). I interpreted all imaging and proteomics data with guidance from Dr. Carmen Faso, Dr. Richard Kahn, and Dr. Joel Dacks.

The population genomics and bioinformatic analyses performed in Chapter 6 are under preparation for a first-author manuscript with co-supervisors (and co-authors) Dr. Matthew Croxen and Dr. Joel Dacks. The project was initially conceptualized between Dr. Kinga Kowalewska-Grochowska, Dr. Matthew Croxen, Joel Dacks, and myself. MC, JBD, and I designed the experiments, whereas I performed the *de novo* genome assembly and reference-based gene predictions for all investigated BCCDC PHL *Giardia* isolates under the supervision of MC. All data analyses were also performed by me with guidance from MC.

Dedicated to my beloved parents.

“Wisdom begins in wonder.”

-- Socrates

ACKNOWLEDGEMENTS

Although my passion for biomedical and infectious diseases research began early on as an undergraduate student, it was only after joining the Dacks lab in 2017 that I was motivated to pursue a graduate degree. This career trajectory was completely unanticipated yet incredibly rewarding. To misquote the famous African proverb, “it takes a village to raise a Ph.D.,” and I am beyond grateful for the mentorship and support I have received over the past four and a half years. This cross-disciplinary Ph.D. would not have been possible without the financial support I have received over the years from the Canadian Institutes for Health Research (CIHR), Alberta Innovates, Faculty of Graduate Studies and Research, and Faculty of Medicine and Dentistry.

First and foremost, I would like to thank my supervisor Dr. Joel Dacks. Joel's enthusiasm for protistology and evolutionary biology is infectious. He fosters an environment that is conducive to learning and independent exploration. Joel's encouragement and mentorship towards various facets of my research projects have been pertinent to my success as a graduate student. I am especially thankful for his kindness during periods of academic and personal setbacks. Next, I would like to thank my co-supervisor, Dr. Matthew Croxen, for taking me on as his student during the later stages of my degree. Thank you, Matthew, for helping ignite my newfound passion for infectious diseases genomics and for advising me and helping shape my future post-doctoral journey. Your ongoing guidance, fervent excitement towards my projects, and words of encouragement are really motivating, and I appreciate our open and light-hearted conversations about science and beyond. I am also fortunate to have had an incredibly supportive supervisory committee. Thank you, Dr. Kinga Kowalewska-Grochowska, for your enthusiasm for *Giardia* and parasites in general and for encouraging me to explore the translational aspects of my projects. I am also grateful to Dr. Paul Melançon for his wisdom and insightful discussions on general membrane trafficking and ARF biology in the early periods of my degree. I would also like to express gratitude to my Ph.D. candidacy examination committee, particularly Drs. Maria Ioannou and Paul Stothard for making the examination process an enjoyable learning experience. Thank you also to my Ph.D. defense committee, especially Drs. Naomi Fast and Emmanuelle Cordat for taking the time to read my thesis critically and for the constructive discussions. I also appreciate the assistance I have received from the Department of Medicine over the past couple of years and the graduate student administrative staff, particularly Julian Schulz and Dr. Nadia Jahroudi, for being patient with me and helping in the smooth transition of programs and departments during tumultuous periods of my degree.

Collaborative research has been fundamental throughout my Ph.D., and I have had the pleasure of learning from brilliant minds in the fields of evolutionary protistology, parasitology, molecular biology, and *Giardia* biology. First, I would like to thank Drs. Andrew Roger and Dayana Salas-Leiva at Dalhousie University (Nova Scotia, Canada) for their valuable mentorship over the years and enthusiasm towards my research endeavours, especially in all things *Carpodidomonas*. I am also grateful to Dr. Richard Kahn at the Emory University (United States of America), whose passion for ARF biology and open-mindedness towards exploratory research is unparalleled. Finally, I am incredibly thankful to Drs. Adrian Hehl and

Carmen Faso for supporting the *Giardia* benchtop projects and their lab members (Rahel, Lenka, Jackie, Chandra, Julian, Jessie, Saša, Rui, Erina, Timothé, and Corina) for welcoming me into their groups at the University of Zurich and the University of Bern (Switzerland). Thank you, especially to Carmen, for wearing the unofficial third supervisor hat, and being patient and teaching me numerous molecular functional techniques. I have had the privilege of learning so much about *Giardia* membrane trafficking from the best in this field.

My experiences in the Dacks lab have been enjoyable, largely thanks to my colleagues with whom I have had the pleasure of working. The past and current Dacklings (Alex (x2), Alexa, Raegan, Inma, Kira, Kika, and many more) have been essential in this academic journey. I have learned so much from each of you and continue to do so every day. I am especially grateful to Dr. Lael Barlow, Dr. Chris Klinger, and my spooky sisters of science, Dr. Beth Richardson, Dr. Emily Herman, and Laura Lee, who welcomed me into the group when I first joined as an undergraduate student. Your mentorship and support have been vital to my success as a student in the lab, and I am forever in awe of your wit and wisdom. Thank you for always lifting me in the good and bad times; I deeply cherish our friendships. My academic journey has taken me abroad to Switzerland on multiple occasions. While being away from home, I was fortunate to have the company of some of the warmest group of individuals with who I have fostered life-long friendships. Thank you, Atitbhai and Aartiben, for welcoming me into your home in Zurich and for becoming my second family. Thank you, Aldo and Peter, for feeding me delicious meals, watching terrible reality singing shows with me, and teaching me German over the many months in Bern. Thank you, Rahel, for all the spontaneous hiking trips to the Alps and for being my personal tour guide to your beautiful country. Thank you, Matt, for the books, the desserts, the laughs, and the companionship. Thank you, Erina, for your unconditional support and our sisterhood; you have the best of energies, and your kind-heartedness is inspirational. I am so grateful to all of you for especially brightening my days when I felt homesick. My friends in Edmonton have also been an essential source of strength over many years. Riya, Himani, Dhruti, Miraben, Grishmaben, Michelle, Vivian, Isha, Alisha, Sparsh, and Chintan, you have been my personal cheerleaders, and I am very appreciative of your ongoing encouragement, especially during the thesis-writing process. Thank you for believing in me, especially on days when I did not.

Finally, my family has been monumental to my success as a scientist and a person. Thank you, Mummy, Papa, and Dhruvin, for being my pillars and strength in all my academic and non-academic pursuits. Thank you for pushing me to be the best version of myself but with kindness, grace, and humility. Your unwavering zeal, love, and support are what enrich me every day; I am who I am because of you.

TABLE OF CONTENTS

1. Introduction	1
1.1 Overview	1
1.2 Evolution of parasitism: A road well-traveled within the existing eukaryotic diversity	1
1.2.1 Apicomplexan parasites – Parasitism within TSAR	5
1.2.2 Trypanosomatids and <i>Naegleria fowleri</i> – Parasitism within Discoba	6
1.2.3 <i>Entamoeba histolytica</i> – Parasitism within Amoebozoa	7
1.2.4 Parasitism within Opisthokonta	8
1.2.5 <i>Prototheca</i> – Algal parasitism within Archaeplastida	9
1.2.6 <i>Giardia</i> and <i>Trichomonas</i> – Parasitism within Metamonada	10
1.3 A panoramic tour of the endomembrane system	11
1.3.1 Endoplasmic reticulum and the <i>trans</i> -Golgi network – early secretory pathway	15
1.3.2 Post-Golgi endo-lysosomal trafficking	16
1.3.3 Key vesicle formation machinery	18
1.3.4 Key vesicle fusion machinery	22
1.3.5 Endocytic systems of key protist parasites and their roles in the pathogenesis	23
1.4 <i>Giardia intestinalis</i> and Giardiasis – brief history and clinical disease	25
1.5 Ecology and phylogenetic relationships within the subphylum Fornicata	26
1.6 An 'omics perspective on the <i>Giardia</i> spp. and <i>G. intestinalis</i> assemblages and implications on the clinical disease	30
1.7 <i>Giardia</i> lifecycle and pathogenesis	35
1.8 <i>Giardia</i> 's enigmatic cell biology with a spotlight on its endomembrane system	38
1.8.1 Peripheral vacuoles: constitutive secretion in trophozoites	38
1.8.2 Endoplasmic reticulum: protein synthesis and possible roles in constitutive secretion	40
1.8.3 Encystation-specific vesicles: <i>bona fide</i> Golgi-like compartments for regulated secretion	41
1.8.4 Mitosomes	42
1.8.5 Other structural and unusual aspects of <i>Giardia</i> cell biology	43
1.9 Current state of molecular parasitological and genomic research in <i>Giardia</i>	45
1.10 Scope and organization of the thesis	45
2. A minimal vesicle coat protein complement in <i>Giardia</i> is shaped by lineage-specific evolutionary modulations	49
2.1 Overview	49
2.2 Introduction	49
2.3 Materials and Methods	55
2.3.1 Dataset and query sequence retrieval	55
2.3.2 Homology Searching	55
2.3.3 Phylogenetic analyses with heterotetrameric adaptor subcomplexes and COPII-Sec24	56
2.4 Results	57
2.4.1 Ancestral reduction likely shapes a minimal endocytic system within the fornicates	57
2.4.2 Losses along the path to parasitism	61
2.4.3 Lineage and parasite-specific expansions within the adaptin and COPII machinery	64
2.5 Discussion	67
2.5.1 Reductive evolution in fornicates may be indicative of redundant or dispensable aspects of the early and late endosomal system	67
2.5.2 Duplications within the existing machinery for a diversified role at the ER-Golgi	69
2.5.3 Known functions of the vesicle coats within <i>Giardia</i>	71

2.6 Conclusions	72
3. The reduced ARF regulatory system in <i>Giardia intestinalis</i> pre-dates the transition to parasitism in the lineage Fornicata.....	73
3.1 Preface.....	73
3.2 Introduction	74
3.3 Materials and Methods.....	80
3.3.1 Dataset mining of pan-eukaryotic representatives and Fornicata genomes and transcriptomes	80
3.3.2 Comparative genomics of ADAP across select pan-eukaryotic representatives	80
3.3.3 Comparative genomics of ARF, ARF GAPs, and ARF GEFs in Fornicata	80
3.3.4 Phylogenetic analyses	81
3.3.5 Data accessibility	82
3.4 Results	82
3.4.1. Resampling of key protist lineages identifies ADAP in the LECA ARF GAP complement.....	82
3.4.2 Comparative genomics of ARFs in <i>Giardia</i> and its fornicate relatives reveals an expansion within the ARF GTPases.....	85
3.4.3 Fornicates underwent extreme streamlining within the ARF GAP complement while retaining the most conserved members.....	88
3.4.4. Fornicates underwent a loss within the Golgi ARF GEF GBF1 but retain BIG and Cytohesin	92
3.5 Discussion.....	96
3.6 Afterword.....	101
4. Unexpected organellar locations of ESCRT machinery in <i>Giardia intestinalis</i> and complex evolutionary dynamics spanning the transition to parasitism in the lineage Fornicata.....	103
4.1 Preface.....	103
4.2 Introduction	104
4.3 Materials and Methods.....	109
4.3.1 Taxa studied.....	109
4.3.2 Translation of nucleotide transcriptomes belonging to <i>Carpodionomonas</i> -like organisms	109
4.3.3 Comparative genomics and homology searching	109
4.3.4 Phylogenetic analyses of the ESCRTIII and ESCRTIIIA-SNF7 family proteins	110
4.3.5 <i>Giardia</i> cell culture and transfection.....	111
4.3.6 Construction of expression vectors	111
4.3.7 Immunofluorescence Assays	112
4.3.8 Fluid phase marker uptake.....	112
4.3.9 Microscopy and Image Analysis.....	112
4.3.10 Co-immunoprecipitation with limited cross-linking	112
4.3.11 Protein analysis and sample preparation for mass spectrometry (MS)-based protein identification	113
4.3.12 Crude subcellular fractionation.....	113
4.3.13 <i>in silico</i> co-immunoprecipitation proteomics analyses	113
4.4 Results	114
4.4.1 ESCRT losses in Fornicata are gradual and represent a slow transition leading to parasitism	114
4.4.2 Losses correlating with parasitism and inter-strain variation.....	117
4.4.3. Localization of <i>Giardia</i> ESCRTII-Vps25 and newly identified ESCRTII-Vps36 at peripheral vacuoles	120
4.4.4 Characterization of ESCRTIII-Vps20L and ESCRTII components at the endoplasmic reticulum	124

4.4.5 Evolutionary and protein analyses of the newly identified <i>Giardia</i> ESCRTIII-CHMP7 reveal unsuspected ancient origins and a novel ER-mitosomal localization.....	128
4.5 Discussion.....	133
4.5.1 Gradual reductive evolution of ESCRTs and MVBs in the Fornicata	133
4.5.2 ESCRT promiscuity at <i>Giardia</i> PVs, ER, and mitochondria	136
4.5.3. A comprehensive appreciation of ESCRT evolution and distribution in <i>Giardia intestinalis</i> ..	137
4.6 Conclusions	138
4.7 Afterword.....	138
5. The giardial ARF regulatory system localizes to the peripheral vacuoles and associates with vesicle formation and fusion machinery	140
5.1 Introduction	140
5.2 Materials and Methods.....	144
5.2.1 <i>Giardia</i> trophozoite culture and transfection	144
5.2.2 Genomic DNA isolation and cloning for episomal vector construct generation.....	144
5.2.3 Indirect immunofluorescence and fluid-phase uptake assays.....	145
5.2.4 Laser Scanning Confocal Microscopy and image analyses.....	146
5.2.5 Limited cross-linked co-immunoprecipitation using anti-HA agarose beads.....	146
5.2.6 Western Blotting and Coomassie Staining.....	147
5.2.7 Liquid Chromatography Mass Spectrometry (LC/MS) for quantitative proteomics	147
5.2.8 <i>in silico</i> evaluation of LC/MS hits	148
5.3 Results	149
5.3.1 Canonical ARF1 localizes to the cortical PV populations in vegetative trophozoites.....	149
5.3.2 ARF1FA and ARF1FB1 localize to the peripheral vacuoles in a pattern similar but not identical to ARF1	154
5.3.3 ARFGAP1 and CYTHa are also PV localized	158
5.3.4 ARF1-HA associates with <i>Giardia</i> 's vesicle formation machinery at the PVs	162
5.3.5 ARF1FA-HA and ARF1FB1-HA protein interactions are similar to <i>Gi</i> ARF1 at the peripheral vacuoles	165
5.3.6 ARFGAP1 and CYTHa also associate with the PV machinery	171
5.3.7 Intersection of binding partners confirm the association of ARF-regulatory system proteins at the PVs.....	172
5.4 Discussion.....	174
5.4.1 Paralogous expansions within fornicate ARF regulatory system likely accommodates existing trafficking complexity.....	174
5.4.2 Conventional roles of the <i>Giardia</i> ARF regulatory system to functionally diversify the PVs ...	175
5.4.3 Possibility of ARF regulation in a cascade-like mechanism in crosstalk with the giardial Rab GTPases	178
5.5 Conclusions	178
6. Population-level survey maps differences in the encoded complement of vesicle formation machinery between the two human-infecting <i>Giardia intestinalis</i> assemblages	180
6.1 Introduction	180
6.2 Materials and Methods.....	183
6.2.1 Paired-end read information and data retrieval	183
6.2.2 Read quality assessment, taxonomic classification, and de-contamination using Kraken2 ...	183
6.2.3 MaSuRCA <i>de novo</i> genome assembly	184
6.2.4 Genome assembly completeness and contiguity evaluations.....	185
6.2.5 Reference- mapped gene predictions and functional annotation using Liftoff and TBLASTN	186
6.3 Results	187

6.3.1 Genome contiguity and completeness analyses yield uniform results across all isolates	187
6.3.2 Overall GC content and genome sizes are comparable to other isolates of assemblages A and B	190
6.3.3 Gene predictions and functional annotations suggests close similarity to reference genomes	195
6.3.4 ESCRT repertoire varies between assemblage A and B isolates	195
6.3.5 Vesicle coat complexes follow a similar pattern of inter-assemblage molecular differences ..	198
6.3.6. Paralogues of the ARF regulatory system proteins continue to differ between the two assemblages	201
6.4 Discussion	203
6.4.1 Genome size differences and implications for parasite cell biology	203
6.4.2 Potential impacts of missing vesicle formation machinery on secretory processes within parasites of the two assemblages	204
6.4.3 General implications on giardial parasitism and species complex	206
6.5 Conclusions	208
7. General Discussion	209
7.1 Overview	209
7.2 Cellular implications of conservations and dispensability in the fornicate vesicle formation machinery	209
7.3 An integrated model for peripheral vacuoles as the singular yet multidimensional hub for endo-lysosomal processes	213
7.4 Limitations and future perspectives	215
7.5 Clinical and public health translational significance	218
7.6 Conclusions	219
References	221
Online supplemental material	272

LIST OF FIGURES

Figure 1.1. A schematic tree depicting eukaryotic relationships and the evolution of parasitism within major taxonomic groups.....	3
Figure 1.2. Overview of the eukaryotic membrane trafficking system and vesicle budding and fusion processes.....	14
Figure 1.3. Phylogenetic relationships within Fornicata, <i>Giardia</i> , and with other metamonads.....	29
Figure 1.4. The simple lifecycle of <i>Giardia intestinalis</i>	36
Figure 1.5. Illustration of a <i>Giardia intestinalis</i> trophozoite cross-section.....	44
Figure 2.1. Depiction of known intracellular localizations of the vesicle coat complexes within a typical endomembrane system and their structural configurations.....	53
Figure 2.2. Coulson plot depicting the distribution of endocytic vesicle coat proteins within fornicates compared to various eukaryotic and metamonad representatives.....	59
Figure 2.3. Phylogenetic analyses with fornicate AP- β and β -COP subunits.....	62
Figure 2.4. Phylogenetic analyses with the fornicate Sec24 paralogues against pan-eukaryotic Sec24I-III.....	65
Figure 2.5. Reductive and lineage-specific evolution of the vesicle coats in metamonads.....	70
Figure 3.1. The eukaryotic ARF GTPases and their previously elucidated intracellular roles.....	78
Figure 3.2. Comparative genomic analyses of ADAP in previously analyzed and newly available, genomes from representative eukaryotic lineages.....	84
Figure 3.3. Dot plot depiction of all identified and classified ARF GTPases, ARF GEFs, and ARF GAPs in fornicate genomes and transcriptomes.....	86
Figure 3.4. Phylogenetic investigation assessing ancient versus independent duplications of the fornicate ARF1F paralogs.....	87
Figure 3.5. Phylogenetic analyses of fornicate ARF GAPs homologues.....	90
Figure 3.6. Phylogenetic characterization of the identified <i>Carpediemonas membranifera</i> BIG orthologue.....	94
Figure 3.7. Summary of patterns of gains and losses in ARF, ARF GEF, and ARF GAPs, identified, focusing on the fornicate lineages.....	99
Figure 4.1. Overview of the eukaryotic ESCRT machinery.....	107
Figure 4.2. Distribution of the ESCRT components within Fornicata.....	115
Figure 4.3. Phylogenetic analyses of the SNF7 domain-containing ESCRTs in Fornicata.....	118
Figure 4.4. Characterization of <i>GIVps25</i> -HA subcellular location.....	122
Figure 4.5. Characterization of <i>GIVps36A</i> -HA subcellular location.....	123

Figure 4.6. Co-labelling of <i>GiVps25</i> -HA, <i>GiVps36A</i> -HA and <i>GiHA-Vps20L</i> with ER membrane marker <i>GiPDI2</i>	125
Figure 4.7. Co-expression of epitope-tagged <i>GiVps25</i> with either <i>GiVps20L</i> or <i>GiVps36</i>	127
Figure 4.8. <i>Ab initio</i> homology-based structural analysis of the CHMP7 N-terminus.....	130
Figure 4.9. Characterization of <i>GiCHMP7</i> subcellular location.....	131
Figure 4.10. Proposed ESCRT evolution in Fornicata.....	135
Figure 5.1. Sequence alignment of the three <i>Giardia intestinalis</i> AWB ARF1 paralogues for biochemical activity assessment.....	143
Figure 5.2. Population-level expression analysis of epitope-tagged ARF regulatory system proteins.....	150
Figure 5.3. Characterization of <i>GiARF1</i> -HA subcellular location.....	153
Figure 5.4. Characterization of <i>GiARF1FA</i> -HA subcellular location.....	155
Figure 5.5. Characterization of <i>GiARF1FB1</i> -HA subcellular location.....	157
Figure 5.6. Characterization of <i>GiARFGAP1</i> -HA subcellular location.....	160
Figure 5.7. Characterization of <i>GiCYTHa</i> -HA subcellular location.....	161
Figure 5.8. Co-immunoprecipitation and mass spectrometry analyses to determine <i>GiARF1</i> -HA protein-protein interactions.....	164
Figure 5.9. Co-immunoprecipitation and mass spectrometry analyses to determine <i>GiARF1FA</i> -HA protein-protein interactions.....	167
Figure 5.10. Co-immunoprecipitation and mass spectrometry analyses to determine <i>GiARF1FB1</i> -HA protein-protein interactions.....	168
Figure 5.11. Co-immunoprecipitation and mass spectrometry analyses to determine <i>GiARFGAP1</i> -HA protein-protein interactions.....	170
Figure 5.12. Co-immunoprecipitation and mass spectrometry analyses to determine <i>GiCYTHa</i> -HA protein-protein interactions.....	171
Figure 5.13. Network analysis of the PV-associated interacting partners.....	173
Figure 6.1. Workflow for <i>de novo</i> assembly and reference-based gene predictions.....	189
Figure 6.2. Percent GC content comparisons between assemblage A and B BCCDC PHL genome assemblies.....	191
Figure 6.3. Genome size comparisons between assemblage A and B BCCDC PHL genome assemblies.....	193
Figure 6.4. Tile-plot depictions of the ESCRT repertoire in the BCCDC PHL isolates.....	197
Figure 6.5. Tile-plot depictions of COPII and retromer repertoire in the BCCDC PHL isolates.....	199

Figure 6.6. Tile-plot depictions of heterotetrameric adaptor complexes and clathrin repertoire in the BCCDC PHL isolates.....**200**

Figure 6.7. Tile-plot depictions of the giardial ARF regulatory system protein repertoire in the BCCDC PHL isolates.....**202**

LIST OF ABBREVIATIONS

ADAP – ArfGAP with dual PH domains	factor
AP – adaptins; adaptor protein complexes	GTP – guanosine triphosphate
ARF – ADP-ribosylation factor	GFF – general feature format
ARL – ARF-like	HA – hemagglutinin
BAR – Bin, Amphiphysin, and Rvs	HMM – hidden markov model
BCCDC PHL – British Columbia Centre for Disease Control Public Health Laboratory	HTAC – heterotetrameric adaptor complexes
BFA – brefeldin-A	ILV – intraluminal vesicles
BIG – Brefeldin-inhibited GEF	IBAQ – intensity-based absolute quantification
BiP – binding immunoglobulin protein	LECA – last eukaryotic common ancestor
BLAST – basic local alignment search tool	LSCM – laser scanning confocal microscopy
CHMP – charged multivesicular body protein	LC/MS – liquid-chromatography mass- spectrometry
CIPRES – cyberinfrastructure for phylogenetic research	MLST – multi-locus sequence typing
CMC – central mitochondrial complex	MRO – mitochondria-related organelle
CME – clathrin-mediated endocytosis	MTC – multi-subunit tethering complexes
Co-IP – co-immunoprecipitation	MTS – membrane trafficking system
COP – coatamer protein	MUSCLE – multiple sequence comparison by log-expectation
CRuMs – collodictyonids, rigifillids and mantamonadids	MVB – multivesicular body
CWP – cyst-wall protein	OPH – organelle paralogy hypothesis
EGT – endosymbiotic gene transfer	ORF – open reading frame
ER – endoplasmic reticulum	PIP – phosphatidylinositol phosphate
ERGIC – ER/Golgi intermediate compartments	PV – peripheral vacuoles
ESCRT – endosomal sorting complexes required for transport	PX – phox homology
ESV – encystation-specific vesicle	SDS-PAGE – sodium dodecyl sulphate- polyacrylamide gel electrophoresis
FYVE - Fab1p, YOTBm Vac1p, EEA1	SM – sec1/munc-18
GAP – GTPase activating protein	SNARE - soluble N-ethylmaleimide- sensitive factor attachment protein receptor
GBF1 - golgi-specific brefeldin resistance factor	SNF – sucrose non-fermenting protein
GDP – guanosine diphosphate	SNX – sorting nexins
GEF – guanine exchange nucleotide	SRP – signal recognition particle

TBC – Tre2/Bub2/Cdc16

TGN – *trans*-Golgi network

TRAPP – transport protein particle

TSAR – Telonemia, Stramenopila,

Alveolata, and Rhizaria

VPS – vacuolar protein sorting

CHAPTER 1

Introduction

1.1 Overview

The goal of this thesis was to uncover the molecular evolution and roles of trafficking system proteins within the existing endomembrane complexity of the microbial parasite, *Giardia intestinalis*. In order to gain a comprehensive appreciation of the work presented in the subsequent chapters, it is necessary to understand how eukaryotic parasitism is generally defined and the patterns of evolution it follows across the tree of life. This introduction chapter begins by providing examples of important human-infecting parasites nested within different eukaryotic supergroups that have taken interesting evolutionary paths to reach pathogenicity as an end-state. It also provides a contextual overview into the best-known mechanisms and players that underpin trafficking processes, and how parasites have evolved these in comparison to their free-living relatives for a host-adapted lifestyle. Because the parasite of focus in this thesis is *Giardia intestinalis*, this introductory chapter also discusses the clinical and epidemiological relevance of the disease it causes. Known aspects of *Giardia* biology and the underlying molecular mechanisms for disease establishment, especially the role of membrane trafficking system within those processes, are also covered. One of the objectives in this thesis was to uncover the precise evolutionary mechanisms by which trafficking pathways have diverged in this pathogen in comparison to its relatives. Therefore, taxonomic descriptions and an overview into the ecological niches that are occupied by the members of the Fornicata lineage, to which *Giardia* belongs, are also provided. Known morphophysiological and cellular adaptations that follow a transition to parasitism are described as well. Later chapters also investigated trafficking complement differences within the *Giardia* genus itself, and so genomic and phylogenetic perspectives on the molecular and cellular divergences between the different *Giardia intestinalis* species, and which possibly translate to differences in host-specificity and disease manifestation, are also discussed.

1.2 Evolution of parasitism: A road well-traveled within the existing eukaryotic diversity

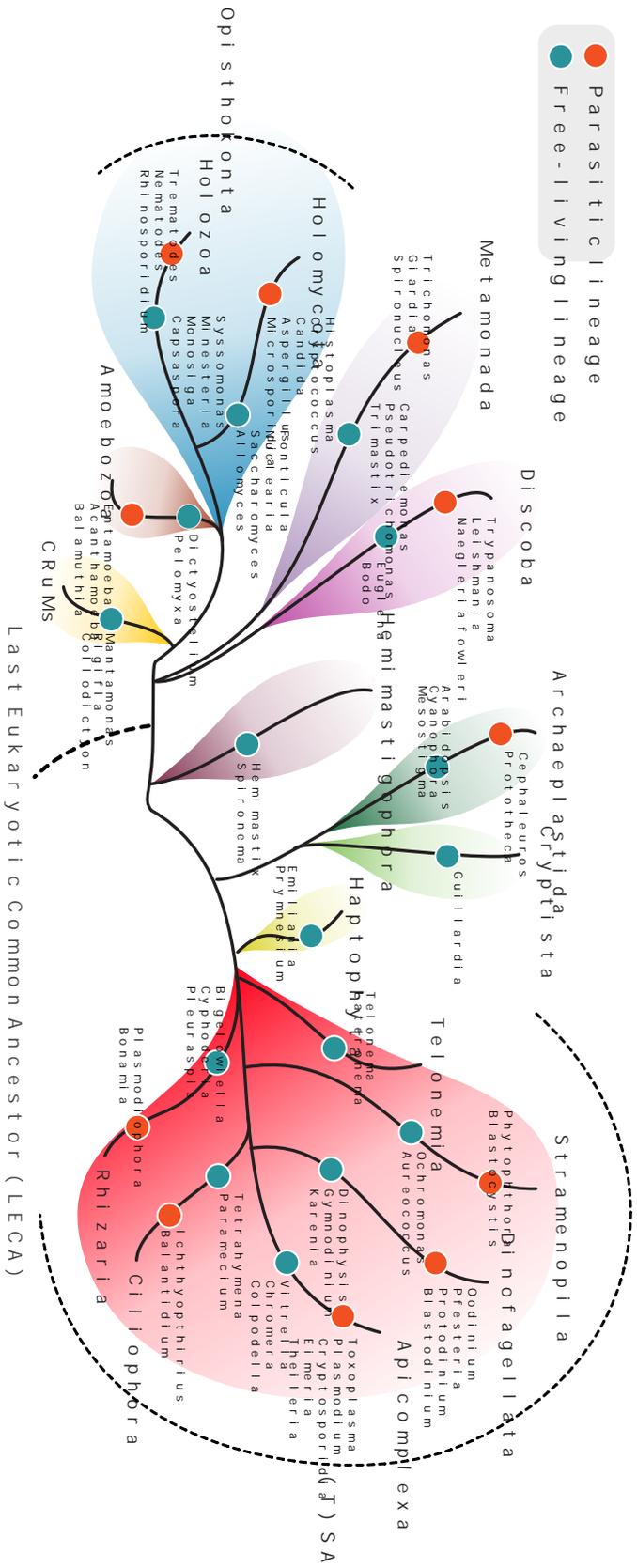
Infections caused by protist (*i.e.*, single-celled eukaryotes) parasites pose significant public health and financial burden and are consequential contributors to high rates of global disability-adjusted life years (Kirk et al., 2015). A considerable number of these are characterized as neglected diseases that affect populations inhabiting parts of the world that are disproportionately faced with increased levels of socioeconomic disparities (World Health Organization, 2013). Many of these regions also have temperate and subtropical climates that house habitats ideal for vector-mediated disease propagation (Githeko et al., 2000). A paradigm shift towards globalization and ease in international travel has increased global interconnectedness. However, it has also led to the emergence of new diseases and their accelerated spread, prevalence, and endemicity in parts of the world that were once unaffected (Wilson, 1995). From a

clinical perspective, some of these problems are further substantiated by over-prescription of antimicrobials, giving rise to multidrug resistance in pathogens and 'superbugs' (Llor & Bjerrum, 2014). When combined, these multifactorial variables have exacerbated the growing inability to treat infectious diseases and eliminate their contribution towards global healthcare and socioeconomic burdens. Many clinically implemented pharmacological agents target critical aspects of parasite biology that have evolved due to reciprocal host-pathogen evolutionary 'arms' race known as the Red Queen Hypothesis, a theory first proposed by biologist Leigh Van Valen (Papkou et al., 2019; Van Valen, 1973). Therefore, understanding parasite biology and cellular systems through an evolutionary lens has significant clinical utility and can underpin drug treatment efficacy in distantly related microbes. This insight also becomes valuable for developing and clinically implementing new therapeutics, or repurposing treatments that may otherwise be unconventional (Klug et al., 2016).

The lay audience often associates parasites as biological entities with shared ancestry, just as animals, plants, or fungi. However, a more complex definition of parasitism and its evolution is necessary to accurately reflect distribution of these organisms amongst other eukaryotes. Parasitism is a life strategy employed by any organism through obligate association with another (*i.e.*, a host) for survival and one that results in a decreased fitness of the host (Stevens, 2010). This definition allows parasites to be interspersed with free-living organisms and within most major lineages of eukaryotes (Figure 1.1). Eukaryotic parasitism is, therefore, a by-product of independent convergence. In fact, transition to parasitism as a trophic lifestyle is a major evolutionary shift within eukaryotes and perhaps the most successful life strategy on Earth, with more parasites populating different ecosystems than free-living species and for every potential organism to serve as a host (Adl et al., 2019; Poulin and Randhawa, 2015).

Many previous notions, such as the now-dismantled Archaezoa Hypothesis first proposed by Cavalier-Smith in 1983, presumed numerous protist parasites to be primitive eukaryotes (Cavalier-Smith, 1989). The Archaezoa Hypothesis was an early model for eukaryogenesis that placed highly reduced and once-thought-to-be 'amitochondriate' parasitic lineages such as *Giardia*, *Trichomonas*, *Entamoeba*, and microsporidia as an early-branching monophyletic group of organisms that branched off before the endosymbiotic event that gave rise to mitochondria-possessing eukaryotes (Cavalier-Smith, 1989, 1993). The absence of canonical mitochondria paired with a high degree of genomic and protein sequence divergence resulted in incorrect phylogenetic placements of these lineages and assumptions about the evolution of eukaryotic protist parasitism (Keeling, 1998). With the continual discovery of new free-living protist lineages combined with advancements in sensitive phylogenomic methods and computational models to assess protein and nucleotide evolution, the modern view of the road to protist parasitism is one that has been traveled in a multitude of ways and independently across different eukaryotic supergroups (Poole & Penny, 2007) (Figure 1.1).

Figure 1.1. A schematic tree depicting eukaryotic relationships and the evolution of parasitism within major taxonomic groups. Ten consensus supergroups (*i.e.*, Opisthokonta, Amoebozoa, Metamonada, Discoba, Archaeplastida, Cryptista, Hemimastigophora, Haptophyta, CRuMs (collodictyonids, rigifillids, and mantamonadids), and TSAR (Telonemia, Stramenopila, Alveolata, and Rhizaria)) are depicted and highlighted in different colours. Most supergroups can be subdivided into additional phyla, which consist of both parasitic and free-living members. A selection of those is shown here. These relationships and terminology are based on the most recent nomenclature and phylogenetics determined by Adl et al. (2019) and Burki et al. (2020). This figure was reproduced from Rueckert et al. (2019) and modified for the inclusion of several additional supergroups.



Many primeval evolutionary assumptions also incorrectly categorized protists as a single group of organisms basal to multicellular eukaryotes (Rothschild, 1989; Scamardella, 1999). Instead, protists comprise a significant portion of the eukaryotic diversity on Earth and fall roughly within ten major 'supergroups'. These are Archaeplastida, TSAR (Telonemia, Stramenopila, Alveolata, and Rhizaria), Amoebozoa, Opisthokonta, Haptista, Cryptista, CRuMs (collodictyonids, rigifillids, and mantamonadids), Hemimastigophora, Discoba, and Metamonada, where a number of these comprise both multicellular and unicellular members (Adl et al., 2019; Burki et al., 2020) (Figure 1.1). The advent of better sampling and improved metagenomics and phylogenomics tools have paved the way for the discovery of new lineages and permits regular additions and revisions to these taxonomic classifications (Burki et al., 2020). To better appreciate the evolution of protist parasitism across the eukaryotic tree of life, the following sections will discuss examples of parasites from six of these supergroups responsible for causing human diseases and are well-known to clinicians and biologists alike. By no means is this a comprehensive list, but instead, these examples serve as illustrative representatives that have notable evolutionary histories. Many of these arose from free-living ancestors, highlight trends of secondary cellular streamlining, and are shaped by an accrument of parasite-specific innovations.

1.2.1 Apicomplexan parasites – Parasitism within TSAR

Within TSAR, apicomplexan parasites, which belong to the phylum Alveolata, are a group of spore-forming intracellular pathogens characterized by the presence of a relic plastid known as the apicoplast. They are best known for causing numerous well-recognized diseases that have high morbidity and mortality in human and animal populations across the globe and, as a result, also have lasting socioeconomic impacts in resource-poor settings. Key examples of human-infecting apicomplexan parasites are *Toxoplasma gondii*, *Plasmodium* spp. and *Cryptosporidium* spp.

Toxoplasma gondii is responsible for the foodborne illness Toxoplasmosis, a disease that has a high global prevalence with seropositivity rates in human populations ranging between 30% to 90%, according to epidemiological surveys conducted in Central America, South America, Europe, and the United States (Dubey & Jones, 2008; Wilking et al., 2016). The parasite lifecycle is complex involving definitive and intermediate hosts. In humans, pathogenesis begins when an oocyst is transmitted from a feline to a human host to develop acute or latent toxoplasmosis caused by the parasite tachyzoite or bradyzoite stages, respectively (Blader et al., 2015; Dubey et al., 1998). While most immunocompetent individuals do not suffer from serious disease outcomes, the infection can be especially problematic during pregnancy and can result in a maternal transmission to the developing fetus where severe complications in the newborn infant, such as congenital toxoplasmosis, can manifest (McAuley, 2014).

Perhaps the best-known parasitic infections are those caused by the *Plasmodium* spp., which are responsible for the highly prevalent vector-borne disease Malaria. The World Health Organization's World Malaria Report estimates the annual burden of disease to be 229 million global infections and 409,000 deaths (World Health Organization, 2020). *Plasmodium* has a temporally regulated lifecycle in the human

host that involves infection of the hepatocytes and red blood cells, each consisting of different parasite forms (e.g., merozoites, sporozoites, and gametocytes) (Phillips et al., 2017). This complexity paired with continuous surface antigen variation adapted for immune evasion allows for rapid parasite-specific microevolution at the genetic and protein-complement level (Recker et al., 2011).

Lastly, the enteric *Cryptosporidium* spp. are responsible for the gastrointestinal, or less frequently respiratory, disease cryptosporidiosis, which makes up approximately 20% of all diarrheal infections in children in developing countries such as India, China, and Nigeria, where the burden of disease is high (Shrivastava et al., 2017). Unlike *Toxoplasma* or *Plasmodium*, the parasite lifecycle is comparatively simpler requiring only a single host for transmission. However, like all apicomplexans, within the human host, *Cryptosporidium* has three developmental stages consisting of meronts, gamonts, and oocysts, where the latter is necessary for a host to host transmission and can survive prolonged periods of environmental stress (Current and Garcia, 1991).

The path to parasitism within Apicomplexa has occurred through numerous evolutionary courses, including parallel evolution, as has become evident through continual single-celled genomics, transcriptomics, and phylogenomics with newly discovered apicomplexan-like and gregarine lineages (Mathur et al., 2019, 2021). Most notably, however, apicomplexans have their roots from free-living ancestors. Environmental sampling and genome sequencing efforts have characterized free-living photosynthetic chromerid algae as basal relatives of the apicomplexan parasites (Moore et al., 2008; Oborník et al., 2011; Woo et al., 2015) (Figure 1.1). Comparative genomics between the free-living algal lineages and apicomplexan parasites uncovered progressive losses to have occurred in the cellular machinery necessary for metabolism and homeostasis, photosynthesis and sterol biosynthesis, cargo trafficking, and motility, as the transition to an obligate lifestyle ensued (Woo et al., 2015). Aside from general overall cellular streamlining, their successful intracellular lifestyle is also shaped by novel lineage-specific innovations. A notable example in *Plasmodium* is the neogenesis of *Plasmodium*-specific efflux pumps (e.g., SERCA-type ATPase and Chloroquine-Resistance Transporter (*PfCRT*)). These are expressed by the intraerythrocytic parasites to facilitate ion transport in the parasite-induced parasitophorous vacuole for survival and to confer antimalarial drug resistance (Chinappi et al., 2010; Shandilya et al., 2013).

Altogether, reduced cellular complexity paired with parasite-specific novelties allowed for evolution towards successful pathogenic life strategies within Apicomplexa from their early origins as free-living and benign marine algae.

1.2.2 Trypanosomatids and *Naegleria fowleri* – Parasitism within *Discoba*

Parasites belonging to the Euglenozoa and Heterolobosea, which fall within the *Discoba* supergroup, are etiological agents for numerous fatal neglected tropical diseases that have also hampered economic development in Sub-Saharan and South American rural communities (Figure 1.1). Within Euglenozoa, the blood-borne trypanosomatid parasites, *Trypanosoma cruzi*, *Trypanosoma brucei*, and

Leishmania spp. cause Chagas' Disease, African Sleeping Sickness, and Leishmaniasis, respectively. Meanwhile, the heterolobosean soil and freshwater-dwelling *Naegleria fowleri* causes primary amoebic meningoencephalitis, a fatal waterborne infection.

The unifying feature of the kinetoplastids is the presence of a ramified mitochondrial organelle known as the kinetoplast that contains kDNA, a network of circular DNA complexed in the form of minicircles and maxicircles (Lukeš et al., 2002; Shapiro & Englund, 1995). Several first-line anti-trypanosomatid drugs, such as Suramin and Pentamidine, directly target this kDNA to hinder parasite metabolism, replication, and progeny growth (Alsford et al., 2012). Kinetoplastida is a subgroup that falls within the larger phylum Euglenozoa, which consists of parasitic (e.g., trypanosomatids), commensal (e.g., bodonids), and free-living (e.g., euglenids and diplomonads) genera (Simpson et al., 2006). The recent availability of genomes and transcriptomes from the free-living representatives (*Euglena gracilis*, *Bodo saltans*, and diplomonads) permitted comparative analyses with previously published genomes of the trypanosomatid parasites (Ebenezer et al., 2019; Jackson et al., 2016; Opperdoes et al., 2016; Záhonová et al., 2021). These studies provide evidence for streamlining and divergence in endomembrane systems, cytoskeletal and cellular structural components, and macromolecular degradation as a transition to parasitism occurred. Like Apicomplexa, trypanosomes encode parasite-specific expansions that are absent from the rest of their relatives. These are within amino acid transporters, mucins, cathepsins, cysteine peptidases, and surface glycoproteins, all of which are necessary to establish pathogenicity and evade the host immune responses (Berriman et al., 2005; Camacho et al., 2019).

Naegleria fowleri is the other major parasite within the Discoba lineage and belongs to the Heterolobosea, the second major phylum within this supergroup (De Jonckheere, 2011). Members of this lineage consist of enigmatic free-living organisms studied primarily for their unique shape-shifting amoeboid to flagellate morphologies (Tikhonenkov et al., 2019). Although there are several insect-infecting pathogens within Heterolobosea (e.g., *Vahlkampfia* sp.), *Naegleria fowleri* represents the only known opportunistic pathogen of humans or animals (Carter, 1970; Marciano-Cabral, 1988; Visvesvara et al., 2007). In fact, *N. fowleri* is also the only pathogenic member within the *Naegleria* genus, which contains over 47 species, including *Naegleria gruberi*, a protist model organism studied for its relatively complete cellular organization (Marciano-Cabral, 1988; Visvesvara et al., 2007). Therefore, parasitism in *N. fowleri* is a secondarily-derived lifestyle strategy (De Jonckheere, 2014; Fritz-Laylin et al., 2010). Parasite-specific innovations are also present in *N. fowleri*, which has an extended repertoire of pore-forming proteins, proteases, and transcription factors that likely play roles in brain tissue degradation and subsequent trophocytosis for fulminant neuropathogenesis (Herman et al., 2021; Liechti et al., 2019).

1.2.3 Entamoeba histolytica – Parasitism within Amoebozoa

Amoebozoa is a lineage consisting of amoeboid protists characterized by their pseudopodia membrane projections used for phagocytosis of food and prey (Adl et al., 2019; Smirnov et al., 2011). Members have diverse trophic lifestyles that range from free-living and commensal to parasitic (Cavalier-

Smith et al., 2015). *Entamoeba histolytica* is a popular representative and an intestinal microaerophilic parasite of humans responsible for causing amoebic dysentery (also known as amoebiasis) (Shirley et al., 2018). Amoebiasis is endemic in low socioeconomic regions that have limited access to clean water and de-sanitation infrastructures.

Entamoeba and other amoebozoan parasites, such as the opportunistic soil pathogens *Acanthamoeba castellanii* and *Balamuthia mandrillaris*, which are responsible for granulomatous amoebic encephalitis (GAE), follow a similar path to parasitism as apicomplexans and discobids (Clark et al., 2006). Comparative genomics with close free-living relatives, *Dictyostelium discoideum* and *Mastigamoeba balamuthii* illuminate reductive evolution within metabolic pathways, structural components, and endomembrane systems in *Entamoeba* and *Acanthamoeba* (Eichinger et al., 2005; Loftus et al., 2005; Shabardina et al., 2018; Žárský et al., 2021). The *Mastigamoeba* genome is also much larger compared to *Entamoeba*, where the latter has lost approximately 3000 conserved orthologs (Žárský et al., 2021). Aside from general reduction, *Entamoeba*'s protein repertoire also experiences numerous lineage-specific expansions, especially within membrane trafficking system proteins. The most notable example is the presence of an extensive Rab GTPase repertoire with the parasite encoding over 90 Rabs, many of which result from lineage-specific duplications (Saito-Nakano et al., 2005). Their roles within the *Entamoeba* cargo secretory pathways and how they contribute towards establishing pathogenicity are subject to ongoing investigations and are discussed in section 1.3.5.

1.2.4 Parasitism within Opisthokonta

A great diversity of human and animal-infecting parasites occurs within Opisthokonta. Opisthokonta can roughly be divided into Holomycota, which consists of fungi and fungi-like lineages, and the Holozoa, comprising animals and their single-celled relatives (Adl et al., 2019) (Figure 1.1). The evolution of parasitism within this group is vastly complex and deserves a separate comprehensive discussion of its own, which is out of scope for this chapter and thesis. Nonetheless, it is still necessary to briefly recognize key examples with clinical implications or fascinating evolutionary histories.

Within Holomycota, the dominating group of pathogenic fungi are opportunistic yeasts but also include the once-misclassified microsporidia. Fungal pathogens are responsible for immense global health burdens, agricultural damage, and overall loss of life (see Brown et al., 2012 and Kim, 2016 for review on this topic). Pathogenesis within fungi is multiply-derived and has been a product of vertical and convergent evolution (Moran et al., 2011; Taylor, 2014). Prominent examples include *Candida* spp. where many species are considered as commensals, while some are opportunistic pathogens (Brown et al., 2012; Moran et al., 2011). *Candida albicans* is a well-known opportunistic yeast that causes Candidiasis infections that can remain superficial (*i.e.*, vaginal candidiasis, thrush, and onychomycosis) or become systemic or invasive (Candidemia sepsis) (Achkar & Fries, 2010; Nobile & Johnson, 2015). *Cryptococcus* spp. represent another recognizable group of opportunistic pathogens that cause cryptococcosis, fatal infection of the lungs and the brain, which is especially problematic in HIV/AIDS patients and immunocompromised

individuals (Maziarz & Perfect, 2016). Like *Candida* and *Naegleria*, numerous species have been recognized as soil-dwelling benign yeasts. However, only four have been implicated in human disease (*C. neoformans*, *C. gattii*, *C. albicans*, and *C. uniguttulatus*), suggesting pathogenicity to be a secondarily and multiply derived trait within this genus (Cogliati, 2013).

The enigmatic spore-forming microsporidia is another group of fungal parasites and hyperparasites (*i.e.*, parasites of parasites) that infect all major groups of animals, including humans, where they cause microsporidiosis. This opportunistic infection results in diarrheal, lung, kidney, and keratitis outcomes in the immunocompromised and HIV/AIDS patients (Didier, 2005). So far, at least 15 different genera have been implicated in causing the human disease (Didier & Weiss, 2006). Historically, their streamlined biology (*i.e.*, lack of canonical mitochondria and endomembrane organelles), paired with bacterial characteristics (*i.e.*, ribosomes with low sedimentation coefficients) and highly compact and divergent genomes characterized these as an early-emerging clade from other eukaryotes (Cavalier-Smith, 1993; Keeling & Fast, 2002; Keeling & Slamovits, 2005; Vossbrinck et al., 1987; Williams et al., 2002). However, since then, more robust phylogenetic and phylogenomic assessments have redefined their taxonomic position within the tree of life to be sister to modern Fungi (Capella-Gutiérrez et al., 2012; Gill & Fast, 2006; James et al., 2013; Keeling et al., 2005; Thomarat et al., 2004).

Within Holozoa, polyphyletic metazoan parasites comprising platyhelminths, cestodes, trematodes, nematodes, acanthocephalans, and crustaceans are amongst the best known to cause superficial and systemic human infections, which are especially concerning in resource-poor settings of the world. Recent characterizations and revised taxonomic classification have also implicated single-celled holozoan species in causing human diseases. A prominent example includes *Rhinosporidium* sp., which is implicated in causing rhinosporidiosis, a rare but chronic granulomatous human disease of the nasopharyngeal tract and the genitalia (Arseculeratne, 2005; Herr et al., 1999; Reynolds et al., 2017). This parasite was once considered to be fungi-like, but revised phylogenetic placements groups it within the Ichthyosporea, a basal group of unicellular organisms closely related to sponges and animals.

1.2.5 Prototheca – Algal parasitism within Archaeplastida

Archaeplastida is a major supergroup well-known for its great diversity of photosynthetic organisms (Figure 1.1). Until recently, it was considered that human parasites did not originate from this lineage. However, improved phylogenetics placed the reduced opportunistic and chlorophyll-lacking *Prototheca* spp., parasites within this lineage. *Prototheca* is responsible for the disease protothecosis (*syn.* algaemia), a rare infection that can result in the development of cutaneous lesions, Olecranon bursitis, or systemic infections in both immunocompetent and immunocompromised individuals. They were once thought to be related to fungi due to their yeast-like morphology, but the availability of genomic data revised their placement to instead be within Chlorophyta, a green algal phylum within Archaeplastida (Kano, 2020; Yan et al., 2015). Secondary reductive evolution in its plastid organelles and genome compactness have shaped

the transition from a photoautotrophic to an obligate heterotrophic lifestyle, a scenario that draws parallels to the path taken by apicomplexans (Yan et al., 2015).

1.2.6 *Giardia* and *Trichomonas* – Parasitism within Metamonada

Representative metamonad parasites are well-known to clinicians, molecular cell biologists, and evolutionary biologists. These are *Trichomonas vaginalis* and *Giardia intestinalis* (Figure 1.1). Both are microaerophilic flagellated parasites devoid of canonical mitochondria, and like microsporidia, once classified within the primitive Archaezoa kingdom.

Trichomonas vaginalis is the etiological agent for human and animal trichomoniasis, the most common non-viral sexually-transmitted infection with a reported incidence of 276 million annual cases (Kissinger, 2015). Both males and females are affected and, if left untreated, can cause many chronic secondary complications (Kissinger, 2015). These can be increased susceptibility to other types of genitourinary infections (e.g., bacterial vaginosis), risk of cervical cancer, infertility in men, complications during pregnancy, and congenital disabilities in infants (Rathod et al., 2011; Tao et al., 2014; Wiringa et al., 2019; Zhang & Begg, 1994). *T. vaginalis* infects the human urogenital tract and can exist in an amoeboid or flagellated form (Mercer & Johnson, 2018). Within Metamonada, *Trichomonas* belongs to the lineage Parabasalia which additionally consists of representatives with free-living (e.g., *Monotrichomonas*, *Honigbergiella*, and *Anaeramoeba*) and symbiotic lifestyles (e.g., hypermastigotes) (Cepicka et al., 2010; Keeling et al., 2005; Malik et al., 2011; Noda et al., 2012; Ohkuma et al., 2005). Unlike most lineages where the transition to a parasitic lifestyle is defined by secondary cellular reduction, *Trichomonas vaginalis* is unique in this case, where extraordinary paralogous expansions in gene families belonging to many of its cellular systems have enabled its virulence. Though there are notable molecular absences in comparison to its free-living relatives, such as the lack of a glycosylphosphatidylinositol (GPI) anchor biosynthetic pathway necessary to produce GPI-anchors which are essential constituents of the eukaryotic plasma membrane, the *Trichomonas* genome is sculpted by high levels of transposable elements and bacterial lateral gene transfer (Carlton et al., 2007; Hirt et al., 2007; Singh, 1993). A highly expanded protein-coding repertoire in genes belonging to the membrane trafficking system, transporter families, kinases, and surface adhesion proteins is also evident (Carlton et al., 2007). These are integral for *Trichomonas* cytoadherence to the host epithelia and for phagocytosis and tissue destruction.

Apart from *Trichomonas*, the other best-known member within Metamonada is *Giardia intestinalis*, a diplomonad parasite within the subphylum Fornicata and a causative agent for the food and waterborne-illness Giardiasis. Compared to other eukaryotes, this parasite has a significantly reshaped cellular landscape, especially its endomembrane system (McCaffery & Gillin, 1994). *Giardia* serves as an interesting model for understanding the molecular evolution that follows a path to parasitism and adaptation to a host-dependent lifestyle. This thesis takes a close look at the mode and tempo of these processes, specifically within the membrane trafficking system of this parasite. The following section will highlight the diversity of eukaryotic trafficking processes and their evolution within some of the abovementioned

parasites. Section 1.4 and onwards is devoted to a thorough discussion on the clinical manifestation and epidemiology of Giardiasis, known aspects about *Giardia*'s endomembrane organization and trafficking processes, its phylogenetic relationship within the Fornicata and ecological niches of the comprising members, and diversity of the *Giardia intestinalis* species itself. Each of these aspects will signify the need to investigate this parasite's trafficking system from a multitude of perspectives and scientific approaches.

1.3 A panoramic tour of the endomembrane system

As compared with their prokaryotic relatives, eukaryotes are uniquely defined by the presence of a comparatively complex cellular system. One of these chief salient features is the compartmentalization of cytoplasmic content into distinct membrane-bound organelles that arose by curving membranes to efficiently separate cellular functions (Blobel, 1980). Together with the protein machinery that governs their biogenesis and function, these endomembrane organelles comprise the eukaryotic membrane trafficking system (MTS) (Figure 1.2A). In general, the MTS is a functionally interconnected highway within the cell that synchronously performs functions of bi-directional transport of cargo between the external environment and the intracellular milieu through organellar co-participation, a concept first integrated by Morr  and Mollenhauer (Morr  & Mollenhauer, 1974). Within human cells, breakdown of these processes leads to diverse array of neurodegenerative diseases (e.g., Parkinson's disease, Alzheimer's disease, Huntington's disease, Charcot-Marie-Tooth disease, Pettigrew syndrome), multi-systemic disorders (e.g., acute myeloid leukemia, Griscelli syndrome, Hermansky-Pudlak syndrome), cancer, and diabetes, to name a few (see review by Yarwood et al., 2020). The biomedical importance of the membrane trafficking system is paramount. It has garnered the interest of cell biologists for many decades in order to understand the mechanistic intricacies that underpin human diseases that arise as a result of defects in cargo transport pathways. From an evolutionary perspective, understanding the diversity and organization of this system across the breadth of extant eukaryotes from different supergroups can help map the time points at which those various complexities arose during eukaryogenesis approximately 1.6 billion years ago.

When examining eukaryogenesis, the engulfment of an α -proteobacterium (and its eventual evolution into a modern mitochondrion) by an Asgard archaeal host is one influential event that defines this evolutionary transition to yield complex cells (Eme et al., 2017; Mu oz-G mez et al., 2019; Sagan, 1967; Sicheritz-Pont n et al., 1998). Others include primary and serial acquisition of green and red plastids for the evolution of photoautotrophic eukaryotes (Archibald, 2015; Gray, 1992; Zimorski et al., 2014). Both scenarios required endosymbiosis and endosymbiotic gene transfer (EGT) between an archaeal host and a bacterium, which gave rise to existing metabolic complexities observed in extant eukaryotes (Martin et al., 1993; Spang et al., 2019; Stiller, 2007; Timmis et al., 2004). However, unlike endosymbiosis, compartmentalization and origins of the endomembrane system have autogenous origins, which is explained by the organelle paralogy hypothesis (OPH) (Dacks et al., 2008; Dacks & Field, 2007). As first postulated by Dacks and Field, OPH is a mechanistic model that hypothesizes that the existing complexity within the membrane trafficking machinery is a by-product of paralogous expansions and

neofunctionalization of protein families present in the proto-eukaryotic cell (Dacks & Field, 2018). These gene duplications and their subsequent accumulation of new functions, in turn, may yield new organelles or organellar roles (Dacks et al., 2016; Dacks & Field, 2018; Mani & Thattai, 2016). By this virtue, the time points at which these duplication events took place can then be used to recapitulate the evolutionary histories of extant endomembrane organelles. The OPH has been used to describe the complexity and evolution of numerous trafficking machinery that otherwise function at distinct endomembrane interfaces (e.g., proto-coatomer coat complexes, Rab GTPases, and SNAREs) (Dacks et al., 2008; Dacks & Robinson, 2017; Elias et al., 2012).

The Nobel prize-winning work put forward by cell biologists Palade, Rothman, Schekman, and Südhof, established that the eukaryotic trafficking system facilitates bi-directional transport of macromolecules via exocytic (*i.e.*, secretion of material outside of the cell) and endocytic (*i.e.*, cellular uptake of material) pathways (Balch et al., 1984; Jahn & Südhof, 1999; Novick et al., 1980; Palade, 1975; Rothman, 1994) (Figure 1.2A). Secretion and cargo uptake require vesicle carriers whose kinetics, directionality, and fidelity are regulated by distinct sets of protein machinery. Their functions can be divided into two categories: vesicle formation and vesicle fusion (see Bonifacino & Glick, 2004 for a thorough review). A synchronized association between both types of machinery is pertinent to ensure correct trafficking dynamics. The entire process is initiated upon cargo recognition, which results in adaptor and coat protein recruitment for nucleation at donor membranes (Figure 1.2B). These then facilitate membrane budding and scission to form cargo-containing vesicles that are transported to their destination, where the transport material is released upon fusion with the donor endomembrane compartment (Figure 1.2B). This final process is orchestrated by tethering and fusion proteins (Bonifacino & Lippincott-Schwartz, 2003; Kirchhausen, 2000; Wickner & Schekman, 2008) (Figure 1.2B). In most eukaryotic cells, the major endomembrane organelles comprise the endoplasmic reticulum (ER), the Golgi apparatus and the *trans*-Golgi network, distinct endosomal compartments (early, late, recycling), lysosomes and lysosome-related organelles, and multivesicular bodies (MVBs) (Figure 1.2A). Although not technically an organelle, the plasma membrane is also a vital member of this system. Others include autophagosomes and peroxisomes, which are also of equal importance but are seldom involved in the active transport of material to other compartments within the cell. Instead, they are terminal destinations for material degradation. Emerging evidence from studies conducted in many eukaryotes prompt the inclusion of lipid droplets as *bona fide* endomembrane organelles that serve as dynamic reservoirs of fat and act as regulators of cellular metabolism, signaling, and buffers for oxidative stress and lipotoxicity (Lundquist et al., 2020; Olzmann & Carvalho, 2019; Welte & Gould, 2017). The discovery of new eukaryotic lineages has also shed light on the existence of several other enigmatic membrane-bound organelles (e.g., mucocysts in ciliates, acidocalcisomes in trypanosomatids, and contractile vacuoles in a variety of freshwater and marine protistan taxa). However, the scope of discussion in this chapter will be limited to the best-conserved trafficking organelles and known cargo transport processes at each that are mediated by the vesicle formation and fusion machinery. Only machinery whose roles have been comprehensively investigated

through molecular functional approaches for their specific mechanisms and uniformly distributed across most major taxonomic groups will be outlined. Comparative genomics and phylogenetic investigations with MTS proteins across the eukaryotic diversity have allowed to recapitulate the ancestral complement and diversity of trafficking processes that likely existed within the last eukaryotic common ancestor (LECA). As a result, the expectation is that these will be conserved in a given eukaryotic lineage as the starting null hypothesis and, if absent, is reasoned as a secondary loss. Based on this notion, the LECA-specific complement of trafficking proteins will specifically be assessed for their conservation in *Giardia* and its relatives in the subsequent chapters.

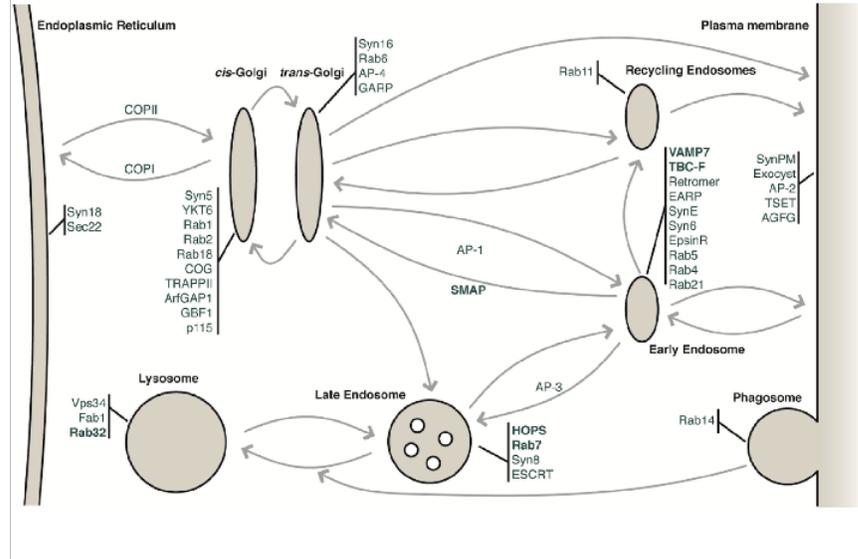
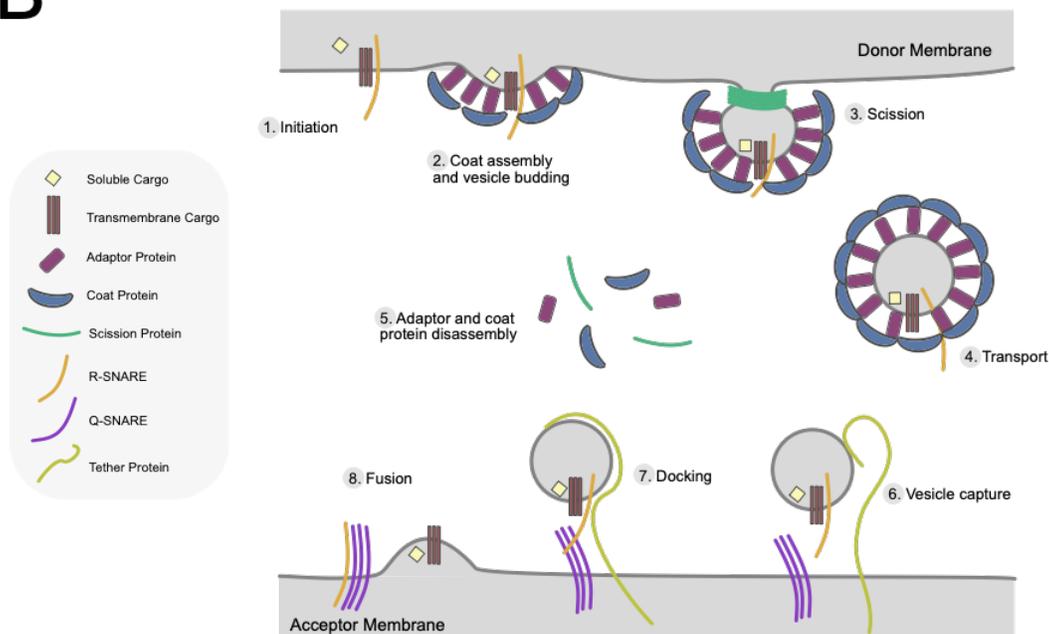
A**B**

Figure 1.2. Overview of the eukaryotic membrane trafficking system and vesicle budding and fusion processes. (A) depicts endomembrane compartments that mediate canonical secretory and endocytic processes. The diversity of cargo trafficking proteins necessary for vesicle formation and fusion processes between these different organelles are also shown here. This panel was reproduced from Karnkowska et al. (2019). (B) illustrates the specific vesicle budding, scission, and fusion steps at donor and acceptor compartments. Cargo-containing vesicles are formed by protein complexes belonging to the vesicle formation machinery that first recognize and bind soluble or transmembrane cargo, which results in the assembly of adaptor and coat proteins at donor membranes for membrane deformation (steps 1, 2, and 3). Vesicle budding takes place with the help of scission proteins and GTPases that catalyze this process, after which the coat proteins disassemble and are recycled back into the cytoplasm (step 4). The mature vesicle traverses to its final destination, where tethering proteins and partner SNAREs bind to one another for vesicle fusion with the acceptor compartment to release the enclosed cargo (steps 5,6,7). The figure is adapted from Bonifacino and Glick (2004).

1.3.1 Endoplasmic reticulum and the trans-Golgi network – early secretory pathway

Biogenesis of proteins targeted for extracellular secretion or integration into the plasma membrane occurs at the endoplasmic reticulum (ER), an interconnected compartment consisting of membranous cisternae that emerge from the nuclear envelope (Palade, 1975; Redman et al., 1966). The ER can be sub-compartmentalized into the rough ER (RER) and the smooth ER. In many cells, the smooth ER is a site for lipid biosynthesis, detoxification processes, carbohydrate metabolism, and intracellular calcium storage (Fagone & Jackowski, 2009; Olzmann & Carvalho, 2019; Vidugiriene & Menon, 1993). The RER is a dynamic structure ornate with ribosomes, where mRNA transcripts of proteins with secretory and endocytic fates are further translated and folded (Braakman & Hebert, 2013; Palade, 1975; Reid & Nicchitta, 2015). Briefly, mRNA translation and polypeptide synthesis begin in the cytosol with the help of cytosolic ribosomes. A signal recognition particle (SRP) is then recruited and recognizes signal peptide sequences located at the amino terminus of nascent polypeptides to form a complex consisting of mRNA, ribosome, SRP, and the nascent polypeptide (Braakman & Hebert, 2013). This complex is targeted to the ER membranes, where the remainder of the translation is completed. Proteins destined for extracellular secretion are subject to additional post-translational modifications, such as N-linked glycosylation and the formation of disulfide bonds (Aebi, 2013). Families of resident ER lumen chaperones and co-factors such as heat shock proteins (HSPs) and binding immunoglobulin protein (BiP) perform the final quality control to ensure maturation and correct protein folding prior to further transport (Ellis et al., 1993). In many organisms, the ER has distinct domains such as the mitochondria-associated membranes (MAMs) required for mitochondrial tethering and sites for biogenesis of lipid droplets and peroxisomes (Glick, 2014; Joshi et al., 2018; Lynes & Simmen, 2011; Ohsaki et al., 2017; Raychaudhuri & Prinz, 2008; Vance, 2014). Importantly, the ER also has specialized regions known as the transitional ER (tER), where coatamer protein complex II (COPII) assembly for vesicle biogenesis and anterograde transport of newly synthesized material to the Golgi takes place (Bannykh et al., 1996; Palade, 1975) (Figure 1.2A) (COPII assembly at the ER is discussed in detail in section 1.3.3).

The Golgi apparatus, named after its discoverer Camillo Golgi, is a critical endomembrane compartment where post-translational modification of proteins through glycosylation and phosphorylation occurs (Drouin et al., 2015; Rothman & Fine, 1980). Like the ER, the Golgi is often a distinguishable compartment defined by the presence of multiple cisternae in many organisms, however, it can also exist as inconspicuous morphologies (*e.g.*, in *Saccharomyces cerevisiae*, *Plasmodium falciparum*, and *Entamoeba histolytica*) (Bredeston et al., 2005; Hallée et al., 2018; Herman et al., 2018; Struck et al., 2005; Tamaki & Yamashina, 2002) (Figure 1.2A). Whether Golgi stacking is a multiply-derived convergent trait or a feature present in the common ancestor of eukaryotes has been subject to comprehensive comparative genomic investigations using mammalian Golgins and stacking factors involved in the reassembly and maintenance of the cisternae (Barlow et al., 2018). Based on the conservation in the repertoire of Golgi-associated proteins across 75 different taxa, suggest LECA to have possessed a complex and a differentiated Golgi, but that mechanism of cisternal organization may be an emergent property (Barlow et

al., 2018). In organisms with a stacked Golgi, internal protein trafficking within the Golgi occurs by way of cisternal maturation to yield three distinct phases: *cis*, *medial*, and *trans* (Papanikou & Glick, 2014). The *cis* phase is closest to the ER and traffics cargo to and from this organelle. In contrast, at the *trans* phase, the *trans*-Golgi network (TGN) presides as a network of tubules and vesicles and is a critical juncture for anterograde and retrograde trafficking of material to and from distal endo-lysosomal organelles (Papanikou & Glick, 2014).

Carrier vesicles mediate anterograde transport of proteins from the ER to the *cis*-Golgi with the help of the COPII proteins (Sato & Nakano, 2007). COPII-coated vesicles bud from the ER membranes, which then tether and fuse with the *cis*-Golgi lumen (Sato and Nakano, 2007). Material with extracellular fates, other intracellular destinations such as endosomes, lysosomes, and multivesicular bodies, or for retrograde transport back to the ER are packaged in clathrin-coated vesicles (CCV) at the TGN (Guo et al., 2014). CCV formation occurs with the help of evolutionarily-related heterotetrameric adaptor complexes (HTACs) that consist of coatamer protein complex I (COPI), five adaptins (AP1-5), and TSET (Hirst et al., 2011, 2014). COPII and HTAC components possess α -solenoid and β -propellor secondary structural configurations, and so together with other complexes with similar architectures, such as the nuclear pore complex, SEA complexes, components of the multi-subunit tethering complex, intraflagellar transport proteins, and retromer, are thought to have proto-coatomeric origins (Dacks & Field, 2007; Devos et al., 2004; Field et al., 2011; Koumandou et al., 2011; Rout & Field, 2017). The OPH postulates ancient pre-duplicate proto-coatamer components to have been present in the Asgard archaeal host and the first eukaryotic common ancestor, which likely gave rise to the existing repertoire of proteins detailed above (Dacks & Field, 2018). Individual components of the proto-coatomeric vesicle formation machinery and the detailed mechanisms by which they assemble for vesicle formation dynamics are discussed in greater detail in section 1.3.3.

1.3.2 Post-Golgi endo-lysosomal trafficking

Post-Golgi, the plasma membrane and the endo-lysosomal organelles constitute a highly dynamic tubulovesicular network that mediates both exocytic and endocytic processes (Figure 1.2A). It comprises numerous biochemically and morphologically distinct organelles that collectively facilitate a vectorial transfer of material from one compartment to the next while preserving organellar identity (Rothman, 1994). The best characterized endo-lysosomal compartments are the early endosomes, late endosomes/multivesicular bodies, recycling endosomes, lysosomes, and lysosome-related organelles (Huotari & Helenius, 2011; Schmid et al., 1988).

Typically, material enters the cells through clathrin-mediated endocytosis (CME) in primary vesicles that fuse and mature into early and recycling endosomes (Kaksonen & Roux, 2018). Internalized receptors are sorted and recycled back to the plasma membrane, whereas other cargoes are targeted for sorting at the *trans*-Golgi network or breakdown at the acidic lysosomes (Heuser & Reese, 1973). Endosomes undergo maturation and a series of fusion events through “kiss-and-run” interactions and are marked with

specific phosphatidylinositol phospholipids (PIPs) for targeting to different intracellular destinations (Corvera et al., 1999; Huotari & Helenius, 2011; Kjaerulff et al., 2002). Increasing evidence suggests that these PIPs and other lipid profiles, which change and differ between distinct endosomal stages, are recognized by specific vesicle formation and fusion proteins that recruit additional adaptors and accessory machinery (Whitley et al., 2003). These recruitment cascades dictate the maturation and directionality of an endosomal carrier. For example, adaptor protein complex 2 (AP-2) and dynamin recognize and bind phosphatidylinositol 4,5-bisphosphate (PI(4,5)P₂) in the plasma membrane to form early endosomes. FYVE (Fab1p, YOTBm Vac1p, EEA1) and Phox homology (PX) domain proteins preferentially recognize and bind to PI(3)P on early endosomes for their targeting to multivesicular bodies and lysosomes (Elkin et al., 2016).

Acidification and endosomal pH also influence cargo directionality. Endosomal pH of 6.5 is associated with recycling of transferrin and low-density lipoprotein receptors to the plasma membrane via their transport in recycling endosomes (Elkin et al., 2016). Endosomes destined for the TGN or downstream maturation as multivesicular bodies have lower pH (4.5-5.5) (Elkin et al., 2016). Apart from pH, the size and shape of endosomes also determine specific transport mechanisms at distinct cellular locations. The presence of sorting signals such as PX and Bin, Amphiphysin, and Rvs (BAR) domains on proteins can drive endosomal tubulation. These are present in the sorting nexin (SNXs) family of proteins of the retromer complex, which are discussed later in section 1.3.3. SNX and BAR domain-containing proteins recognize and interact with PI(3)P on late endosomes to recycle cargo, such as mannose-6-phosphate receptors, back to the plasma membrane or the TGN (van Weering et al., 2012). Finally, cargo sorting and endosomal maturation are also steered by reversible post-translational modifications such as ubiquitination. Proteins are conjugated with single (mono-ubiquitination) or chain (poly-ubiquitination) of ubiquitin molecules with the help of ubiquitin-activating (E1), -conjugating (E2), and -ligating (E3) enzymes (Pickart, 2001). Ubiquitin-tagged cargo recruit membrane deformation machinery such as the endosomal sorting complexes for transport (ESCRTs), which induce negative membrane curvature in early endosomes to form intraluminal vesicles (Raiborg & Stenmark, 2009). Assembly and membrane dynamics of the ESCRTs subcomplexes and their roles in multivesicular body (MVB) formation are discussed in detail in section 1.3.3 and Chapter 4.

Altogether, these biochemical and morphological configurations lead to the maturation and acidification of a primary vesicle or early endosome to yield morphologically distinct late endosomes/MVBs that fuse with lysosomes. Soluble acid hydrolases, which are active in the low pH lysosomal environment, mediate proteolytic breakdown of cargo (Ohkuma & Poole, 1978). Many eukaryotic lineages and mammalian cell types have derived forms of these organelles that perform similar functions and can exist in a myriad of morphologies. Collectively, they are referred to as lysosome-related organelles (LROs) (Bowman et al., 2019). Examples from mammalian systems include melanosomes, *Drosophila* pigment granules, and granules from polymorphonuclear leukocytes. In protists such as *Tetrahymena* and *Dictyostelium*, compartments such as mucocysts and post-lysosomes, respectively, are also postulated to

be LRO in nature (Bowman et al., 2019). In whatever shape they may be, it is nonetheless clear that lysosomes and LROs are important within eukaryotic cells as they maintain functions of autophagy, apoptosis, and general cellular homeostasis (Figure 1.2A).

1.3.3 Key vesicle formation machinery

The formation of vesicles requires a donor compartment and a suite of protein complexes recruited sequentially for the recognition and loading of cargo into nascent vesicles for transport to their final destination (Bonifacino & Glick, 2004) (Figure 1.2B). Previous molecular functional studies and detailed comparative genomic investigations have played fundamental roles in uncovering these protein functions and evolutionary distribution at distinct endomembrane interfaces. Evolutionary analyses have also illuminated the conservation in each machinery and what is likely a core functional unit that makes up each complex across most eukaryotes. The focus of this thesis is to investigate the evolution and role of the vesicle formation machinery in *Giardia*, and therefore, this section will familiarize the reader with the essential protein complexes that fall within this category and contextualize their importance for being chosen as targets for investigations in the later chapters.

As previously discussed in section 1.3.1, within the ER-Golgi pathway, COPII and coatamer complex I (COPI) are necessary for vesicle formation and cargo transport between the ER and the Golgi (Barlowe et al., 1994; Orci et al., 1986). Here it was also highlighted that the anterograde transport of proteins from the ER destined for further modification within the Golgi is COPII-dependent (Sato & Nakano, 2007) (Figure 1.2A). COPII is a multimeric complex consisting of seven subunits that assemble sequentially onto the transitional ER membrane to facilitate vesicle budding, cargo loading, and vesicle scission (Sato & Nakano, 2007) (See Figure 2.1 in Chapter 2 for the specific organization and location of this machinery). Evolutionary analyses have determined the ancestral LECA COPII subunits to consist of SAR1, Sec12, Sec16, Sec23, three paralogs of Sec24, Sec13, and Sec31, and therefore, these represent the core machinery likely to be present in most eukaryotes (Schlacht & Dacks, 2015) (Figure 2.1B). Generally, the process begins with the recruitment of the Sec12 guanine exchange nucleotide factor (GEF), which exchanges bound GDP to GTP on the small GTPase SAR1 for its activation and ER membrane association (Sato & Nakano, 2007). This triggers the assembly of the heterodimeric GTPase activating proteins, Sec23/Sec24, for cargo binding and organization of the heterotetrameric Sec13 and Sec31 to form a polyhedral cage complex and for membrane deformation and vesicle budding. Sec16 is an accessory protein that regulates the specific timing of the SAR1-GTP hydrolysis and is also thought to have additional scaffolding functions (Sato & Nakano, 2007). Altogether, this results in the formation of a COPII-coated vesicle, after which coat disassembly follows, and fusion of the newly freed vesicle occurs at the *cis*-Golgi with the help of SNARE complexes and other vesicle fusion machinery (e.g., multi-subunit tethering complex TRAPP1 and Rab1 GTPase)(Allan et al., 2000). Recent microscopic studies in mammalian systems have identified COPII to be forming ER membrane protrusions to drive tubulation along actin

filaments for cargo transport in conjunction with COPI (Weigel et al., 2021). These new findings redefine our understanding of COPII-mediated trafficking and provide new models for testing in other eukaryotes.

COPI is an evolutionarily conserved complex belonging to the paralogous family of heterotetrameric adaptor complexes (HTACs) that have proto-coatomeric origins (Devos et al., 2004; Schledzewski et al., 1999). COPI is required for vesicle formation and retrograde transport of material from the *cis*-Golgi to the ER/ER-Golgi intermediate compartments (ERGIC), as well as for intra-Golgi cisternal transport (Orci et al., 1986) (Figure 1.2A; Figure 2.1A). Like COPII, COPI assembly and kinetics are regulated by the small GTPase, ADP-ribosylation factor (ARF) 1. ARF1 activation is mediated by Sec7 domain-containing guanine exchange nucleotide factors (ARF GEFs) that exchange GDP to GTP, which results in downstream myristoylation of the N-terminal hydrophobic tail for membrane recruitment to initiate coat protein assembly (see Kahn, 2009 for a review on this topic). ARF GTPase activating proteins (ARF GAPs) hydrolyze the bound GTP back to GDP, which allows for cytosolic recycling of the ARF GTPase (a detailed discussion on the ARF cycle, the molecular complement of ARF GTPases, their regulators, and diverse functions in eukaryotic cells are presented in Chapter 3). In the case of COPI, ARF1 activity results in the assembly of COPI at the Golgi membranes (Beck et al., 2009). At the *cis*-Golgi, a specific ARF GEF, Golgi-specific brefeldin resistance guanine exchange nucleotide factor (GBF1), promotes the GDP to GTP exchange on the ARF1 GTPase (Niu et al., 2005). Activated ARF1 binds to the *cis*-Golgi membranes via the p24 coiled-coiled transmembrane protein to form a priming complex, which results in the recruitment and binding of COPI subunits. A combination of phylogenetic, comparative genomics, and molecular functional analyses have elucidated COPI to be a multimeric complex. It consists of an F-COP subcomplex that comprises β -COP, γ -COP, δ -COP, and ζ -COP, as well as accessory factors, ϵ -COP, α -COP, β' -COP, which have α -propeller and β -solenoid secondary structures that recognize cargo containing KKxx motifs (Beck et al., 2009; Hirst et al., 2011; Schledzewski et al., 1999) (See Figure 2.1B for the organization of the F-COP complex subunits). Stable association of COPI subunits results in a lattice formation followed by coating and budding of vesicles. Hydrolysis of the ARF1-GTP, catalyzed by ARFGAP1, mitigates COPI disassembly from the newly-budded vesicles (Bigay et al., 2005).

Apart from COPI, the heterotetrameric adaptor complexes also consist of adaptins (also known as adaptor protein complexes; APs) and TSET, critical for post-Golgi and endosomal vesicle formation processes (Hirst et al., 2014) (Figure 2.1A). The ARF GTPase regulatory family also controls the spatial and temporal recruitment and organizational dynamics of adaptins and clathrin, which in turn recognize Yxx Φ and dileucine motifs on cargoes destined for a specific secretory pathway (Sztul et al., 2019). So far, pan-eukaryotic comparative genomics have identified five ancestral adaptins (AP1-5), each of which is a heterotetrameric complex composed of two large (β and γ), medium (μ), and small (ζ) subunits that form the core complex, along with the accessory proteins, β' , α , and ϵ (Hirst et al., 2011) (Figure 2.1B). TSET, first described in plants as the TPLATE complex, is also conserved across many eukaryotes and considered to have been present in the LECA (Hirst et al., 2014). Like adaptins, TSET is also composed of a core complex consisting of two large (TPLATE and TSAUCER), a medium (TCUP), and a small subunit

(TSPOON), as well as accessory components with α -propeller and β -solenoid architectures (TTRAY1 and TTRAY2) (Hirst et al., 2014) (Figure 2.1B).

Although the HTACs share a similar structural composition, they are distinct in their cellular functions and localizations. Adaptors recognize cargo and membrane phospholipids such as PI(4,5)P2 to recruit clathrin triskelion components (clathrin heavy chain and clathrin light chain) and other adaptors to form clathrin-coated vesicles at various endo-lysosomal interfaces (Hirst et al., 2011) (Figure 1.2A; Figure 2.1A). AP-1 mediates clathrin-coated vesicle formation at the *trans*-Golgi network for trafficking to early/recycling endosomes (Kural et al., 2012) (Figure 1.2A; Figure 2.1B). AP-2 is required for clathrin-mediated endocytosis (CME) and cargo selection at the plasma membrane (Conner & Schmid, 2003) (Figure 1.2A; Figure 2.1A). In mammalian cells, AP-3 traffics cargo between the endosomes and the lysosomes, likely in a clathrin-dependent manner (Gupta et al., 2006) (Figure 1.2A; Figure 2.1A). Compared to AP1-3, AP-4 is still much more elusive but is postulated to have roles in retrograde trafficking between the early endosomes and the TGN, according to studies performed in *Arabidopsis thaliana* (Fuji et al., 2016). AP-4 also regulates auto-phagosomal trafficking of ATG9 proteins in mammalian systems (Mattera et al., 2017). The more recently described AP-5 and TSET complexes are enigmatic in their functions and still subject to ongoing cell biological investigations. Recent cellular localization studies point to AP-5 role within the late-endosomal to TGN pathway, whereas TSET is predicted to function in clathrin-mediated endocytosis at the plasma membrane, similar to AP-2 (Hirst et al., 2014, 2018) (Figure 1.2A; Figure 2.1A). Recent investigations in plants implicated TSET role in autophagosomal trafficking as well (Johnson et al., 2021; Wang et al., 2019; Yperman et al., 2021). As more functional studies shed light on these pathways, interchangeable versus exclusive roles of these complexes and their conservation within different eukaryotic endomembrane landscapes will become apparent.

Within the early endosomal pathway, another vesicle formation machinery required for cargo transport is the retromer complex, which is also postulated to have proto-coatomeric origins (Cullen & Steinberg, 2018). Retromer recycles mannose-6-phosphate receptors (or Vps10 in yeast) from endosomes to the TGN or back to the plasma membrane (Burd & Cullen, 2014; Marcusson et al., 1994; Seaman, 2004) (Figure 2.1A). Recent studies in *Drosophila* have also pointed to retromer function in the exocytosis of secretory granules in association with SNARE proteins (Neuman et al., 2021). It is a complex that comprises six evolutionarily conserved components that are subclassified to have membrane deformation, cargo selection, and receptor recognition functions (Cullen & Steinberg, 2018; Koumandou et al., 2011) (Figure 2.1B). Vps29, Vps26, and Vps35 have cargo selection functions; meanwhile, BAR domain-containing Vps5 (or sorting nexins, SNX1, SNX2, SNX5, and SNX6 in mammalian cells) are required for membrane-deformation (Cullen & Steinberg, 2018; Koumandou et al., 2011) (Figure 2.1A). In Opisthokonta, Vps17 is an additional component present that performs similar functions as Vps5 (Koumandou et al., 2011). Finally, Vps10 is often the cargo for retromer and a sorting receptor within the Golgi that sorts vacuolar protein carboxypeptidase Y (CPY) in yeast (Marcusson et al., 1994). Retromer assembly is concomitant to cargo recognition in early endosomes and is orchestrated by the trimer cargo-selection complex consisting of

Vps35, Vps29, and Vps26, which recognize Vps10 (Cullen & Steinberg, 2018). The trimer complex assembles in a Rab7- dependent manner and requires Tre-2/Bub2/Cdc16 (TBC) domain-containing GAPs, which regulate the Rab7 GTPase activity. Vps5 (or SNXs in mammals) also assembles upon recognition of early endosomal PI(3)P phospholipids through recognition via its BAR-PX domain in order to induce vesicle tubulation and cargo sorting into the recycling endosome (Seaman, 2004). Once this process is complete, the machinery dissociates and is recycled into the cytosol. The more elusive and recently described heterotrimeric retriever complex can also perform retromer-independent recycling of surface receptors. Retriever utilizes components similar to retromer, namely DSCR3, which is a paralog of Vps26, Vps29, and a Vps35-like protein (C16orf62) (McNally et al., 2017). Both retromer and retriever bear structural similarities and have identical protein-protein associations, such as with the WASH (WASP and Scar Homologue) complexes, which polymerize actin filaments in an Arp2/3-dependent fashion. Although DSCR3 is evolutionarily conserved, whether C16orf62 is present outside of mammalian systems is unclear, and therefore, additional evolutionary investigations are necessary to pinpoint the origins of retriever and its exact relationship to retromer (Chen et al., 2019; Koumandou et al., 2011).

Finally, the endosomal sorting complexes required for transport (ESCRTs) are a family of proteins implicated within the late endo-lysosomal pathway. ESCRTs, unlike the previously discussed machinery, do not have proto-coatomeric origins (Raiborg & Stenmark, 2009). Instead, subunits of this complex have arisen from progenitor SNF7 domain-containing protein, ESCRTI-Vps28, ESCRTII-Vps22/25, and ESCRTIII-Vps36-like proteins likely inherited from the host Asgard archaeal ancestor (Zaremba-Niedzwiedzka et al., 2017). ESCRTs are necessary for the maturation of early endosomes into multivesicular bodies (MVBs), a class of late endosomes characterized by the presence of numerous intraluminal vesicles (ILVs) that are 40-50 nm in size (Schmidt & Teis, 2012). This machinery comprises five multimeric subcomplexes: ESCRT0 (or the analogue Tom1-Esc), ESCRTI, ESCRTII, ESCRTIII, and ESCRTIII-A (Raiborg & Stenmark, 2009; Schmidt & Teis, 2012) (See Figure 4.1 for the organization of this machinery). Assembly of these subcomplexes occurs sequentially upon recognition of ubiquitin-tagged cargo, as well as endosomal phospholipid PtdIns(3)p, by the ESCRT0 subcomplex, which initiates this process (Raiborg & Stenmark, 2009; Schmidt & Teis, 2012) (Figure 4.1). In yeast, ESCRT0 consists of Vps27 and signal-transducing adaptor molecule (STAM), which recognize ubiquitin via the Vps27, Hrs, STAM (VHS) domain (Figure 4.1). Subsequently, ESCRTI, a heterotetrameric complex consisting of Vps23 (also known as TSG101), Vps28, Vps37, and Mvb12 (the latter present only in opisthokonts), recognizes ubiquitin and prosaposin motifs in ESCRT0-Vps27 for endosomal binding (Figure 4.1). This triggers cascade recruitment of ESCRTII, which consists of Vps36, Vps22, and Vps25 dimers, that function to bridge ESCRTI with ESCRTIII and III-A subcomplexes (Figure 4.1). ESCRTI-ESCRTII interactions occur between the Gram-like Ubiquitin-binding (GLUE) domain of Vps36 and the Vps28 carboxy-terminal domain (Figure 4.1). Vps36 also binds to ubiquitin on cargo through the GLUE domain and to the endosomal membrane lipids (Figure 4.1). Recruitment of ESCRTII results in the downstream assembly of ESCRTIII subcomplex, which comprises the paralogous SNF7 domain-containing family of proteins, Vps20, Vps32, Vps2, Vps24,

and CHMP7 (Figure 4.1). These trigger inward membrane deformation and abscission to form intraluminal vesicles. This subcomplex is recruited upon assembly and nucleation of Vps20, which interacts with ESCRTII-Vps25, resulting in an oligomeric assembly and filament formation of the Vps32 subunits that are capped off by Vps24 monomers (Figure 4.1). This filament constricts membrane invaginations into the endosome for ILV formation. Vps2 recognizes the Vps24 cap and facilitates the recruitment of the AAA-ATPase family protein, Vps4, which enters the pores to depolymerize and recycle the ESCRTIII subunits back into the cytosol, a process that relies on ATP hydrolysis. This latter recycling function is also performed by the ESCRTIII-A subcomplex consisting of Vps60, Vps46, IST1, and VTA1. Altogether, the assembly and inter-subunit interactions are critical for ILV formation for the maturation of early endosomes to multi-vesicular bodies.

1.3.4 Key vesicle fusion machinery

Although the focus of this thesis is on the vesicle formation machinery, it is important to briefly discuss the other arm of membrane trafficking proteins that perform vesicular tethering and fusion at acceptor membranes (Figure 1.2B). This is because the vesicle fusion machinery is integral to providing fidelity and spatiotemporal regulation of trafficking processes and dictates the evolutionary dynamics that define extant endomembrane complexity, as per the OPH (Dacks et al., 2008). Indispensable players that belong to this family include the multi-subunit tethering complexes, soluble N-ethylmaleimide-sensitive factor (NSF) attachment protein receptor (SNAREs) and Sec1/Munc-18 (SM) proteins, and the Rab GTPases and their regulators.

The MTCs are a large family of coiled-coil proteins necessary for the long-range capture of the vesicles by tethering donor and acceptor SNARE molecules for fusion with acceptor membrane (Bröcker et al., 2010) (Figure 1.2B). Evolutionarily conserved families present in most eukaryotes are CATCHR, HOPS/CORVET, and TRAPP complexes (Klinger et al., 2013).

Like the MTCs, SNAREs are a family of short coiled-coil transmembrane proteins that play fundamental roles in vesicle capture and fusion to the receiving organelle (Jahn & Scheller, 2006). Briefly, the process entails interaction between a v-SNARE (vesicle SNARE) and a t-SNARE (target SNARE) at the donor and acceptor membranes, respectively, to form a helical bundle complex that leads to a fusion event between the membrane bilayers belonging to the two compartments. N-ethylmaleimide-sensitive factor (NSF) then disassembles the SNARE helical bundle, a process that is ATP-dependent, for reuse for subsequent cycles of vesicle fusion (Bonifacino & Glick, 2004). Currently accepted nomenclature for the different v- and t-SNAREs classifies them into Q-SNAREs (Qa, Qb, Qc) or R-SNAREs based on the presence of glutamine or arginine, respectively, in the zero ionic layer (*i.e.*, the primary interaction site within the helical bundle) (Fasshauer et al., 1998). SNAREs themselves require other accessory proteins such as Sec1/Munc18-like (SM) proteins and α -SNAP, which stabilize the conformational dynamics of the helical bundle (Söllner et al., 1993; Tognneri et al., 2006).

Small GTPases, such as the Ras-like proteins from rat brain (Rab) GTPases, are essential catalysts of the vesicle fusion process within the cell (Stenmark, 2009). A given Rab cycle, comprising a distinct Rab and its partner GEF and GAP, defines the specificity and spatiotemporal regulation of various trafficking processes. The Rab GTPases are a member of the Ras superfamily, which consists of the previously discussed ARF1 and SAR1 but also ARF-like (Arl) small GTPases (Stenmark & Olkkonen, 2001). Rabs are extensively proliferated compared to ARFs and SAR1, which constitute only one or two members, where comparative genomic studies have reconstructed the LECA repertoire of Rabs to constitute at least 23 different proteins (Elias et al., 2012; Schlacht & Dacks, 2015; Vargová et al., 2021b). Like other GTPases, Rabs cycle between GTP- and GDP-bound states. Nucleoside binding induces conformational changes within the switch I and switch II regions of the protein (Stenmark, 2009). This dictates the GTPase in an 'on' or 'off' state for subsequent prenylation of the hypervariable Rab C-terminal lipid-tail cysteine residues, which results in effector-mediated recruitment onto membranes (Pereira-Leal et al., 2001). Rab GTPases cycle between membrane interfaces for vesicle docking and recycling back into a cytoplasmic pool upon enzymatic conformational changes. Rab binding to GTP or GDP is coordinated by GEFs or GAPs that cycle or hydrolyze the bound GTP, respectively. Of the identified and functionally best-understood RabGAPs are the Tre-2/Bub2/Cdc16 (TBCs) proteins, marked by the presence of a TBC domain (Pan et al., 2006). Phylogenetic investigations have identified at least ten ancient TBC RabGAP families across the eukaryotic diversity (Gabernet-Castello et al., 2013). RabGEFs, on the other hand, constitute TRAPP complexes, Mon1/Ccz1, Vps9 domain-containing proteins, DENN domain-containing proteins, and Ric1/Rgp1 (Cai et al., 2008; Delprato & Lambright, 2007; Hegedűs et al., 2016; Siniosoglou et al., 2000; Yoshimura et al., 2010).

1.3.5 Endocytic systems of key protist parasites and their roles in the pathogenesis

Due to the biomedical implications of the trafficking system, especially within the context of human health, these networks and their constituents have been an intense area of research for numerous decades in yeast, plants, and metazoan model systems. However, with the emergence of new molecular toolkits, previously unexplored avenues have opened up for parasitologists to probe and investigate these pathways in historically challenging to study protist parasites. This has allowed for the discovery of a tantalizing array of specialized endo-lysosomal systems and redefined our models by which trafficking processes can exist. Shedding light onto these systems has proven to be a cornerstone in our ability to assess mechanisms that underpin parasite pathogenicity. Many of these studies have demonstrated a central role of functional repurposing of traditional trafficking proteins at parasite-specific organelles that are otherwise involved in canonical secretion in model systems.

The best-studied parasitic endo-lysosomal systems are ones that have received considerable attention from biomedical and clinical perspectives. For example, research investigating the trafficking systems and organelles of the apicomplexan parasites has proliferated in the past few decades due to their roles in diseases that have tremendous global morbidity and mortality. Apicomplexans are characterized

by the presence of invasion organelles consisting of micronemes, rhoptries, and inner membrane complex, which are collectively termed the apical complex and used for host cell penetration and intracellular motility (Carruthers, 2002). Ongoing evolutionary and molecular functional investigations have illuminated these as endo-lysosomal secretory organelles (Klinger et al., 2013). Notably, these evolutionary connections, namely their characterization as lysosome-related organelles (LROs), can be made due to associations with specific endo-lysosomal markers such as the late endosomal Rab5 and Rab7, components of the retromer complex, and AP-1, all of which are postulated to be involved in trafficking at the micronemes and rhoptries (Kremer et al., 2013; Ngô et al., 2003). Genetic disruptions of few of these proteins caused aberrant microneme morphology, which corroborates these notions of functional homology to LROs (Sparvoli & Lebrun, 2021).

Like apicomplexan parasites, kinetoplastid *Trypanosoma brucei* has been subject to ardent molecular functional investigations of its trafficking system. Compared to other trypanosomatids, which are intracellular, *T. brucei* resides extracellularly and therefore is a tractable model for this group of parasites to study using *in vitro* and *in vivo* methods. Trypanosomatids heavily rely on secretory processes to express and undergo antigen switching of their variant surface glycoprotein coat (McConville et al., 2002). Comparative genomic and functional studies show expansions in families of exocytic SNAREs, namely Qb-NPSN and Qc-Syp7, and point towards a potential lineage-specific neofunctionalization of these paralogs for their role in VSP secretion (Venkatesh et al., 2017). *T. cruzi* and *Leishmania* species also possess contractile vacuoles and acidocalcisomes, which are thought to be LROs that facilitate osmoregulation and cation storage. Like the apical complex, these structures associate with numerous endo-lysosomal trafficking complexes such as Rab11, Rab2, Rab1, and SNARE-VAMP7 (Ulrich et al., 2011).

Pathogens, where trafficking plays a central role, are the amoebozoan parasites that rely on phagocytosis for overall cell survival. Of these, *Entamoeba* is the representative lineage where most of the transcriptomic and molecular functional studies have been performed and have highlighted the need for a whole array of trafficking complexes for pathogenesis (Herman et al., 2017). Phagocytosis in *Entamoeba* is vital for overall cellular homeostasis and nutrient exchange, colonization of the gut epithelium, evasion of the host necrotic factors, and erythrophagocytosis to establish dysentery (Boettner et al., 2008). The identified *Entamoeba* MTS is also necessary to the parasite as it transitions from trophozoite to cyst for subsequent propagation. Transcriptomic investigations showed upregulation in several Rab GTPases, components of the COPII complex, several SNARE proteins, and components of the ESCRT complexes as the parasite underwent encystation (De Cádiz et al., 2013; Herman et al., 2017). *Entamoeba* also expresses stage-specific pre-phagosomal vacuoles (PPV) that arise *de novo* and fuse with phagosomes upon extracellular material uptake (Saito-Nakano et al., 2005). Maturation of these compartments is tightly regulated by the late endosomal Rab5 and Rab7, where mutations in these genes impaired phagocytosis and erythrocyte engulfment by the parasites (Saito-Nakano et al., 2005).

Though comparatively little is known about the cell biology and endomembrane compartmentalization of the intracellular obligate microsporidian parasites, a few existing studies have

implicated the role of trafficking proteins in these pathogens. Previous comparative genomics has determined a highly reduced complement of vesicle coats that consists of AP-1, COPI, and COPII (Barlow et al., 2014). Parallels can be drawn between the endomembrane organization of *Giardia* and species of microsporidia, as both are devoid of canonical Golgi and endosomal vesicle carriers. Electron tomography studies and immunoprobings with existing COPI and COPII components in *Paranosema grylli* and *Paranosema locustae* has suggested spores as well as the intracellular stages (*i.e.*, meronts and sporoblasts) to possess an alternate Golgi morphology, namely one that presents itself as an interconnected tubular and varicose network stemming from regions of the ER (Beznoussenko et al., 2007). This network contacts the plasma membrane and the parasite polar tube, which is a major virulence factor used to inject the sporoplasm to the neighbouring host cell cytosol (Beznoussenko et al., 2007). Hence, it is likely that this organelle represents a major and possibly the only highway for trafficking and secretory processes to the parasite polar tube and plays essential roles in neighbouring-cell propagation and infection. More studies are necessary to illuminate the precise mechanisms by which this may be occurring.

Like microsporidia, another parasite with a largely enigmatic endomembrane organization and a minimal endo-lysosomal trafficking system is *Giardia intestinalis*, which will be the focus of the remainder of this chapter and thesis.

1.4 *Giardia intestinalis* and Giardiasis – brief history and clinical disease

First identified by Antonie van Leeuwenhoek in 1681 from his own stool and viewed under a rudimentary light microscope, van Leeuwenhoek described *Giardia intestinalis* as an animalcule in a letter to Robert Hooke. *Giardia intestinalis* (*syn.* *G. lamblia* and *G. duodenalis*) is a well-recognized enteric protist parasite responsible for the food and waterborne diarrheal disease, Giardiasis, which is colloquially also known as ‘Beaver Fever’ in North America (Dobell, 1920; Esch & Petersen, 2013). First taxonomic descriptions of *Giardia* were provided by physicians Vilém Dušan Lambl and Alfred Mathieu Giard, who were commemorated for their efforts through the parasite’s namesake (Lipoldová, 2014).

The World Health Organization approximates 300 million global cases of Giardiasis per year (Lanata et al., 2013). However, this is likely an underestimation as a large number of cases go unreported. Due to the considerable morbidity and the disease’s direct impacts on socio-economic developments in resource-poor settings of the world, Giardiasis has been included in the World Health Organization’s Neglected Disease Initiative (WHO NDI) (Savioli et al., 2006). The WHO NDI aims to develop and implement infection control and mitigation strategies to combat neglected infectious diseases by supporting biomedical research that investigates parasite biology, zoonoses, epidemiology, and host-parasite interactions (Savioli et al., 2006). The incidence of this disease is global and occurs in most industrialized and developing countries. However, it is much more endemic in resource-poor settings that face healthcare inequities, as compared to other populations, where overall prevalence can range between 10-20% and can be as high as 50% (Feng & Xiao, 2011). This percentage reduces to 2-5% in developed countries (Lanata et al., 2013). According to the Canadian Notifiable Disease Surveillance System, the most recent

estimates placed the Canadian incidence of Giardiasis equaling to roughly 10.42 per 100,000, making it one of the top five infectious diarrheal disease in Canada (Health Canada, 2019). Compared to most provinces, prevalence and potential for Giardiasis outbreaks are also much higher in Alberta and British Columbia, where watershed testing quantifies 2,500-8,700 cysts per 100 liters during peak spring runoff events (Health Canada, 2019). Global epidemiological studies show a bimodal distribution of Giardiasis in different population ages with peaks corresponding to children aged 0-9 and adults between 45 and 49 years old (Painter et al., 2015). When assessing specific demographics, children, daycare workers, travelers, and outdoor enthusiasts are the main groups where the annual incidence of the disease is highest. Infections in these populations can be attributed to poor hand hygiene, barriers to water sanitation infrastructure, or frequent exposure to parasites.

Giardiasis can be asymptomatic, which is the case for approximately 10% of infections. However, the disease primarily manifests as symptomatic gastroenteritis after one to two weeks of exposure to infectious cysts, where as few as ten are sufficient to establish an infection (Hellard et al., 2000; Rendtorff, 1954). The most common symptoms consist of watery diarrhea, nausea, vomiting, and epigastric cramps, which appear 6-15 days post-infection (Flanagan, 1992). The clinical diagnoses of Giardiasis involve fecal immunoassays (e.g., enzyme-linked immunosorbent assay (ELISA)), serum IgG detection, and counter immune electrophoretic (CIE) techniques. Alternative methods such as molecular testing using polymerase chain reaction is also a popular method that is rapid and efficient (Flanagan, 1992). In cases where the disease is not self-resolving, tinidazole, nitazoxanide, and metronidazole are used as primary pharmacotreatments in conjunction with oral rehydration therapy if the diarrheal symptoms are especially severe (Gardner & Hill, 2001; Watkins & Eckmann, 2014). In rare instances, Giardiasis can progress to more chronic outcomes. The most common long-term consequence of disease in most adults is the potential development or exacerbation of irritable bowel syndrome (Halliez & Buret, 2013). In children, developmental issues such as failure to thrive, impaired cognitive functions, and growth stunting can pervade due to sustained malnutrition and dehydration caused by persistent or recurring infections (Halliez & Buret, 2013).

1.5 Ecology and phylogenetic relationships within the subphylum Fornicata

Within Metamonada, *Giardia* belongs to the subphylum Fornicata consisting of free-living, commensal, and parasitic members (Kolisko et al., 2008) (Figure 1.3). Phylogenetic placement of fornicates within Metamonada places them sister to Preaxostyla and Parabasalia. Like fornicates, both have members with free-living, commensal, and endobiotic/parasitic lifestyles. Well-known representatives from Preaxostyla and Parabasalia are the mitochondria-lacking *Monocercomonoides exilis* and the previously discussed STI-causing parasite, *Trichomonas vaginalis*, respectively (Karnkowska et al., 2019; Leger et al., 2017) (Figure 1.3). Recent environmental surveys have identified a group of marine sediment-dwelling protists termed the barthelonids that share morphological and metabolic characteristics with the fornicates (Yazaki et al., 2020). Of these, representative *Barthelona* sp. was subject to transcriptomics for phylogenomic characterizations, which confirmed the phylogenetic placement of this lineage as close

relatives of the fornicates (Yazaki et al., 2020) (Figure 1.3). Within Fornicata itself, free-living lineages are heterotrophic flagellates that consist of *Carpediemonas membranifera* and a morphologically similar group of organisms, collectively referred to as *Carpediemonas*-like organisms (CLOs) (Kolisko et al., 2010) (Figure 1.3). Commensal fornicates are termed the retortamonads, which are gut endobionts of animals. The parasitic fornicates, to which *Giardia* belongs, are the diplomonads that also include the fish pathogen *Spironucleus salmonicida* and the secondarily free-living *Trepomonas* sp. PC1 (Takishita et al., 2012; Xu et al., 2014, 2016) (Figure 1.3).

Free-living *Carpediemonas membranifera* and *Carpediemonas*-like organisms were first isolated from marine sediments and are microaerophilic (Kolisko et al., 2010; Simpson & Patterson, 1999). All lineages are characterized by the presence of two or more flagella and have cytostome-like structures. Phylogenetics has placed *C. membranifera* and *Carpediemonas*-like organisms to be paraphyletic in relation. So far, CLOs consist of the following members: *Aduncisulcus paluster*, *Ergobibamus cyprinoides*, *Kipferlia bialata*, and *Dysnectes brevis* (Figure 1.3). Nonetheless, multigene phylogenies and phylogenomics consistently retrieve *Carpediemonas membranifera* to be the most basal member within this sub-phylum (Leger et al., 2017; Takishita et al., 2012).

Closest to the CLOs are the retortamonads, a group of commensal flagellates that reside within the guts of numerous animals ranging from mammals to amphibians, including humans (e.g., *Retortamonas intestinalis*) (Kulda et al., 2017). Representative members that have sequencing data available include *Chilomastix* spp. (Figure 1.3). Morphologically, like the CLOs, retortamonads possess one or more flagella and are characterized by the presence of cytostomes. Unlike CLOs, however, retortamonads undergo encystation and have lifecycle stages consisting of an environmentally dormant cyst stage and a trophozoite feeding stage that resides within the animal gut where propagation between the host and the environment occurs fecal-orally (Kulda et al., 2017).

The diplomonads make up the third and final major lineage within the Fornicata and within which *Giardia* is sub-classified (Figure 1.3). Diplomonadida is essentially a parasitic clade with representatives that infect various animals, except for *Trepomonas*, which secondarily reverted to a free-living lifestyle after its evolution within this lineage (Xu et al., 2016). *Spironucleus salmonicida*, a freshwater salmon pathogen and a threat to the aquaculture industry, is also a well-studied parasite within diplomonads apart from *Giardia* (Stairs et al., 2019). Almost all diplomonads, including *Giardia*, are diplozoic in their cellular features and hence, as their namesake implies, are characterized by two nuclei. Unlike the CLOs, diplomonads possess a set of four or more flagella. Members of the sub-group Hexamitinae, which includes *Trepomonas* and *Spironucleus*, also use alternative genetic codes, TAA and TAG, to encode glutamine rather than a stop codon compared to the rest of the diplomonads and other fornicates (Keeling & Doolittle, 1997). Recent studies have performed metabolic reconstructions from genomes available from this lineage and have confirmed that the diplomonad ancestor was parasitic and that host-adapted lifestyle is an ancestral trait that predates *Giardia* and all extant diplomonads (Jiménez-González et al., 2020).

The Fornicata lineage is exemplary of gradual and reductive evolution leading to parasitism. Extensive ultrastructural studies have revealed varying complexities within the cellular organization across members, including the previously discussed endomembrane organelles. *Carpediemonas membranifera* is the only member within this lineage that possesses a stacked Golgi, which is absent from the rest of the CLOs, retortamonads, and the diplomonads (Simpson & Patterson, 1999). Dynamic endosome-like compartments have been observed in *Carpediemonas*, CLOs, and *Spironucleus* sp., but are missing from *Giardia* (Park et al., 2010; Simpson & Patterson, 1999; Yubuki et al., 2016; Santos et al., in preparation). All fornicates also lack typical cristate mitochondria and instead possess simplified forms of this organelle referred to as mitochondria-related organelles (MROs), which vary in their metabolic capacities. The diversity of MROs in Fornicata ranges from hydrogenosomes in CLOs, *Trepomonas*, and *Spironucleus* to the further reduced mitosomes in *Giardia* (Leger et al., 2017). Recent transcriptomic efforts and comparative genomic investigations into the mitochondrial structural proteins and metabolic pathways reveal streamlining in this system as transition to parasitism occurred (Leger et al., 2017). This was especially evident in the repertoire of mitochondrial membrane import proteins and in proteins required for glycine metabolism, NADH reduction-oxidation pathway, and ATP-generation enzymes, which were present in some free-living lineages but absent from most diplomonad genomes and transcriptomes (Leger et al., 2017).

Reductive evolution in other cellular systems is also persistent within this lineage. Genome sequencing efforts have especially shed light on cases of losses that occurred before the transition to parasitism. One such example is the reduced DNA segregation machinery in *Carpediemonas membranifera*. Comparative genomics with the genome of this organism shows the absence in the highly conserved origin recognition complex, Cdc6 division proteins, numerous replisome machinery, and parts of the kinetochore machinery (*i.e.*, Ndc80 complex) (Salas-Leiva et al., 2021). These patterns were also evident when the genome of the CLO, *Kipferlia bialata*, was comparatively examined against the diplomonads (Tanifuji et al., 2018). Functionally annotated clusters of orthologous groups of proteins (COGs) reconstructed from *K. bialata*, *S. salmonicida*, and *G. intestinalis* were compared to one another (Tanifuji et al., 2018). These analyses determined that the overall number of COGs belonging to cellular processing and signaling, information storage and processing, and metabolism pathways were pared down in the parasites (Tanifuji et al., 2018).

Genome streamlining paired with neogenesis of novel machinery are salient features of parasitism and a by-product of host-pathogen co-evolution. These are apparent in other protist parasites, as discussed previously, but also in the fornicate parasites. Virulence genes such as the variant surface proteins (VSPs) are encoded in *Giardia*, abundant in these parasite genomes but absent elsewhere (Adam et al., 2010). Similarly, though not identical, families of cysteine-rich proteins have been identified in *Spironucleus salmonicida*, which are neither encoded in *Giardia* nor the other fornicates (Xu et al., 2014).

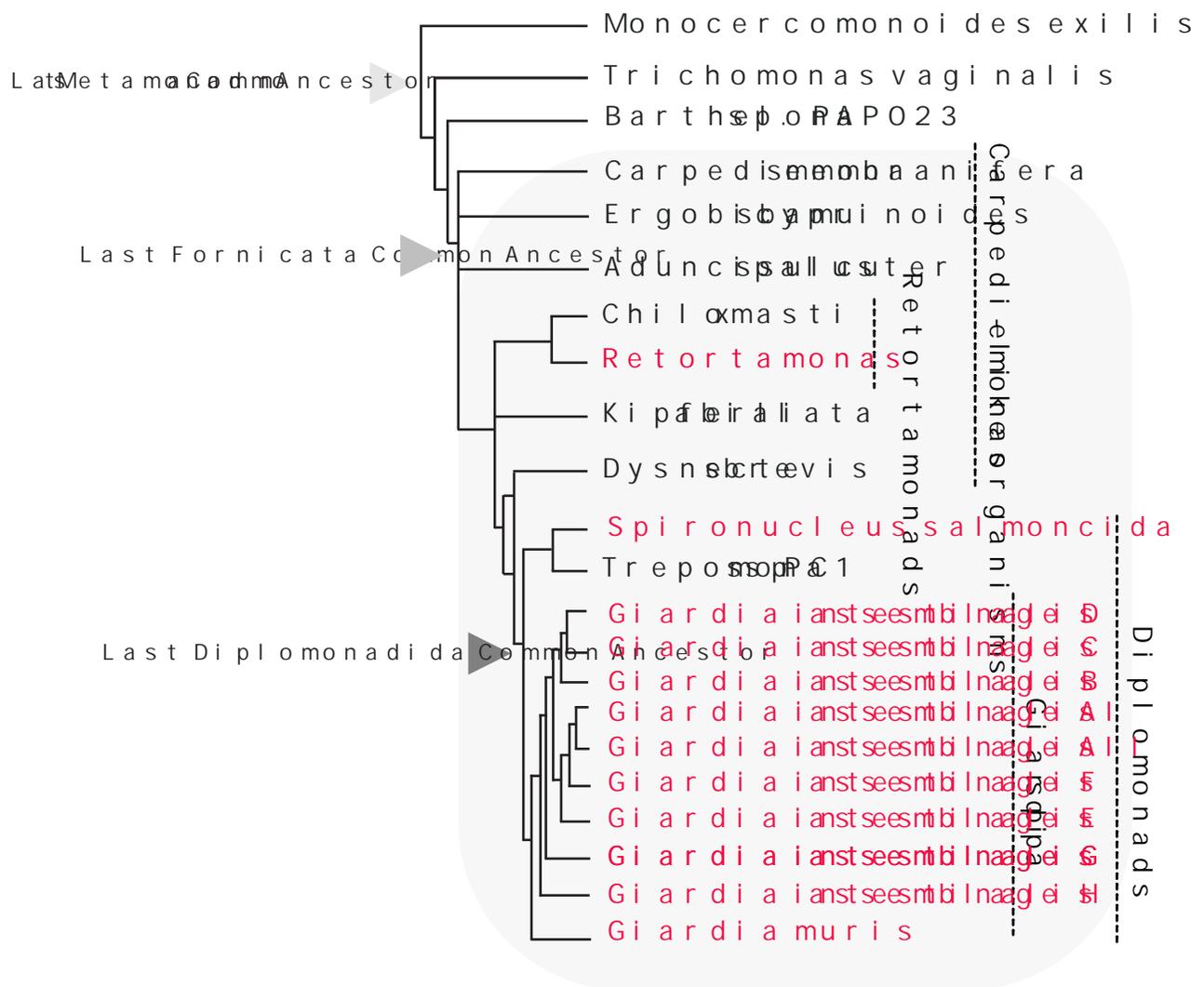


Figure 1.3. Phylogenetic relationships within Fornicata, *Giardia*, and with other metamonads. Free-living and parasitic fornicates are enclosed within the grey box and depicted in relation to one another and to other close metamonad representatives from the lineages Parabasalia (*i.e.*, *Trichomonas vaginalis*) and Preaxostyla (*i.e.*, *Monocercomonoides exilis*). The recently discovered basal branching lineage to fornicates known as barthelonids is also shown through the representative *Barthelona* sp. PAP020. Fornicata comprises three sub-groups: the free-living *Carpediemonas*-like organisms, the commensal or parasitic retortamonads, and the parasitic diplomonads. Overview of the relationships between the eight known assemblages of *Giardia intestinalis* is also depicted in comparison to the outgroup and mice pathogen, *Giardia muris*, and other diplomonads (*i.e.*, *Spiroucleus salmonicida* and *Trepomonas* sp. PC1). Important to notice here is that the two-human infecting *Giardia intestinalis* assemblages A (AI and AII) and B are paraphyletic. Fornicate relationships are based on phylogenomic reconstructions by Leger et al. (2017) and Yazaki et al. (2020), and *Giardia intestinalis* assemblage branching orders are based on the latest MLST analyses conducted by Feng & Xiao (2011). Parasitic species are indicated in red font, while those in black font are free-living or commensal.

1.6 An 'omics perspective on the *Giardia* spp. and *G. intestinalis* assemblages and implications on the clinical disease

The *Giardia* genus comprises numerous species that can infect a broad range of animal hosts (Adam, 2001). *Giardia muris* is a pathogen of rodents and is used as a model to investigate *Giardia* infections in murine systems. Other species include *Giardia microti*, a pathogen of voles and muskrats, *G. psittaci*, which infects budgerigars, *G. ardeae*, which parasitizes numerous bird species, *G. agilis*, a parasite of amphibians, and finally, *G. intestinalis*, which infects numerous mammalian hosts (Heyworth, 2016). Of these, *G. intestinalis* is best-known and is well-researched due to its implications in causing human Giardiasis (Heyworth, 2016).

Although morphologically identical, *G. intestinalis* can be further subdivided into different assemblages (*i.e.*, genotypes with attributable differences in marker proteins used for phylogenetic classification) (Figure 1.3). Parasite genotyping is performed using several molecular assays such as genomic PCR amplification, restriction fragment length polymorphism (RFLP), and more recently, whole-genome sequencing and comparative genomics (Bertrand et al., 2005). Taxonomic classification of the assemblages is based on multi-locus sequence typing (MLST), which is performed using four *Giardia* genes: β -giardin (*bg*), triosephosphate isomerase (*tpi*), small ribosomal subunit (*ssu-rRNA*), and glutamate dehydrogenase (*gdh*) (Ryan & Cacciò, 2013). So far, these methods have classified *Giardia intestinalis* isolates into eight distinct assemblages (A-H) (Heyworth, 2016). Assemblages A and B have the greatest host ranges and the only two of the identified eight capable of causing disease in humans. These also have the highest potential for zoonoses and anthroponosis. Other hosts for these include companion and domesticated farm animals such as dogs, cats, cattle, horses, pigs, sheep, chickens, alpaca, and guinea pigs. Wild animals such as deer, marsupials, muskrats, non-human primates, seals, moose, reindeers, porpoises, and gulls have also been identified to be infected by assemblage A and B parasites (Heyworth, 2016). In companion animals, namely dogs and cats, assemblages C and D are also prevalent, while assemblage E infects pigs and cattle. Reports of isolates belonging to all three of these assemblages have been identified in human fecal samples and pose a concern for potential zoonosis (Fantinatti et al., 2016). Finally, the least understood are assemblages F, G, and H, which reportedly infect cats, murines, and grey seals, respectively. Because of their veterinary and clinical importance, of all the identified assemblages, isolates belonging to A and B have been subject to intense genomics investigations for more than a decade (Heyworth, 2016).

Giardia intestinalis trophozoites possess two transcriptionally active nuclei and have a tetraploid genome that is encoded in five chromosomes that are 1.49, 1.54, 1.97, 2.76, and 4.02 Mbp in size (Bernander et al., 2001; Upcroft et al., 2010). The trophozoite stage can cycle between a ploidy of 4N to 8N in G1, S, and G2 phases; meanwhile, the cysts have a ploidy of 16N (Bernander et al., 2001). Due to this ploidy, *Giardia* genomes are challenging to assemble. Unexpectedly, diplomixis (*i.e.*, exchange of DNA between the two nuclei) is limited, and therefore, allelic sequence heterozygosity is unelevated in *Giardia* compared to other polyploid eukaryotes (Bernander et al., 2001).

The first sequenced genome from any *Giardia* species was of the human-infecting assemblage A isolate, WB (clone C6), published in 2007 and a cornerstone to assess the parasite's evolutionary origins and underpinnings of pathogenesis (Morrison et al., 2007). This availability also lent insight into the substantial levels of bacterial lateral gene transfer that have shaped the parasite genome. Genome sequencing and assembly also revealed extreme sequence divergence that has prevented accurate *in silico* orthology assignment to approximately 40% of the encoded genes, which remain annotated as hypothetical proteins. Apart from the reference genome, WB (C6) is also the model isolate subject to mass-spectrometry and RNA-sequencing studies to study *Giardia* biology. As a result, a wealth of proteomics and transcriptomics data has also been generated over the years and is an essential resource to *Giardia* biologists to assess the biological relevance of the genomically encoded machinery (Emery et al., 2016).

Because the reference genome of the WB isolate is a vital asset to molecular and evolutionary parasitologists alike, an improved and more contiguous chromosome-scale assembly was generated using reads from long-read genome sequencing technology in early 2020 (Xu et al., 2020). Long-read sequencing also provided better estimates of the percent GC content, which decreased from 49.0% to 46.3%. It also improved genome contiguity and allowed for the removal of approximately 1000 fragmented genes. Importantly, intron predictions and the number of ribosomal RNA genes also doubled as compared to previous estimates. Apart from WB, few other isolates from assemblage A were also recently sequenced. These are isolates, DH, AS175, and AS98 (Adam et al., 2013; Ankarklev et al., 2015). MLST and comparative genomics with these additional isolates allowed for further sub-classification of assemblage A into AI and AII lineages, wherein the WB isolate belongs to the AI lineage, while DH, AS175, and AS98 belong to AII (Adam et al., 2013; Ankarklev et al., 2015).

Parasites belonging to assemblage B have also been sequenced. Unlike assemblage A, *in vitro* axenic cultures of assemblage B parasites are harder to establish as they are slow-growing compared to assemblage A isolates. They are also difficult to genetically manipulate to generate transgenic variants to study genes of interest. Therefore, genomic data is the current best arsenal to explore parasite biology in this particular assemblage. Isolate GS serves as the representative of this group and has been subject to both short-read and long-read genome sequencing, which has permitted the assembly and availability of both short-read and hybrid genomes (Franzén et al., 2009; Pollo et al., 2020). Two more isolates, GS_B and BAH15c1, have also been sequenced in recent years (Adam et al., 2013; Wielinga et al., 2015). The availability of sequencing data from several assemblage A and B isolates has broadened the scope for *Giardia* genomicists and evolutionary biologists to perform comparative analyses between the two-human infecting assemblages.

These studies have revealed significant differences to exist at the genomic level between isolates of assemblages A and B, such as in the average genome sizes, which ranges between 10-11 Mbp and 11-13 Mbp, respectively (Adam et al., 2013; Ankarklev et al., 2015; Franzén et al., 2009). Notable differences also exist in the encoded virulence genes such as VSPs. For example, DH and WB encode approximately 190 and 244 candidate VSPs, respectively, whereas B_GS is predicted to have over 700 putative VSPs

(Jerlström-Hultqvist, Franzén, et al., 2010). Apart from the overall complement, differences in the sequence composition of VSPs have also been noted. VSPs from isolates of WB and GS are highly divergent when compared to, as sequence similarity reaches only as high as 55% (Adam et al., 2013). Differences in repertoires of other virulence genes such as high cysteine membrane proteins (HCMPs) and cellular machinery required for scaffolding and flagellar functioning also exist (Adam et al., 2013). Examples include numerous NimA-related kinases (NEK) families that are absent from one assemblage or differed in their architecture of ankyrin-repeats (Adam et al., 2013). Variabilities in sequence-identity between annexin-related α -Giardins, which provide parasite cytoskeletal and structural integrity, were also evident (Adam et al., 2013). Finally, whole-genome synteny analyses between the five chromosomes from WB and GS confirm dissimilarities in locally collinear blocks (LCBs) belonging to the two isolates. Although no major chromosome rearrangements were reported, DNA sequence identities across all LCBs was a maximum of 70% (Adam et al., 2013).

Apart from assemblages A and B, genomes from a few animal-infecting strains have also become available in recent years. Isolate P15 from assemblage E was assembled and used for comparative genomic comparisons, which revealed it is closely related to assemblage A (Jerlström-Hultqvist, Franzén, et al., 2010). Although all three assemblages share a core set of approximately 4000 open reading frames (ORFs) comprising housekeeping genes and regulatory enzymes, whole-genome analyses consistently identified key differences in the families and repertoires of virulence genes, where only 27 VSPs were shared between assemblage A, B, and E isolates (Jerlström-Hultqvist, Franzén, et al., 2010). New genomes from assemblages C and D, and the murine-infecting *Giardia muris* have also been sequenced and published (Kooyman et al., 2019; Xu et al., 2020). The continual availability of new *Giardia* genomes permits ongoing and improved comparative genomics between the different assemblages to elucidate variabilities in their underlying biology. Findings from these investigations have raised an important question within the field. Is *Giardia intestinalis* a species complex comprising multiple biologically distinct lineages? Moreover, do divergences within parasite biology contribute to Giardiasis disease outcomes, epidemiology, and transmission between hosts?

Clinical cross-sectional studies have attempted to correlate assemblage prevalence with clinical cases and instances of heightened disease severity when symptomatic Giardiasis persists. A yearlong survey from a cohort of 400 Egyptian children and adults with diarrheal disease was undertaken, where 60 cases were attributable to Giardiasis (El Basha et al., 2016). Genotyping of cysts from these fecal samples distributed 63% of the infections were caused by assemblage B isolates, while the remaining 37% were assigned as assemblage A infections. A similar trend was observed in a more extensive cohort study undertaken in the United Kingdom (UK) with 406 Giardiasis cases. Once again 64% of the cases were attributed to assemblage B. By contrast, 33% were caused by assemblage A (Minetti et al., 2015). An epidemiological survey of Giardiasis outbreak in British Columbia (Canada) mapped a predominant number of clinical cases to assemblage B (65%) (Tsui et al., 2018). Surface water sampling of non-outbreak-

associated isolates also primarily classified to assemblage B (76%), indicating that assemblage B may also be generally greater in environmental prevalence (Prystajek et al., 2015).

It is important to note that Giardiasis is only clinically diagnosed when diarrheal symptoms are present and when molecular diagnostics on stool samples is performed. Whether disease manifestation and symptom severity are an underlying cause of differences in the virulence capacities of assemblage A and B, have also been investigated, albeit to a lesser degree. Self-reported questionnaires from the previous UK study reported that patients infected with assemblage B were more likely to experience symptoms of vomiting and an overall longer duration of infection persisted compared to those caused by assemblage A (Minetti et al., 2015). Interestingly, studies conducted in Saudi Arabia, where both symptomatic and asymptomatic Giardiasis infections in children were accounted for, reported that 62% of the symptomatic cases were attributable to isolates of assemblage B. In contrast, a greater likelihood of asymptomatic disease was in children who harbored assemblage AI or AII isolates (Al-Mohammed, 2011). Additional clinical and cell biological investigations are required to determine whether there is an interplay between parasite biology and disease and whether genetic differences between strains underpin disease severity and persistence.

1.7 *Giardia* lifecycle and pathogenesis

Unlike other gastrointestinal parasites such as *Entamoeba*, *Giardia* is a non-invasive pathogen that does not penetrate the gut mucosa but instead resides extracellularly by attaching to the intestinal epithelial cells (IECs) (Figure 1.4). Parasite transmission occurs fecal-orally via a simple lifecycle that consists of two stages, infectious cysts and vegetative trophozoites (Adam, 2001) (Figure 1.4). It begins with the ingestion of cysts through contaminated food or water, which are induced to excystation upon encountering physiological triggers such as the acidic stomach milieu and duodenal bile and trypsin (Ankarklev et al., 2010) (Figure 1.4). A single excyzoite (*i.e.*, newly excysted cell) undergoes several rounds of cytokinesis to give rise to four bi-nucleated trophozoites (Einarsson et al., 2016). Trophozoites then attach to and colonize the cholesterol-rich duodenum and upper jejunum using an adhesive ventral disk, a complex microtubular structure that resembles a suction cup, and surface lectins (Adam, 2001; Einarsson et al., 2016) (Figure 1.4; Figure 1.5). As the parasites continue to proliferate on top of the microvilli, a concomitant uptake of host luminal contents and secretion of parasite substrates takes place. A complex interplay between host and parasite factors underpins disease manifestation and pathophysiology. Hence, Giardiasis is a multifactorial disease (Cotton et al., 2011).

Trophozoites express a suite of virulence factors to evade surveillant intestinal intraepithelial lymphocytes, disrupt the gut brush border and tight junctions, induce enterocyte apoptosis and villous atrophy, and promote anion hypersecretion (Buret & Cotton, 2011; Cotton et al., 2011). Direct epithelial damage and host dysbiosis can be ascribed to parasite releasing enzymatic excretory-secretory products, cysteine proteases, cathepsin-like proteases, and tenascins (Liu et al., 2018). However, the most consequential and comprehensively investigated family of virulence genes are the group of cysteine-rich

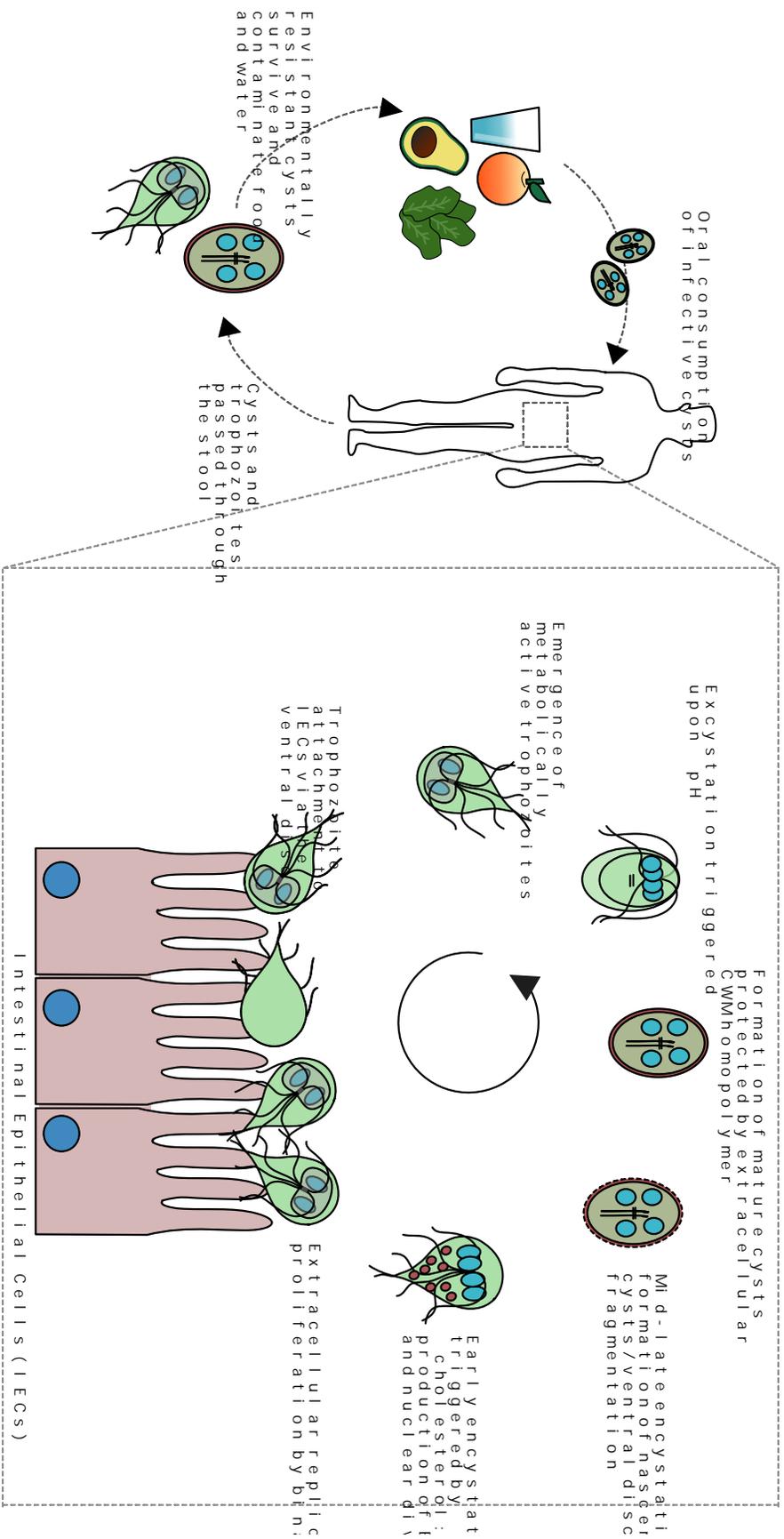
variant-specific surface proteins (VSPs) and high cysteine membrane proteins (HCMPs) (Davids et al., 2006; Liew et al., 1997). Together they account for roughly 4% of the overall protein-coding genes encoded in *Giardia* genomes. Of the two, VSPs are the best characterized and defined by the presence of CxxC tetrapeptide motifs (Prucca & Lujan, 2009). Although the parasite encodes hundreds of these proteins, the surface antigen profile only consists of a single VSP member at any given time (Adam et al., 2010). During an infection, the trophozoites undergo VSP antigenic switching in an epigenetic mechanism that involves histone acetylation and post-transcriptional regulation through an RNA-interference-like pathway (Prucca et al., 2008). As discussed in section 1.6, intense genomic evaluations have shed light on the differences in the repertoire and biochemical profiles of VSPs and have played a critical role in the postulation of the hypothesis that *Giardia intestinalis* lineage may be an amalgamation of genetically different species (Jerlström-Hultqvist et al., 2010). *In vivo* studies in rodents allude to the role of VSPs in the modulation of humoral and cytotoxic immune responses and in prolonging gut infection (Faubert, 2000; Heyworth, 1986; Singer et al., 2001). Less is known about the role of HCMPs; however like the VSPs, these cysteine-rich proteins are expressed on trophozoite surfaces and implicated in immuno-modulatory functions and host-parasite interactions (Peirasmaki et al., 2020).

Apart from surface proteins, metabolic enzymes such as arginine deiminase (ADI) are also secreted by the parasites as virulence factors for immune system modulation (Banik et al., 2013). ADI depletes the host arginine that is essential for nitric oxide production and inhibits adaptive and innate responses. Investigations in murine models demonstrate that giardial ADI hampers dendritic cell functioning to create a disbalance in the types of inflammatory cytokines that are ultimately produced (Banik et al., 2013). Other secretory enzymes released by trophozoites during an infection include ornithine carbamoyltransferase (OCT), fructose-1,6-bisphosphate aldolase (FBPA), and enolase. However, their exact roles in parasite virulence are still being investigated (Ma'ayeh et al., 2017; Balmer and Faso, personal communication).

After some period of residing in the host gut, trophozoites undergo encystation. Peristaltic movements and downstream sweeping of the *Giardia* trophozoites play a role in the transition to the second part of its lifecycle (Lauwaet et al., 2007) (Figure 1.4). As the trophozoites travel down to the distal ileum and the colon, physiological changes such as an increase in bile concentrations, change to an alkaline pH, and cholesterol starvation triggers encystation, which is routinely replicated *in vitro* (Boucher & Gillin, 1990). Some studies have implicated an increase in trophozoite density also to play a role (Barash et al., 2017). Encystation in *Giardia* can be subdivided into two phases, early and late, and takes between 15 to 24 hours to complete cyst morphogenesis (Einarsson & Svärd, 2015) (Figure 1.4). The process starts with transcriptional changes in the early stage to trigger the expression of stage-specific organelles known as encystation-specific vesicles (ESVs) (Figure 1.4; Figure 1.5). These are vesicle carriers for cyst-wall material (CWM), a mechanically and chemically resistant complex biopolymer that protects the dormant parasites from environmental assaults (Reiner et al., 1989). Constituents of the cyst-wall material are three sets of cyst-wall proteins (CWP1-3) intercalated into networks with β -1,3-linked N-acetylgalactosamine

(GalNAc) polysaccharides (Erlandsen et al., 1996; Hehl et al., 2000; Luján et al., 1995). Post-translational modification and accumulation of CWP1-3 at the ER-exit sites drive a concerted neogenesis of ESVs, followed by their transport to the cell surface (Faso et al., 2013). Several MTS proteins are implicated in ESV biogenesis and ESV-mediated trafficking of CWM, discussed in more detail in the following section. Encysting cells also undergo other morphological changes such as rounding of the cell and internalization of the ventral disk and flagella (Figure 1.4). In the late phase, the nuclei undergo DNA replication and division to yield an octaploid cyst that has completed depositing and polymerizing its cyst-wall (Einarsson & Svärd, 2015) (Figure 1.4). Mature cysts, as well as trophozoites that failed to undergo encystation, are then fecally transmitted into the environment (Figure 1.4). But only cysts survive in the environment where they remain dormant and can survive up to three months in low-temperature waters.

Figure 1.4. The simple lifecycle of *Giardia intestinalis*. *Giardia* parasites can manifest in two forms: an infective cyst stage and a gut-infecting trophozoite (feeding) stage. The lifecycle begins upon the oral consumption of cysts, which undergo excystation upon entering the host gastrointestinal tract to yield metabolically active trophozoites. This occurs upon making contact with the acidic stomach milieu. Trophozoites attach to the host microvilli, where they remain extracellular and proliferate by binary fission. Increased trophozoite density, peristaltic movements, increased pH, and higher cholesterol levels in the lower gut triggers encystation which consists of reversible (early) and irreversible (mid-late) stages. Encystation is characterized by the synthesis of cyst-wall material carried in encystation-specific vesicles and deposited onto the trophozoite surface. During this time, bi-nucleated trophozoites also undergo nuclear replication to yield four nuclei. As encystation progresses, the ventral disc and the flagella are fragmented, and its constituents are recycled back into the cytoplasm. Upon complete transport and polymerization of cyst-wall material, an environmentally resistant mature cyst, along with trophozoites that did not encyst, are fecally excreted into the environment. Cysts can survive prolonged periods of desiccation and cold temperatures in the environment, contaminating food and water for subsequent parasite propagation to the next host. This figure was based on Ankarklev et al. (2010).



1.8 *Giardia*'s enigmatic cell biology with a spotlight on its endomembrane system

Secretion and uptake of material are critical for disease establishment by the *Giardia* vegetative trophozoites and for encystation. As discussed in the previous sections, exocytosis and endocytosis of material require harmonized cargo transport by the membrane trafficking system. Unlike many eukaryotes, *Giardia* lacks many of the defining compartments that are a hallmark of the endomembrane system, such as a stacked Golgi apparatus or any of the endo-lysosomal carriers (Benchimol & De Souza, 2011; Friend, 1966) (Figure 1.5). Instead, microscopy studies have identified *Giardia*-specific organelles: the peripheral vacuoles, encystation-specific vesicles, and a highly expanded network of the tubulovesicular endoplasmic reticulum (Figure 1.5). Below is a brief discussion on the biology of these organelles and protein machinery that may play roles in intracellular and extracellular trafficking within this parasite. Additionally, some unconventional aspects of other cellular systems and parasite biology are also highlighted.

1.8.1 Peripheral vacuoles: constitutive secretion in trophozoites

The peripheral vacuoles (PVs) are considered a major interface for cargo uptake, recycling, and sorting within *Giardia* trophozoites (Feely & Dyer, 1987; Zumthor et al., 2016). The PVs are a set of single membrane-bound polarized vacuolar organelles 100-200 nm in size and situated ca. 50 nm from the plasma membrane (Faso & Hehl, 2011; Touz et al., 2012) (Figure 1.5). Microscopic cross-section of a trophozoite identifies two sets of PV populations: 1) cortical and dorsal PVs distributed at the cell periphery, and 2) a circular bare zone, which is a densely aggregated region of PVs between the two nuclei (Figure 1.5). Both populations are easily distinguishable when labeled with a luminal marker for endosomes, Dextran, and other PV-associated proteins (Cernikova et al., 2020; Zumthor et al., 2016). Early electron microscopy and more recently focused ion beam scanning electron microscopy (FIB-SEM) and 3D reconstructions show these compartments to be heterogeneous in shape and size (Friend, 1966; Midlej et al., 2019; Zumthor et al., 2016; Santos et al., in preparation). Notably, these studies also show some populations of the PVs to have membrane contiguity with tubules of the endoplasmic reticulum (Abodeely et al., 2009; Zumthor et al., 2016). This raises the possibility for these regions to be involved in cargo trafficking between the ER and PVs in the absence of dynamic vesicular carriers.

PVs are also postulated to be functionally fused forms of late endosomes and lysosomes. This notion has been reasoned with numerous molecular functional investigations. Biochemical studies with proteolytic lysosomal enzymes, cathepsin B, and soluble acid phosphatase (AcPh) showed distinct localization to the PV lumen, lending weight to the hypothesis that these are acidified organelles with endo-lysosomal properties (Feely & Dyer, 1987; Touz et al., 2004). Recent investigations have also shed light on the unconventional mechanisms underpinning the uptake of extracellular milieu into the *Giardia* trophozoites, which occurs at the PVs in conjunction with the plasma membrane (PM). A combination of immunofluorescence confocal microscopy, FIB-SEM, and transmission electron microscopy (TEM) analyses demonstrated fusion dynamics between PV and PM membranes (Zumthor et al., 2016). This fusion results in a steady-state pore formation mediated by the giardial clathrin heavy chain and dynamin,

which has led to a working model of cellular import that relies on bulk flow uptake and inward flushing of material into the PV lumen (Zumthor et al., 2016). These PV dynamics have been demonstrated through time-series experiments with fluorescently labeled Cholera toxin B (CtxB) fragment and Dextran (Zumthor et al., 2016). Dextran is a small glycan polymer that preferentially binds to endosomal compartments in mammalian cells and is used as a tractable dye for fluid-phase endocytosis (Oliver et al., 1984). CtxB binds to the giardial GM1 gangliosides and cholesterol-lipid rafts present on the plasma membrane of *Giardia* trophozoites and cysts (Ljungström et al., 1985; McCardell et al., 1987). Both of these stain PV compartments, but in two different ways. Experiments with fluorescently labeled CtxB demonstrate direct lipid exchange between the organellar membranes of the PVs, which results in CtxB marking PV-membranes over time. On the other hand, extracellular fluorophore-conjugated Dextran is taken up into the lumen of the PVs. It remains there without any cytosolic diffusion, therefore acting as a dye for PV-lumen staining (Zumthor et al., 2016).

Because PVs are at the primary interface between host and parasite, it makes them attractive compartments to study from a trafficking perspective. Previous comparative genomic investigations and findings presented in this thesis in the subsequent chapters conclude the *Giardia* trafficking system to be reduced in its repertoire (Marti et al., 2003). These studies have also identified numerous components of the vesicle formation and fusion machinery to be encoded in this parasite, some of which have been probed through molecular functional investigations to elucidate their intracellular roles. These include ESCRTs, adaptins, SNAREs, Rab GTPases, SAR1, ARFs, and clathrin heavy chain. A thorough discussion of their roles within *Giardia* is presented in Chapters 2, 3, 4, and 5.

Recent studies have also investigated the roles of lipid-binding proteins and markers of canonical endosomal carriers in the endosome-lacking *Giardia*. As discussed in section 1.3.2, phosphatidylinositol phosphate (PIP) residues mark the fate of vesicles and determine endosomal identity in traditional endocytic pathways. Recent molecular functional studies with *Giardia* trophozoites have examined the role of several newly identified PX and FYVE domain-containing proteins that generally recognize and bind PIP residues to recruit adaptors and membrane deformation proteins for endosomal maturation in model systems (Cernikova et al., 2020). In *Giardia*, these proteins associate with PV membranes and interact with numerous PV-associated machinery, including trafficking proteins such as clathrin heavy chain, dynamin, components of AP-2 subcomplex, and select SNARE proteins (Cernikova et al., 2020). Based on the collective findings with PV-localizing proteins and proteomics investigations, the proposed model is that PX and FYVE domain-containing proteins serve as membrane adaptors for recruitment and assembly of clathrin and dynamin to catalyze fusion dynamics between the PVs and the plasma membranes in an AP-2 and SNARE-mediated manner (Cernikova et al., 2020; Zumthor et al., 2016).

1.8.2 Endoplasmic reticulum: protein synthesis and possible roles in constitutive secretion

In the absence of a stacked Golgi, secretion of parasite material into the host environment is a process that remains puzzling. Trafficking during the trophozoite stages is not simply restricted to an “outside-in” process. Exocytosis also takes place to secrete and express virulence factors, as previously discussed in section 1.7. The primary player suspected to be involved in this process is the parasite’s endoplasmic reticulum (ER). The ER is a highly visible organelle in *Giardia* trophozoites as it is a labyrinthine network that ubiquitously spans much of the cytosol (Abodeely et al., 2009) (Figure 1.5). As in other eukaryotes, ER in *Giardia* has regions decorated with ribosomes, consistent with a rough ER morphology, and is a site for protein synthesis (Benchimol et al., 2004; Mccaffery & Gillin, 1994). Another typical characteristic that corroborates the ER’s role in protein synthesis is the identification of the luminal chaperone, BiP. Immunogold microscopy studies show giardial BiP to localize to the ER and is biochemically characterized by the presence of a classical C-terminal KDEL ER-retention signal, which posits a canonical role in protein folding within this compartment (Abodeely et al., 2009; Soltys et al., 1996). Other ER luminal proteins have also been identified. These are protein disulfide isomerase (PDI), three of which possess a single catalytic site and were localized to the ER and suggested for their involvement in tertiary protein structure folding and quality control of the secreted variant surface proteins (McArthur et al., 2001).

Apart from protein synthesis, the *Giardia* ER is postulated to be necessary for cargo packaging and secretion with possible *trans*-Golgi network-like functions. In live and fixed cells, the Golgi is routinely stained and visualized with a ceramide analog, NBD-C6 (Pütz & Schwarzmann, 1995). In *Giardia* trophozoites, NBD-C6 accumulates in sub-domains of the perinuclear ER, which are suggested to be functionally analogous to the TGN (Luján et al., 1995b; Zamponi et al., 2017). Canonical post-ER trafficking was corroborated through functional studies with *Giardia* COPII components. GFP and HA-tagged COPII-Sec23 were localized to the parasite’s ER-exit sites (ERES) (Faso et al., 2013; Zamponi et al., 2017). Dominant-negative mutagenesis in the GTP binding residues (*i.e.*, H74G) of the SAR1 GTPase, which controls recruitment and assembly of the COPII-coat proteins, abolished accumulation of Sec23 at the ERES (Stefanic et al., 2009). Assembly of COPII at ERES is necessary for ESV biogenesis and cargo loading of CWM during encystation (detailed discussion on mechanisms of ESV formation is done in section 1.8.3).

Experimental investigations have also shed light on the *Giardia* ER acting as a potential sorting factory with regions that have an endo-lysosomal biochemical profile. Co-localization studies with giardial cathepsin-B-like proteases and ER-tracker, a selective dye that binds to sulfonylurea receptors prominently present on the ER, show co-translocation to the ER perinuclear regions (Abodeely et al., 2009; Liu et al., 2018). Albumin uptake experiments and co-localization assays with the *Giardia* cathepsins also localized to the perinuclear ER, and time-series experiments demonstrate the eventual breakdown of albumin within these regions (Abodeely et al., 2009). These findings have raised the possibility that the *Giardia* ER may also serve as a site for protein sorting and cargo breakdown of material from the PVs trafficked to this

compartment in a retrograde fashion. However, more investigations are imperative to consolidate these possible endo-lysosomal roles.

Finally, the *Giardia* ER is also postulated as a site for mitosomal replication and fission. In yeast and mammalian cells, ER and the mitochondria are physically tethered at the MAMs (discussed in section 1.3.1) by series of protein machinery collectively termed the ER-mitochondria organizing network (ERMOINE) that facilitate mitochondrial division and maintain its morphology (Hoppins et al., 2011; van der Laan et al., 2012). Live-cell microscopy experiments have shown *Giardia* mitosomes to undergo division during mitosis in a synchronized but dynamin-independent mechanism (Voleman et al., 2017). Co-localization studies with the ER marker, PDI2, showed close associations between the mitosomal compartments and the ER tubules. Further assessments illuminate the co-localization of PDI2 and components of the mitochondrial-associated membrane proteins, such as the *Giardia* LACS4, to regions corresponding to the ER (Voleman et al., 2017). MAMs are integral within the ERMIONE pathway in canonical mitochondria-bearing organisms for ER-tethering functions, which raises the possibility that in this parasite, mitosome-ER contacts and dynamics are potentially similar to those observed in yeast and animals.

1.8.3 Encystation-specific vesicles: bona fide Golgi-like compartments for regulated secretion

Perhaps the best-studied trafficking organelles in *Giardia* are the encystation-specific vesicles (ESVs), which have been comprehensively investigated over several decades. ESVs are stage-specific carriers for cyst-wall proteins (CWP1-3) that are deposited extracellularly on the trophozoite surface as the parasite is undergoing encystation (Einarsson & Svärd, 2015; Reiner et al., 1989). *In vitro*, ESV neogenesis and encystation proceed upon cholesterol depletion and switch to an alkaline pH. All biosynthesized CWPs accumulate in ca. 30 ESVs which migrate to the cell periphery over a 10 to 14 hour period (Boucher & Gillin, 1990).

ESVs are large vesicles that are approximately one μm in diameter (Figure 1.5). Their biogenesis occurs at specific foci of the perinuclear ER and is driven by both trafficking factors and the cargo they enclose (Faso et al., 2013) (Figure 1.5). ESV cargoes are newly-synthesized CWPs that aggregate in membrane-bound clefts within the ER. Biochemical properties of the CWPs orchestrate dynamics of coat assembly, vesicle budding, and scission. CWP2 possesses a 13 kDa basic tail that reroutes CWP1 and CWP3 away from cytosolic secretion and instead for sorting into the nascent ESVs, a process that is temperature-sensitive (Konrad et al., 2010). CWP1 sorting is also targeted via its N-terminal (II) and middle (III) domains (Hehl et al., 2000). The role of cyst-wall proteins, particularly that of CWP1, is crucial to ESV biogenesis. Knockout studies where all four copies of *cwp1* were ablated using a Cre/loxP homologous recombination system exhibited a pseudocyst phenotype as ESV biogenesis and CWM transport were abolished upon functional deletion of this protein (Ebnetter et al., 2016). CWP1 also interacts with the GalNAc homopolymer during cyst-wall material polymerization and is therefore necessary for the correct architecture of the final mature cyst-wall (Chatterjee et al., 2010).

Apart from CWM, numerous trafficking proteins are crucial to ESV biogenesis. These are the COPI and COPII machinery typically required for canonical retrograde and anterograde trafficking, respectively, in organisms with distinct ER and Golgi compartments. In the early stages of ESV biogenesis, COPII components, SAR1, Sec23, and Sec24, perform functions of CWM loading into nascent ESVs (Faso et al., 2013). Once the cargo is packaged, mature ESVs are trafficked in a COPI and ARF1-dependent manner (Marti et al., 2003; Stefanic et al., 2009) (discussion on the role of ARF1 in ESV biogenesis is discussed in detail in Chapters 3 and 5). β '-COP and clathrin heavy chain localization to nascent and mature ESVs have also been confirmed through immunofluorescence microscopy (Marti et al., 2003). Components belonging to the vesicle fusion machinery, namely Rab1, Rab11, and Syntaxin1, also associate with ESVs (Castillo-Romero et al., 2010; Stefanic et al., 2009; Wampfler et al., 2014).

These molecular associations that typically occur to facilitate ER-Golgi trafficking, and biophysical characteristics of ESV biogenesis and maturation, are remarkably similar to ones observed in Golgi cisternal progression in model organisms with a stacked Golgi. Over the years, this has led to a working hypothesis by numerous groups within the *Giardia* field that ESVs represent transiently expressed Golgi compartments (Luján et al., 1995). In model systems, the Golgi's morphology and function are routinely disrupted using Brefeldin-A (BFA), a lactone antiviral that inhibits ARF1 activation and COPI-mediated transport (Helms & Rothman, 1992). Interestingly, *Giardia* ESVs are sensitive to BFA, and their maturation is abolished upon BFA-treatment, which supports the notion of ESV functional homology to the Golgi (Luján et al., 1995). The separation of ER-resident proteins from ESVs also indicates that these compartments are not offshoots of the ER but, instead, morphologically and biochemically distinct structures. However, these postulations are still debated as many Golgi-structure and assembly proteins such as Golgi reassembly-stacking protein (GRASPs) and other coiled-coil golgins are not encoded in *Giardia*, momentarily leaving ESV identity as analogs rather than homologs of the Golgi (Barlow et al., 2018). Additionally, neither ESV-like compartments nor a stacked Golgi is present in the vegetative trophozoites. This implies that in the absence of a canonical secretory pathway, atypical mechanisms for exocytosis are suggested to be at play in the trophozoite (Balmer & Faso, 2021).

1.8.4 Mitosomes

Although once thought to be an amitochondriate organism, *Giardia* possesses reduced mitochondria-related organelles (MRO) known as the mitosomes. Though these are not a primary focus in this thesis, the biology of these compartments is still worth a brief discussion. That is because the discovery of mitosomes in *Giardia* once created significant debates in the field of evolutionary biology, as it fundamentally redefined our understanding of the order and timing of mitochondrial endosymbiosis for eukaryogenesis and its and other MRO's relationships to the canonical mitochondrion (Tovar et al., 2003). Overwhelming evidence has since then emerged from *Giardia* and other protist lineages to support mitosomes as secondarily reduced forms of the metabolically complex cristate-mitochondria (see Gray, 2012 for a review on this topic).

Mitosomes in *Giardia* are 100 nm-sized double-membrane organelles that lack a mitochondrial genome (Martincová et al., 2015). *Giardia* mitosomes are not involved in ATP synthesis or hydrogen production, with the latter process occurring in several of *Giardia*'s close relatives that possess hydrogenosomes, as previously discussed. Identification of and microscopy studies with classical mitochondrial proteins such as chaperonin 60, mtHSP70, and iron-sulfur (Fe-S) cluster assembly machinery allowed for the discovery and characterization of the *Giardia* mitosomes (Roger et al., 1998; Tachezy et al., 2001). Localization studies with mitosomal markers have elucidated two distinct patterns of their cellular distribution within *Giardia* trophozoites. First is the presentation of an aggregated rod-like cluster between the parasite nuclear regions referred to as the central mitosomal complex. The remainder organellar population is dispersed throughout the cytosol and is termed peripheral mitosomes (Regoes et al., 2005) (Figure 1.5). Since these early discoveries, subsequent bioinformatic and protein-tagging studies have also identified numerous outer membrane and inner membrane translocases such as TOM40, TIM23, and components of the PAM complex (Voleman et al., 2017). Organellar proteomics investigations with the *Giardia* mitosomes have also uncovered a suite of mitosomal lumen proteins (e.g., ferredoxin, IscU, IscA, IscS, Cpn60, and others)(Jedelský et al., 2011).

1.8.5 Other structural and unusual aspects of *Giardia* cell biology

Giardia trophozoites are symmetrical pear-shaped cells 12 to 15 μm in length and 5 to 10 μm in width (Adam et al., 2001) (Figure 1.5). Each trophozoite possesses four pairs of flagella, subdivided into anterior, ventral, posterior-lateral, and caudal flagella, which are necessary for parasite movement, attachment to the host epithelia, and cell division (Dawson & House, 2010) (Figure 1.5). A collection of microtubular structures, known as the median body, are centrally located within the cell (Friend, 1966). Its exact functions currently remain unelucidated but have been suggested to be sites for ventral disc biogenesis and to potentially act as a pool of polymerized microtubules (Woessner & Dawson, 2012).

Notable absences in a typically conserved eukaryotic machinery and cellular processes have also been observed. For example, sexual reproduction is not present in *Giardia*, but remnants of the meiotic (e.g., Mnd11, Dmc1, and Mer3) and DNA recombination (e.g., Spo11, Hop1, Hop2) machinery have been identified, though whose roles within the parasite remain poorly understood (Malik et al., 2007). *Giardia* also lacks myosin, typically required by eukaryotes for actin-based motility and cytokinesis (Hardin et al., 2017).

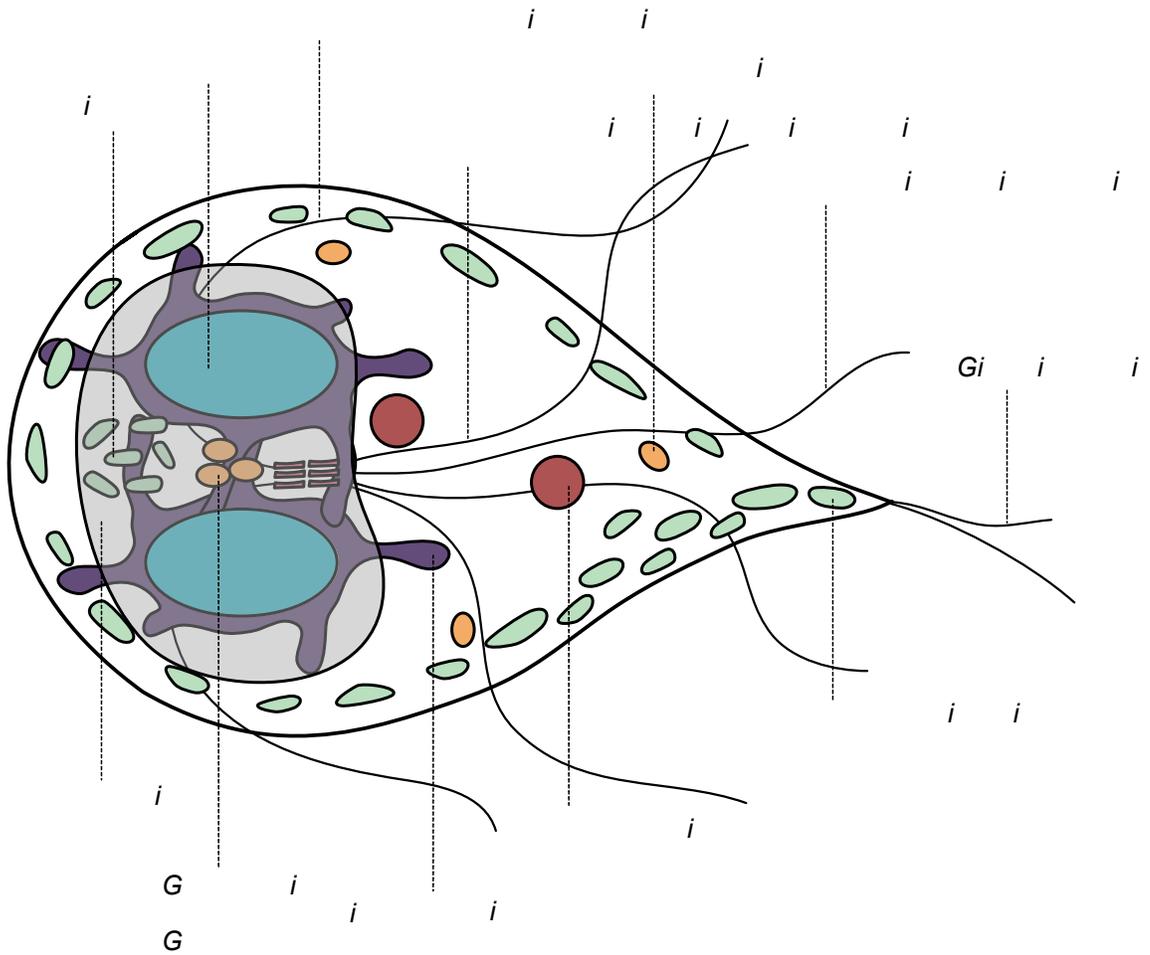


Figure 1.5. Illustration of a *Giardia intestinalis* trophozoite cross-section. *Giardia intestinalis* trophozoites are simplified tear-drop-shaped diplozoic cells characterized by two nuclei, eight flagella, a ventral disc cytoskeletal structure, an endoplasmic reticulum network, mitosomes (peripheral and central mitosomal complex), and static peripheral vacuoles. During encystation, encystation-specific vesicles (ESVs) emerge from perinuclear ER domains and transport cyst-wall material to the parasite surface for cyst morphogenesis. Both stages are devoid of a canonical stacked Golgi apparatus.

1.9 Current state of molecular parasitological and genomic research in *Giardia*

Elucidating cellular functions and uncovering biological complexities is dependent on the discovery and appraisal of the underlying protein machinery. Compared to many model organisms, including protist parasites subject to intense cell biological investigations (e.g., apicomplexans and trypanosomatids), traditional strategies for *in silico* and functional characterization of proteins have failed in *Giardia*.

The parasite's divergent genome has historically hampered efforts at determining its protein repertoire to reconstruct the complexity of existing cellular pathways. Sensitive homology searching tools and sophisticated methods for phylogenetic inferences for gene classifications were also lacking when the first *Giardia* genomes and molecular sequence data became available. A recent proliferation in genomic, transcriptomic, and proteomic data from new isolates of *Giardia*, parasitic diplomonads, and free-living fornicates paved the way for fine-grained comparative genomic assessments in this lineage. Additionally, rapid and ongoing improvements in the methods and models for homology searching (e.g., Hidden Markov Model-based), phylogenetics (e.g., Bayesian and Maximum Likelihood inferences), and genome assembly (e.g., ploidy-aware genome assemblers) are allowing *Giardia* computational biologists to overcome some of those early obstacles faced during *in silico* investigations.

From a molecular functional standpoint, determining protein functions in *Giardia* through traditional gene silencing or knockout approaches requires mutagenesis or deletion of all four copies of the gene of interest, which is an arduous and often unsuccessful process. Approaches such as CRISPR/Cas9-mediated deletion strategies, as well as morpholino and RNAi-mediated gene silencing, have been attempted, all with variable and limited success (Carpenter & Cande, 2009; Lin et al., 2019; McInally et al., 2019). So far, the Cre/loxP homologous recombination system is the only successful example that has achieved complete gene knockout but is arduous to implement for routine use (Ebnetter et al., 2016). Because many of these methods are time-intensive with lamentable outcomes, *Giardia* biologists have instead opted for microscopy-based investigations as a first-line approach to assess intracellular protein functions. Of these, immunoprobng for protein visualization through fluorescence microscopy is widely used, and, therefore, protocols for performing these experiments are benchmarked and well-established in this parasite. In *Giardia*, immunofluorescence assays are often conducted in two ways: development and probing of *Giardia*-specific antibody raised against the protein of interest or expression and visualization of a *Giardia* recombinant protein engineered with an epitope tag. Recombinant proteins are also targetable as affinity handles for co-immunoprecipitation and assessment of protein-protein interactions. Nonetheless, both methods have been used to study protein localization in trophozoites and encysting cells and are essential to the molecular toolkit in order to perform *in vitro* studies with this parasite.

1.10 Scope and organization of the thesis

The membrane trafficking system is essential to *Giardia* in order for it to establish infection in its host and for its continual propagation and survival in the environment. Previous studies have mainly investigated the role of trafficking processes primarily in the context of encystation, and so many questions

remain regarding the evolution and molecular mechanisms of the trafficking organelles that are present in the vegetative trophozoites. The absence of dynamic vesicle carriers but the presence of static-state PVs in the trophozoite endosomal system is especially mystifying. The vesicle formation machinery is critical for the neogenesis of vesicle carriers in a typical eukaryotic endomembrane organization and marks the complexity of secretion and endocytic pathways within a eukaryotic cell. Therefore, this machinery is pertinent to understanding the mode and tempo by which *Giardia* trophozoites' endomembrane system has been shaped to yield the existing trafficking architecture.

Based on previous findings from other cellular systems investigated within fornicates, as well trends followed by other parasites across the eukaryotic tree of life, the hypothesis for this thesis is that a simplified and aberrant *Giardia* endomembrane landscape is a by-product of both a transition to a host-dependent lifestyle and ancestral evolutionary modulations that pre-date *Giardia*. Additionally, the existing trafficking proteins in this parasite have diversified their roles to accommodate lineage-specific organellogenesis and cellular functions. The investigations presented in the subsequent chapters aimed to address the evolution of the vesicle formation trafficking machinery within the Fornicata-lineage as a whole, the *Giardia* genus, and the human-infecting assemblages. A subset of these proteins were also studied through molecular functional investigations to assess intracellular roles in the lab strain of *Giardia intestinalis*, assemblage A, isolate WB (C6). *Giardia*-specific differences were probed further through genome assembly and population-level surveys of the trafficking machinery belonging to the human-infecting isolates. Altogether, this thesis took a systems approach to bridge evolutionary bioinformatics, molecular parasitology, and population genomics to study the integral families of proteins pertinent to vesicle formation processes in a canonical eukaryotic landscape. The specific aims were as follows:

1. Determine the molecular evolution of the heterotetrameric adaptor complexes, retromer, COPII, and the ARF regulatory system across Fornicata (Chapters 2 and 3)
2. Investigate the molecular evolution of the late endo-lysosomal ESCRT complexes in Fornicata and elucidate functions of essential ESCRT subunits in the MVB-lacking *Giardia* (Chapter 4)
3. Elucidate molecular localization and interaction partners of the ARF GTPase regulatory system proteins within the ESV-lacking *Giardia* trophozoites (Chapter 5)
4. Trace differences in the repertoire of the early secretory and endo-lysosomal vesicle formation machinery in the human-infecting *Giardia* isolates through a population-level survey (Chapter 6)

In Chapter 2, the molecular evolution of the vesicle coat protein complexes required for biogenesis and cargo transport between the ER and Golgi and endo-lysosomal pathways was investigated. To do so, comparative genomics and phylogenetic analyses with the heterotetrameric coat complexes (AP1-5, COPI, TSET), retromer, COPII, clathrin subunits, and SPG11/SPG15 was performed. This allowed for mapping the timepoints at which losses and innovation in this machinery occurred and their correlation with existing endomembrane organization within Fornicata members.

Chapter 3 undertook similar investigations as Chapter 2 but with the ARF regulatory system proteins (ARF1 GTPase, ARF GAPs, and ARF GEFs) that temporally regulate the fidelity and organization of the heterotetrameric proteins at various endomembrane interfaces within the cell. Comparative genomics and phylogenetics determined tight evolutionary modulation within this system before the fornicate common ancestor.

Chapter 4 examined the repertoire and evolution of the ESCRT machinery necessary for MVB-biogenesis across fornicates. Through comparative genomics and phylogenetics, the time points at which evolutionary gains and losses occurred within subunits of ESCRT0, ESCRTI, ESCRTII, ESCRTIII, and ESCRTIIIA were determined. Lineage-wide, parasitism-specific, and *Giardia*-specific transitions could be correlated with subunit losses and a lack of MVB-morphology in *Giardia*. Newly identified and previously uninvestigated components belonging to *Giardia* ESCRTII and III complexes were also assessed for their intracellular localization and molecular interactions using confocal immunofluorescence microscopy, co-immunoprecipitation, and proteomics in the lab strain, WB(C6). These studies unveiled unexpected ESCRT localizations at several parasite organelles.

Like Chapter 4, fluorescent microscopy and proteomics investigations with select giardial ARF regulatory system proteins, identified in Chapter 3, were pursued in Chapter 5. This was performed to assess the roles of these proteins in the ESV-lacking trophozoites. Here it was demonstrated that the *Giardia* ARF-regulatory system localizes to the parasite PVs in conjunction with several vesicle formation proteins described in Chapter 2. Additionally, crosstalk between endo-lysosomal vesicle scission proteins was also evident.

Chapter 6 built on the bioinformatic findings from Chapters 2, 3, and 4 to investigate differences in the molecular complement of the vesicle formation machinery in the isolates of human-infecting *Giardia intestinalis* assemblages A and B through a population-level survey. This was done by first performing *de novo* genome assembly using previously sequenced and published short reads from British Columbia Centre for Disease Control Public Health Laboratory (BCCDC PHL) assemblages A and B isolates. The resulting genomes were then used for comparative genomic analyses with HTACs, COPII, retromer, ESCRTs, and the ARF regulatory system proteins from the preceding chapters. These investigations confirmed differences in the repertoire of the machinery identified in the pan-global isolates of these two assemblages in Chapters 2, 3, and 4 to exist at a larger population level.

Finally, Chapter 7 provided a general discussion on the findings from each preceding chapter and bridged them to develop a broader narrative on the evolutionary implications and cellular mechanisms that underpin the organization of the giardial endomembrane system, especially the PVs. This discussion also highlighted novel discoveries identified in this thesis and the advancements it provides towards an improved understanding of the membrane trafficking system in this parasite and its evolution within the context of general eukaryotic parasitism. The results from the population survey with the human-infecting *Giardia* isolates were also discussed and what implications similar and continual studies would have on clinical and public health outcomes. Finally, this chapter also examined some of the limitations of the methodologies

used in this thesis but also future investigations that could be pursued as a direct follow-up to the works presented in this thesis.

CHAPTER 2

A minimal vesicle coat protein complement in *Giardia* is shaped by lineage-specific evolutionary modulations

2.1 Overview

In order to thoroughly reconstruct the evolution and complexity of the endocytic pathway in *Giardia*, investigations with the evolutionarily related family of vesicle coats with proto-coatomeric origins that assemble and recognize cargo within the early and late secretory pathways were pursued. Based on these proteins' molecular complement and with underlying assumptions that molecular functions are conserved across eukaryotes, predictions regarding the diversity of endosomal trafficking pathways within a cell can be made. The heterotetrameric adaptor complexes, COPII, and retromer have distinct roles for cargo trafficking between the ER and the Golgi and between the plasma membrane and the endo-lysosomal compartments. Hence, these complexes were used as markers for assessing which of these trafficking aspects are conserved or lost within fornicates.

Previous published evolutionary investigations of the vesicle coats in fornicates were restricted to *Giardia intestinalis* assemblage AI, isolate WB, which performed BLASTP analyses to conclude the presence of a minimal repertoire in this parasite (Hirst et al., 2011, 2014; Koumandou et al., 2011; Marti et al., 2003; Schlacht & Dacks, 2015). Though these early *in silico* studies have been pertinent for downstream molecular investigations to illuminate protein functions in *Giardia*, there still remain several gaps. From an evolutionary perspective, it was still necessary to deduce how this reductive evolution took place. Therefore, this chapter took a lineage-wide analysis to uncover the molecular evolution within the vesicle coat proteins by assessing their conservations and losses across free-living and parasitic fornicates, as well as several strains of *Giardia*. It also recapitulated which aspects of the endosomal machinery are dispensable versus essential on the path to parasitism in this lineage. Additionally, the genomes and transcriptomes from close fornicate relatives were also critical stepping-stones to help identify highly divergent orthologues in *Giardia* which may have remained unidentified in the previous studies. Altogether, using improved and sensitive homology searching tools (*i.e.*, Hidden Markov Model-based searching) and new genomic and transcriptomic data, a more comprehensive bioinformatic survey of the vesicle coat proteins to determine the evolution of the endocytic pathways in *Giardia* and its relatives was performed in this chapter.

2.2 Introduction

Vesicle coats are an integral family of trafficking proteins that regulate neogenesis and vesicle budding dynamics for bi-directional transport of material between different sub-cellular interfaces and the extracellular environment (Bonifacino & Glick, 2004). The major families identified in the last eukaryotic common ancestor and those that are distributed across most eukaryotic supergroups consist of the

evolutionarily related vesicle adaptor coat complexes that share proto-coatomeric origins (Hirst et al., 2011, 2014; Koumandou et al., 2011; Schlacht & Dacks, 2015). These are the heterotetrameric adaptor protein complexes (HTACs), which themselves comprise adaptins, TSET, and coatomer protein complex I (COPI), coatomer protein complex II (COPII), and retromer (Dacks & Field, 2007).

These proteins initiate retrograde and anterograde trafficking of material between the endoplasmic reticulum (ER) and the Golgi for early secretion and within the endo-lysosomal pathways for extracellular import and export (Figure 2.1A). As previously discussed in Chapter 1, anterograde trafficking of newly synthesized material from the ER to the *cis*-Golgi relies on COPII-coated vesicles for additional post-translational modifications and sorting at the Golgi for downstream endosomal export (Sato & Nakano, 2007) (Figure 2.1A). On the other hand, COPI facilitates inter-Golgi cargo transport and retrograde trafficking of material back to the ER (Beck et al., 2009) (Figure 2.1A). Many post-Golgi endocytic processes are clathrin-dependent and occur in concordance with adaptins, retromer, and TSET. Adaptins are a family of five subcomplexes (AP1-5) that were conserved in the LECA, which commences endosome biogenesis and maturation between the plasma membrane and the *trans*-Golgi network (TGN) (Kirchhausen, 1999; Robinson & Bonifacino, 2001) (Figure 2.1A). To briefly remind the reader, AP-1 and AP-4 are requisite for TGN to early endosomal traffic, with the latter also participating in autophagosomal maturation, while AP-2 is a driver of clathrin-mediated endocytosis at the plasma membrane (Conner & Schmid, 2003; Hirst, Irving, et al., 2013; Kural et al., 2012b; Mattera et al., 2017) (Figure 2.1A). AP-3 and the newly discovered AP-5 mediate bi-directional cargo transport between the late endosomes and the TGN (Hirst et al., 2013; Hirst et al., 2018; Kural et al., 2012) (Figure 2.1A). Like adaptins, retromer and TSET are also critical players for endocytic and secretion dynamics with a eukaryotic cell. Retromer is implicated in the recycling of mannose-6-phosphate receptors to and from the plasma membrane and the TGN, whereas the newly discovered TSET complex is involved in autophagosomal trafficking and clathrin-mediated endocytosis at the plasma membrane in conjunction with AP-2 (Cullen & Steinberg, 2018; Gadeyne et al., 2014; Hirst et al., 2014; Wang et al., 2021; Wang et al., 2019) (Figure 2.1A).

Each of these is a multimeric complex that has modular configurations for cargo binding and recognition. As the name indicates, the heterotetrameric complexes consist of four components: two large subunits, each approximately 100 kDa, a medium subunit that is 50 kDa, and a small subunit that is 20 kDa in molecular weight (Robinson & Bonifacino, 2001) (Figure 2.1B). Adaptin large subunits are named β 1-5 and γ 1, α 2, δ 3, ϵ 4, and ζ 5, while the medium and small subunits are μ 1-5 and σ 1-5, respectively (Hirst et al., 2011) (Figure 2.1B). Similarly, COPI also comprises two large subunits termed β -COP and γ -COP, a medium δ -COP, and a small ζ -COP subunit (Beck et al., 2009) (Figure 2.1B). Counterparts of the large subunits in TSET are TPLATE and TSAUCER, while the medium and small subunits are TCUP and TSPOON, respectively (Hirst et al., 2014) (Figure 2.1B). These different components aggregate at donor membranes and complex together to recognize cargo for sorting into nascent vesicles. They also initiate membrane deformation by recruiting vesicle scaffolding proteins such as clathrin and SPG11/15 (see Chapter 1 for a thorough mechanistic overview of this process).

Unlike HTACs, the ancestral repertoire of the COPII-coat constitutes heterodimeric Sec23 and Sec24 and heterotetrameric Sec13 and Sec31, which recognize cargo upon recruitment of the SAR1 GTPase and its activation by the transmembrane Sec12 (Figure 2.1B). These latter two proteins trigger sequential assembly of the Sec23/Sec24 and Sec31/Sec13 pre-budding complex onto ER membranes for a cuboctahedron lattice formation to coat the vesicle (Sato & Nakano, 2007) (Figure 2.1B). Recent investigations in mammalian cells and using high-resolution microscopy have pointed to COPII's role in driving ERES-membrane tubulation for cargo trafficking to the Golgi (Weigel et al., 2021). Whether or not ER-tubulation is a eukaryote-wide conserved mechanism for COPII-mediated cargo delivery should be investigated in other systems for a broader perspective into the mechanistic intricacies of COPII assembly and dynamics. In either case, the accessory Sec16 component is necessary for this process as it is speculated to stabilize the pre-budding complexes by directly interacting with Sec24 (Sprangers & Rabouille, 2015) (Figure 2.1B).

Retromer is also an evolutionarily conserved hetero-hexameric subcomplex with probable proto-coatomeric origins. The components of this machinery bear cargo selection, membrane deformation, and cargo receptor functions (Seaman, 2004; Seaman et al., 1997). Carboxypeptidase Y receptor Vps10 in yeast or mannose-6-phosphate receptors (M6PR) in mammals is the cargo that is recognized by Vps29, Vps26, and Vps35 that make up the cargo-selective complex and assemble onto the early endosomes (Seaman, 2004; Seaman et al., 1997) (Figure 2.1B). BAR-domain containing sorting nexins (SNXs) in Metazoa or Vps5 in other eukaryotes recognize endosomal phospholipids and recruit clathrin and clathrin adaptors to initiate endosomal curvature and drive membrane tubulation (Figure 2.1B). Altogether, this allows for the trafficking and recycling of Vps10/M6PR to the TGN and plasma membrane via early or recycling endosomes (Burd & Cullen, 2014; Temkin et al., 2011) (Figure 2.1A). Although not investigated here, the related retriever complex performs similar roles for retromer-independent recycling and uses paralogues of the cargo-selective proteins, namely DSCR3, Vps29, and C16orf62 (which bears sequence similarity to Vps35). Though DSCR3 was determined to be evolutionarily ancient, it is unclear whether C16orf62 is distributed outside humans (Koumandou et al., 2011).

Because these coat proteins are diversely routed within the endosomal system, by tracing the molecular complement and evolution of this machinery in Fornicata will help recapitulate the complexity of this pathway in this lineage. It will also help evaluate the path taken to yield a minimal trafficking pathway in *Giardia*. This is important because different fornicate lineages have morphologically diversified endomembrane compartmentalization. As discussed in Chapter 1, *Carpodomonas membranifera* is the only lineage that maintains a stacked Golgi (Simpson & Patterson, 1999). On the other hand, the *Carpodomonas*-like organisms (e.g., *Kipferlia bialata*) lack a discernible Golgi but are characterized by the presence of numerous endosomal and vacuolar compartments (Yubuki et al., 2013). The endo-lysosomal organization within diplomonads also appears to be variable in comparison to one another where *Spironucleus* spp. seem to possess dynamic endosome-like organelles that are absent from *Giardia* (Morecki & Parker, 1967; Sterud et al., 1997). In comparison to its relatives, *Giardia* has a

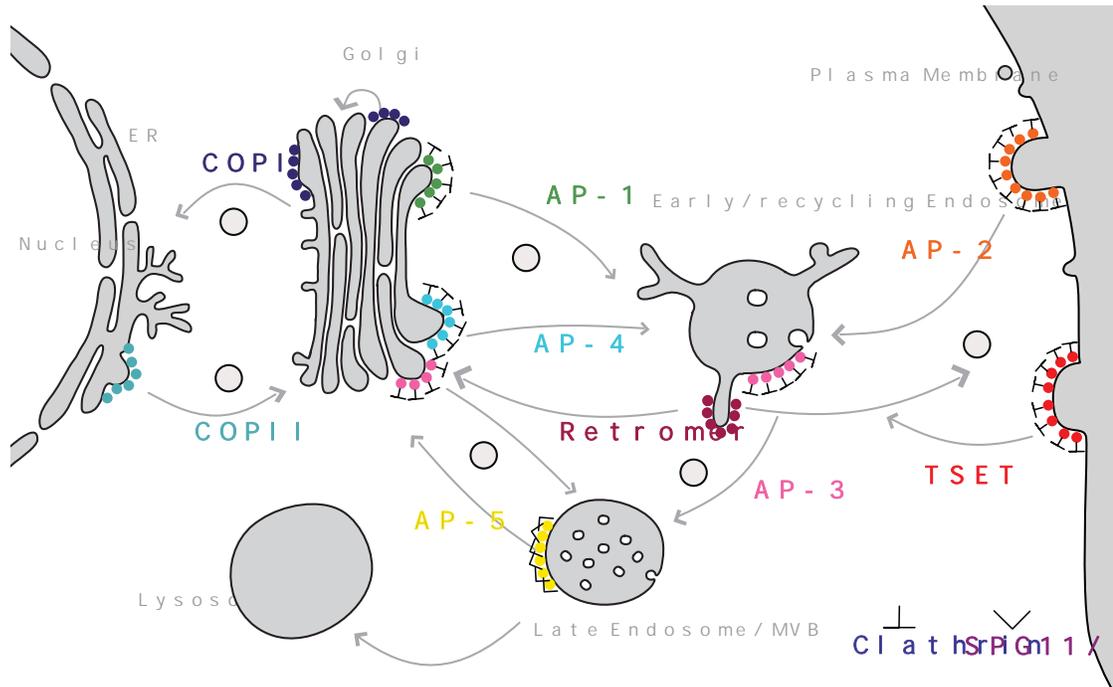
distinct endomembrane organization devoid of endosomes and a stacked Golgi but instead is defined by the presence of PVs and transient-state encystation-specific vesicles. Altogether, these morphological variations within the fornicate lineages prompt investigations into the evolutionary dynamics of the endosomal trafficking system proteins that would otherwise mark and define the complexity that *Giardia* is lacking. Furthermore, by tracing the time-points at which these modifications occurred, one can predict which aspects of this machinery are dispensable versus conserved.

The scope of this chapter was to perform comparative genomics with the HTACs, COPI, COPII, and retromer, as well as with the membrane deformation proteins that are utilized by these complexes (*i.e.*, clathrin and SPG11/SPG15) using BLAST and Hidden Markov Model-based (HMMER) searching. Previous comparative genomic reconstructions have implicated HTAC subunits to have emerged from a precursor protein with shared domain architectures consisting of α -solenoids, β -helices, longin, and μ -homology domains, and so phylogenetic analyses were performed for orthology consolidation and to trace any lineage-specific duplications when paralogs were present (Dacks & Robinson, 2017; Kirchhausen, 1999; Lewin & Mellman, 1998; Robinson & Bonifacino, 2001). Previous pan-eukaryotic analyses determined the presence of at least three ancestral paralogues of COPII-Sec24 in the LECA (Sec24I-III) (Schlacht & Dacks, 2015). Therefore, the identified fornicate Sec24 proteins were also subject to phylogenetic investigations to classify them within these LECA families. Bayesian inference and maximum likelihood-based phylogenetics were opted as the methods of choice, as both use sophisticated statistical models to reconstruct protein evolutionary relationships accurately. These approaches are especially optimal when dealing with divergent sequences, such as those present in *Giardia*.

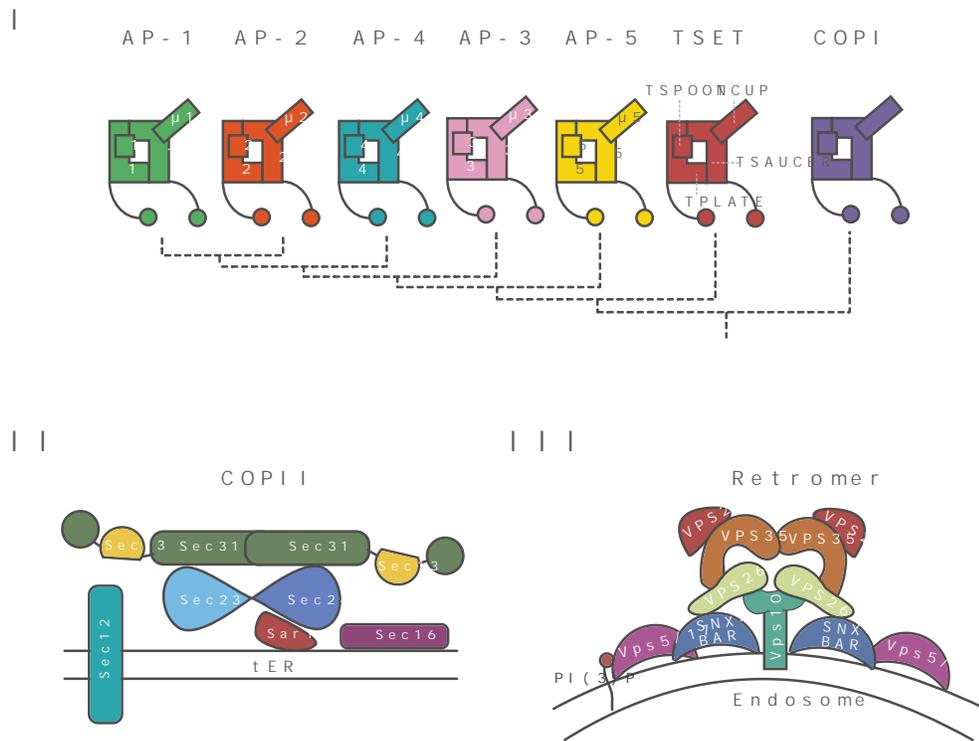
This issue of sequence divergence in diplomonads and to identify any novelties that arose within this system was also addressed by employing a two-part strategy for homology searching and phylogenetics. First, the free-living *Carpodiemonas* and *Kipferlia* were subject to BLAST and HMMER homology searching to identify vesicle coat proteins. Orthologs from these two organisms were then used to build more sensitive HMM profiles for iterative searching into the diplomonads to capture and characterize divergent orthologs. Phylogenetic analyses were performed using the *Carpodiemonas membranifera* sequences as the fornicate representative for characterization against pan-eukaryotic sequences. Then, metamonad-specific phylogenetic analyses were performed using these newly classified *Carpodiemonas* sequences along with orthologs from oxymonad and parabasalid representatives, *Monocercomonoides* and *Trichomonas*, respectively, for classifying *Kipferlia* and diplomonad HTACs.

Figure 2.1. Depiction of known intracellular localizations of the vesicle coat complexes within a typical endomembrane system and their structural configurations. (A) depicts intracellular locations for each vesicle coat complex at canonical membrane trafficking organelles. COPI and COPII mediate retrograde and anterograde trafficking, respectively, between the ER and the Golgi. Adaptins, retromer, and TSET are necessary within the endo-lysosomal pathway for inter-organellar and endo- and exocytic cargo transport. AP-1 and AP-4 function between the TGN and early endosomes, while AP-2 and TSET perform clathrin-mediated endocytosis at the plasma membrane. AP-3 and AP-5 are required in the bi-directional transport between late endosomes and early endosomes/lysosomes. Retromer is necessary for receptor recycling to the TGN or the plasma membrane. AP1-4 use clathrin for membrane deformation, while AP-5 uses SPG11 and SPG15. This figure was adapted from Hirst et al. (2011, 2014). (B) shows the individual modules that make up each of these complexes. (I) represents the heterotetrameric adaptor complexes and their evolutionary relationships to one another. Each consists of two large, one medium, and one small subunit. (II) depicts the COPII complex at the transitional ER membranes and consists of Sec23/Sec24 and Sec31/Sec13 coats as well as the SAR1 GTPase, Sec12 GEF, and Sec16 accessory protein. (III) illustrates organization of the retromer complex and interactions between the different subunits, consisting of Vps10 cargo, Vps5 and SNX-BAR domain-containing membrane deformation proteins, and Vps29, Vps35, and Vps26 coat recognition components. Figures in these panel were based on Sato and Nakano (2007), Hirst et al. (2014), and Suzuki et al. (2019).

A



B



2.3 Materials and Methods

2.3.1 Dataset and query sequence retrieval

Basic local alignment search tool (BLAST) and Hidden Markov Model (HMMER) searching were used for pairwise and profile-based sequence comparison, respectively (Altschul et al., 1990; Eddy, 1998). HMM profiles were built using previously characterized vesicle coat protein sequences from pan-eukaryotic representatives, *Arabidopsis thaliana* (Archaeplastida), *Homo sapiens* (Opisthokonta), *Bigelowiella natans* (SAR), *Dictyostelium discoideum* (Amoebozoa), *Naegleria gruberi* (Discoba), *Monocercomonoides exilis* (Metamonada-Oxymonada), and *Trichomonas vaginalis* (Metamonada- Parabasalia) as queries (Hirst et al., 2011, 2014). Sequence accessions corresponding to each query are summarized in the Online Appendix Table 2.1.

Searches were performed in representative fornicate genomes and transcriptomes as per the strategy outlined in the introduction. The short and long-read genome assemblies of *Carpediemonas membranifera* (strain BICM) were provided by Andrew Roger and Dayana Salas-Leiva (Dalhousie University, Canada)(Salas-Leiva et al., 2021). The genome of the *Carpediemonas*-like organism, *Kipferlia bialata*, was initially provided by Goro Tanifuji (National Museum of Nature and Science, Tokyo, Japan). A polished and updated version of the short-read assembly was then acquired from the National Center for Biotechnology Information (NCBI) after its public availability to cross-validate any gene absences (Tanifuji et al., 2018). Predicted proteins and nucleotide scaffolds belonging to the diplomonads were also retrieved from public databases. Specifically, genomes corresponding to *Giardia intestinalis* assemblages A, isolates DH and WB, *Giardia intestinalis* assemblage B, isolates GS and GS_B, and *Giardia intestinalis* assemblage E, isolate P15, were obtained from GiardiaDB (<http://giardiadb.org/giardiadb/>)(Adam et al., 2013; Franzén et al., 2009; Jerlström-Hultqvist et al., 2010; Morrison et al., 2007). The genome of *Giardia muris* was provided by Feifei Xu (Uppsala University, Sweden), and the genome of *Spironucleus salmonicida* (strain ATCC 50377) was downloaded from GiardiaDB (Xu et al., 2014; Xu, Jiménez-González, et al., 2020). Lastly, the transcriptome of the secondarily free-living diplomonad *Trepomonas* sp. PC1 was obtained from NCBI (Xu et al., 2016).

2.3.2 Homology Searching

BLASTP searches were initially performed using *Arabidopsis thaliana*, *Naegleria gruberi*, and *Homo sapiens* queries for forward and reciprocal searching against the Fornicata genomes and transcriptomes (Altschul et al., 1990). Next, more sensitive HMMER searching was performed by first aligning pan-eukaryotic protein sequences corresponding to individual HTAC, retromer, COPII, clathrin, and SPG11/SPG15 subunits using MUSCLE v. 3. 8.31 (Edgar, 2004) (Online Appendix Table 2.1). The resulting alignments were then used to build Hidden Markov Model profiles using the hmmbuild program available through the HMMER v. 3.1.b1 package, which were used to perform HMMER searches using the hmmsearch tool (Eddy, 1998; Prakash et al., 2017). For both BLAST and HMMER searching, an e-value threshold was set to ≤ 0.01 , where only hits meeting this cut-off were kept for analyses. Non-redundant

forward hits were considered true orthologs if reciprocal BLASTP searching retrieved the correct *Homo sapiens* ortholog with an e-value ≤ 0.05 and was two-orders or greater in magnitude than the next best non-orthologous hit. Reciprocal hits were extracted and sorted using an in-house Perl script (Herman, 2018). Forward TBLASTN and reverse BLASTX searches into the nuclear scaffolds of all fornicates were also performed for instances where components were unidentified in the initial BLASTP and HMMER searches to rule out possible false negatives. All identified fornicate orthologs with their corresponding forward, and reciprocal e-values are summarized in the Online Appendix Table 2.2.

2.3.3 Phylogenetic analyses with heterotetrameric adaptor subcomplexes and COPII-Sec24

Orthology of the HTAC subunits was additionally validated through phylogenetics. Homology searching identified numerous Sec24 paralogues across all fornicates and, therefore, a phylogenetic assessment was performed to determine whether these classified as the LECA-specific Sec24I, II, or III.

To do so, newly identified fornicate sequences with HTAC subunits from query organisms were aligned using MUSCLE v. 3.8.31 (Edgar, 2004). For classification of fornicate Sec24 sequences, previously published backbone protein alignment consisting of eukaryotic Sec23 and Sec24 was used (Schlacht & Dacks, 2015). This was performed by iteratively aligning the fornicate Sec24 sequences to the backbone alignment using the profile option in MUSCLE v. 3.8.31. The resulting alignments were visualized and manually adjusted using Mesquite v. 3.61 (Maddison & Maddison, 2019). Maximum likelihood (ML) and Bayesian inference (BI) phylogenetics were performed using RAxML-HPC2 on XSEDE v8.2.10 for non-parametric bootstrapping, and MRBAYES on XSEDE v. 3.2.1 were used. ProtTest v. 3.4.2 was used to determine the best-protein model, which computed LG+I+F for all HTAC alignments and LG+F for the COPII alignment (Darriba et al., 2011). For ML analyses, bootstrapping was performed by generating 100 non-parametric pseudo-replicates under the default tree faster hill-climbing method (-f b, -b, -N 100) (Felsenstein, 1985). These were used to build a consensus tree using the Consense program available through the Phylip v. 3.66 package (Felsenstein, 1989). MRBAYES v. 3.2.1 was used to perform Markov Chain Monte Carlo analysis to estimate the posterior distribution of model parameters (Gupta & Rawlings, 2014; Huelsenbeck & Ronquist, 2001). Specific run settings included 10 million MCMC generations under a mixed amino acid model (LG4X) and gamma rate categories set to 4. MCMC sampling frequency was set to occur every 1000 generations, and tree convergence was ensured when the average standard deviation of split frequency fell below 0.01. Additionally, the burnin value was set to 0.25 to discard the first 25% of the samples from the cold chain. A random starting seed value of 12345 was chosen for all phylogenetic analyses.

Both programs were run on the publicly available portal, Cyberinfrastructure for Phylogenetic Research (CIPRES) (Miller et al., 2010). RAxML bootstrap values were overlaid onto the MRBAYES tree-topology alongside BI posterior probability values. Tree visualizations and rooting were performed in FigTree v. 1.4.3 and annotated in Adobe Illustrator CS4 (Rambaut, 2010). HTAC subunit trees were rooted at COPI and TSET sequences, while the COPII-Sec24 tree was rooted at Sec23. A combined BI posterior

probability and ML bootstrap values of 0.80 and 50, respectively, were required for a node to be considered statically supported. Multiple rounds of tree-building and pruning were performed for the removal of long branching clades or sequences. Phylogenetic analyses were also repeated with or without TSET sequences to improve node support and clade resolution.

2.4 Results

2.4.1 Ancestral reduction likely shapes a minimal endocytic system within the fornicates

Evolutionary bioinformatic investigations were performed with the vesicle adaptor coat proteins that share proto-coatomeric origins, which determined the molecular complement and evolution of this machinery in the fornicates. Modulation in this system has occurred in several different ways. First, and perhaps the most prominent pattern observed, was an ancestral streamlining in the molecular machinery predating parasitism and the Fornicata lineage altogether.

Within the HTACs, three adaptin subcomplexes (AP1-3) and COPI were conserved across free-living and some diplomonad lineages and consequently concluded to be present in the Last Fornicata Common Ancestor (Figure 2.2; Online Appendix Table 2.2). Notably, though, within the HTACs, a universal loss of the entire TSET, AP-5, and AP-4 subcomplexes was observed across all sampled fornicates (Figure 2.2; Online Appendix Table 2.2). In comparison, the other metamonad relatives at least partially retained one or more of these subcomplexes. The parabasalid representative, *Trichomonas vaginalis*, encodes $\beta 5$ and $\mu 5$ from the AP-5 subcomplex, while the oxymonad *Monocercomonoides exilis* possesses TPLATE, TTRAY1, and TSPOON from TSET (Figure 2.2; Online Appendix Table 2.2). Both lineages also retain a complete repertoire of AP-4. The orthology of the individual subunits that make up this subcomplex was also phylogenetically validated (Figure 2.3A; Online Appendix Figures 2.1-2.4; Online Appendix Table 2.2). On the other hand, pan-eukaryotic and metamonad-specific phylogenetics confirmed that none of the HTAC subunits identified in *Carpodiemonas* belonged to AP-5 or TSET, and instead clustered within clades corresponding to $\beta 1-3$, $\gamma 1$, $\alpha 2$, $\delta 3$, $\sigma 1-3$, β -COP, γ -COP, δ -COP, or ζ -COP (Figure 2.3B; Online Appendix Figures 2.1-2.4; Online Appendix Table 2.2). Similarly, metamonad specific-phylogenetics with *Kipferlia*, *Spiroplasma*, and *Trepomonas* classified the HTAC subunits within $\beta 1-3$, $\gamma 1$, $\alpha 2$, $\delta 3$, $\sigma 1-3$, β -COP, γ -COP, δ -COP, or ζ -COP clades. In comparison, the identified *Giardia* components only clustered with subunits belonging to the AP-1, AP-2, or COPI subcomplexes (Figure 2.3B, Online Appendix Figures 2.1-2.4; Online Appendix Table 2.2).

AP1-4 and retromer associate, either partially or entirely, with clathrin components (clathrin heavy chain (CHC) and clathrin light chain (CLC)) to initiate triskelion cage formation around the newly budding vesicle (Barois & Bakke, 2005; Kirchhausen, 1999; Kural et al., 2012b; Zwiewka et al., 2011). AP-5, on the other hand, interacts with SPG15 and SPG11 for membrane deformation (Hirst, Borner, et al., 2013; Hirst et al., 2018). Therefore, clathrin components and SPG15 and SPG11 were also subject to comparative analyses to assess if these proteins were subject to any evolutionary co-modulation with their adaptors. These investigations revealed a synchronous loss within the auxiliary membrane deformation proteins to

have also occurred within the fornicates. Although all members retained the clathrin heavy chain, the highly conserved light chain was unexpectedly absent (Figure 2.2; Online Appendix Table 2.2). Clathrin light chain loss in *Giardia* has been reported in the literature previously (Marti et al., 2003; Zumthor et al., 2016). However, these findings extend this to be a fornicate-wide trait. SPG15 and SPG11 were also unidentified, which is present in the AP-5 bearing *Trichomonas* (Figure 2.2; Online Appendix Table 2.2). In comparison to CLC, this was anticipated due to the lack of AP-5 in fornicates.

Examination of the retromer machinery revealed patterns of conservations, but like AP-4, AP-5, TSET, and CLC, ancestral losses within components of this machinery were also evident. Although the three core cargo-selection subunits, Vps26, Vps29, and Vps35, were identified in nearly all fornicates, including diplomonads, the membrane deformation subunit, Vps5, and the traditional retromer cargo Vps10, were lost entirely (Figure 2.2; Online Appendix Table 2.2). Previous pan-eukaryotic investigations determined this absence be restricted to *Giardia* (Koumandou et al., 2011). However, the inclusion of these expanded sampling points confirms that Vps10 and Vps5 absences occur across all investigated fornicates.

Comparative genomics discerned an identical pattern of ancestral reductive evolution within the highly conserved COPII machinery. Although all lineages encode Sec23, Sec24, Sec13, Sec31, and the SAR1 GTPase, the transmembrane Sec12 GEF was lost not only in the fornicates but across the entire Metamonada, wherein both *Trichomonas* and *Monocercomonoides* lack this component (Figure 2.2; Online Appendix Table 2.2). Previous comparative genomic investigations determined Sec12 absence in *Trichomonas*; however this phylum-wide survey with the inclusion of various fornicates and *Monocercomonoides* suggests this loss occurred before the Last Metamonada Common Ancestor (Schlacht & Dacks, 2015). Sec12 is patchily conserved across eukaryotes, so this is not entirely surprising (Schlacht & Dacks, 2015).

Altogether, these findings signify that some of the previously observed absences in *Giardia*'s vesicle coat complement are due to ancestral pre-adaptations prior to the Metamonada and Fornicata common ancestors.

Figure 2.2. Coulson plot depicting the distribution of endocytic vesicle coat proteins within fornicates compared to various eukaryotic and metamonad representatives. Colored sectors indicate the presence of a specific subunit identified using comparative genomics, and in the case of HTACs, additionally validated using phylogenetics. Numbers within each sector indicate the presence of multiple paralogues. Complete losses within AP-4, AP-5, TSET, SPG11/SPG15, and partial losses in clathrin, COPII, and retromer subunits were evident across all fornicates. Parasitism-specific absences were noted within the COPI and retromer subunits, while *Giardia*-specific loss of AP-3 was also determined.

2.4.2 Losses along the path to parasitism

Apart from ancestral losses, comparative genomics also highlighted a trend of molecular streamlining concurrent with host-parasite adaptations and general endomembrane remodeling. This was apparent as losses within individual subunits, as well as in entire subcomplexes. These especially followed once diversification from the *Carpodiemonas membranifera* lineage ensued.

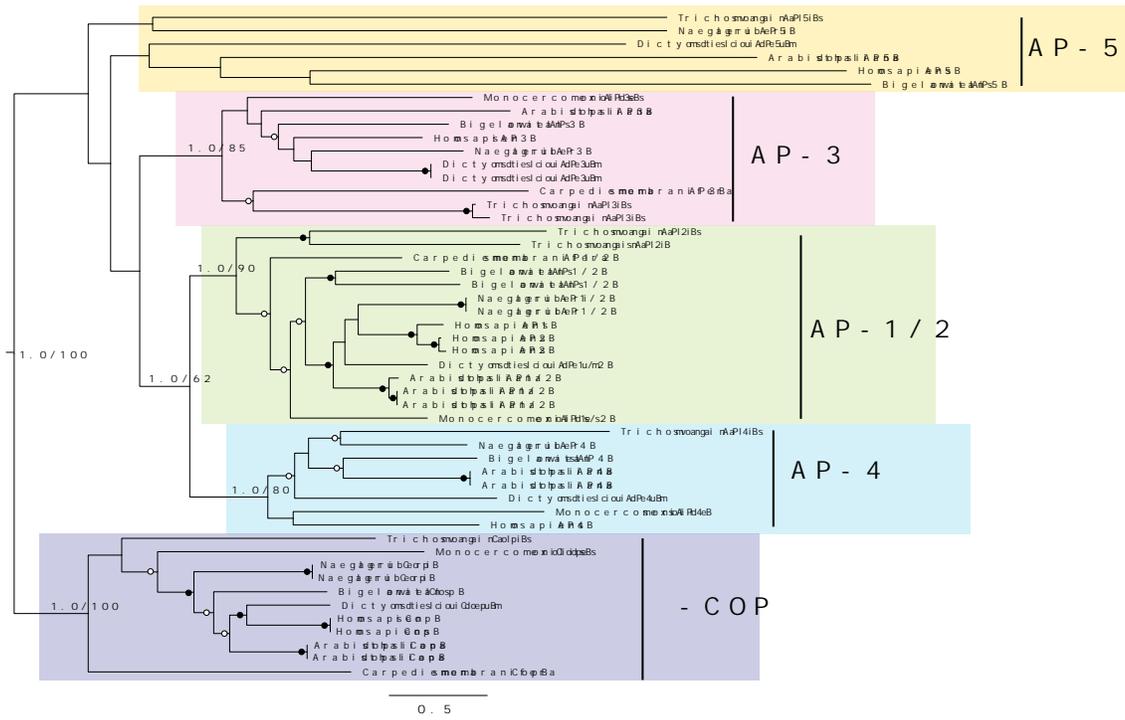
Subtle differences within individual subunits of COPI and COPII machinery highlight a pattern of progressive loss to exist prior to the diplomonads and the free-living *Kipferlia bialata*. Within the COPI machinery, the accessory ϵ -COP subunit is present in the free-living *Carpodiemonas* but is subsequently lost from the diplomonads and *Kipferlia* (Figure 2.2; Online Appendix Table 2.2). A similar trend exists within the COPII component, Sec16, which though present in *Carpodiemonas*, is absent from *Kipferlia* and the diplomonads, coinciding with a slow transition into parasitism.

Most conspicuous parasitic adaptations were evident in *Giardia*, where lineage-specific absences occurred within its adaptin repertoire. Although both *Spironucleus* and *Trepomonas* encode either complete or partial AP-3 subcomplex, it was entirely absent in all examined assemblages of *Giardia intestinalis* as well as from *Giardia muris* (Figure 2.2; Online Appendix Table 2.2). Phylogenetics additionally confirmed that none of the identified *Giardia* adaptin subunits clustered with the metamonad or diplomonad AP-3 sequences (Figure 2.3B; Online Appendix Figures 2.1-2.4). Therefore, the Fornicata repertoire of adaptins reduced from three to just two upon transition into the *Giardia* genus. Though *Spironucleus* possessed most AP-3 components, a β 3 subunit remained unidentified (Figure 2.2; Figure 2.3B; Online Appendix Table 2.2). Once again, phylogenetics ascertained confidence in this absence as all other β subunits clustered with AP-1 or AP-2 sequences (Figure 2.3B). *Spironucleus* also lacked the highly conserved COPII subunit, Sec13 (Figure 2.2). Absences could be due to genome-assembly-related false-negative artifacts and therefore should be validated in other *Spironucleus* genomes or transcriptomes once those become available. Lineage-specific absences were also evident in *Trepomonas*, which was devoid of Vps35, albeit could also be a gene expression artifact (Figure 2.2; Online Appendix Table 2.2).

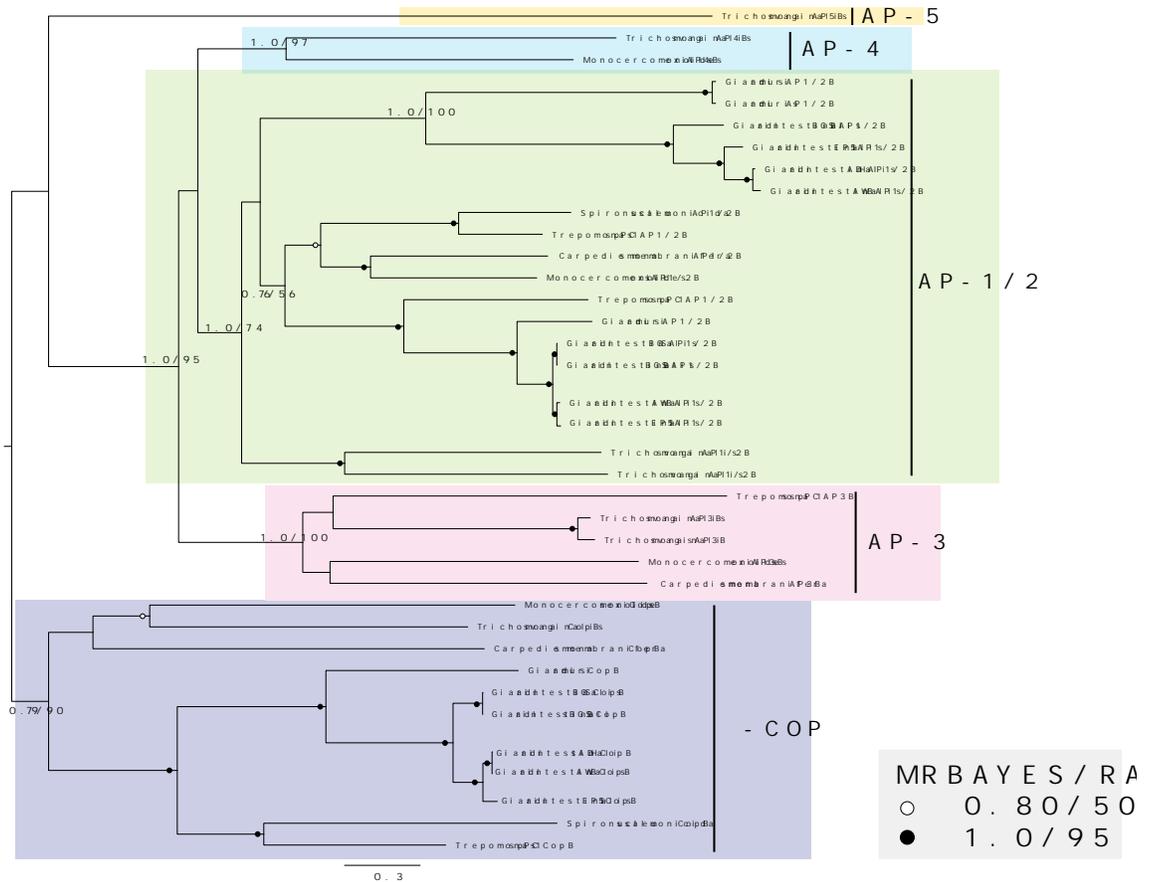
Altogether, these results confirm that although there were numerous fornicate-wide absences, losses continued to persist well within the transition into the lineage where some preceded parasitism, some coincided with the diversification into the Diplomonadida lineage (e.g., COPII-Sec16, COPI- ϵ), and some were organism- or genus-specific adaptations (e.g., AP-3 in *Giardia*, AP-3 β and Sec13 in *Spironucleus*, and Vps35 in *Trepomonas*).

Figure 2.3. Phylogenetic analyses with fornicate AP- β and β -COP subunits. (A) represents phylogenetic analyses with *Carpediemonas membranifera* AP1-3 β and β -COP subunits, which grouped within resolved pan-eukaryotic clades corresponding to these subunits with posterior probability and bootstrap support values of 1.0/90 (AP1/2 β), 1.0/85 (AP3 β), and 1.0/100 (β -COP). (B) depicts a metamonad-specific tree of the AP1-5 β and β -COP subunits, where *Kipferlia*, *Spironucleus*, *Trepomonas*, and *Giardia* sequences were classified against the newly characterized *Carpediemonas*, as well as *Trichomonas* and *Monocercomonoides* sequences. These analyses determined *Kipferlia*, *Spironucleus*, and *Trepomonas* sequences to fall within AP1/2 β (1.0/96), AP3 β (1.0/100), and β -COP (0.97/90) clades. None of the *Giardia* AP- β subunits are classified within AP-3 β . Instead, two separate clades of AP1/2 β were present, one of which was *Giardia*-specific (1.0/100), suggesting for a genus-specific duplication event to have occurred within this subunit.

A



B



2.4.3 Lineage and parasite-specific expansions within the adaptin and COPII machinery

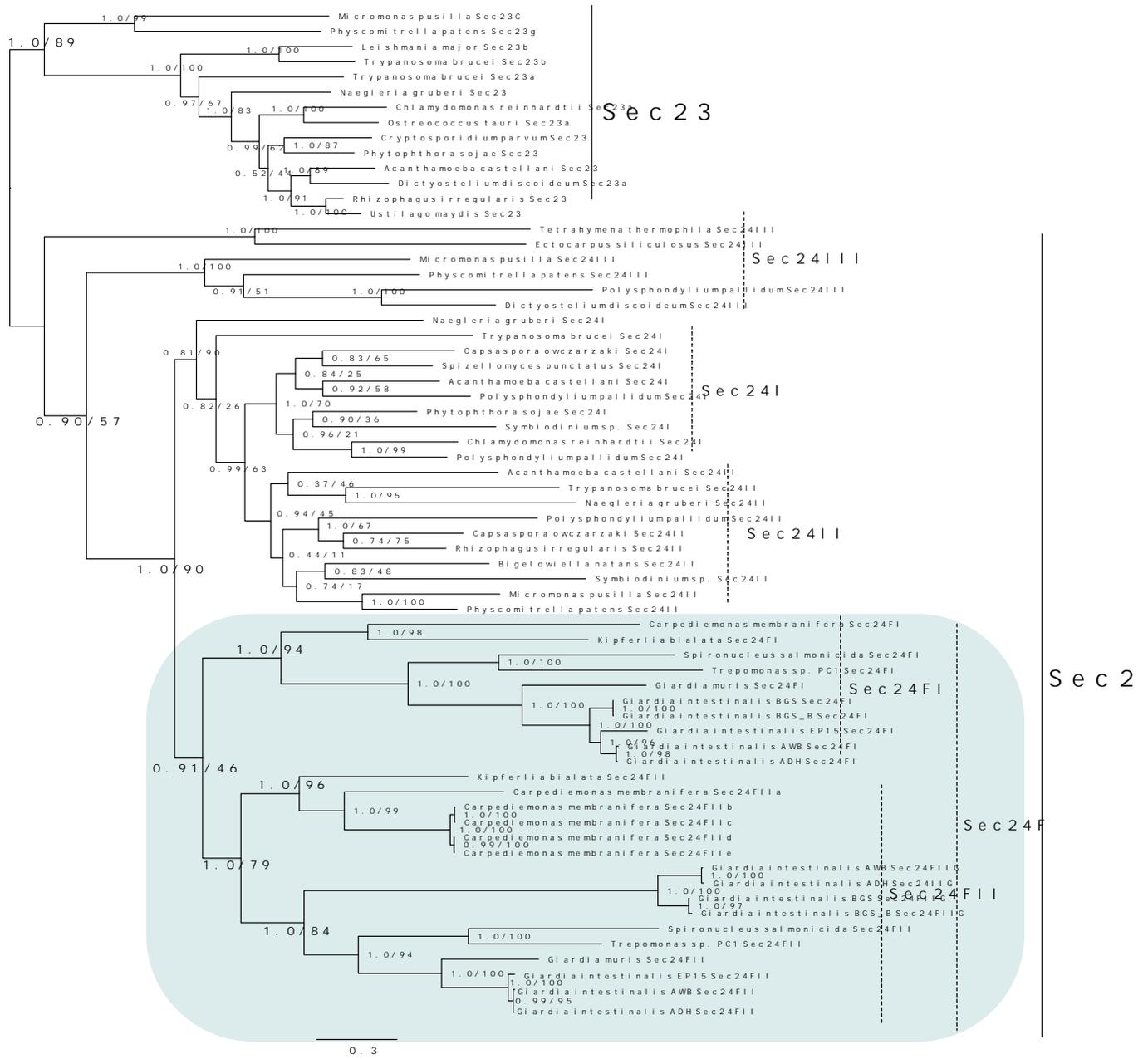
An unanticipated finding of these surveys was that evolutionary plasticity existed within the vesicle coats to accommodate expansions aside from molecular pruning. Duplications occurred within the COPII-Sec24 and AP- β 1/2 to yield fornicate- and *Giardia*- specific paralogues, respectively, confirmed through phylogenetics.

Homology searching identified numerous copies of COPII-Sec24 across all fornicates (Figure 2.2; Online Appendix Table 2.2). Previous pan-eukaryotic investigations elucidated Sec24 to have undergone pre-LECA expansion events to yield three ancestral paralogs termed Sec24I-III (Schlacht & Dacks, 2015). In order to assess how the fornicate Sec24 paralogues are classified within these three ancient families, a phylogenetic investigation using previously characterized pan-eukaryotic Sec23 and Sec24 sequences was performed (Figure 2.4). These determined fornicate Sec24 sequences to be related to pan-eukaryotic Sec24I/II clades (Figure 2.4). Notably, the fornicate Sec24I/II sequences underwent duplication to yield two fornicate-specific paralogues, termed Sec24FI and Sec24FII (*i.e.*, Sec24 in Fornicata I and II) that clustered as two independent clades nested within pan-eukaryotic Sec24I-III sequences (Figure 2.4). Interestingly, these analyses also determined several of the *Carpodiemonas* Sec24 sequences resulting from independent organism-specific expansion events within its Sec24FII (Figure 2.4). This was not only restricted to *Carpodiemonas* as both *Giardia intestinalis* and *Giardia muris* expanded their repertoires to yield *Giardia*-specific paralogs termed Sec24FIIG (*i.e.*, Sec24FII in *Giardia*) (Figure 2.4). This was evident by the presence of two separate clusters of *Giardia* sequences within the diplomonad clade of Sec24FII (Figure 2.4). These analyses also discerned distinct differences within the *Giardia* repertoire of Sec24FIIG. Notably, assemblage B isolates, BGS and BGS_B, lacked one of two Sec24FII paralogues, suggesting that a secondary loss within this protein likely occurred in this assemblage.

Phylogenetic investigations also unveiled that COPII-Sec24 is not the only protein illustrative of *Giardia*-specific expansions. Another included duplication within the giardial AP- β 1/2 subunit, a component that is typically shared between AP-1 and AP-2 subcomplexes in most eukaryotes. Homology searching revealed the presence of at least two paralogs across both *Giardia muris* and all *Giardia intestinalis* isolates (Figure 2.2). However, through phylogenetics, it was determined that apart from encoding a canonical β 1/2 that is present in all fornicates, *Giardia* spp. underwent a duplication event to yield additional *Giardia*-specific paralogues that cluster separately within the metamonad β 1/2 clade (Figure 2.3B)

These results demonstrate that though the endocytic machinery is gradually dispensed, the lineage was also subject to expansions in subunits that are typically conserved across other eukaryotes.

Figure 2.4. Phylogenetic analyses with the fornicate Sec24 paralogues against pan-eukaryotic Sec24I-III. Previous evolutionary bioinformatic investigations across the eukaryotic diversity elucidated the presence of three ancestral Sec24 paralogues (Sec24I-III). Comparative genomic investigations in this survey identified several copies of Sec24 sequences to be encoded within all fornicates, and therefore, were subject to phylogenetic classifications against the pan-eukaryotic paralogs. Fornicate sequences clustered as a monophyletic group closely related to the LECA-specific Sec24I/II (highlighted in turquoise lozenge). A combined posterior probability and bootstrap node support value of 1.0/90 was noted for this fornicate-specific clade. Closer examination of this clade reveals that lineage-specific duplications additional occurred to yield fornicate-specific paralogues, termed Sec24FI and Sec24FII (*i.e.*, Sec24 in Fornicata I and II). *Carpediemonas* was identified to possess numerous copies of Sec24FII, while *Giardia* spp. also underwent additional duplication events within its Sec24FII repertoire to yield *Giardia*-specific Sec24FIIG (*i.e.*, Sec24FII in *Giardia*). Differences were also noted in the Sec24II complement encoded between the two *Giardia* assemblages wherein assemblage B isolates lacked one of the two paralogues.



2.5 Discussion

The molecular evolution of the protein machinery that governs coated vesicle formation within the ER-Golgi secretory pathway and between the post-Golgi endo-lysosomal compartments was discerned through this study. These vesicle coats have proto-coatomeric origins and are a family of evolutionarily related proteins that arose by mechanisms dictated by the organelle paralogy hypothesis (Dacks et al., 2008; Dacks & Field, 2007). As discussed in Chapter 1, the OPH describes autogenous origins of the eukaryotic endomembrane complexity from an undifferentiated organelle through incremental gene accumulation followed by their sub-specialization for the emergence of distinct membrane-bound trafficking compartments (Dacks et al., 2008; Dacks & Field, 2007). The vesicle coats have endosomal and secretory functions between functionally distinct organellar interfaces. Therefore, by recapitulating the molecular complement of the players involved, one can predict the diversity of the trafficking processes that likely occurs within that cell. The findings from this investigation can be used to assess the evolution of the endocytic system in fornicates and that which underpins *Giardia's* minimal trafficking landscape. Overall, this bioinformatic survey demonstrated three discrete evolutionary patterns to have shaped the existing complement of the vesicle coats in fornicates: 1) ancestral reduction, 2) parasitism-associated molecular streamlining, and 3) lineage-specific duplications (Figure 2.5). Based on this protein repertoire, it can be determined that losses are greatest within protein complexes that participate within the endo-lysosomal pathways. In contrast, conservation and even expansions exist in machinery required in early exo- and endocytic transport (Figure 2.5).

2.5.1 Reductive evolution in fornicates may be indicative of redundant or dispensable aspects of the early and late endosomal system

Loss and retention of several HTAC families and retromer components in fornicates resemble trends of dispensability that exist in other parts of the eukaryotic tree of life. Previous comparative genomics unveiled TSET and AP-5 to be most likely lost within Opisthokonta (e.g., *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster*) and green algal representatives (i.e., *Volvox carteri* and *Micromonas pusilla*) (Hirst et al., 2014). AP-4 is the next member of this family that is subject to molecular streamlining in many organisms, as is the case in fungi (e.g., *S. cerevisiae*, *S. pombe*, and microsporidia) and metazoan species (e.g., *Drosophila* and *C. elegans*). Finally, though AP-3 is comparatively conserved, instances of reductive evolution within this subcomplex are also evident such as in green algae (i.e., *Volvox carteri*) and parasites (e.g., *Plasmodium* and microsporidia) (Barlow et al., 2014; Hirst et al., 2014; Nevin & Dacks, 2009). These previous investigations also revealed a high degree of conservation within AP-1, AP-2, and COPI subcomplexes. Only a few instances of AP-1 and AP-2 loss have been reported and so far restricted to kinetoplastid parasites (e.g., loss of AP-2 in certain species of *Trypanosoma*) and obligate endosymbionts or intracellular parasites (e.g., loss of AP-1 in *Perkinsela* sp. and loss of AP-2 in microsporidia) (Barlow et al., 2014; Manna et al., 2013; Herman and Dacks, personal communication). Retromer components, Vps5 and Vps10, are also lost frequently within the SAR,

Cryptophyta, Discoba, and Amoebozoa supergroups, and hence have been suggested to be 'patchy' in their conservation (Koumandou et al., 2011).

The absence in trafficking complexes may be indicative of several scenarios. The first is that the missing machinery was redundant in cellular function and therefore subject to molecular sweeping. Another is the lack of an endocytic organelle that may be present in model eukaryotes with specialized or complex endo-lysosomal pathways where these proteins have been demonstrated to function. The final is a combination of these possibilities, wherein secondary remodeling within trafficking pathways followed by a redirection of the existing machinery for lineage-specific functions could have occurred. Examples for each can be represented by eclectic functions of adaptins and retromer that continue to be unveiled through functional studies in eukaryotes previously not studied. It is feasible for these possibilities to also exist within the fornicates.

In metazoan systems, both AP-1 and AP-4 seem to facilitate clathrin-dependent transport between the TGN and endosomes for polarized sorting within the cell upon ARF1 recruitment and PI(4)P binding (Boehm et al., 2001; Burgess et al., 2011; Simmen et al., 2002; Starnes & Rothman, 1993). Several studies investigating TSET function in plants have implicated its role in clathrin-mediated endocytosis at the plasma membrane in concert with AP-2, but to also be involved in autophagosome biogenesis and trafficking (Gadeyne et al., 2014; Wang et al., 2021; Wang et al., 2019). The absence of TSET in fornicates could suggest that clathrin-mediated endocytosis at the plasma membrane is mitigated by AP-2 alone or a lack of autophagosomes, or both.

A minimal retromer complex in *Trypanosoma brucei* can perform endosomal functions even in the absence of Vps10-sortillin (Koumandou et al., 2011). Existing cargo-selection retromer components in plants and Metazoa also demonstrate plasticity in their ability to recognize and traffic various cargoes at lineage-specific compartments. Vps35 in plants interacts with and recycles vacuolar sorting receptors (VSRs) at the pre-vacuolar compartment (PVCs), while in Metazoa, sorting nexins mediate lysosomal targeting of epidermal growth factor receptors (EGFR) (Cavet et al., 2008; Oliviussen et al., 2006). Similarly, it is possible that in the absence of Vps10, encoded retromer subunits may be functioning to transport alternative receptors at fornicate-specific organelles.

AP-3 and the more recently elucidated AP-5 share overlapping roles in LRO biogenesis and in lysosomal trafficking and homeostasis (Hirst et al., 2018; Odorizzi et al., 1998). It is possible that AP-5 functions in the ancestor of fornicates were redundant to the roles performed by AP-3, and therefore, dispensed from the molecular repertoire. In fornicate-lineages bearing distinct endosomes or multivesicular bodies such as *Carpodomonas*, *Kipferlia*, and *Spironucleus*, AP-3 alone may be sufficient to facilitate trafficking functions at these organelles (Simpson & Patterson, 1999; Yubuki et al., 2013; Santos et al., in preparation). Finally, the absence of AP-4 across all free-living and parasitic fornicates could be ancillary to the absence of a defined stacked Golgi compartment. Although Golgi dictyosomes are visible in *Carpodomonas*, the loss of a structurally defined TGN across the rest of the fornicates signals ongoing dissipation within the molecular complement that assembles at that interface.

2.5.2 Duplications within the existing machinery for a diversified role at the ER-Golgi

One of the more unexpected findings of this investigation was the numerous lineage-specific duplications within the fornicate COPII-Sec24 and giardial adaptin subunits.

Sec24 forms heterodimers with Sec23 to form pre-budding complexes by recognizing cargo to recruit Sec13 and Sec31 to form a polyhedral coat for vesicle budding (Sato & Nakano, 2007). Pan-eukaryotic investigations have determined the presence of ancient paralogues of Sec24 to have been encoded in the LECA (Schlacht & Dacks, 2015). Cellular roles and differences in their molecular functions have been tested through mutagenesis experiments in mammalian systems. These analyses have indicated distinct and overlapping functions for the different Sec24 paralogues at the ERGIC and for each to have variable affinities towards different cargoes (Adams, 2014; Wendeler et al., 2007). Though the fornicate sequences did not classify within the different Sec24I-III clades, they still represent novel-lineage duplications. Like the mammalian Sec24 paralogues, these may perform diversified roles at the ER membranes by recognizing and trafficking different types of cargoes. It also may be that these proteins functionally replaced the missing Sec16 or Sec12 that typically scaffold or stabilize the rest of the machinery upon its assembly at the ER-exit sites (Sprangers & Rabouille, 2015; Whittle & Schwartz, 2010). Both of these proteins are also involved in the regulation of the SAR1 GTPase activity. Therefore, the possibility exists that one or more of the encoded Sec24F paralogues perform these roles, as has been shown with yeast Sec24 (Kung et al., 2012).

In this study, phylogenetics also pinpointed giardial expansions within the AP- β 1/2 subunit. In most eukaryotes, this component complexes with both AP-1 and AP-2 for clathrin recruitment and binding (Sosa et al., 2012). However, exceptions have been noted in vertebrates, embryophytes, and trypanosomatids which underwent lineage-specific duplications to give rise to additional β 1/2 subunits (Dacks et al., 2008; Larson et al., 2019; Manna et al., 2013). In humans, β 1/2 paralogues vary in their appendages that differentially recognize and bind to different adaptin subcomplexes (*e.g.*, AP-1 and AP-2) and accessory proteins such as AP180, Epsin, and Eps15 (Owen et al., 2000). Similar subcomplex specificity of the different giardial β 1/2 paralogues may exist towards its AP-1 and 2. It could also suggest a molecular interplay involving the recruitment of unique repertoires of accessory proteins to perform trafficking functions at distinct giardial compartments.

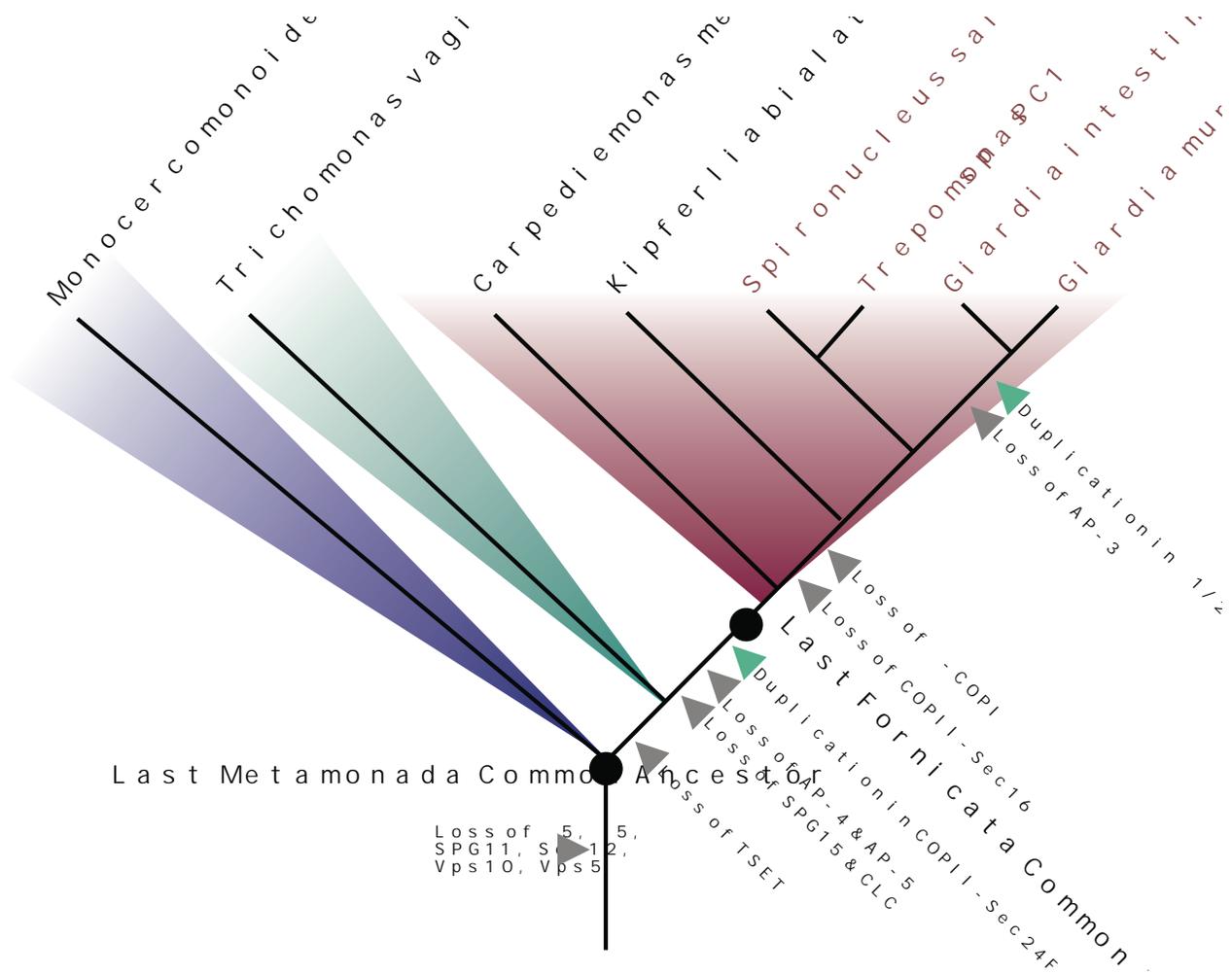


Figure 2.5. Reductive and lineage-specific evolution of the vesicle coats in metamonads. Streamlined vesicle coat repertoire in *Giardia* is a product of ancestral and lineage-specific losses. These were metamonad-wide (e.g., components of the AP-5 and TSET subcomplex, SPG11, COPII-Sec12, and Retromer-Vps10 and Vps5), Fornicata-specific (e.g., loss of AP-4 and AP-5, SPG15, and CLC), parasitism-specific (e.g., COPII-Sec16 and ϵ -COP), and *Giardia*-specific (e.g., AP-3). Aside from losses, this vesicle coat complement is also defined by lineage-specific duplications (e.g., Sec24F in fornicates to yield Sec24FI and FII and additional duplication of Sec24FII in *Giardia* to yield Sec24FIIG; *Giardia*-specific duplication in AP- β 1/2 to yield two paralogues of β 1/2). Diplomonads are indicated in red font.

2.5.3 Known functions of the vesicle coats within *Giardia*

Some of the machinery investigated in this study has been subject to *in vitro* probing in the lab strain of *Giardia* to better understand their roles in this Golgi- and endosome-lacking parasite. Additionally, the clathrin heavy chain functions in *Giardia* and its assembly mechanisms in the absence of clathrin-light chain and clathrin-coated vesicles were elusive. Combining these previous findings can help postulate the homology of *Giardia*-specific organelles and contribute to our general understanding of the plasticity and conservation of vesicle formation processes in eukaryotes.

Immunofluorescence microscopy and co-localization analyses have determined components of the *Giardia* AP-1 to be implicated in the trafficking of lysosomal enzymes to the peripheral vacuoles in the vegetative trophozoites, and for transport of encystation-specific cysteine proteases (ESCP), also to the PVs, during encystation (Touz et al., 2004). The specific mechanism for cargo recognition was also elucidated, wherein AP-1 μ recognizes a tyrosine-based motif on ESCPs at specific ER-domains (Touz et al., 2004). These biochemical and microscopy studies concluded the necessity for *Giardia* AP-1 in anterograde transport of endo-lysosomal proteins to the PVs. AP-2, on the other hand, was localized to the cell periphery and to be in association with the giardial clathrin heavy chain (Rivero et al., 2010; Zumthor et al., 2016). Uptake experiments with horseradish peroxidase and Dextran confirmed that cells subject to μ 2 repression failed to recycle BODIPY-low density lipoprotein membrane receptors and internalize them within the PVs. This suggests that AP-2 in *Giardia* participates in surface membrane receptor recycling (Rivero et al., 2010, 2012). AP-2 is also posited to be required during encystation as μ 2 gene silencing also disrupted cyst-wall formation (Rivero et al., 2010).

Similar to AP-2, the role of retromer in *Giardia* is critical for retrograde trafficking of acid hydrolase receptors from the PVs through specific interactions with the cytosolic domains of the giardial Vps35 (Miras et al., 2013). Unlike AP-1, which mediates anterograde transport from ER to PVs, *Giardia* retromer functions in the opposite direction as determined through co-localization microscopy investigations (Miras et al., 2013). Although *Giardia* lacks the BAR domain-containing Vps5 or SNX proteins, the rest of the cargo-selective subunits, Vps26, Vps35, and, Vps29 were membrane-associated at the ER and the PVs as distinct puncta (Miras et al., 2013). Yeast-two hybrid investigations determined direct and tight interactions between the *Giardia* retromer and its cargo at these compartments (Cai et al., 2008; Delprato & Lambright, 2007; Hegedűs et al., 2016; Miras et al., 2013; Siniosoglou et al., 2000; Yoshimura et al., 2010).

Unlike adaptins and retromer, COPI and COPII in *Giardia* are primarily needed during encystation, especially for the neogenesis and maturation of encystation-specific vesicles (ESVs). The giardial COPII and its subunits have been implicated in early encystation and for the production of nascent ESVs at the ER. The SAR1 GTPase, Sec23, and Sec24, all showed punctate localization in regions corresponding to the ER-exit sites in the vegetative trophozoites (Faso et al., 2013; Stefanic et al., 2009; Zamponi et al., 2017). Once *in vitro* encystation was triggered, these components recognized newly synthesized cyst-wall proteins as their cargo to initiate nucleation of the rest of the COPII machinery for ESV biogenesis and scission from ERES (Faso et al., 2013). COPII-Sec31 colocalized with the nascent ESVs during the early

stages but is later replaced by COPI and ARF1. Like COPII, biochemical and localization studies with the *Giardia* β -COP subunit showed association to the perinuclear ER regions in conjunction with ARF1 during the trophozoite stages, whereas during late encystation, COPI localization switches to instead associate with mature ESVs (Luján et al., 1995; Marti et al., 2003). COPI assembly at the ER during the trophozoite stage and associations with ESVs during encystation, were Brefeldin-A sensitive (Luján et al., 1995).

Finally, as previously discussed in Chapter 1, the role of clathrin heavy chain in *Giardia* has been subject to comprehensive high-resolution microscopy and proteomics investigations (Zumthor et al., 2016). These have concluded its involvement in bulk-flow uptake at the cell periphery in association with dynamin and AP-2 (Gaechter et al., 2008; Zumthor et al., 2016). Unlike canonical clathrin functions, *Giardia* has abandoned the use of this protein for triskelion cage formation and instead forms static assemblies in the intervening PV-PM spaces for a steady-state pore formation between the two compartments, likely to allow for an inward uptake of host milieu (Zumthor et al., 2016).

Overall, a working model derived from these findings suggests that although *Giardia* is missing many canonical features of what constitutes a typical endomembrane system, the vesicle coats demonstrate plasticity in their functions at the parasite-specific organelles. Functional parallels also exist with canonical trafficking pathways. PVs represent a minimal endo-lysosomal system based on their association with clathrin heavy chain, AP-1, AP-2, and retromer. COPI and COPII interactions with the ESVs and assembly at the ERES suggest ESVs have Golgi-like characteristics. Whether these features and functions are analogous or homologous in status remains unknown. Nonetheless, it is evident that the simplified protein complement continues to perform comparable functions in the absence of a typical endocytic complexity and represents the core set of machinery that is necessary within this lineage.

2.6 Conclusions

This comparative genomic and phylogenetic survey of the vesicle coats in Fornicata has allowed for the identification of the molecular complement and its evolution in this lineage. Altogether, this investigation determined that the minimal vesicle coat repertoire in *Giardia* is due to pre-fornicate losses, parasitism-specific streamlining, and lineage-specific expansions. These findings stipulate reductive evolution within the giardial endomembrane system to have been neither sudden nor exclusive.

CHAPTER 3

The reduced ARF regulatory system in *Giardia intestinalis* pre-dates the transition to parasitism in the lineage Fornicata

3.1 Preface

The previous chapter explored the molecular evolution of the vesicle coat proteins to assess how the early and late endosomal system has evolved in *Giardia* and its relatives. However, the coats are not sufficient to initiate vesicle formation processes. Active GTP-bound small GTPases regulate nucleation and assembly of adaptor and membrane deformation proteins onto endomembranes and, therefore, are also vital players in endocytic trafficking (Donaldson & Jackson, 2011). Small GTPases, particularly those belonging to the Ras-superfamily, act as molecular switches to control the tempo, mode, and fidelity of vesicle formation processes throughout the cell (Donaldson & Jackson, 2011). These 'on' and 'off' states are controlled by guanine exchange factors (GEFs) and GTPase activating proteins (GAPs), respectively, to promote GTPase cycling between membranes and cytoplasm (Sztul et al., 2019). GEFs exchange the bound GDP to GTP to induce conformational changes to activate the GTPase, permitting membrane recruitment. GAPs hydrolyze a phosphate group from the GTP back to GDP for GTPase inactivation for membrane dissociation and cycling back into the cytosol (Donaldson & Jackson, 2011; Sztul et al., 2019).

Each of the coat families discussed in the previous chapter requires recruitment onto their respective membranes (*i.e.*, ER, Golgi, or endosomes) with the help of a small GTPase belonging to the ARF/SAR/ARL family. As discussed previously, COPII requires SAR1 for ER-assembly upon GEF-mediated activation by Sec12 and dissociation by the Sec23 GAP. On the other hand, assembly dynamics of heterotetrameric complexes and retromer are mediated by ARF1 and ARF6 and their GEFs and GAPs, which altogether comprise the ARF regulatory system. ARF1 GTPase regulates COPI, AP-1, AP-3, and AP-4, while AP-2 and retromer are ARF6-dependent (Boehm et al., 2001; Goldberg, 1999; Marquer et al., 2016; Ooi et al., 1998; Paleotti et al., 2005; Ren et al., 2013). The more recently discovered AP-5 and TSET do not have their regulation mechanism determined as of yet. Nonetheless, the spatiotemporal dynamics of the majority of HTACs and retromer rely on the ARF regulatory system.

Though the evolution and molecular complement of SAR1, Sec12, and Sec23 in fornicates were mapped in the previous chapter, proteins belonging to ARF regulatory system were not assessed. Therefore, this chapter performed an identical comparative genomic and phylogenetic survey of the ARFs, ARFGAPs, and ARFGEFs in Fornicata to assess how this endosomal machinery evolved.

The work presented in this chapter was published as a research article: Pipaliya, S. V., Thompson, L. A., & Dacks, J. B. (2021). The reduced ARF regulatory system in *Giardia intestinalis* pre-dates the transition to parasitism in the lineage Fornicata. *International Journal for Parasitology*, 51(10), 825-839. <https://doi.org/10.1016/j.ijpara.2021.02.004>. I performed all bioinformatic analyses (comparative

genomics and phylogenetics) with the ARFs, ARF GAPs, and ARF GEFs in fornicates. Comparative genomic analyses to update the distribution of ADAP across eukaryotes were surveyed by L. Alexa Thompson. Some redundancies pertaining to the descriptions of *Giardia*'s biomedical relevance, its endomembrane features, and background on the fornicate lineage, that were previously described in Chapter 1 are also in the following introductory section.

3.2 Introduction

Giardia intestinalis is a gut pathogen responsible for the diarrheal disease giardiasis, which according to the World Health Organization, affects approximately 300 million individuals across the globe annually (Lanata et al., 2013). As a result, the parasite is of tremendous global health importance. *Giardia* relies heavily on its endomembrane organelles and the membrane trafficking system for pathogenesis and disease establishment throughout its lifecycle, both as a trophozoite and the environmentally-resistant cyst stage (Faso & Hehl, 2019). During the trophozoite stage, the parasites secrete virulence factors such as cysteine proteases for gut brush border disruption and variant surface protein and arginine deiminase for immune evasion. During encystation, the parasites secrete and deposit cyst-wall proteins onto the exterior of the cell membrane (Faso et al., 2013). These processes would typically require coordinated intracellular processing of virulence genes at the endoplasmic reticulum, and the Golgi followed by subsequent cargo transport within endocytic carriers. Notably, however, *Giardia* is entirely devoid of a canonical stacked Golgi apparatus, as well as any distinguishable canonical endosomal, lysosomal, or peroxisomal compartments (Marti et al., 2003).

Instead, microscopy analyses have revealed this parasite to possess unique *Giardia*-specific organelles, namely the peripheral vacuoles, encystation-specific vesicles, mitosomes, and a highly expanded network of the tubulovesicular endoplasmic reticulum (Faso & Hehl, 2011; Konrad et al., 2010). Subsequently, molecular functional studies elucidated the roles of these endomembrane organelles in the context of the lifecycle and pathogenicity mechanisms of *Giardia*. Collective results of these studies have proposed the peripheral vacuoles to be modified fused forms of the late endosomes and lysosomes, which are a major interface for cargo uptake and sorting within the parasite (Feely & Dyer, 1987; Zumthor et al., 2016). Encystation-specific vesicles, on the other hand, have been suggested as analogs of the Golgi compartment due to their association with protein markers that normally associate with this organelle and are necessary for the transport of cyst-wall material to the cell surface prior to encystation (Faso et al., 2013; Stefanic et al., 2009). Finally, the endoplasmic reticulum (ER) in *Giardia* is also thought to have functions beyond protein synthesis. Instead, it is suggested to be dynamic in its capacity to sort cargo taken up by parasites (Abodeely et al., 2009; Zamponi et al., 2017). Assessing the dynamics and the timepoints of the modifications for the *Giardia* membrane trafficking system is one crucial aspect of understanding the cell biology of *Giardia* and its evolution to a parasitic lifestyle.

Giardia belongs to the phylum Metamonada and subphylum Fornicata (Kolisko et al., 2008). Fornicata encompasses both free-living and parasitic lineages (Kolisko et al., 2008). These include free-

living heterotrophs, *Carpodiemonas membranifera*, and close relatives collectively known as *Carpodiemonas*-like organisms (CLOs), as well as retortamonads which are gut endobionts of animals (Takishita et al., 2012). Fornicata also includes a largely parasitic clade known as the diplomonads to which *Giardia* belongs, together with the fish pathogen *Spironucleus salmonicida* and the secondarily free-living organism, *Trepomonas* sp. PC1 (Kolisko et al., 2008; Xu et al., 2014, 2016). Recently, a clade of heterotrophic flagellates, the barthelonids, have been placed as sister to the entire fornicate lineage based on the analysis of a single representative transcriptome (Yazaki et al., 2020). Information about the ultrastructure of this lineage is sparse, but they are an intriguing sample point for evolutionary analysis. By contrast, rather extensive ultrastructural analysis has revealed varying complexity within the cellular organization across the Fornicata, including in membrane-trafficking organelles. While *Carpodiemonas membranifera* is the only member possessing a classical stacked Golgi apparatus, distinct endosomal and vesicular bodies are frequently observed in many organisms within the Fornicata (Einarsson et al., 2016; Park et al., 2010; Simpson & Patterson, 1999; Yubuki et al., 2007, 2013, 2016).

Regulation of directional cargo movement is a key function of the membrane trafficking system and is coordinated by a variety of protein complexes (Bonifacino & Glick, 2004). This includes biogenesis of vesicles, cargo loading, and fusion at donor compartments, all of which are complex and intricate processes relying on membrane deformation and fusion protein complexes (Bonifacino & Glick, 2004). The function and regulation of this machinery are further reliant on molecular switches, particularly families of small GTPases. One such family necessary for regulating cargo vesicle trafficking is the ADP ribosylation factor (ARF) GTPase family, a member of the larger ADP ribosylation factor /secretion-associated Ras-related protein/ ADP ribosylation factor-like (ARF/SAR/ARL) family of GTPases (Kahn et al., 2006). ARF activity is regulated through the cycling of GTP and GDP. When GTP binds, ARF undergoes a conformational change into an 'active' membrane-bound state. Following GTP hydrolysis to GDP, the ARF-GDP complex is inactive and dissociates from the membrane (Donaldson & Jackson, 2011) (Figure 3.1A). The GTP and GDP cycling is regulated by effector and activating proteins, where Sec7 domain-containing guanine nucleotide exchange factors (GEFs) replace bound GDP with a GTP while GTPase-activating proteins (GAPs) hydrolyze GTP back to GDP to inactivate ARF activity (Sztul et al., 2019) (Figure. 3.1A).

Phylogenetic and *in vitro* studies have classified ARF proteins into Class I (ARF1, 2, and 3), Class II (ARF4 and 5), and Class III (ARF6) (Li et al., 2004). In mammalian cells, different ARFs have distinct intracellular localization within the membrane trafficking system and regulate the activity of different coat proteins, tethers, and enzymes, although many of these ARFs arose from lineage-specific expansions in animals and their single-celled relatives (*i.e.*, holozoans) (Jackson & Bouvet, 2014) (Figure 3.1B). However, across many eukaryotes, ARF1 is the best described. ARF1 is considered to be primarily Golgi-associated and necessary for the regulation of several Golgi and clathrin-associated coat protein complexes such as COPI, COPII, and adaptin proteins, although ARF1 is also thought to be a generalist (Jackson & Bouvet, 2014). The other studied ARF is ARF6, which primarily regulates endosomal and plasma membrane proteins (Montagnac et al., 2009). Outside of mammalian systems, many of these functions have been

elucidated in other eukaryotic cells, most notably *Saccharomyces cerevisiae*, *Trypanosoma brucei*, *Toxoplasma gondii*, and *Dictyostelium discoideum* (Chen et al., 2010; Liendo et al., 2001; Price et al., 2007; Yahara et al., 2001). Examination of the published literature and previous comparative genomic and phylogenetic work with the ARF/SAR/ARL family point towards ARF1 as conserved across many eukaryotes and likely to have been present in the Last Eukaryotic Common Ancestor (LECA) (Li et al., 2004).

Arf GTPase-activating proteins (ARF GAPs) play an important role within the ARF GTPase cycle. They are necessary since ARFs lack intrinsic GTPase activity that allows for the hydrolysis of the bound GTP to GDP and thus the termination of the GTP-effector interactions (Spang et al., 2010) (Figure 3.1A). Outside of the GAP domain, there is extensive variation and additional diagnostic domains that define each of the ARF GAP paralogues. Comprehensive bioinformatic analyses to examine the distribution and evolution of the ARF GAPs have shown that at least six ARF GAPs (Stromal Membrane Associated Protein (SMAP), ARFGAP1, ARFGAP2/3, ArfGAP with GTPase, ANK repeat and PH domain-containing protein (AGAP), ArfGAP with FG repeats (AGFG), and ARFGAP_C2) were present in the LECA (Schlacht et al., 2013). However, there remained uncertainty around the taxonomic distribution of the ARF GAP subfamily ArfGAP with dual PH domains (ADAP) and whether it is also an ancient component is unclear (Schlacht et al., 2013).

Another important set of ARF regulators are the Sec7 guanine exchange nucleotide factors (ARF GEFs). Three ARF GEF families are ancestral and common to the larger eukaryotic diversity, Brefeldin-inhibited GEF (BIG), Golgi Brefeldin A-resistant factor 1 (GBF1), and Cytohesin (Pipaliya et al., 2019). These have non-overlapping intracellular functions and localizations within the cell (Sztul et al., 2019). In mammalian cells, GBF1 is necessary to cycle class I and class II ARF proteins on COPI-coated vesicles between the *cis*-Golgi and the ER-Golgi Intermediate Compartment (ERGIC), although alternative roles have been observed in plants (Quilty et al., 2014; Richter et al., 2012). Disruption in GBF1 activity via Brefeldin-A often results in inhibition of ARF1-mediated trafficking, resulting in an aberrant Golgi morphology (Niu et al., 2005). BIG is necessary for the trafficking within the Golgi cisternae and *trans*-Golgi trafficking and facilitates vesicle scission and fusion processes for clathrin- and adaptin-associated vesicles (Kitakura et al., 2017; Shin et al., 2004). Finally, Cytohesin and its taxon-specific additional paralogues are necessary for diversity of endocytic trafficking (Sztul et al., 2019). Aside from the presence of a Sec7 domain, ARF GEF proteins possess a variety of additional characteristic domains, useful for further verifying the identities of candidate proteins (Pipaliya et al., 2019). A shortcoming of previous investigations of the evolutionary perspective of ARF GAP and ARF GEF complement and evolution in metamonads was the limited taxon sampling, with only *Trichomonas vaginalis* and a single isolate of *Giardia intestinalis* having been investigated (Pipaliya et al., 2019; Schlacht et al., 2013).

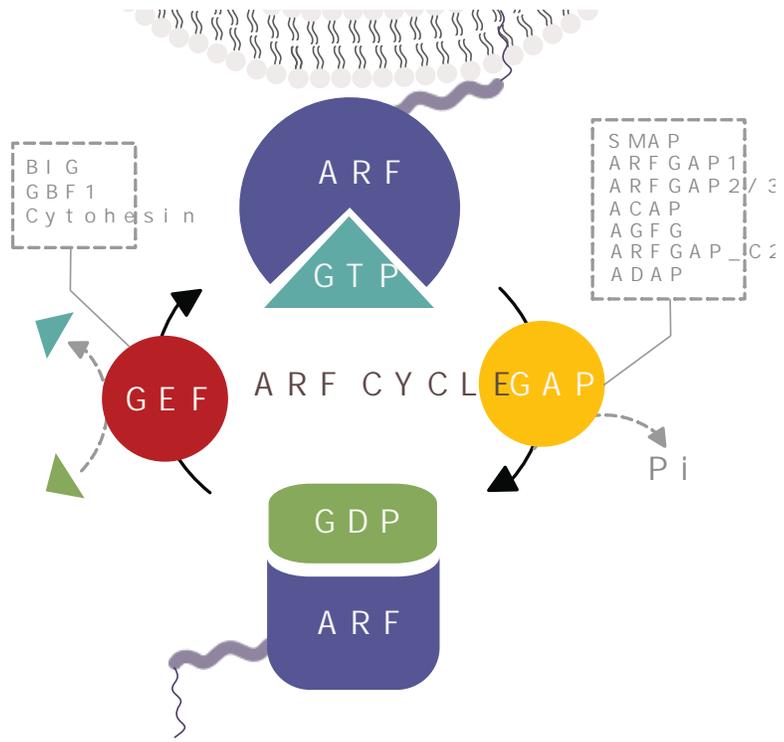
Especially for parasitic lineages, where tools for *in vivo* experimentation are generally limited, genomic and phylogenetic methods serve as important starting points towards the identification of protein complement and prediction of intracellular organization, complexity, and machinery. Furthermore, there is

both a general theoretical expectation of simplification correlated with parasitism, and past observations of reduced ARF regulatory complements, not only in *Giardia* but also in other parasitic representatives (Pipaliya et al., 2019; Schlacht et al., 2013). This correlation of parasitism and reduction sets up a null hypothesis of the transition to parasitism as the driver of simplification, but this need not be the case. For this reason, an evolutionary perspective and more intensive taxon sampling are key to understanding the timepoints and dynamics for how the trafficking system in the fornicate lineage underwent modifications and streamlining to what is now observed in *Giardia*, and whether these truly correlate with adoption of a parasitic lifestyle.

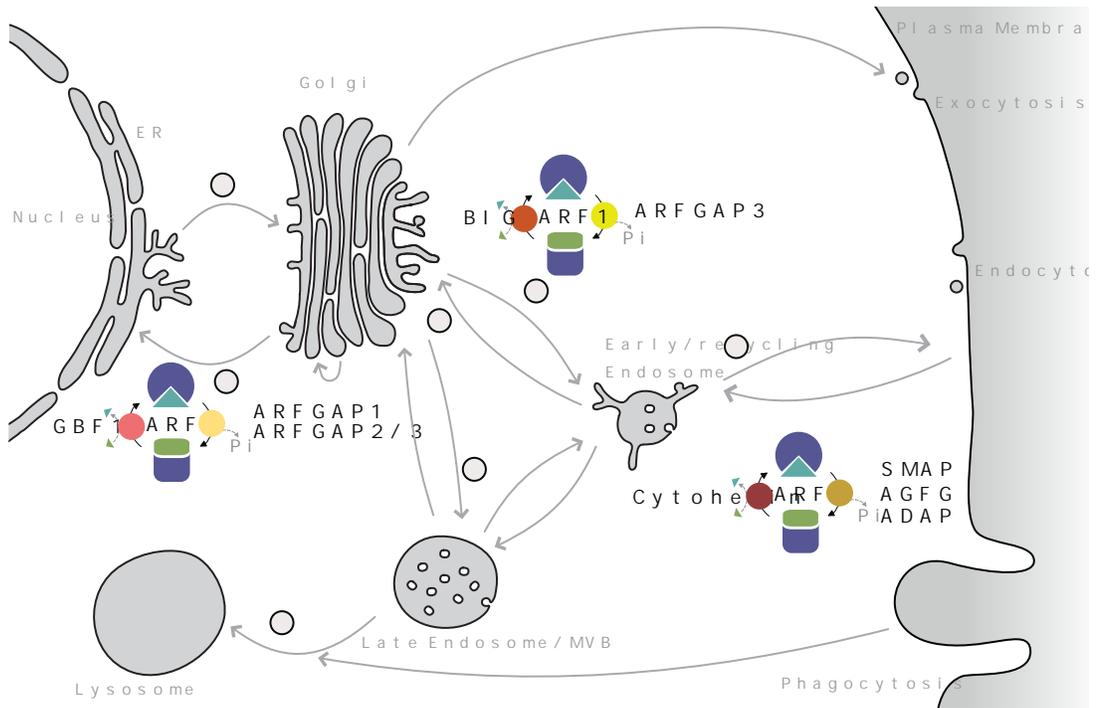
We, therefore, have performed comprehensive comparative genomic and phylogenetic analyses of the ARF regulatory system proteins in the fornicates, allowing us to understand how this aspect of vesicle trafficking changed in fornicates and *Giardia* over time.

Figure 3.1. The eukaryotic ARF GTPases and their previously elucidated intracellular roles. (A) represents a simplified depiction of the ARF GTPase cycle. All ARF GTPases are required to be recruited to the donor membrane compartment for vesicle budding. This process is facilitated by binding of ARF GEFs which exchange bound GDP to GTP, that in turn induce conformational change within the GTPase switch regions. This results in the myristoylated N-terminal amphipathic helix being exposed for recruitment and insertion into the outer leaflet of the recipient organelle membrane. Once the cargo loading and vesicle budding process is complete, bound GTP are hydrolyzed back to GDP by ARF GAPs to inactivate and recycle the ARF GTPase back into the cytosol. The ARF GAPs and ARF GEFs deduced as being present in the LECA are shown in the relevant dashed boxes. (B) is a model for ARF cycle and functions in the endomembrane landscape of model eukaryotes. The ARF regulatory cycle occurs at various interfaces within the eukaryotic cell. ARF1 participates in pre- and post- Golgi trafficking in association with BIG and GBF1, and ARFGAP1 and ARFGAP2, to mediate trafficking of clathrin-coated vesicles in association with heterotetrameric adaptor complexes. These functions have been derived from studies in mammalian model cells. In Metazoa, ARF6 is also necessary for early endocytic and plasma membrane trafficking, in conjunction with ARF GEF Cytohesin and ARF GAPs, SMAP, AGFG, and ADAP, to promote clathrin-mediated endocytosis, actin-mediated phagocytosis, and sorting of early to late endosomal fusion.

A



B



3.3 Materials and Methods

3.3.1 Dataset mining of pan-eukaryotic representatives and Fornicata genomes and transcriptomes

Updated genome assemblies of *Bigelowiella natans*, *Bodo saltans*, *Chromera velia*, *Dictyostelium discoideum*, *Entamoeba histolytica*, *Fonticula alba*, *Guillardia theta*, *Naegleria gruberi*, *Phytophthora ramorum*, *Phytophthora sojae*, *Stentor coeruleus*, and *Vitrella brassicaformis* were obtained and used for ADAP analyses. Draft genomes of *Kipferlia bialata*, *Spironucleus salmonicida*, *Giardia intestinalis* assemblage AI, isolate WB, *G. intestinalis* assemblage AII, isolate DH, *G. intestinalis* assemblage B, isolate GS, *G. intestinalis* assemblage B, isolate BAH15c1, *G. intestinalis* assemblage B, isolate GS-B, *G. intestinalis* assemblage E, isolate P15, *G. intestinalis* assemblage C, *G. intestinalis* assemblage D, and *Giardia muris* were used for Fornicata analyses of the entire ARF regulatory system proteins. Additionally, the transcriptome of *Trepomonas* sp. PC1, *Barthelona* sp. PAP020, *Carpediemonas membranifera* and *Carpediemonas*-like organisms (CLOs), *Aduncisulcus paluster*, *Ergobibamus cyprinoides*, *Dysnectes brevis*, *Chilomastix cuspidata*, and *Chilomastix caulleryi* were also analyzed. Transcriptomes were additionally translated using *ab initio* gene prediction program, GeneMarkS-T under the default parameters (Tang et al., 2015). In addition to the published transcriptome of *Carpediemonas membranifera*, the newly available genome was also used to cross-validate gene presence and absence.

Database sources and corresponding references from which all genomes and transcriptomes were obtained are summarized in Online Appendix Table 3.1.

3.3.2 Comparative genomics of ADAP across select pan-eukaryotic representatives

Using the ARF GAP domain from pan-eukaryotic orthologs, previously curated by Schlacht et al. (2013), a protein alignment was generated using MUSCLE v. 3.8.31 (Edgar, 2004; Schlacht et al., 2013). Alignments were used to generate Hidden Markov Models (HMMs) using the hmmbuild option available through the HMMER v. 3.1.b1 (Eddy, 1998). The resulting HMM profiles were used for forward HMMER searching using the hmmsearch tool available through the HMMER3 package (Prakash et al., 2017). All identified forward hits were subject to reverse homology search using BLASTP into the *Homo sapiens* protein database (Altschul et al., 1990). Both forward and reciprocal searches were performed with an e-value threshold of 0.05. Forward protein hits that retrieved *H. sapiens* ADAP homologs were considered putative positives if the e-value was two-orders or higher in magnitude than the next best non-orthologous hit. Additional assessment of correct orthology was performed using domain similarity searches using the NCBI conserved domain database and *de novo* protein modelling using Phyre2, under the intensive setting, to identify relevant ARFGAP and dual PH domains (Kelley et al., 2015). Additional reciprocal BLASTP analyses were also performed against the NCBI Metazoa non-redundant database. Finalized hits are summarized in Online Appendix Table 3.2.

3.3.3 Comparative genomics of ARF, ARF GAPs, and ARF GEFs in Fornicata

Published query ARF sequences were obtained from Li et al. and Vargová et al., ARF GAP sequences from Schlacht et al., and ARF GEF Sec7 sequences from Pipaliya et al. (Li et al., 2004; Pipaliya et al., 2019; Schlacht et al., 2013; Vargová et al., 2021). Forward HMMER searching was performed against fornicate genomes and transcriptomes in the same manner as described above. Reciprocal BLASTp analyses of forward hits were performed against *C. membranifera* and *H. sapiens* predicted proteins, as described above. To identify divergent protein orthologs, the e-value stringency was lowered to 0.01 for both hmmsearch and BLASTP. Hits were deemed putative positives using the same forward and reverse BLAST criteria as described above and were subjected to additional domain analyses using InterProScan (Quevillon et al., 2005). To rule out false-negative artifacts resulting from improper protein mispredictions, we additionally performed TBLASTN searches followed by reverse BLASTX with an e-value cut-off set to 0.05 (Altschul et al., 1990). Upon consolidation of *Carpediemonas* orthologs, HMM profiles were regenerated with the inclusion of *Carpediemonas* sequences to identify any remaining highly divergent *Carpediemonas*-like organism (CLO) and diplomonad orthologs of ARF regulatory proteins. Finally, if any of the CLO and diplomonad ARF GEF and ARF GAP proteins remained unclassified with the approaches used above, we removed the Sec7 or the ARF GAP domain and attempted to classify the remainder protein sequences using BLASTP searches against the NCBI non-redundant database, *C. membranifera* predicted proteins, and *H. sapiens* predicted proteins. All results from the Fornicata analyses are summarized in Online Appendix Tables 3.3 and 3.4

3.3.4 Phylogenetic analyses

Identified *Carpediemonas*, CLO, diplomonad, and *Giardia* ARF sequences were classified against metamonad ARF sequences. ARF GAP phylogenetic analyses were restricted to SMAP, ARFGAP1, and ARFGAP2/3, with all fornicate sequences directly classified against pan-eukaryotic orthologs. Finally, ARF GEF analyses were performed only with putative *Carpediemonas* BIG sequence against a pan-eukaryotic BIG and GBF1 resolved phylogenetic backbone (Pipaliya et al., 2019).

Maximum likelihood analyses were performed using non-parametric and ultrafast bootstrapping using RAxML-HPC2 on XSEDE v. 8.2.10 and IQTREE, respectively (Nguyen et al., 2015; Stamatakis, 2014). The best fit model was determined using IQTREE-ModelFinder, and the suggested model according to the Bayesian Information Criterion was used (Kalyaanamoorthy et al., 2017). For RAxML analysis, 100 non-parametric bootstrap replicates were generated using the default tree faster hill-climbing method (-f b, -b, -N 100), and a consensus tree from the 100 pseudoreplicates were determined using Consense from the Phylip v. 3.66 package (Felsenstein, 1989). Ultrafast bootstraps (1000) and the resulting consensus tree was obtained using IQTREE v. 2.0.6 under the default parameters (Nguyen et al., 2015). Bayesian inference phylogenetics was performed using MRBAYES v. 3.0.6 on XSEDE v. 3.2.6. Ten million Markov Chain Monte Carlo generations were specified and simulated using an LG4X mixture model, and the number of gamma rate categories was set to 4 (Huelsenbeck & Ronquist, 2001). Sampling frequency was set to occur every 1000 generations with the first 25% of the runs discarded from the cold chain. Tree

convergence was ensured, with a standard deviation of split frequency reaching below 0.01. A random seed value of 12345 was chosen for all phylogenetic analyses. Non-parametric and ultrafast bootstraps were overlaid onto the MRBAYES tree topologies with posterior probabilities. RAxML and MRBAYES analyses were performed on the CIPRES portal (<http://www.phylo.org>), while the IQTREE2 package was installed and run locally (Miller et al., 2010). Trees were visualized in FigTree v1.4.4, and branch and node value annotations were prepared in Affinity Designer. All protein substitution models determined by ModelFinder and used for phylogenetic analyses are provided through Online Appendix Table 3.5 (Kalyaanamoorthy et al., 2017).

3.3.5 Data accessibility

Masked alignments for phylogenetic analyses are made publicly available through Figshare (<https://doi.org/10.6084/m9.figshare.13673587.v2>) (Supplementary Files 1-5).

3.4 Results

3.4.1. Resampling of key protist lineages identifies ADAP in the LECA ARF GAP complement

To appropriately investigate the diversity of ARF regulatory proteins in *Giardia* and its fornicate relatives, it was first important to establish the pan-eukaryotic complement of this system as a null hypothesis for what could have been present in the fornicate ancestor. LECA complements for the GTPases, GEFs, and GAPs have been recently established (Li et al., 2004; Pipaliya et al., 2019; Schlacht et al., 2013). However, for the latter, uncertainty remains regarding the ADAP subfamily of ARF GAPs. These proteins in Holozoa are classified by the ARF GAP domain followed by a dual PH domain and arginine fingers and zinc-binding sites (Schlacht et al., 2013). Outside the Holozoa, candidate orthologues were tentatively identified in *Naegleria gruberi*, *Dictyostelium discoideum*, and *Giardia intestinalis*, and the authors left open the possibility that ADAP could be ancient (Schlacht et al., 2013). Here, we have pursued this possibility with a targeted approach, taking advantage of improved gene assemblies and protein models for a few of the genomes previously sampled and the addition of several key taxonomic sampling points.

Comparative genomic searches using an HMM of the ARF GAP domain into selected eukaryotic genomes (see section 3.3.3) identified candidate ADAP orthologues in *Bigeloviella natans*, *Dictyostelium discoideum*, and *Naegleria gruberi*. These candidate orthologues were subsequently used as queries for BLAST analyses against *H. sapiens* predicted proteins and holozoan databases in NCBI and successfully retrieved ADAP as the top orthologous hit at two orders or higher magnitude. The orthology of these candidates was further confirmed based on the presence of C-terminal dual PH domains that are characteristically present in the holozoan ortholog of this protein. Additionally, similar to the holozoan ADAP, these proteins also possessed a zinc finger or an arginine binding site downstream of the PH domains (Figure 3.2A; Online Appendix Table 3.2). In place of an arginine binding site, *Naegleria gruberi* ADAP was shown to possess a coiled-coil domain. Lastly, we performed *de novo* protein threading to compare homology at a greater structural level with all three orthologs retrieving similarity to the human centaurin,

alpha-1 (*i.e.*, metazoan alias for ADAP) at 90% or higher similarity (Figure 3.2B). The collective reciprocal BLAST, domain analyses, and structural homology increase our confidence from the previous findings that these lineages, distributed within different eukaryotic supergroups, in fact, do possess ADAP. ADAP, therefore, represents the seventh family of ARF GAP to have been present in the LECA and to be patchily distributed amongst eukaryotes.

Having now established the LECA complement for the ARF regulatory pathway, we shifted our focus to examine this system in the context of the evolution of parasitism by analyzing genomes and transcriptomes of *Giardia* and its free-living Fornicata relatives.

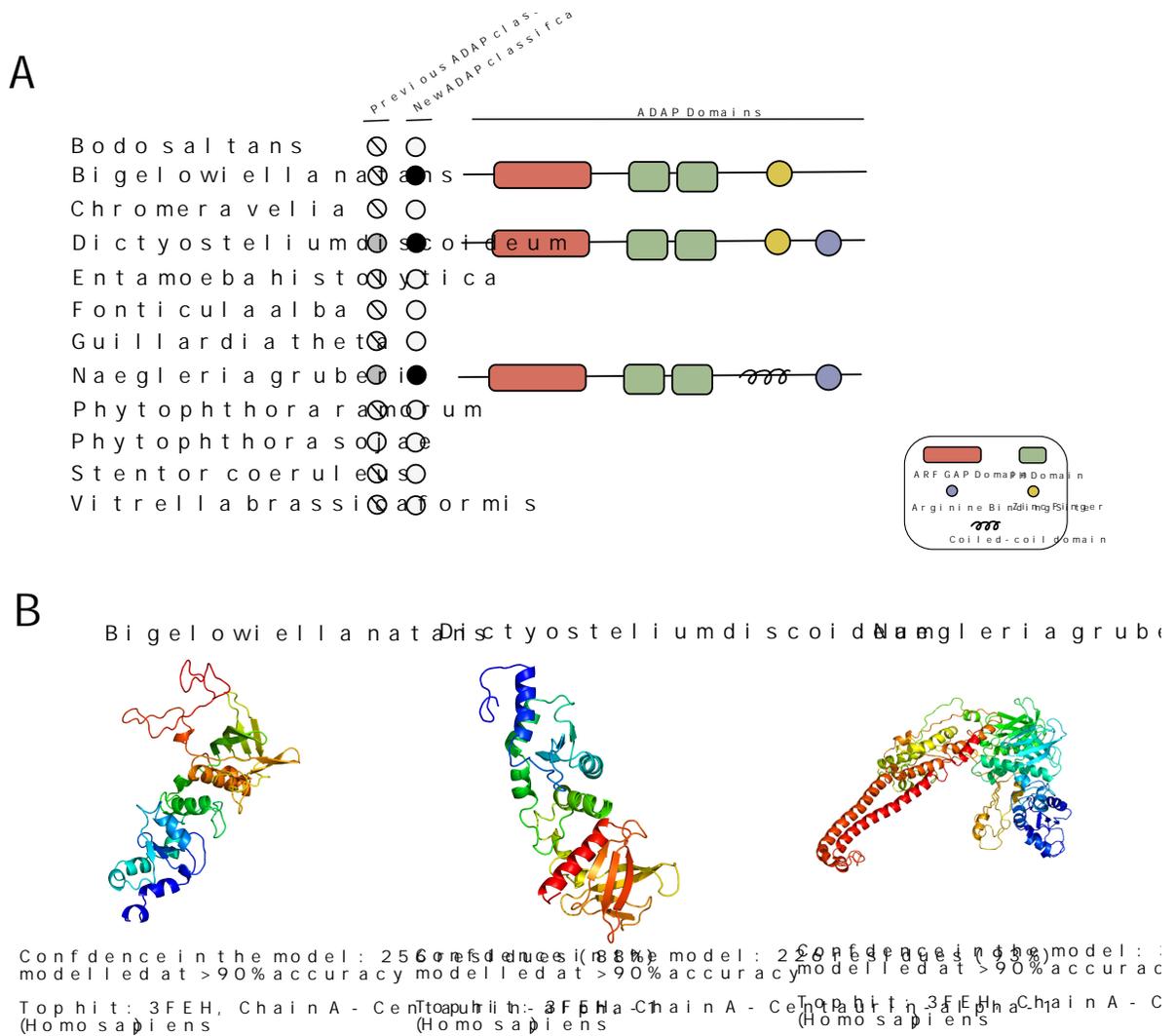


Figure 3.2. Comparative genomic analyses of ADAP in previously analyzed and newly available, genomes from representative eukaryotic lineages. (A) Dot plot summary depicting gene presence/absence. Previous investigation of ARF GAPs by Schlacht et al. (2013) putatively classified ADAP in two lineages outside of Metazoa, *Dictyostelium discoideum* and *Naegleria gruberi*, based on weak reciprocal BLAST e-value results. We re-investigated these lineages in newer/improved versions of these genomes as well as other genomes closely related to those organisms and other major representatives from TSAR, Fungi, and Amoebozoa, which have become newly available since that previous study. Based on both reciprocal BLAST analyses and domain architecture investigation, we identified the presence of ARF GAPs possessing ARF GAP domains, C-terminal dual Pleckstrin homology (PH) domains, and either zinc fingers or arginine binding sites in at least three protists from three major eukaryotic lineages — *Bigelowiella natans* (SAR), *Dictyostelium discoideum* (Amoebozoa), and *Naegleria gruberi* (Discoba), suggesting that ADAP was likely present in the LECA with multiple subsequent losses in descendent lineages. Circles with lines indicate lineage was previously unanalyzed. Grey circles indicate putative homology. White circles indicate gene absence, and black circles indicate gene presence. **(B)** *De novo* protein structural modelling of the three identified ADAP homologs. Protein threading using PHYRE2 further confirmed the closest structural homology of the ADAP identified in *Bigelowiella*, *Dictyostelium*, and *Naegleria*, using comparative genomics, to be with alpha-centaurin from *Homo sapiens* (PDB: 3FEH, Chain A) with 90% or greater accuracy.

3.4.2 Comparative genomics of ARFs in *Giardia* and its fornicate relatives reveals an expansion within the ARF GTPases

Previous examination of ARFs in eukaryotes concluded LECA to possess a single ARF1 which is therefore expected to be present across the breadth of eukaryotic diversity (Li et al., 2004). However, only *G. intestinalis* AWB was investigated and shown to possess a single ARF1 protein (Li et al., 2004). Since that study, three ARF isoforms in *Giardia* have been described on GiardiaDB (<https://giardiadb.org/>) but not formally introduced or investigated in-depth for their evolutionary histories. With the availability of sequence data from new fornicates, we aimed to build on these previous findings to widen the scope of ARF-related investigations in fornicates (Leger et al., 2017).

Our initial HMMER searches identified numerous ARF1 GTPases in all fornicate lineages including all investigated strains of *G. intestinalis* and *G. muris* (Figure 3.3; Online Appendix Table 3.3), plus in *Barthelona* sp. PAP020. Phylogenetic analysis of the ARFs, outgroup rooted at the Preaxostyla member, recovered three separate clades of ARFs in fornicates, albeit with minimal backbone support (Figure 3.4). Specifically, we observed a short-branching clade of ARF1 and two separate clades of more divergent paralogues, which we name as here ARF1 in fornicates, abbreviated as ARF1Fs. We noted that the paralogues of diplomonad and *Carpediemonas* proteins identified in our comparative genomics fell within two clades herein termed ARF1FA and ARF1FB. Although we identified numerous ARF1F proteins in the *Barthelona* and CLO transcriptomes, these produced long branches even upon taxonomic filtering and so were removed to improve the resolution of the diplomonad paralogues. Based on reciprocal BLAST results with *H. sapiens* and *Carpediemonas* predicted proteins, these were still putatively classified as ARF1F based on their retrieval of Class I ARFs, in the case of humans, and ARF1F, in the case of *Carpediemonas*.

Within diplomonads, *Spironucleus* and *G. muris* possessed only ARF1FA, likely suggesting that ARF1FB was secondarily lost (Figure 3.4). *Trepomonas*, on the other hand, surprisingly possesses a large complement of ARF1 and ARF1F proteins (Figure 3.3; Online Appendix Table 3.3). However, these numbers are tentative, given that issues with paralogue numbers in the *Trepomonas* transcriptome data have been raised previously (Xu et al., 2016). Notably, within *G. intestinalis*, it is interesting to note that isolates from assemblages A and E, particularly WB, DH, and P15, possess two paralogues of ARF1FB compared to assemblage B, C, and D isolates. This indicates a duplication event in the ancestor of assemblages A and E, yielding an additional paralog of ARF1FB, and suggests there are differences at both protein-complement and sequence levels between the three human-infecting and zoonotic isolates (Figure 3.4; Online Appendix Table 3.3).

Overall, these collective findings suggest that a lineage-specific expansion within the ARFs seemed to have occurred prior to fornicates. Additionally, although these paralogues were previously annotated in *G. intestinalis* as ARF1, 2, and 3, here we confirm that one of these is a more canonical ARF1 paralogue, resembling pan-eukaryotic ARFs, while the remaining two are more divergent lineage-specific paralogues ARF1FA and ARF1FB.

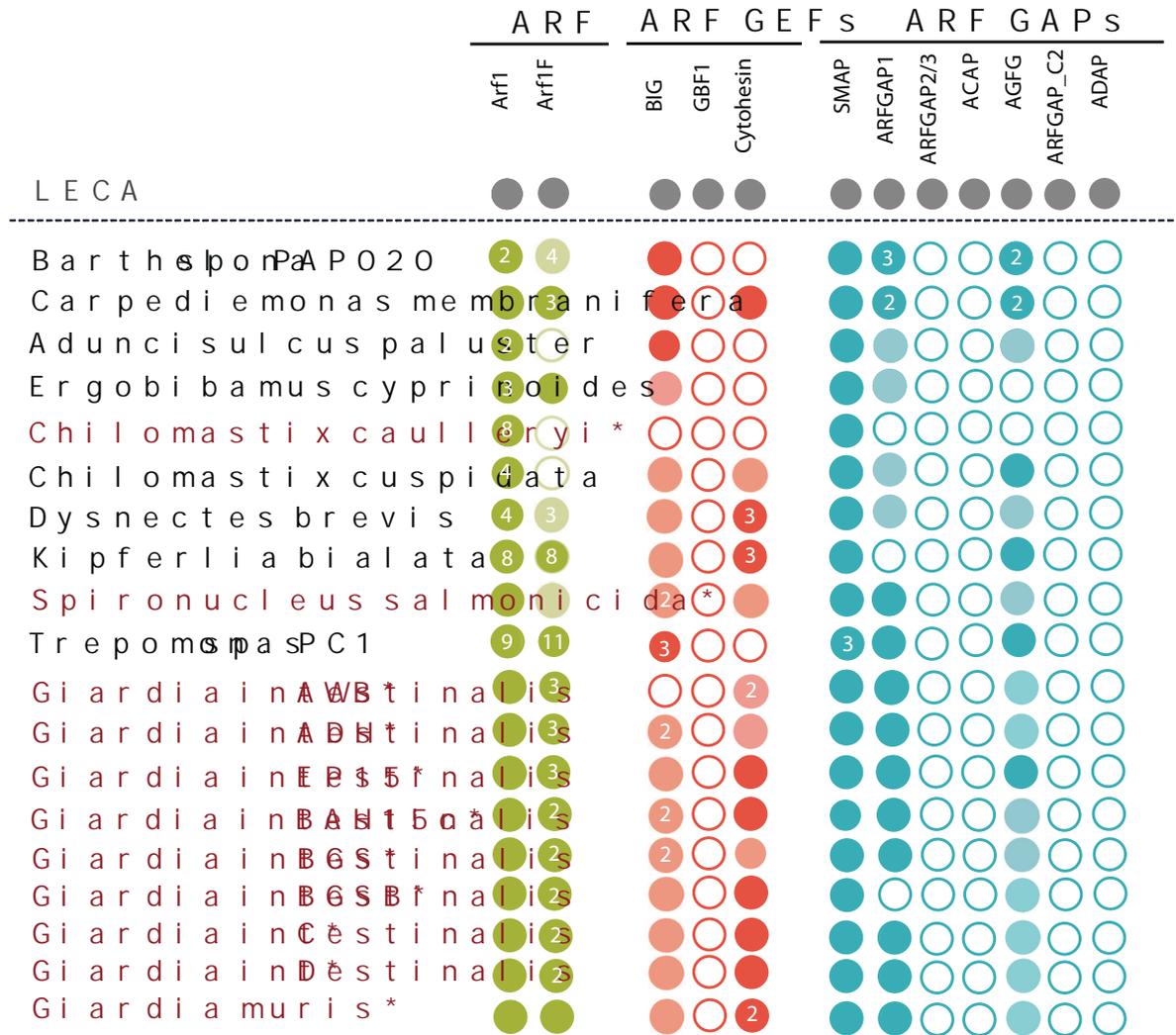


Figure 3.3. Dot plot depiction of all identified and classified ARF GTPases, ARF GEFs, and ARF GAPs in fornicate genomes and transcriptomes. Upon investigation of ARF1, BIG, GBF1, Cytohesin, SMAP, ARFGAP1, ARFGAP2/3, ACAP, AGFG, ARFGAP_C2, and ADAP, we identified that all fornicates possess the highly conserved ARF1 ortholog as well as numerous ARF1L paralogues closely related to ARF1. Within the ARF GEFs, we identified the presence of BIG and Cytohesin but not GBF1 in any lineages. Finally, of the ARF GAPs, only SMAP, ARFGAP1, and AGFG were identified of the seven LECA orthologs. Within *Giardia*, *Giardia intestinalis* AWB lacks an identifiable BIG orthologue. Differences in the number of ARF1F orthologs are also present between assemblage A and assemblage E isolates which possess three ARF1F paralogs, while assemblage B isolates were identified to have two ARF1F paralogs. *Giardia muris*, on the other hand, was only identified to have one ARF1F. Parasitic and endobiont lineages are indicated with an asterisk. Lineages for which genomes were available and analyzed are indicated in boldface font, while transcriptomes are indicated in regular font. Dark shading indicates robust assignment based on RBH retrieval of this paralogue with two orders of magnitude e-value to the next best scoring paralogue and/or classification by phylogenetics. Light shading indicates reciprocal best hit (RBH) retrieval of the assigned paralogue but not meeting the more stringent criteria.

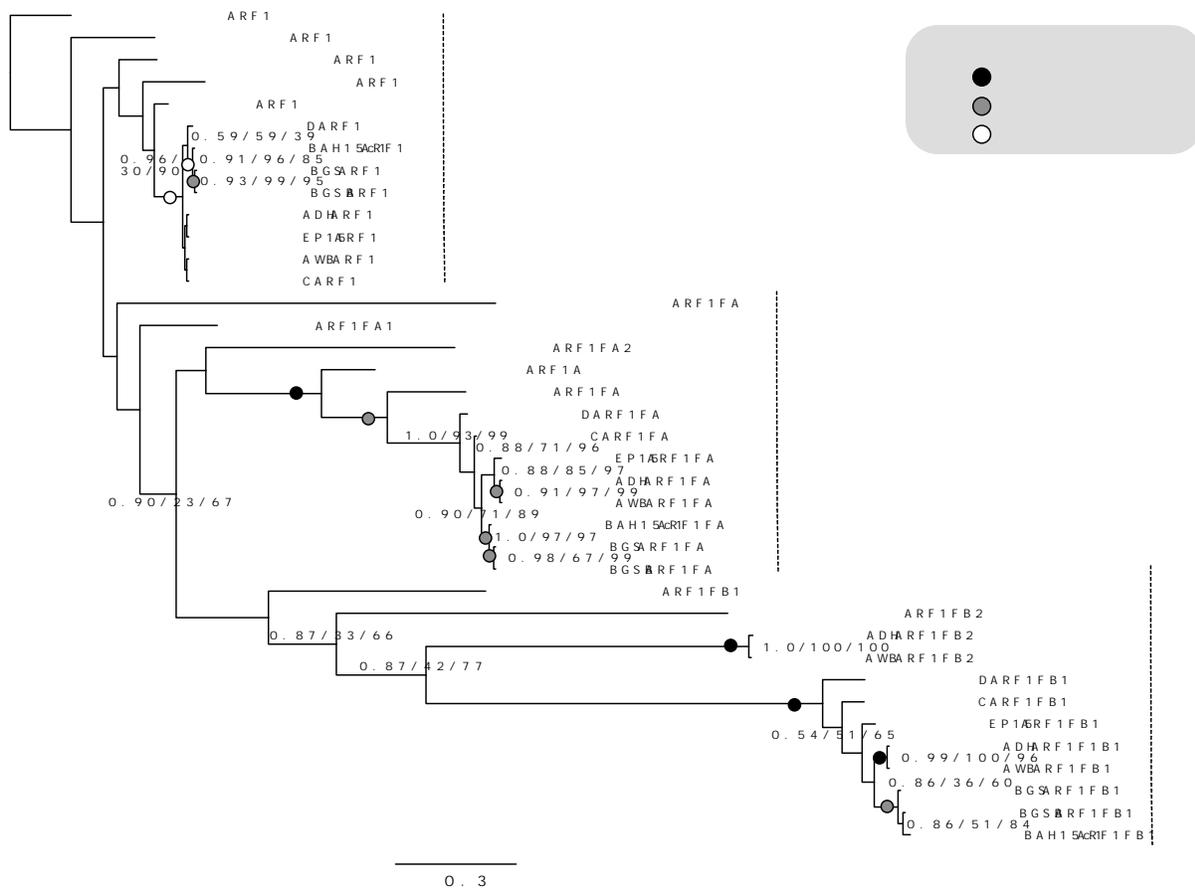


Figure 3.4. Phylogenetic investigation assessing ancient versus independent duplications of the fornicate ARF1F paralogs. We took a fine-grained investigation into the multiple ARF1F paralogs identified where our analyses retrieved two separate clades, herein named ARF1FA and ARF1FB, each consisting of ARF1F paralogs from *Carpodomonas* and the diplomonads. *Giardia* assemblage A isolates underwent an additional duplication within ARF1FB to produce ARF1FB1 and ARF1FB2. In this and all subsequent phylogenies, numerical values represent Bayesian posterior probabilities (MRBAYES)/Maximum-Likelihood non-parametric bootstrap values (RAxML)/ultrafast bootstrap values (IQTREE). Nodes of interest are indicated in boldface. Values for other supported nodes have been replaced by symbols: black circles $\geq 0.99/95/99$, grey circles $\geq 0.95/75/97$, and white circles $\geq 0.80/50/95$. Node support values are overlaid onto a MRBAYES tree topology.

3.4.3 Fornicates underwent extreme streamlining within the ARF GAP complement while retaining the most conserved members

In the LECA, six ancestral ARF GAPs were identified by previous comparative genomics analyses (Schlacht et al., 2013). However, in those analyses, only *G. intestinalis* AWB was included as the sole representative of the Fornicata. Based on these limited data, the authors inferred the putative presence of SMAP, ARFGAP2, AGFG, and even ADAP (Schlacht et al., 2013). Having confirmed the ancient nature of ADAP above, we, therefore, expanded our analysis to determine if these putative ARF GAPs can be confirmed in *Giardia* and whether the remaining fornicates have a similar or different ARF GAP complement.

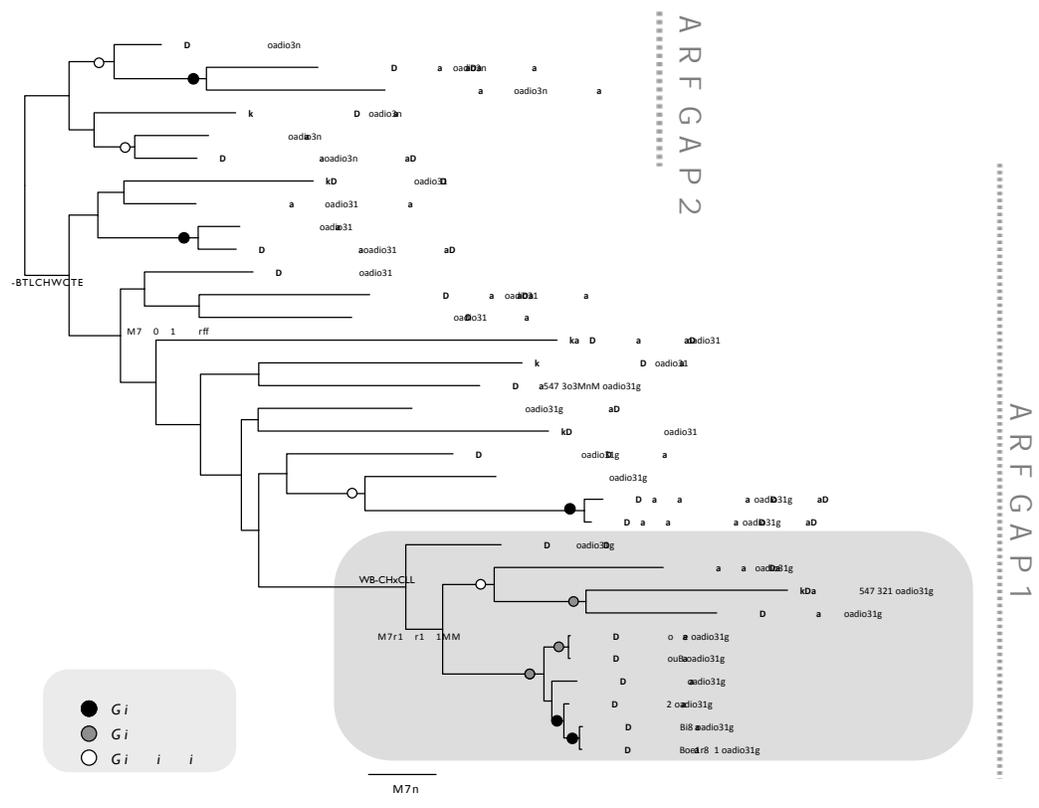
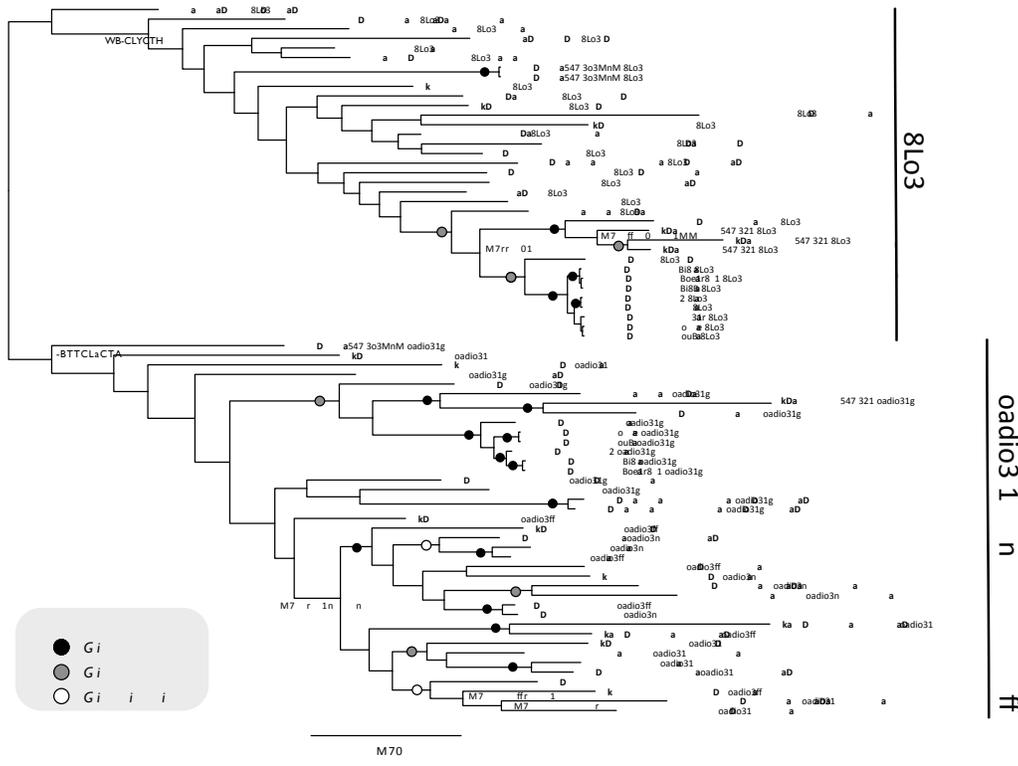
Homology searches identified candidate ARF GAP proteins in all genomes and transcriptomes analyzed. Initial reciprocal BLAST analyses were often insufficient to robustly classify SMAP, ARFGAP1, and ARFGAP2, as often the e-value for the next best non-orthologous ARF GAP was within two orders of magnitude of the e-value of the top hit (Online Appendix Tables 3.3 and 3.4). Therefore, we performed phylogenetic analyses of the Fornicata sequences of these proteins against pan-eukaryotic representative sequences of SMAP, ARFGAP1, and ARFGAP2 (Figure 3.5A). This approach allowed us to consolidate and robustly classify SMAPs from all lineages, including the diplomonads and all isolates of *G. intestinalis* (Figure 3.5A). However, in this same analysis, we noted that ARFGAP1 and 2 still did not form two different, distinguishable clades.

To better resolve these clades and classify our candidate sequences, we performed an ARFGAP1 and ARFGAP2-specific phylogeny (Figure 3.5B) with SMAP, as well as the metazoan-specific ARFGAP3, sequences omitted, and including only the shortest branching ARFGAP1 and ARFGAP2 representatives as landmark sequences (Figure 3.5B). This resulted in resolving two separate clades of ARFGAP1 and ARFGAP2, with *Barthelona* and Fornicata sequences falling exclusively within the ARFGAP1 clade with a robust posterior probability, non-parametric, and ultrafast bootstrap values of 0.98/68/96, respectively. This suggests that the fornicates lost ARFGAP2 at the base of this lineage and possibly earlier. *Dysnectes brevis* and the diplomonad sequences also formed an internal subclade with pan-eukaryotic ARFGAP1 with a support value of 1.0/96/99, suggesting that these were also orthologs of ARFGAP1 (Figure 3.5B).

Many of the remaining ARF GAPs were robustly classified as AGFGs based on their retrieval of classified *Carpodemonas* and *H. sapiens* AGFG sequences with reciprocal BLAST e-values of two orders or higher magnitude (Figure 3.3; Online Appendix Tables 3.3 and 3.4). However, there remained a few ARF GAPs that were ambiguously assigned as either SMAP or AGFG. To classify these, we performed domain analyses using InterProScan. Based on the presence of an ARF GAP AGFG domain and Zn²⁺-finger/arginine binding sites, we classified all remaining proteins as putative AGFGs (Online Appendix Tables 3.3 and 3.4). All putatively classified AGFG sequences were also subject to BLAST searches against the NCBI non-redundant database, which retrieved AGFG orthologs from various pan-eukaryotic representatives. This further increased our confidence in this orthology assignment.

Taken together, our findings demonstrate that of the seven ancestral ARF GAPs, only three were kept by the ancestor of fornicates, with losses occurring in ARFGAP2, ACAP, ARFGAP_C2, and ADAP prior to all lineages and with limited variation in the complement between the different *Giardia* strains.

Figure 3.5. Phylogenetic analyses of fornicate ARF GAPs homologues. (A) Classified fornicate orthologs of ARFGAP1/2 versus SMAP. Phylogenetic analysis robustly reconstructs a SMAP clade, separate from either ARFGAP1 or ARFGAP2, confirming the classification of fornicate SMAP orthologues. (B) Classification of fornicate ARFGAP1 and not ARFGAP2. Phylogenetic analysis robustly supports the separation of ARFGAP1 and ARFGAP2 clades, with all fornicate sequences placed within the ARFGAP1 clade. Diplomonads were grouped together with a combined nodal support value of 1.0/96/99 (grey lozenge). In all instances, numerical values represent Bayesian posterior probabilities (MRBAYES)/Maximum-Likelihood non-parametric bootstrap values (RAxML)/ultrafast bootstrap values (IQTREE). Nodes of interest are indicated in boldface. Values for other supported nodes have been replaced by symbols: black circles $\geq 0.99/95/99$, grey circles $\geq 0.95/75/97$, and white circles $\geq 0.80/50/95$. Node support values are overlaid onto a MRBAYES tree topology.



3.4.4. Fornicates underwent a loss within the Golgi ARF GEF GBF1 but retain BIG and Cytohesin

Our recent investigation of ARF GEF evolution deduced three ARF GEFs in the LECA; BIG, GBF1, and Cytohesin (Pipaliya et al., 2019). However, our investigation was limited to *G. intestinalis* AWB where only two paralogs of Cytohesin were identified (Pipaliya et al., 2019). Here, we performed a more thorough investigation of the ARF GEF complement across the fornicate lineages.

Our initial comparative genomics identified multiple Sec7 proteins in most fornicates, classified as BIGs and Cytohesins, based on reciprocal BLAST analyses. To validate the BIG orthologs, we first classified the putative *Carpediemonas* BIG sequence using phylogenetics against pan-eukaryotic BIG and GBF1 orthologs. The *Carpediemonas* sequence falls robustly within the BIG clade with an overall clade support value of 1.0/94/100 (Figure 3.6; Online Appendix Figure 3.1). Domain analyses also revealed *Carpediemonas* to possess the HDS4 domain, which characteristically distinguishes BIG from GBF1, and the remainder of the BIG domains (Dimerization/Cyclophilin Binding region (DCB), Homology Upstream of Sec7 (HUS), and Homology Downstream of Sec7 1-3 (HDS1-3) that were previously described and phylogenetically classified (Mouratou et al., 2005; Pipaliya et al., 2019) (Online Appendix Table 3.3; Supplementary File 4).

Subsequent attempts to classify the remainder of the putative BIG orthologs from the CLO and diplomonad transcriptomes resulted in unresolved phylogenies, likely due to the extreme long branching by these sequences. Instead, a reciprocal BLAST approach with the *Carpediemonas* BIG sequence was more fruitful. Based on *H. sapiens* and *C. membranifera* reciprocal best hit (RBH) results, we were able to assign robust BIG orthology to identified Sec7 proteins in *Aduncisulcus paluster* and *Trepomonas* sp. PC1. Upon classification, we iteratively added sequences to the previous pan-eukaryotic Sec7 alignments to generate more sensitive HMMs for searching in *Giardia* and the remainder of the CLOs. This approach allowed us to identify numerous additional Sec7 proteins in *Dysnectes brevis*, *Spiroucleus salmonicida*, and all isolates of *Giardia* that were putatively classified as BIG orthologs. Once again, we attempted to classify these proteins through phylogenetics against the BIG/GBF1 pan-eukaryotic backbone; however, a lack of resolution prompted us to rely on domain analyses and RBH criteria (Figure 3.3; Online Appendix Tables 3.3 and 3.4). Based on the retrieval of *Carpediemonas* and *H. sapiens* BIG sequences, these were classified as putative BIGs. Overall, we did not identify any distinct orthologs of GBF1 in any of the lineages, suggesting a loss, at least before the last fornicate common ancestor.

Contrary to this pattern, we note a difference in the distribution of BIG orthologs across *Giardia* and fornicates, with some lineages retaining multiple paralogs and others undergoing complete loss within this protein (Figure 3.3; Online Appendix Tables 3.3 and 3.4). Notably, despite HMM searches using BIG orthologs from other *Giardia* isolates, we did not identify this protein in assemblage AWB for which an improved chromosome level assembly is now available and was surveyed (Xu et al., 2020). This increases our confidence that BIG was independently lost in this *Giardia* isolate.

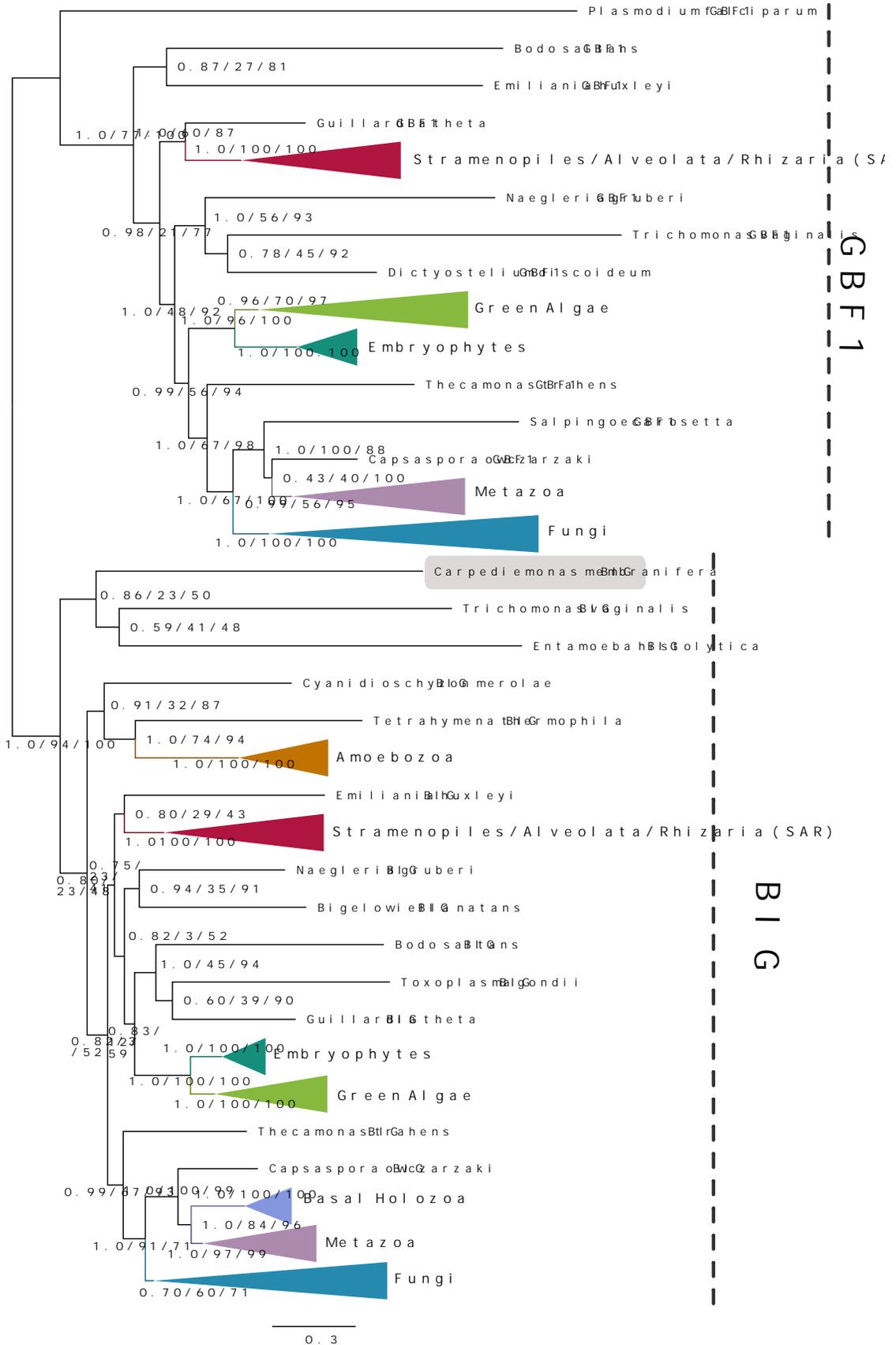
All remaining Sec7 proteins were classified as Cytohesins or putative Cytohesins. Compared with BIG and GBF1, Cytohesin orthologs across eukaryotes are divergent and subject to considerable domain

acquisition and plasticity (Pipaliya et al., 2019). Since our previous work showed a lack of phylogenetic resolution in pan-eukaryotic Cytohesin datasets, we relied on reciprocal BLAST and domain analyses to classify our candidate Cytohesins from fornicates. In *Carpediemonas*, we retrieved a Cytohesin ortholog with an e-value greater than five orders of magnitude better than the next best hit and, therefore, was unambiguously classified as a Cytohesin. This assignment was further verified through BLASTP searches against the NCBI non-redundant database, which retrieved classified Cytohesin and Cytohesin-derived proteins from a diversity of eukaryotes (Online Appendix Tables 3.3 and 3.4). InterProScan domain analyses identified its Sec7 domain as greatest in similarity to one present in Cytohesins, despite the lack of a C-terminal PH domain in this protein (Online Appendix Table 3.3). However, this is not surprising, as the PH domain was frequently lost in Cytohesin orthologs that were identified in other eukaryotic lineages (Pipaliya et al., 2019). Based on these collective findings, we concluded *Carpediemonas* to possess Cytohesin.

We then used this newly classified *Carpediemonas* Cytohesin and *H. sapiens* Cytohesins and Cytohesin-derived proteins for RBH as well as an NCBI non-redundant BLAST search to putatively or definitively classify Cytohesins in the remaining of the fornicates, including *Giardia*. No Cytohesin homologues were identified in *Barthelona* sp. PAP020 which could represent an independent loss in this lineage or a false negative as only transcriptomic data are available for this single lineage representative. Nonetheless, in the CLOs or the diplomonads, as in *Carpediemonas*, we did not detect a C-terminal PH domain in any of the identified Cytohesins, suggesting loss of the PH domain took place prior to the divergence of the fornicates but after the split with parabasalids, based on the presence of a C-terminal PH domain in the Cytohesin homologue of *T. vaginalis*. Overall, we conclude that of the three LECA ARF GEFs, only two were present in fornicates, and similar to ARF GAPs, streamlining has occurred within this group of proteins. Further loss was not associated with a parasitic lifestyle as both BIG and Cytohesin were identified in all diplomonads and *Giardia*.

In the previous pan-eukaryotic analyses of ARF GEFs, only *Giardia* AWB was examined as a representative of fornicates. Here we demonstrate that in fact, aside from Cytohesin, BIG is also present in *Giardia* (Figure 3.3; Online Appendix Tables 3.3 and 3.4), revealing inter-isolate differences in the complement of ARF GEFs and that independent loss likely occurred within isolate WB.

Figure 3.6. Phylogenetic characterization of the identified *Carpediemonas membranifera* BIG orthologue. Phylogenetic analyses of a pan-eukaryotic dataset containing BIG and GBF1 orthologs from various eukaryotic lineages. The *Carpediemonas* sequence fell robustly within the BIG clade (highlighted in the grey box). In all instances, numerical values represent Bayesian probabilities (MRBAYES)/Maximum-Likelihood non-parametric bootstrap values (RAxML)/ultrafast bootstrap values (IQTREE). Node support values are overlaid onto a MRBAYES tree topology. Clusters of BIG and GBF1 sequences belonging to defined eukaryotic lineages were collapsed into clades. These were Stramenopiles/Alveolata/Rhizaria (SAR), green algae belonging to the Archaeplastida, embryophytes belonging to Archaeplastida, Holozoa, Metazoa, Amoebozoa, and Fungi as indicated by the clade labels.



3.5 Discussion

In this study, we investigated the molecular complement and evolution of the ARF regulatory system in Fornicata to understand the evolutionary streamlining and modification observed in the endomembrane compartmentalization of *Giardia*. We first established that ADAP is indeed an ancient but sparsely retained ARF GAP paralogue. We then determined that the ARF GTPases underwent multiple duplication events, yielding multiple paralogues in fornicates. The ARF GAP and ARF GEF complements, by contrast, were streamlined early in the fornicate lineage, consistent with a possible repurposing of machinery to function in one secretory and one endocytic pathway.

Previous evolutionary comparative genomics putatively identified ADAP in lineages outside of Holozoa (*Dictyostelium* and *Naegleria*), however, with little confidence. We re-investigated this distribution into these same genomes with updated gene models and added selected sampling points that were unavailable at the time of the original analysis. We confirmed that ADAP is, in fact, encoded in taxa from four major and disparately related eukaryotic lineages, all with the conserved ADAP domain structure. Our investigation was by no means intended to be an exhaustive survey but rather a targeted analysis to assess the presence of this protein in key sampling points. The result is the identification of ADAP as a sparsely conserved paralogue but still as the seventh ARF GAP to have been present in the LECA (Figure 3.7).

We next performed comparative genomics and phylogenetics with the ARF1 small GTPase. Previous work by Li et al. demonstrated that ARF1 is ancestral and distributed across a diversity of eukaryotes (Figure 3.7) (Li et al., 2004). Sensitive HMMER searching allowed us to identify multiple ARF proteins in each fornicate lineage, with phylogenetics showing that aside from canonical ARF1 orthologs, all fornicate lineages and the sister taxon *Barthelona* additionally possess one or more ARF1 paralogues that are specific to fornicates, which we have now termed as ARF1F. These proteins are distinctly related to ARF1 rather than Arls, the closest outgroups to the ARF subfamily, based both on Reciprocal Best Hit criteria and by the conservation of the canonical ARF N-terminal NH₂ motif (Supplementary File 1). Expansion within ARF proteins is not a novel phenomenon. It has been observed in other eukaryotic lineages, with the most notable being within the Opisthokonta, where at least two ancient pre-duplicates of Class I/II and Class III ARFs were present (Li et al., 2004). In animals, more specifically vertebrates, this resulted in additional duplication events leading to ARF1-6 (Sztul et al., 2019). Extensive molecular functional investigation into the roles of these proteins has revealed that different ARF paralogues have distinct functions within the endocytic and secretory pathways of animal cells (Sztul et al., 2019).

The presence of a canonical ARF1 within all fornicates is unsurprising since ARF1 has, so far, not been lost in any investigated eukaryotic lineage, including highly reduced parasites such as *Plasmodium* (Cook et al., 2010). Functional investigation of ARF1 in plants, animals, and protist lineages such as *Trypanosoma brucei*, *Toxoplasma gondii*, and *Saccharomyces cerevisiae* have demonstrated ARF1 having functions at the Golgi as well as within the early endosomal -plasma membrane vesicle pathway (Liendo et al., 2001; Price et al., 2007; Yahara et al., 2001). The functional role of the canonical ARF1 has also been elucidated in the laboratory strain of *G. intestinalis* assemblage A, isolate WB (C6), where it has stage-

specific functions associated with both peripheral vacuoles and Golgi-analogous compartments. In the trophozoites, HA-tagged ARF1 primarily produced punctate localization at the peripheral vacuoles when examined using immunofluorescence microscopy (Stefanic et al., 2009). However, this localization changes as the parasites undergo encystation, with ARF1 aggregating to encystation-specific vesicles and associates with cyst-wall material and β 1-3 GalNAc homopolymer (Stefanic et al., 2009). This role has been further confirmed through later studies which have utilized dominant-negative mutants of *Gi*ARF1 and resulted in dysfunctional transport and cargo-loading of cyst-wall material (Ebnetter et al., 2016). Association of both ARF1 and other small GTPases, such as SAR1 and Rab1, that generally localize to the Golgi, with encystation-specific vesicles suggests them to be stage-specific analogs of the Golgi. Additionally, it should also be noted that ARF1 localizes to both the peripheral vacuoles during the trophozoite stage, which indicates that the post-Golgi endocytic functions usually occurring at early endosomal and plasma membrane interfaces likely take place similar to that observed in model eukaryotes where ARF1 also functions (Stefanic et al., 2009). A transmission electron microscopy study showed the presence of vesicular structures that were consistent in size and morphology to endosomes in *Carpediemonas*, and unlike *Carpediemonas*, all identified *Carpediemonas*-like organisms have no identifiable stacked Golgi but still retain numerous vesicular organelles (Park et al., 2010; Simpson & Patterson, 1999; Yubuki et al., 2007, 2013, 2016). The presence of ARF1FA and ARF1FB proteins across all fornicates opens the possibility for roles of ARF1F proteins in post-Golgi trafficking, specifically to and from early and late endosomes and plasma membrane trafficking in association with endocytic adaptin complexes.

By contrast to the pattern of expansion observed in the ARFs, ARF GAPs and GEFs show a contraction in the fornicate complement from that in the LECA. While we recognize that the dataset available and used for many of these lineages is transcriptomic-only, the complete absence of the proteins deduced as lost in lineages for which genomes are available (*i.e.*, *Carpediemonas*, *Kipferlia*, *Spironucleus*, and multiple from *Giardia* sp.) suggest that in fact, these likely represent true absences. Furthermore, given the pattern, it is unlikely that the consistent absence of these proteins is a by-product of multiple independent secondary losses or an equivalent lack of gene expression across all 17 lineages. The lack of 'randomness' in this pattern suggests that these were products of ancestral loss. In all fornicates and *Barthelona*, only SMAP, AGFG, and ARFGAP1 were identified. Previous comparative genomic analyses showed SMAP, AGFG, and ARFGAP1 are well retained across the eukaryotic diversity, hinting at universal and comparable cellular roles within most eukaryotic cells and membrane trafficking systems (Schlacht et al., 2013). In mammalian cells, SMAP tightly regulates clathrin-dependent endocytosis at the plasma membrane through its interaction with ARF6 by aiding in the catalysis and budding of vesicles from the plasma membrane (Tanabe et al., 2005, 2006). AGFG is also required for AP-2 mediated clathrin-mediated endocytosis (CME) from the plasma membrane (Dergai et al., 2016). While *Giardia* does not perform CME for cargo uptake, it does perform bulk flow uptake, a process mediated by numerous trafficking proteins (clathrin heavy chain, dynamin, and AP-2) that assemble between the plasma membrane and the peripheral vacuole interface and function altogether to fuse two organellar membranes (Zumthor et al., 2016). By

contrast, ARFGAP1 regulates ARF1 at the Golgi in model systems by direct interaction with the δ -COP subunit of COPI for retrograde trafficking (Weimer et al., 2008).

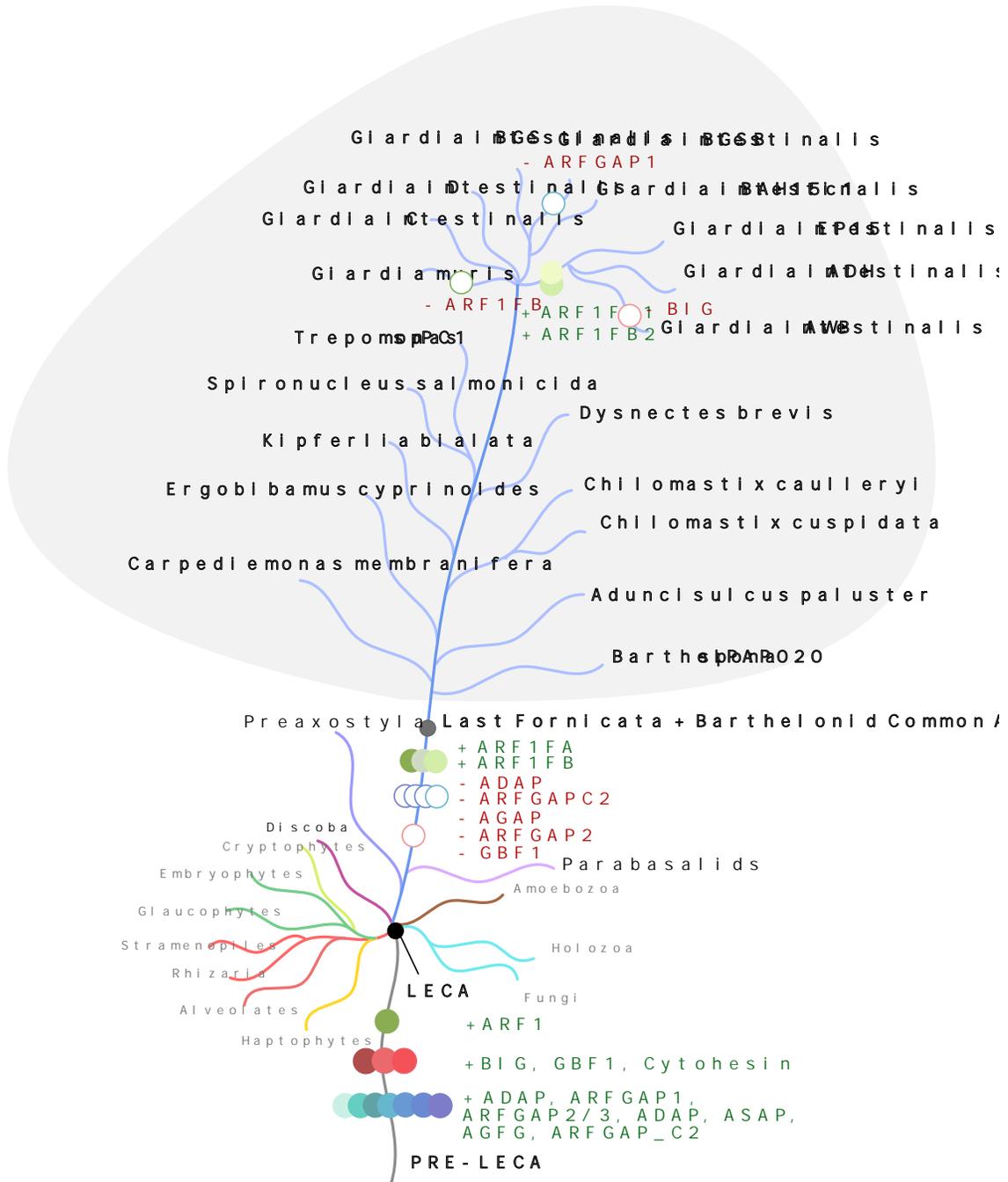


Figure 3.7. Summary of patterns of gains and losses in ARF, ARF GEF, and ARF GAPs, identified, focusing on the fornicate lineages. Our investigation, as well as a combination of results from previous evolutionary comparative genomic studies, identified the presence of at least one ARF1, three ARF GEFs, and now seven ARF GAPs in the ancestor of all eukaryotes. Fornicates have undergone duplications within the ARF GTPase to yield at least three copies of ARFs while simultaneously also having lost ARF GEF GBF1 and ARF GAPs ARFGAP_C2, ACAP, and ADAP. Within *Giardia*, assemblage A duplicates its ARF1FB to yield ARF1FB1 and ARF1FB2, while *Giardia muris* undergoes a loss within its ARF1FB complement. Gain and loss of proteins are indicated as closed and open circles, respectively.

Within the ARF GEFs, of the three ancestral proteins, only two are retained in fornicates, namely BIG and Cytohesin. Surprisingly, the pattern remains consistent throughout all fornicates in that BIG and Cytohesin are consistently present in all 14 genera, and GBF1 is universally absent. This indicates that the loss in GBF1 likely occurred within the ancestor of all fornicates and potentially earlier if the absence in *Barthelona* bears out. However, the investigation into the ARF GEFs across numerous isolates, compared with the previous analysis with *Giardia* AWB-only, reveals inter-assemblage differences, specifically in the presence and absence of BIG paralogues. This may have important consequences for the cell biology, cargo secretion processes, and regulatory mechanisms between isolates of different assemblages. These could then further translate to differences in disease establishment and host-pathogen interactions. For example, canonical ARF1 specifically recognizes and facilitates loading cyst-wall proteins at the budding ESVs. The knockout of either cyst-wall material or ARF1 resulted in 'naked' cysts lacking a cyst-wall (Ebnetter et al., 2016). If a specific ARF-GEF-GAP cycle was necessary for recognition and transport of a given virulence gene or signalling molecule, then the absence of one or more of the GTPase regulatory cycle components could mean reduced transport of a component that may aid with infection establishment, immune evasion, or brush border disruption. Alternately, strains missing a given component could have compensated with novel mechanisms or transport pathways. Overall, there seems to be more intra-strain variation in the membrane-trafficking pathways in *Giardia* than previously anticipated.

GiardiaDB (<https://giardiadb.org/>) lists gene and protein expression for all of the proteins we identified, consistent with their role in *Giardia* cells. Given the known roles in opisthokont model systems of the paralogues that we identified in fornicates and the limited characterization of the *Giardia* ARF1 protein to date, we hypothesize a regulatory model based on retention of pleisiomorphic function versus neofunctionalization of the ARF paralogues and with streamlining and switching repurposing specificity of the regulatory proteins. We postulate that the BIG and ARFGAP1 act as the ARF GEF and ARF GAP, respectively, and necessary to regulate canonical ARF1 for post-Golgi trafficking, as has been the case in other eukaryotic lineages. We make this speculation with some confidence, given the established role for ARF1 in *Giardia* in Golgi-analogue function. Much more speculatively, we further hypothesize that the ARF1F paralogues may well act in clathrin-mediated bulk flow processes for internalization, regulated by Cytohesins as the ARF GEFs and AGFG and SMAP as the ARF GAPs. In *Giardia*, we specifically expect ARF1FA, ARF1FB1, and ARF1FB2 proteins to be associating with the peripheral vacuoles, the major and the most ubiquitous organelle in this pathogen, that is an interface for material exchange, and potentially with the tubular ER.

Our *in silico* results lay the necessary groundwork for these various hypotheses to be tested through future comprehensive functional investigations in the laboratory strain of *Giardia* and other CLOs once those molecular tools become available. Nonetheless, these findings are essential to understanding the evolution and time-points of modulation in a critical membrane trafficking pathway fundamental to eukaryotes in this parasite and its relatives.

Overall, we have shown that the ARF regulatory system reduced from the LECA complement, which included ADAP, early in the fornicate lineage and prior to any shift to parasitism. The system then remained relatively stable with respect to sub-families across various free-living and parasitic lineages. However, variation in the paralogue numbers and presence of the ARF GEF BIG does exist between *Giardia* strains and requires further investigation. Our collective findings shed new, unexpected light on the evolution of a key cargo regulatory system necessary to all eukaryotes, specifically in a lineage consisting of parasites that affect human health and are of veterinary importance. Although parasites are often viewed as highly reduced lineages that have reached an 'end-state' as a by-product of host-pathogen evolution, a telescopic view for a broader perspective into how this process occurs reveals changes in this system well before parasitism came into play.

3.6 Afterword

Subsequent to the completion of this survey and publication, a recent comparative genomics and phylogenetics study published by Vargová and colleagues mapped the distribution of the ARF/SAR/ARL family and provided an updated view into the evolution of this small GTPase family across eukaryotes using new sampling points that were not available in 2004 when Li et al. undertook similar investigations (Li et al., 2004; Vargová et al., 2021). The results from these analyses concluded the presence of ARF6 within the ancestral complement of ARF GTPases. Vargová and colleagues also included several fornicate and metamonad genomes within their pan-eukaryotic sampling, which determined a secondary loss of ARF6 to have occurred prior to the metamonad ancestor. Therefore, although the LECA complement of the ARF GTPase is now updated from one to two, the conclusions derived in this chapter regarding the evolution of the ARF GTPases and their molecular complement in fornicates remain unchanged.

However, it is essential to acknowledge that these new findings extend and consolidate the notions pertaining to ancestral evolution within the fornicate ARF regulatory system presented in this chapter. Precisely, streamlining observed in the LECA complement of the ARF regulatory system proteins prior to the Last Fornicata Common Ancestor was not only limited to the ARF GEFs (*i.e.*, loss of GBF1) and the ARF GAPs (*i.e.*, loss of ADAP, ARFGAP_C2, AGAP, and ARFGAP2/3), but also within the ARF GTPases (*i.e.*, ARF6). Altogether, combining the findings from this investigation, Chapter 2, and results produced by Vargová et al., suggests that although numerous parasitism-specific adaptations have occurred in secretory and exocytic vesicle formation machinery, ancestral duplications and losses are equally influential in shaping the existing endosomal complexity that occurs within *Giardia* and its relatives.

Very recently, a new ARF GAP with longin and DENN domains (C9orf72:SMCR8) was discovered in humans and coincided with the publication timeline of this survey (Su et al., 2020, 2021). Cryo-electron microscopy determined it to be a regulator of ARF1 and ARF6 for lysosomal trafficking, although as of now, it is unknown whether it is evolutionarily conserved in other eukaryotes. Like the pan-eukaryotic analyses with ADAP performed in this survey and the rest of the ARF GAPs examined by Schlacht and colleagues in 2013, comparative genomics would be necessary to determine the phyletic distribution of this new protein

across different supergroups, including within Metamonada and Fornicata, to assess if the inclusion of an eighth ARF GAP is necessary within the LECA repertoire. Phylogenetic analyses would also be required to determine its evolutionary relationship with the LECA and Metazoa-encoded ARF GAPs, as well as longin domain-containing trafficking proteins (*i.e.*, SNAREs). Nevertheless, these new discoveries are exciting as they continue to shed light on additional players involved in the regulation of the ARF GTPases and demonstrate that the intricacy of this pathway is much greater than first thought.

CHAPTER 4

Unexpected organellar locations of ESCRT machinery in *Giardia intestinalis* and complex evolutionary dynamics spanning the transition to parasitism in the lineage Fornicata

4.1 Preface

The previous two chapters focused on the evolutionary analyses of the vesicle coats with proto-coat-meric origins and that have vesicle formation roles within diverse endocytic pathways. This chapter investigates another family of proteins that, although ancient, do not share the same structural configurations or evolutionary origins as the vesicle coats but are nonetheless critical to the maturation of early endosomes to late endosomes. Within the endo-lysosomal system, late endosomes are important sorting stations for material recycling and lysosomal degradation. Homotypic fusion of vesicles into early endosomes followed by acidification and intraluminal vesicle formation are a few features that morphologically and biochemically define late endosomes (see Russell et al., 2006 for a thorough review). Induction of intraluminal vesicles (ILVs) marks a stage in late endosomal maturation for the formation of multivesicular bodies (MVBs). MVB morphology can be identified in eukaryotes across the tree of life. ILV and MVB biogenesis is performed by the evolutionarily conserved Endosomal Sorting Complexes Required for Transport (ESCRTs) (Leung et al., 2008; Raiborg & Stenmark, 2009). They are an ancient family likely inherited from the Asgard archaeal ancestor of eukaryotes and are the focus of this study (Spang et al., 2015; Zaremba-Niedzwiedzka et al., 2017).

This chapter had two overarching goals. First was examining ESCRT evolution in fornicates for a comprehensive outlook into the conservation of components necessary for MVB-biogenesis. This is important because MVB-like structures are present in *Carpodemonas* and some CLOs but absent from *Giardia*. Examining this machinery's evolution and molecular complement allowed us to pinpoint and correlate the timing of loss within this MVB morphology in fornicates. Although previous investigations by Leung and colleagues determined the ESCRT repertoire in *Giardia* to be minimal, a fornicate-wide perspective was still lacking (Leung et al., 2008). Through these *in silico* investigations, numerous previously unidentified components were characterized within *Giardia*'s ESCRT repertoire. This chapter's second goal was to understand the precise functions of how and where these components function within *Giardia*'s MVB-less endomembrane system. Although some molecular characterization was previously performed with ESCRTIII-A components, the remainder of the subcomplexes and essential parts that bridge them together were not examined and were biologically still elusive (Saha et al., 2018). Therefore, a functional assessment of the newly identified components and those previously uncharacterized was performed to determine cellular localization and protein-protein interactions using immunofluorescence microscopy and proteomics. Overall, the results from this chapter vastly broaden our perception of the

evolutionary dynamics that are interplay, and which shaped the loss of MVBs in *Giardia*. Unexpectedly, the giardial ESCRTs are promiscuous in their localization and hint at unanticipated functions within the parasite cell. These findings also highlight that ESCRTs are generally adaptable to many roles outside of the typical endocytic pathway.

Work presented in this chapter was published as a primary research article: Pipaliya, S.V., Santos, R., Salas, D., Balmer, E., Wirdnam, C., Roger, A.J., Hehl, A.B., Faso, C., Dacks, J.B. (2021). Unexpected organellar locations of ESCRT machinery in *Giardia intestinalis* and complex evolutionary dynamics spanning the transition to parasitism in the lineage Fornicata. *BMC Biology*, 19(1), 167. <https://doi.org/10.1186/s12915-021-01077-2>. I performed all evolutionary bioinformatic analyses (*i.e.*, comparative genomics, phylogenetics, and CHMP7 structural homology assessments), as well as the CHMP7 sub-cellular fractionation experiments and some immunofluorescence microscopy. The remainder of the molecular functional experiments were completed by Rui Santos, Erina Balmer, Corina Wirdnam, and Carmen Faso. All data interpretations were performed by myself, Joel Dacks, Carmen Faso, Adrian Hehl, Rui Santos, and Andrew Roger. Microscopy imaging and image analysis were done using equipment provided by the Center of Microscopy and Image Analysis (ZMB) at the University of Zurich and the Microscopy Imaging Center (MIC) at the University of Bern. The personnel at the Functional Genomics Centre Zurich (FGCZ) are thanked for their assistance with the mass-spectrometry experiments. Redundancies exist in the introduction with respect to background on *Giardia*'s biomedical and public health relevance and its endomembrane system, which have been provided in the previous chapters.

4.2 Introduction

The food and waterborne diarrheal disease known as Giardiasis causes global healthcare and agricultural burden, with approximately 300 million and more than 10 million cases diagnosed in humans and animals every year, respectively (Lanata et al., 2013). The causative agent is the diplomonad *Giardia intestinalis*. This enteric protist parasite has undergone large genome streamlining and modifications in its typical eukaryotic organelles, particularly its endomembrane system and the associated trafficking complement (Faso & Hehl, 2011).

Giardia relies heavily on its endomembrane trafficking system to secrete virulence factors while establishing gut infection, performing antigenic variation for immune system evasion and interfering with immune responses by degrading or reducing the synthesis of signalling molecules (Allain & Buret, 2020; Eckmann et al., 2000; Faso & Hehl, 2019; Gargantini et al., 2016; Stadelmann et al., 2012). Endomembrane trafficking is also required for completion of the life cycle during encystation which features regulated secretion of large amounts of cyst-wall material through COPII- and COPI- associated lineage-specific encystation specific vesicles (ESVs) (Stefanic et al., 2009). *Giardia*'s endomembrane organization is significantly reduced in its complexity, most notably, because it lacks a canonical Golgi apparatus, readily identifiable early and late endosomes, lysosomes, and peroxisomes (Abodeely et al., 2009; Sheffield & Bjorvatn, 1977). Simplification of the endocytic and secretory pathways in this organism is underlined by

complete loss of several protein complexes associated with membrane trafficking such as AP-3, AP-4, AP-5, TSET, and the protein complexes that are present are often reduced in their complement such as Rabs, Rab GEFs, SNAREs, and ARF GEFs (Elias et al., 2012; Herman et al., 2018; Hirst et al., 2014; Pipaliya et al., 2019; Venkatesh et al., 2017). However, *Giardia* does harbour a tubulovesicular endoplasmic reticulum (ER) thought to carry out functions of the late endosomal pathway (Abodeely et al., 2009). *Giardia* also has endocytic organelles called the peripheral vacuoles (PVs), which perform bulk flow uptake of nutrients from the host environment and cargo sorting for retrograde transport (Cernikova et al., 2020; Zumthor et al., 2016).

Endosomal Sorting Complexes Required for Transport (ESCRTs) are evolutionarily ancient machinery composed of five sub-complexes, ESCRT0/Tom1, ESCRTI, II, III, and III-A, and recruited onto the growing late endosomal surface in a sequential manner to induce intraluminal vesicle formation through negative membrane deformation (Figure 4.1) (Raiborg & Stenmark, 2009). In model eukaryotes, the ESCRT machinery is required for the biogenesis of multivesicular bodies (MVBs), which have endocytic characteristics and the ability to mediate exosome biogenesis and release (Vietri et al., 2020). Additional ESCRT roles have been discovered in plasma membrane repair, autophagy functions, post-mitotic nuclear envelope scission, and others with a shared function in membrane abscission (Hurley, 2015; Vietri et al., 2020). This conserved protein complex is never entirely lost by organisms, underlining its importance, and was already elaborated in the LECA, presumably inherited from the Asgard archaea (Leung et al., 2008; Seitz et al., 2019; Spang et al., 2015). Previous bioinformatics studies have shown *Giardia intestinalis* assemblage AI isolate, WB (hereafter shortened to AWB when full name of isolate is not given) to possess patchy ESCRTII, ESCRTIII, and ESCRTIIIA machinery (Leung et al., 2008). However, key components within each of these were reported to be absent (Dutta et al., 2015; Leung et al., 2008; Moyano et al., 2019; Saha et al., 2018).

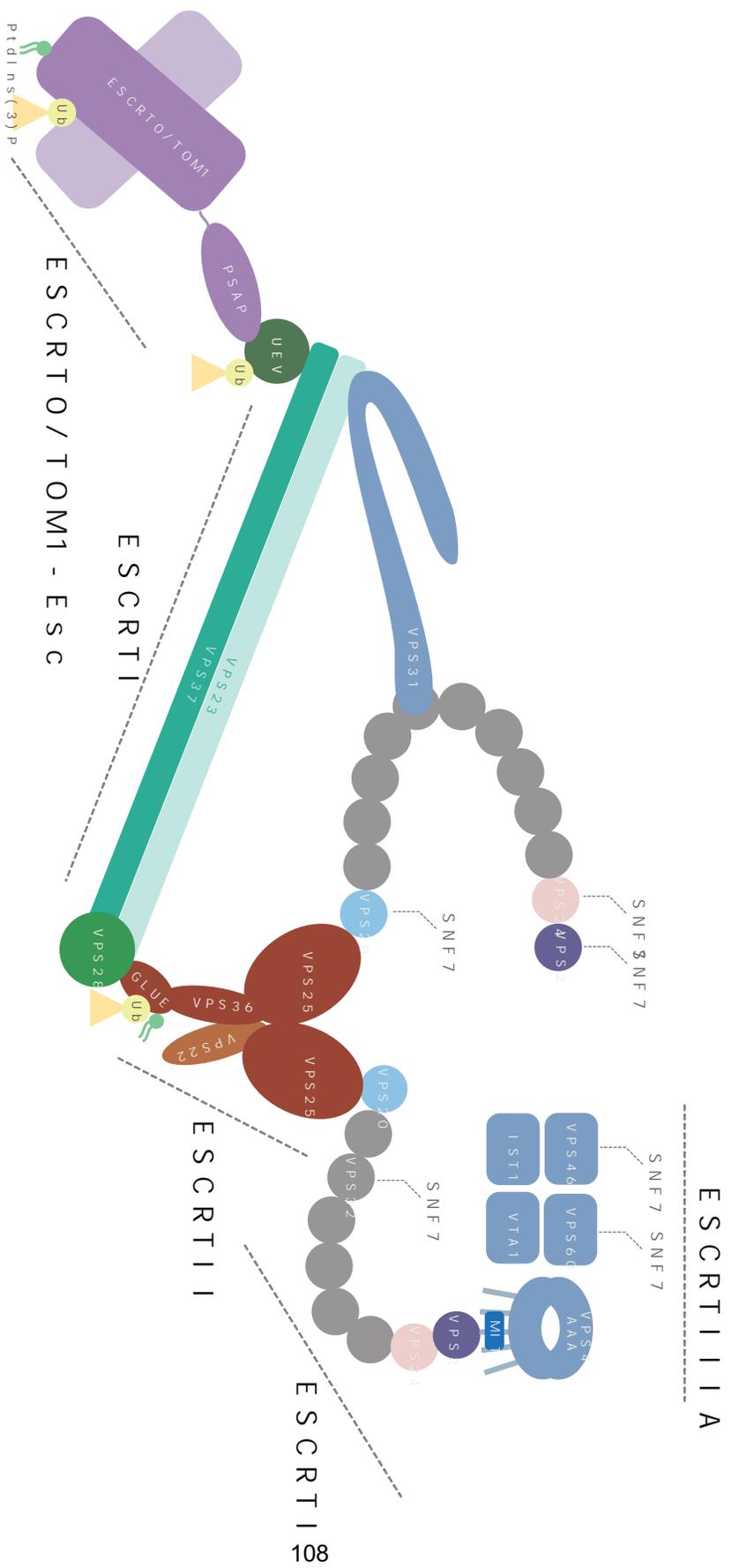
A powerful approach to understanding the evolutionary path to parasitism is to compare protein complements in parasites with those of free-living relatives. *Carpodomonas membranifera* is a small heterotrophic flagellate, and the namesake for the paraphyletic group of free-living organisms (the *Carpodomonas*-like Organisms or CLOs) that diverged basally to the parasitic diplomonads (Takishita et al., 2012). Together the CLOs and diplomonad parasites form the lineage Fornicata, which, in turn, are grouped with other major parasitic groups such as the parabasalids (e.g., *Trichomonas vaginalis*) or anaerobic lineages such as the Preaxostyla in the higher taxonomic ranked Metamonada (Figure 4.2).

To date, although the ESCRT complement of representative metamonads (*Giardia* included) have been reported, no survey has been done of the entire Fornicata lineage nor other *Giardia intestinalis* assemblages, further raising the important evolutionary question of whether the losses reported in *Giardia* evolved concurrently with parasitism or are a product of gradual evolution that predate its movement into this niche.

Our initial approach using bioinformatics traces the evolution of the ESCRT system in the Fornicata, finding losses of ESCRT components across the lineage spanning the transition to parasitism. We also

identified several novel components of the ESCRT machinery in *Giardia* and investigated their subcellular localization, revealing ESCRT association at PVs and other locations. The evolutionary modification of the ESCRT complement spans the transition to parasitism in the lineage leading to *Giardia*, and the modified ESCRT machinery acts at more locations than previously understood in this globally important parasite.

Figure 4.1. Overview of the eukaryotic ESCRT machinery. The ESCRT machinery is composed of five sub-complexes, each functioning consecutively for recruitment of the downstream subcomplex. The process begins with the recruitment of ESCRT0 or its analogue TOM1-esc for recognition of tagged Ubiquitin on cargo, and endosomal membrane phospholipids such as phosphatidylinositol 3-phosphate (PtdIns(3)P) upon which the ESCRTI, composed of Vps23, Vps28, and Vps37, is recruited, with its only known role being ubiquitin recognition via its UIM domain (Raiborg & Stenmark, 2009). The assembly of ESCRTI then leads to nucleation of the heterotetrameric ESCRTII consisting of Vps36, Vps22, and two copies of Vps25, which also bind to PtdIns(3)P via the FYVE domains (Raiborg & Stenmark, 2009). Finally, this leads to the recruitment of the ESCRTIII machinery, a heteropentameric complex consisting of SNF7 domain-containing family proteins, Vps20, Vps32, Vps2, Vps24, and CHMP7 (Raiborg & Stenmark, 2009). A filamentous Vps32 polypeptide capped by Vps2 and Vps24 (also belonging to the paralogous SNF7 domain-containing family of proteins) induces ILV formation by constricting the neck of the budding vesicle, a process which is catalyzed by the ESCRTIII-A-Vps4, an AAA+ ATPase (Raiborg & Stenmark, 2009). It is also hypothesized that ESCRTIII-A components such as Vps31 and Vps46 are required for stabilizing the sub-complexes during the budding processes, while others are needed for recycling the complexes back into the cytosol once the process is complete (Raiborg & Stenmark, 2009). This figure was adapted from Stenmark and Raiborg (2009).



4.3 Materials and Methods

4.3.1 Taxa studied

The previously-published draft genome of *Kipferlia bialata*, the genome of *Spironucleus salmonicida*, the transcriptome of *Trepomonas* sp. PC1, the genome of *Giardia intestinalis* assemblage AI, isolate WB, the genome of *Giardia intestinalis* assemblage AII, isolate DH, the genome of *Giardia intestinalis* assemblage AII, isolate AS175, the draft genome of *Giardia intestinalis* assemblage B, isolate GS, the genome of *Giardia intestinalis* assemblage B, isolate GS-B, the genome of *Giardia intestinalis* assemblage B, isolate BAH15c1, the genome of *Giardia intestinalis* assemblage E, isolate P15, and the genome of *Giardia muris* were obtained from GiardiaDB and National Centre for Biotechnology Information (NCBI) (Adam et al., 2013; Ankarklev et al., 2015; Franzén et al., 2009; Jerlström-Hultqvist et al., 2010; Morrison et al., 2007; Tanifuji et al., 2018.; Wielinga et al., 2015; Xu et al., 2016; Xu et al., 2020a; Xu et al., 2020b). Latest assemblies were used in each case and the database sources and corresponding GenBank assembly accessions have been summarized in Online Appendix Table 4.1.

4.3.2 Translation of nucleotide transcriptomes belonging to Carpediemonas-like organisms

Nucleotide transcriptomes of *Carpediemonas membranifera* and five *Carpediemonas*-like organisms (CLOs), *Aduncisulcus paluster*, *Ergobibamus cyprinoides*, *Dysnectes brevis*, *Chilomastix cuspidata*, and *Chilomastix caulleryi* were obtained from the Dryad Repository and translated using the *ab initio* gene prediction program, GeneMarkS-T under the default parameters (Leger et al., 2017; Tang et al., 2015).

4.3.3 Comparative genomics and homology searching

Query protein sequences for individual subunits from each ESCRT sub-complex belonging to various pan-eukaryotic representatives were obtained and aligned using MUSCLE v. 3. 8.31 (Berriman et al., 2005; Carlton et al., 2007; Curtis et al., 2012, 2012; Edgar, 2004; Eichinger et al., 2005; Eisen et al., 2006; El-Sayed et al., 2005; Fisk et al., 2006; Fritz-Laylin et al., 2010; Gardner et al., 2002; Karnkowska et al., 2019; Lorenzi et al., 2016; Salas-Leiva et al., 2021; The Arabidopsis Genome Initiative, 2000; Tyler et al., 2006; Wheeler et al., 2008) (Query accessions and database sources summarized in the Online Appendix Table 4.1). Resulting alignments were used to generate Hidden Markov Models using the hmmbuild option available through the HMMER 3.1.b1 package to perform HMMER searches into all Fornicata genomes and transcriptomes using the hmmsearch tool with an e-value cutoff set to 0.01 (Eddy, 1998). Non-redundant forward hits were deemed positive if BLASTP reciprocally retrieved the correct ortholog from the *Homo sapiens* protein database with an e-value ≤ 0.05 and were two orders of magnitude better in e-value than the next best hit. Reciprocal hits were extracted and sorted using an in-house Perl script.

Additional analyses of hits that failed to retrieve any reciprocal hits were analyzed by BLASTP against the NCBI non-redundant database (Altschul et al., 1990; Pruitt, 2004). Additional orthology

assessment was performed using the HHPRED suite for an HMM-HMM profile comparison and predicted secondary structure homology comparisons with proteins deposited in the Protein Data Bank (Gabler et al., 2020). In order to rule out any false negatives, additional translated nucleotide (TBLASTN) searches were carried out in the Fornicata nuclear scaffolds for components that remained unidentified in HMMER searches. In cases where diplomonad sequences were unidentified due to extreme sequence divergence, identified *Carpodomonas membranifera* and CLO ESCRT orthologs were used to search the diplomonad predicted protein databases by subsequently adding these sequences into the previously generated HMM profile to build a new HMM matrix. Additionally, in order to maximize robustness of paralogue count and potential strain-specific differences, any *Giardia*-specific paralogues identified were then used as queries for forward BLAST searches into the other *Giardia* genomes. Exhaustive BLASTP and TBLASTN analyses were also performed using *Carpodomonas membranifera* and CLO sequences in the nuclear scaffolds of all diplomonads. All Fornicata ESCRT orthologs identified by this method were subject to domain analyses using the Conserved Domain Database (CDD) with an e-value cut-off first set at 0.01 and then at 1.0 to detect any highly diverged domains (Lu et al., 2020). All confirmed hits are listed in Online Appendix Table 4.2.

CHMP7 structural analyses were performed using HHPRED, as described above for select pan-eukaryotic orthologs, and with the *ab initio* structural prediction tool iTASSER for protein threading and secondary structure prediction (Roy et al., 2010). HHPRED results are summarized in Online Appendix Table 4.3.

4.3.4 Phylogenetic analyses of the ESCRTIII and ESCRTIIIA-SNF7 family proteins

Phylogenetic analyses of the evolutionarily paralogous SNF7 family proteins belonging to ESCRTIII and ESCRTIIIA sub-complexes were carried out using Bayesian and Maximum Likelihood approaches (Leung et al., 2008). Identified *Carpodomonas membranifera* ESCRT genes belonging to the SNF7 family (*i.e.*, Vps2, Vps24, Vps20, Vps32, Vps46, and Vps60) were used as landmark representative sequences. They were aligned to a backbone alignment containing previously characterized pan-eukaryotic SNF7 sequences as resolved and published by Leung et al. (2008). This was performed using the profile option in MUSCLE v3.8.31 (Edgar, 2004). All alignments were visualized in Mesquite v. 3.5 and manually adjusted to remove gaps and regions lacking homology (Maddison & Maddison, 2019). Upon classification of the *Carpodomonas* sequences, a metamonad-specific phylogenetic analysis was undertaken to classify identified *Giardia* and diplomonad SNF7 sequences using the same process, as described above. Additional set of phylogenetic analyses were repeated to generate separate Vps2, Vps24, and Vps46-specific and Vps20, Vps32, and Vps60-specific trees.

Maximum likelihood-based phylogenetics using non-parametric and ultrafast bootstrapping were performed using RAxML-HPC2 on XSEDE v. 8.2.10 and IQTREE2, respectively (Nguyen et al., 2015; Stamatakis, 2014). For RAxML analyses, protein model testing was performed using ProtTest v. 3.4.2 (Darriba et al., 2011). In all cases, the LG + F + Γ model was used. One hundred non-parametric bootstraps

with the default tree faster hill-climbing method (-f b, -b, -N 100) were specified. A consensus tree was obtained using the Consense program, available through the Phylip v. 3.66 package (Felsenstein, 1989). IQTREE2 best protein model selections were determined using the in-built ModelFinder package (Kalyaanamoorthy et al., 2017). In all cases, LG+F+G4 was determined to be the best-fit model according to the Bayesian Information Criterion. Ultrafast bootstrapping with IQTREE v. 2.0.6 was performed using 1000 pseudo-replicates (Nguyen et al., 2015). Bayesian inference phylogenetics were performed using MRBAYES on XSEDE v. 3.2.6 with 10 million Markov Chain Monte Carlo generations under a mixed amino acid model with the number of gamma rate categories set to 4 (Huelsenbeck & Ronquist, 2001). Sampling frequency was specified to occur every 1000 generations, and a burnin value was set to 0.25 to discard the first 25% of samples from the cold chain. Tree convergence was ensured when the average standard deviation of split frequency values fell below 0.01. A random seed value of 12345 was chosen for all phylogenetic analyses. Non-parametric and ultrafast bootstraps obtained from RAxML and IQTREE analyses were overlaid onto the MRBAYES tree topologies with posterior probabilities. RAxML and MrBAYES analyses were performed on the CIPRES portal while IQTREE v. 2.0.6 was installed and run locally (Miller et al., 2010). All trees were visualized and rooted in FigTree v.1.4.4, and annotations were performed in Adobe Illustrator CS4 (Rambaut, 2010).

Trimmed alignments (Supplementary Alignments 1 – 10) used for the phylogenetic investigations have been made publicly available through the Figshare repository (<https://doi.org/10.6084/m9.figshare.14393495.v1>).

4.3.5 *Giardia cell culture and transfection*

Giardia intestinalis strain WB (clone C6; ATCC catalog number 50803) trophozoites were grown using standard methods, as described by Morf et al. (Morf et al., 2010). Episomally-transfected parasites were obtained via electroporation of the circular pPacV-Integ-based plasmid prepared in *E. coli* as previously described (Zumthor et al., 2016). Transfectants were selected using Puromycin (final concentration 50 µg/ml; InvivoGen). Transgenic lines were generated and analyzed at least thrice as soon as at least 20 million transgenic cells could be harvested (*i.e.*, *ca.* 1.5 weeks post-transfection). Based on microscopy analyses of immunofluorescence assays to detect reporter levels, 85-92% of cells expressed their respective transgene(s) (Appendix Figure 4.1). For an extended protocol on *Giardia* cell culture and transfection, see section 5.2.1 in the subsequent chapter.

4.3.6 *Construction of expression vectors*

Oligonucleotide sequences for construct generation are listed in Online Appendix Table 4.4. Open reading frames of interest were cloned in the pPacV-Integ vector under control of their putative endogenous promoters. Putative endogenous promoters were derived 150bps upstream of the predicted translation start codon. ORFs were cloned in a modified PAC vector (Cernikova et al., 2020; Wampfler et al., 2014; Zumthor et al., 2016).

4.3.7 Immunofluorescence Assays

Chemically fixed cells for subcellular recombinant protein localization were prepared as previously described (Konrad et al., 2010). Hemagglutinin (HA) epitope-tagged recombinant proteins were detected using a rat-derived monoclonal anti-HA antibody (dilution 1:200; Roche) followed by a secondary anti-rat antibody coupled to Alexa Fluor 594 fluorophore (dilution 1:200; Invitrogen). For co-localization experiments with ER or mitochondrial markers, samples were incubated with either a mouse-derived anti-GiPDI2 or a mouse-derived anti-GiIscU primary antibody both using a dilution of 1:1000, followed by incubation with anti-mouse antibody coupled to Alexa Fluor 488 fluorophores (dilution 1:200, Invitrogen) (Rout et al., 2016; Stefanic et al., 2009). For the labelling of the V5 epitope, we used an anti-V5 primary antibody (1:1000; Thermofisher) followed by an anti-mouse antibody coupled to Alexa Fluor 594 fluorophores (dilution 1:200, Invitrogen). Samples were embedded in Vectashield (VectorLabs) or Prolong Diamond Mounting medium (Invitrogen) containing 4',6-diamidino-2-phenylindole (DAPI) for nuclear staining.

4.3.8 Fluid phase marker uptake

Dextran uptake assays were performed using Dextran, Texas Red™, 10,000 MW, Lysine Fixable (2 mg/mL; Invitrogen), as described previously (Gaechter et al., 2008; Zumthor et al., 2016). Immunostaining was performed as described above but using lowered TritonX-100 concentration (0.05% in 2% BSA; Sigma) for permeabilization, to prevent leakage and loss of Dextran signal. For extended versions of the immunofluorescence assay and Dextran-uptake protocols, see section 5.2.3 in the subsequent chapter.

4.3.9 Microscopy and Image Analysis

Imaging was performed under an inverted Leica SP8 Laser Scanning Confocal Microscope using appropriate parameters. Confocal images were subsequently deconvolved using Huygens Professional (<https://svi.nl/Huygens-Professional>) and analyzed using Fiji/ImageJ (Rueden et al., 2017; Schindelin et al., 2012). For co-localization analyses, the coloc2 Fiji/ImageJ plugin was used. For this, automatic background subtraction was performed in Fiji/ImageJ, and 100 Costes' iterations were performed (Costes et al., 2004). Three-dimensional analysis and videos were reconstructed in Imaris version 9.5.0 (Bitplane, AG) (Online Appendix Videos 1-11). A macro was developed in Fiji/ImageJ for statistical analysis of signal overlap between ESCRT subunits and specified markers (version 1.53d). This script has been made available through the Online Appendix File 1. Briefly, each channel was thresholded via WEKA segmentation, a machine learning pipeline (Arganda-Carreras et al., 2017). The derived binary image was used as a mask for signal overlap on ≥ 15 cells per sample/line using the Fiji plugin coloc2 (Costes et al., 2004).

4.3.10 Co-immunoprecipitation with limited cross-linking

Co-immunoprecipitation (co-IP) assays on transgenic trophozoites expressing either HA-tagged GiCHMP7, GiVps25, GiVps36A or GiVps20L, were performed in limited cross-linking conditions using 2.25% formaldehyde to stabilize protein complexes and enrich for weaker protein interactions, as previously

described (Rout et al., 2016). For an extended version of the co-IP protocol, see section 5.2.5 in the subsequent chapter.

4.3.11 Protein analysis and sample preparation for mass spectrometry (MS)-based protein identification

SDS-PAGE analyses were performed on 4%-10% polyacrylamide gels under reducing conditions. Blotting was done by using a primary rat-derived anti-HA antibody (dilution 1:500; Roche) followed by an anti-rat antibody coupled to horseradish peroxidase (dilution 1:2000; Southern Biotech), as previously described (Konrad et al., 2010). Gels for liquid-chromatography mass spectroscopy (LC/MS) analyses were stained with Instant Blue (Expedeon) and de-stained with ultrapure water. As previously reported, LC/MS-based protein identification was performed (Cernikova et al., 2020; Konrad et al., 2010; Rout et al., 2016; Zumthor et al., 2016).

4.3.12 Crude subcellular fractionation

Crude subcellular fractionation experiments were performed on HA-GiCHMP7-expressing transgenic trophozoites, as per previously established protocol (Rout et al., 2016). Briefly, transgenic cells were lysed by freeze-thawing using liquid nitrogen. Soluble (supernatant) and membrane-enriched fractions (pellet) were isolated by centrifugation at 14,000x *g* for 10 minutes at 4°C. Both fractions were subject to immunoprobng using a rat-derived monoclonal anti-HA antibody (dilution 1:500; Roche) followed by a secondary anti-rat antibody coupled to horseradish peroxidase (dilution 1:2000; Southern Biotech) and Coomassie staining. Non-fractionated whole cell lysates from both non-transgenic and HA-GiCHMP7-expressing parasites were prepared and analyzed similarly.

4.3.13 *in silico* co-immunoprecipitation proteomics analyses

The co-IP datasets derived from transgenic cells expressing epitope-tagged “baits” as affinity handles were filtered using dedicated control co-IP datasets generated from non-transgenic wild-type parasites to identify candidate interaction partners unique to bait-specific datasets (Cernikova et al., 2020; Rout et al., 2016; Zumthor et al., 2016). This was done using Scaffold4 (<http://www.proteomesoftware.com/products/scaffold/>). Unless otherwise indicated, bait-derived co-IP data were filtered using high stringency parameters (Exclusive Spectrum Counts (ESCs) at 95-2-95, 0% FDR) and manually curated to rank putative interaction partners in a semi-quantitative fashion using ESCs as a proxy for relative abundance.

Access to raw mass spectrometry data is provided through the ProteomeXchange Consortium on the PRIDE platform (Perez-Riverol et al., 2019). Data is freely available using project accession number and project DOI. Individual project DOIs/accessions for datasets derived from bait-specific and control co-IP LC/MS analyses are as follows: 10.6019/PXD016487 (HA-GiCHMP7), 10.6019/PXD016442 (HA-GiVps20L), 10.6019/PXD016448 (GiVps25-HA) and 10.6019/PXD016446 (GiVps36A-HA). LC/MS hits have also been summarized in Online Appendix Table 4.5.

4.4 Results

4.4.1 ESCRT losses in *Fornicata* are gradual and represent a slow transition leading to parasitism

To understand the extent to which the loss of ESCRT components correlates with parasitism versus pre-dating it, we investigated the complement encoded in the transcriptomes of free-living *Carpodiemonas membranifera* and *Carpodiemonas*-like organisms (CLOs) by comparative genomics (Figure 4.2; Online Appendix Table 4.2). In the case of ESCRTIII and ESCRTIIIA SNF7 components (*i.e.*, Vps20, Vps32, Vps60, Vps2, Vps24, and Vps46) which are themselves homologous, we also performed phylogenetic analyses for classification. We took a two-step approach to account for divergent fornicate sequences, first classifying *Carpodiemonas membranifera* sequences and subsequently using these as landmarks to characterize and verify the classification of SNF7 components in the various fornicate representatives.

This analysis allowed us to resolve the presence of nearly all SNF7 components with clear clustering with pan-eukaryotic orthologs (Figure 4.3A). The exception was the lack of clear Vps32 or Vps20 orthologs in *Carpodiemonas*. Instead, multiple Vps20-like proteins were identified (Figure 4.3A; Online Appendix Figure 4.2). This could imply that one of these protein paralogs may carry out the functions of canonical Vps32 or Vps20. We do not rule out the possibility that orthologs of Vps32 and Vps20 are present in the *Carpodiemonas* gene repertoire, which remained unexpressed in standard culturing conditions and, therefore, absent within the assembled transcriptome. Phylogenetic analyses of the identified SNF7 sequences in the remaining CLOs and diplomonads, including the seven *Giardia intestinalis* isolates, further revealed that similar to *Carpodiemonas membranifera*, Vps20 or Vps32 proteins in all Fornicata lineages have diverged to the extent that no apparent clades are resolvable to determine these as canonical Vps20 or Vps32 (Figure 4.3B; Online Appendix Figure 4.3).

We also notably detected a CHMP7 ortholog in several fornicate representatives, including *Giardia* and the free-living *Chilomastix cuspidata* and *Dysnectes brevis* (Figure 4.2). We further examined these proteins through domain analyses, which revealed that the characteristic C-terminal SNF7 domain normally required for the recruitment of downstream ESCRTIII-Vps20 and Vps32 was absent from all identified CHMP7 orthologs. Following the same pattern as the Vps20L protein, this finding implies partial loss of sequence and divergence of the CHMP7 sequences predates the fornicate common ancestor. Overall, our investigation of the free-living fornicate transcriptomes in direct comparison with the parasitic diplomonads and various isolates of *Giardia* has been useful in retracing the timepoints and instances of ESCRT sequence divergence.

Figure 4.2. Distribution of the ESCRT components within Fornicata. This coulson plot depicts the ESCRT complement identified in Fornicata genomes and transcriptomes compared to the repertoire previously identified across pan-eukaryotic representatives. Filled sectors indicate subunits with solidified orthology determined using both comparative genomics and phylogenetics. Numbers within individual sectors represent multiple paralogues. Light coloured sectors indicate ambiguous phylogenetic classification but confirmed reciprocal blast orthology. Taxa for which genomes were examined are indicated in plain text, whereas lineages, where only transcriptomes were available, are indicated with a superscript symbol. Lineages belonging to the paraphyletic group of *Carpediemonas*-like organisms are indicated with asterisks. Parasitic fornicates are outlined in a burgundy-coloured font. Paralogous ESCRTIII and ESCRTIII-A subunits possessing the SNF7 domain with common evolutionary origins and for downstream phylogenetic investigations are underlined. Of important note, only inferences regarding gene presence, not absences, can be made conclusively in the lineages for which only a transcriptome is available.

4.4.2 Losses correlating with parasitism and inter-strain variation

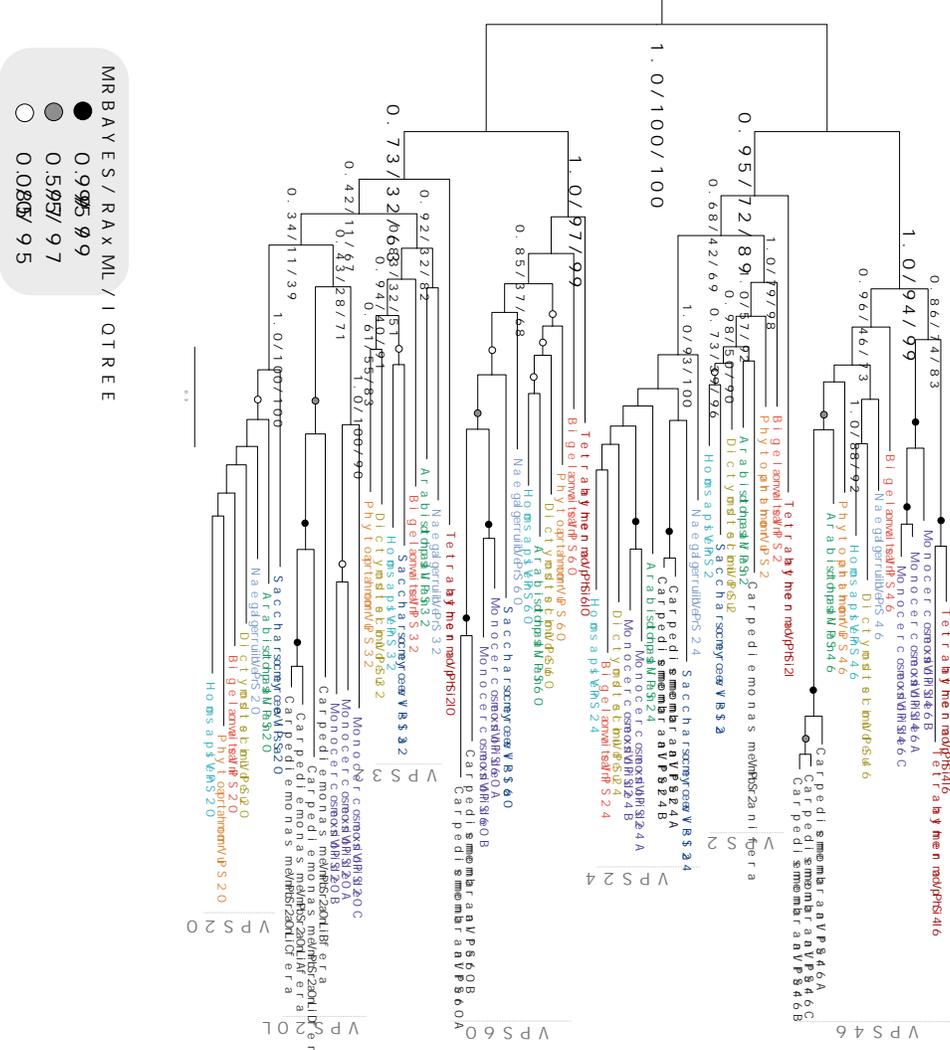
Focusing more specifically on parasitic lineages, including seven *Giardia* isolates, the fish parasite *Spironucleus salmonicida*, and the secondarily free-living *Trepomonas* sp. PC1 shows additional losses when compared to their free-living relatives (Figure 4.2). Within the ESCRTIII machinery, Vps2 and Vps24 are present in *Carpodomonas*; however, we were unable to characterize any of the identified SNF7 proteins as canonical Vps2 in either *Giardia* or *Trepomonas* sp. (Figure 4.3A; Online Appendix Figure 4.4). Instead, phylogenetic classification pointed towards homology to Vps24, and therefore, these proteins were termed Vps24-like (Vps24L) proteins in diplomonads (Figure 4.3B; Online Appendix Figure 4.5). Additionally, the coincident loss of VTA1 and Vps60, which interact to regulate Vps4 oligomerization, in all diplomonads hints at the dispensability of the ESCRTIII-A components and that alternative factors (or potential paralogs of the identified components) may be at play to carry out these functions (Yang et al., 2012) (Figure 4.2; Figure 4.3B; Online Appendix Figure 4.3). Other losses common to all diplomonads include ESCRTI-Vps37 and Vps28 that are not only absent in *Giardia* but also *S. salmonicida* and *Trepomonas* sp. PC1. These indicate adaptive genome streamlining likely occurred in the Last Diplomonadida Common Ancestor (Figure 4.2). By contrast, although greater streamlining has occurred in the diplomonads with respect to other fornicates, the presence of Vps23 in *Trepomonas* sp. PC1 still hints at the capacity of these lineages to form canonical multivesicular bodies.

We also observed unanticipated protein complement differences between the two human-infecting assemblages, A and B. Assemblage A isolates, WB and ADH possess two Vps24 paralogues, with one clustering with other canonical Vps24 orthologs from other excavates, the other forming a clearly separate clade, here termed Vps24L (Figure 4.3B; Online Appendix Figure 4.5). Additionally, we failed to identify any orthologs of Vps20L proteins in assemblage B isolates, BGS and BGS-B. Lastly, we find a similarly encoded ESCRT repertoire between the assemblage A and EP15 strains, as well as phylogenetic clustering of the EP15 sequences with ADH and AWB, consistent with a proposed closer relationship of these strains to one another than to assemblage B (Figure 4.2; Figure 4.3B).

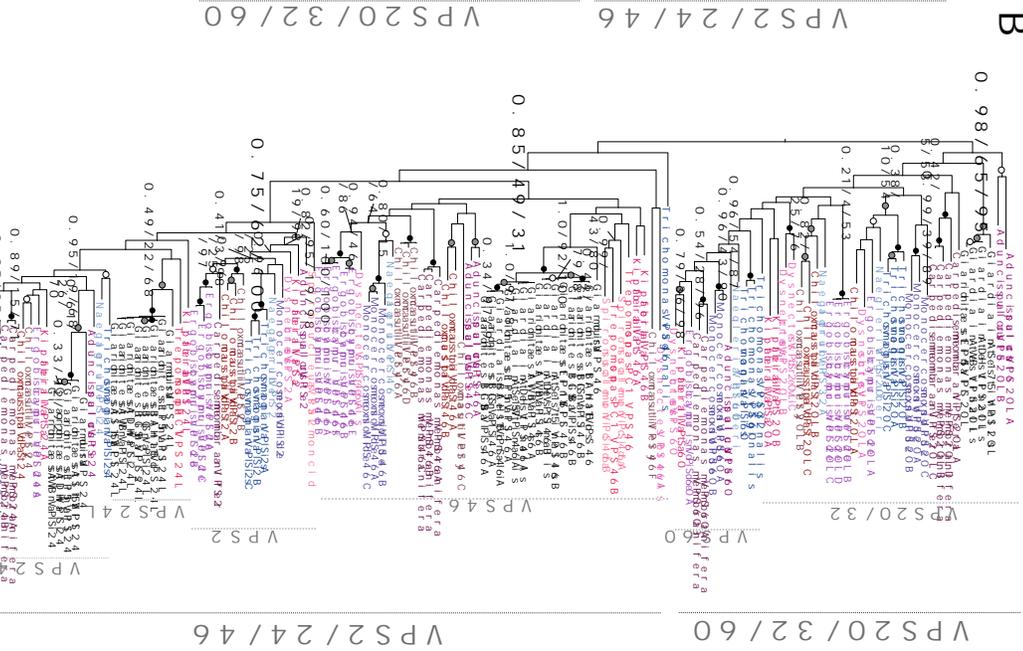
Previous work analyzing only the *Giardia intestinalis* AWB ESCRT machinery reported absences in various components such as ESCRTII-Vps36, ESCRTIII-CHMP7, and ESCRTIIIA subunits (Dutta et al., 2015; Moyano et al., 2019). Here we show these to be present but were not previously detected, probably due to high sequence divergence and the lack of the currently available genomes and or transcriptomes from free-living relatives of *Giardia* (Figure 4.2; Online Appendix Table 4.2).

Figure 4.3. Phylogenetic analyses of the SNF7 domain-containing ESCRTs in Fornicata. (A) depicts phylogenetic analyses of the ESCRTIII/IIIA SNF7 families in Fornicata. Identified ESCRTIII/IIIA SNF7 components from the basal *Carpodiemonas membranifera* as landmark representative sequences for Fornicata were subject to phylogenetic classification. Two of the identified SNF7 sequences from *Carpodiemonas membranifera* clustered clearly with Vps60, whereas the remainder neither strongly grouped with Vps20 or Vps32, and therefore, were determined to be Vps20L proteins in all tree topologies. *Carpodiemonas membranifera* was also determined to have Vps2, Vps24, and Vps46 with strong backbone clade support for two paralogs of Vps24 (1.0/100/100) and three paralogs of Vps46 (1.0/100/100). (B) depicts a Fornicata-specific tree with well-characterized Discoba and metamonad representatives. *Monocercomonoides exilis*, *Trichomonas vaginalis*, and *Naegleria gruberi*, as well as newly characterized sequences from *Carpodiemonas membranifera*, were used to classify SNF7 components in all CLOs and diplomonads. Similar to *Carpodiemonas*, no clear grouping of SNF7 sequences from CLOs within the Vps20 or Vps32 clade was observed, and therefore, were also classified as Vps20L. Only sequences from *Giardia* AWB, ADH, and EP15 formed a group within this clade and therefore were also determined to be Vps20L. Vps2 family proteins identified in the diplomonads grouped with both Vps24 and Vps46 with duplication event pointing in *Giardia* sp. Vps46 to yield two paralogues, Vps46A and Vps46B. An additional set of SNF7 family proteins from *Giardia* AWB, ADH, and EP15 grouped with excavate and CLO Vps24 proteins, therefore were determined to be Vps24-like proteins. However, an additional set of SNF7 proteins from all *Giardia* lineages formed a separate sister clade, and therefore, also termed to be Vps24. Trees were rooted between the Vps20/32/60 and Vps2/24/46 as previously determined by Leung et al. (2008).

A



B



4.4.3. Localization of *Giardia* ESCRTII-Vps25 and newly identified ESCRTII-Vps36 at peripheral vacuoles

Previous molecular cell biological analyses of ESCRTs in *Giardia* have been limited to highly conserved ESCRTIII and ESCRTIII-A components (Saha et al., 2018). The bioinformatic identification of multiple newly described ESCRT components, particularly some with unclear phylogenetic affinity (e.g., Vps20L), make attractive targets for molecular cell biological investigations.

We began by characterizing the ESCRTII component Vps25, consistently identified in previous phylogenetic analyses but never investigated at a subcellular level. Given previous reports on *Giardia* ESCRTIII components and assuming functional homology from model systems, *GiVps25* was predicted to associate with the PVs (Saha et al., 2018). Immunofluorescence assays of standalone staining in transgenic trophozoites expressing *GiVps25* C-terminally HA-epitope-tagged reporters (*GiVps25*-HA) revealed an accumulation in the cell periphery and a punctate cytosolic pattern (Figure 4.4A-I, II; Online Appendix Video 1; Online Appendix Figure 4.1-I). Signal overlap analyses were performed on cells ($N \geq 15$) labelled for *GiVps25*-HA and incubated with the endocytic fluorescent fluid-phase marker Dextran coupled to Texas Red (Dextran-TxR) support a partial association of *GiVps25*-HA to PVs (Figure 4.4B-I-III; Online Appendix Video 2; Online Appendix Figure 4.6-I). The *GiVps25*-HA-derived signal seemed widespread but punctate throughout the cell, suggestive of multiple locations (Figure 4.4B-I-III). By contrast, the Dextran-TxR signal was clearly restricted to the cell periphery, consistent with an exclusive PV location (Figure 4.4B-I-III).

To analyze the potential signal overlap between Dextran-TxR and Vps25, we performed co-localization analysis with the ImageJ/Fiji plugin Coloc2, which evaluates co-localization between two signals based on a series of computed parameters (Rueden et al., 2017; Schindelin et al., 2012). These parameters are the Pearson's coefficient, which computes a relation between the overlapped signal in channels of interest, Manders' coefficients (M1 and M2), which output the percentage of each channel overlapping with the other, and the Costes' p-value, which determines the obtained results being true or false, where a value between 0.95 and 1 denotes true signal overlap (Costes et al., 2004; Li et al., 2004; Manders et al., 1993). Between Vps25 and Dextran signals, the Pearson's coefficient describing overall signal overlap was low, as was the Manders' coefficient 1, quantifying the degree of *GiVps25*-HA overlap with Dextran-TxR (Figure 4.4B-II and III). However, Manders' coefficient 2, describing the degree of Dextran-TxR overlap with *GiVps25*-HA, was high, as was the Costes' value, giving us confidence in our results. Overall, these data show that the *GiVps25*-HA reporter localizes to PVs, consistent with past reports of other ESCRT components functioning at this organellar system. However, *GiVps25*-HA is also found at other locations within the *Giardia* cell.

We proceeded to characterize one of the putative ESCRT components newly identified in our bioinformatic analysis, *GiVps36*, hereafter referred to as *GiVps36A* (Online Appendix Table 4.2). A molecular cell biological approach here was particularly informative, given that only one of the three *GiVps36* paralogues possesses a potential GLUE domain, which was retrieved only weakly. In model systems, this functional module (a type of split Pleckstrin Homology (PH) domain) mediates interactions between the ESCRTI and ESCRTII sub-complexes, which are in turn necessary for ubiquitin-dependent

initiation for ILV biogenesis. Instead, *Giardia* Vps36 paralogues possess an N-terminal PH domain (Online Appendix Table 4.2), raising questions of functional homology of this component with that of other model organisms (Gill et al., 2007). We chose to test the localization of *GVps36A*, as this was readily identified by homology searching and thus likely to be the least divergent in function. As with *GVps25-HA*, a localization pattern associated with the cell periphery and punctate cytosolic foci was observed with the *GVps36A-HA* reporter construct (Figure 4.5A, Online Appendix Video 3; Online Appendix Figure 4.1-II). Signal overlap analyses on cells labelled for *GVps36A-HA* and incubated with Dextran-TxR support partial *GVps36A-HA* association to the PVs (Figure 4.5B - II and III; Online Appendix Video 4; Online Appendix Figure 4.6-III). The Manders' coefficients again suggested PV localization, where M1 represents the signal overlap of *GVps36A-HA* channel with Dextran, and M2 represents the signal overlap between the Dextran and the *GVps36A-HA* channels (Figure 4.5B-II and III). This again suggests that a considerable proportion of the Dextran signal overlaps with Vps36 in these two channels, but that Vps36 is present where Dextran is, as well as elsewhere in the cell.

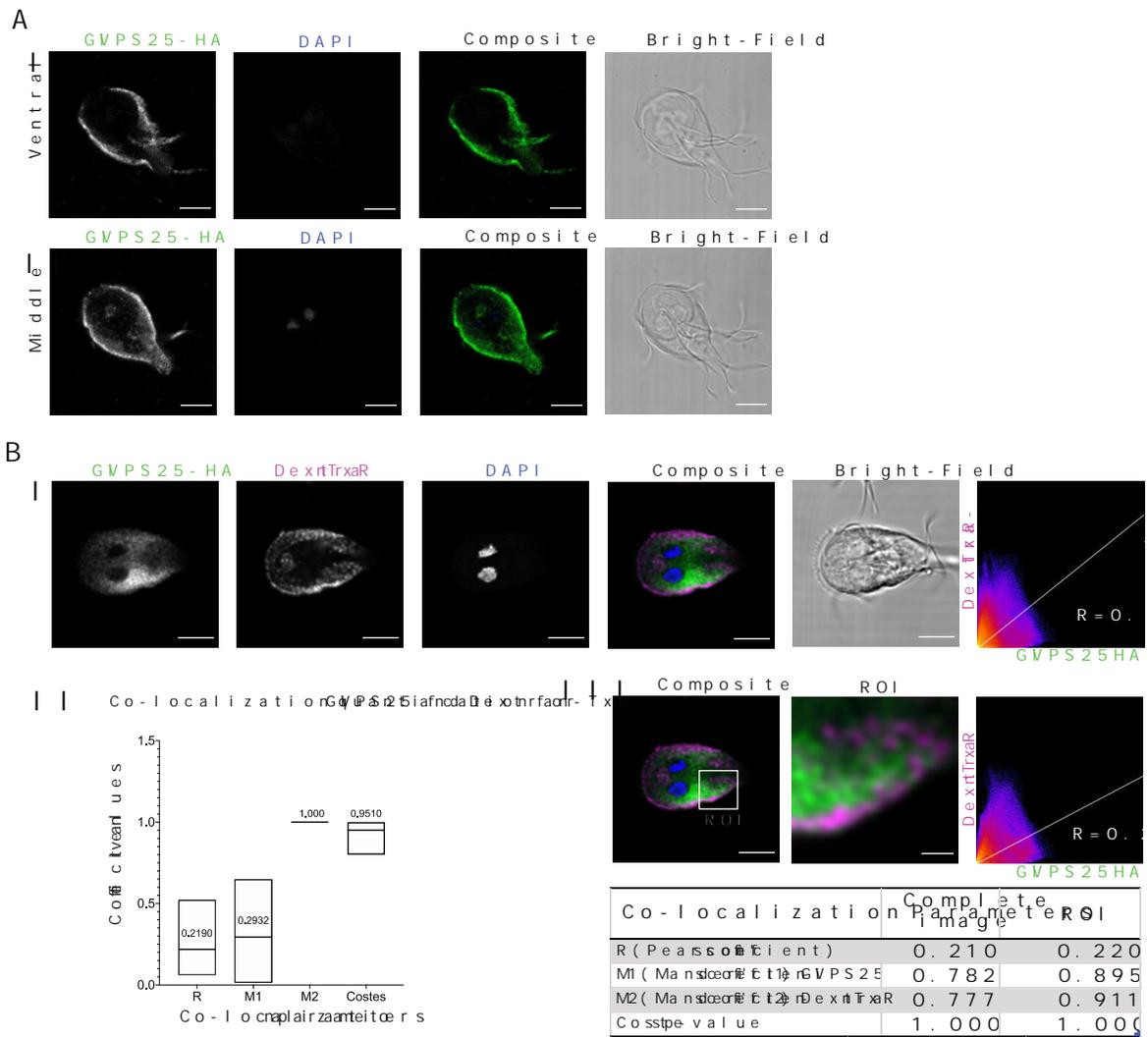


Figure 4.4. Characterization of *GVps25-HA* subcellular location. (A) represents trophozoite cell periphery and cytosol. (I) Ventral and (II) middle optical slices of transgenic *Giardia* trophozoite expressing epitope-tagged *GVps25-HA* (green) are depicted. (B) focuses on PV staining. (I) depicts immunoprobings of transgenic *Giardia* trophozoites labelled for epitope-tagged *GVps25-HA* (green) and Dextran-TexasRed (magenta). (II) is a boxplot depicting the distribution of co-localization parameters for *GVps25-HA* and Dextran-TexasRed labelling from ≥ 15 analyzed cells. Mean values for each parameter are indicated. (III) is signal overlap analyses and co-localization coefficients calculated for all slices of the sample using either whole cell or region of interest (ROI). Scale bars: composite 5 μm and ROI 1 μm . All images were obtained using Laser Scanning Confocal Microscopy.

4.4.4 Characterization of ESCRTIII-Vps20L and ESCRTII components at the endoplasmic reticulum

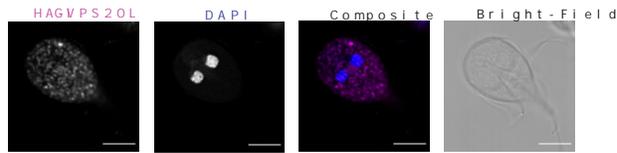
The newly identified ESCRTIII-Vps20L protein family was phylogenetically unresolved in our analyses, and the *G. intestinalis* AWB sequence relatively divergent. Both the novelty and divergence of this protein prompted us to investigate this protein further. *GVps20L* was expressed as an N-terminally epitope-tagged reporter (HA-*GVps20L*) and detected by immunofluorescence localization assay (Figure 4.6A; Online Appendix Video 5; Online Appendix Figure 4.1 - Panel III) where we observed punctate and dispersed cytosolic localization, previously seen with *GVps25*-HA and *GVps36A*-HA and reminiscent of ER association of components (Faso et al., 2013).

This observation, along with the potential for other organellar localization suggested for Vps25 and Vps36A, prompted us to investigate whether all three proteins might be ER-associated. To do this, we proceeded with signal overlap analyses of cells ($N \geq 15$) co-labelled for each epitope-tagged reporter in combination with the ER membrane marker *GiPDI2* (Stefanic et al., 2009) (Figures 4.6B-D). The data shows ESCRT proteins Vps25 (Online Appendix Video 6; Online Appendix Figure 4.6-II), Vps36A (Online Appendix Video 7; Online Appendix Figure 4.6-IV), and Vps20L (Online Appendix Video 8; Appendix Figure 4.6-V) partially associated with the ER (Figures 4.6B-D-II and III). We interpret the low M1 coefficients (*i.e.*, measuring the respective ESCRT components overlap with PDI) but high M2, as most consistent with the ESCRT proteins localized to ER as well as other cellular locations (*e.g.*, the PVs).

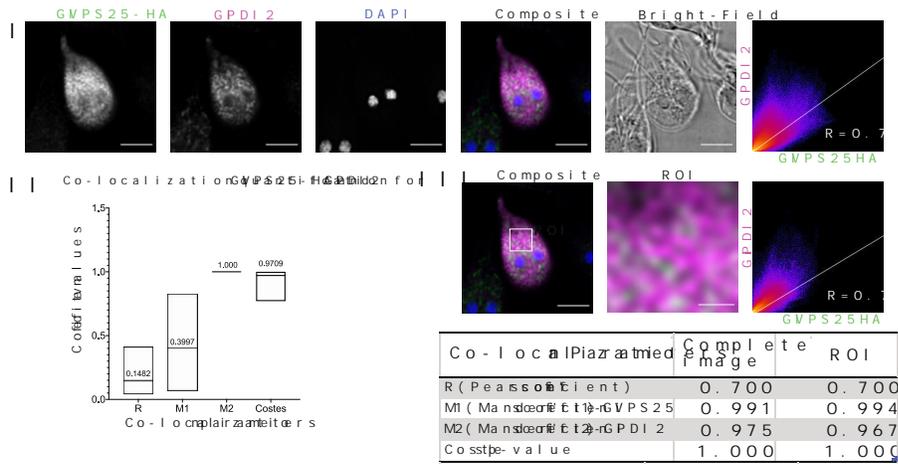
The observation of Vps25 being in ostensibly the same locations as Vps36A and Vps20L, at PVs and ER, respectively, leads to the prediction of overlap in localization for these two proteins. This was assessed by developing and investigating dually-transgenic *Giardia* lines expressing *GVps25*-HA in combination with either *GVps36*-V5 (Figure 4.7A-I) or V5-*GVps20L* (Figure 4.7B-I). Based on signal overlap analysis of the co-labelled cells (≥ 15), there is significant co-localization in subcellular location for both *GVps25*-HA and *GVps36A*-V5, and *GVps25*-HA and V5-*GVps20L* in co-expressing whole cells and highlighted ROIs (Figure 4.7A -II and III; Figure 4.7B-II and III). Notably, the Pearson's coefficients and both Manders' coefficients are substantially higher for ESCRT component overlap (Figure 4.7) than observed for the previous co-localizations against organellar markers (Figures 4.4-4.6). Indeed, these values are higher for the components that in characterized model systems take part in the same sub-complex (ESCRTII) (*i.e.*, Vps25 and 36, than for Vps25 with Vps20, which is predicted to be in the ESCRTIII sub-complex). Together these all suggest a consistent picture of a multi-faceted cellular ESCRT localization in the *Giardia* cell.

Figure 4.6. Co-labelling of *G/Vps25*-HA, *G/Vps36A*-HA and *G/HA-Vps20L* with ER membrane marker *G/PDI2*. (A) HA-*G/Vps20L* is found in the cytosol and as punctate structures. Scale bars: 5 μm . (B-D) Panels I represent co-labelling of PDI2 (magenta) in cells expressing either (B) *G/Vps25*-HA (green), (C) *G/Vps36A*-HA (green) or (D) HA-*G/Vps20L* (green). (B-D) Panels II depict mean values from ≥ 15 analyzed cells for each parameter. (B-D) Panels III represents signal overlap analyses and co-localization coefficients calculated for all slices of the sample, either for the whole cell or ROI. Scale bars: composite 5 μm and ROI 1 μm . All images were obtained using Laser Scanning Confocal Microscopy.

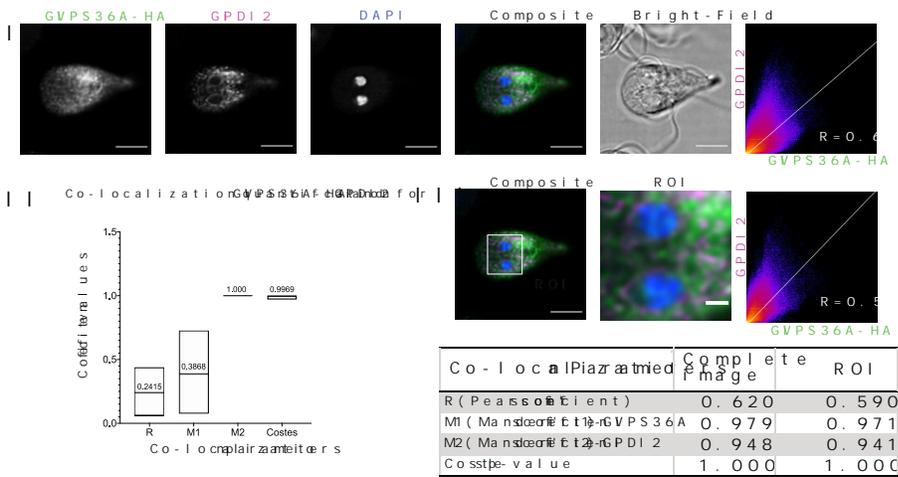
A



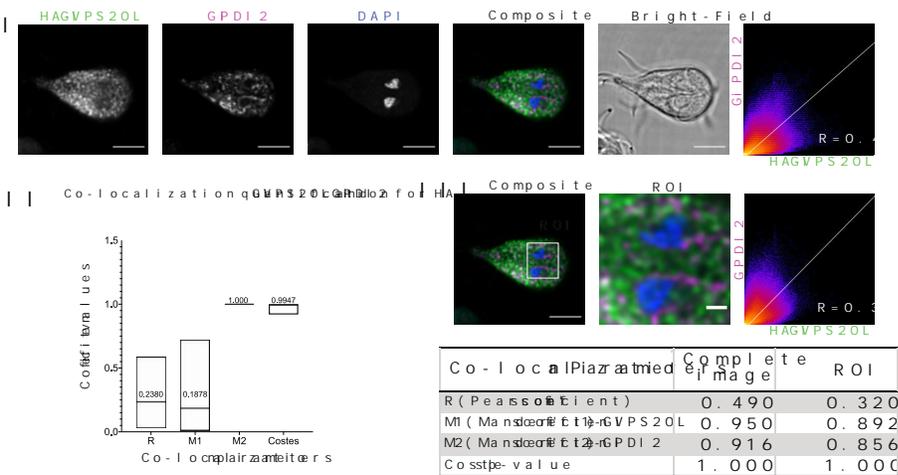
B



C



D



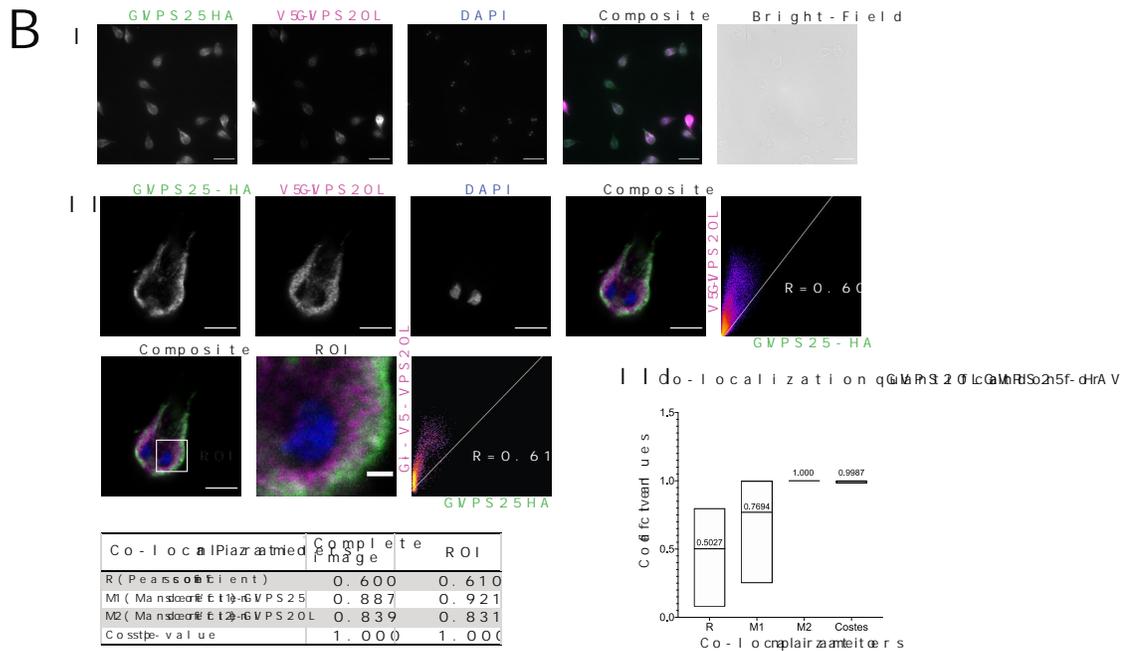
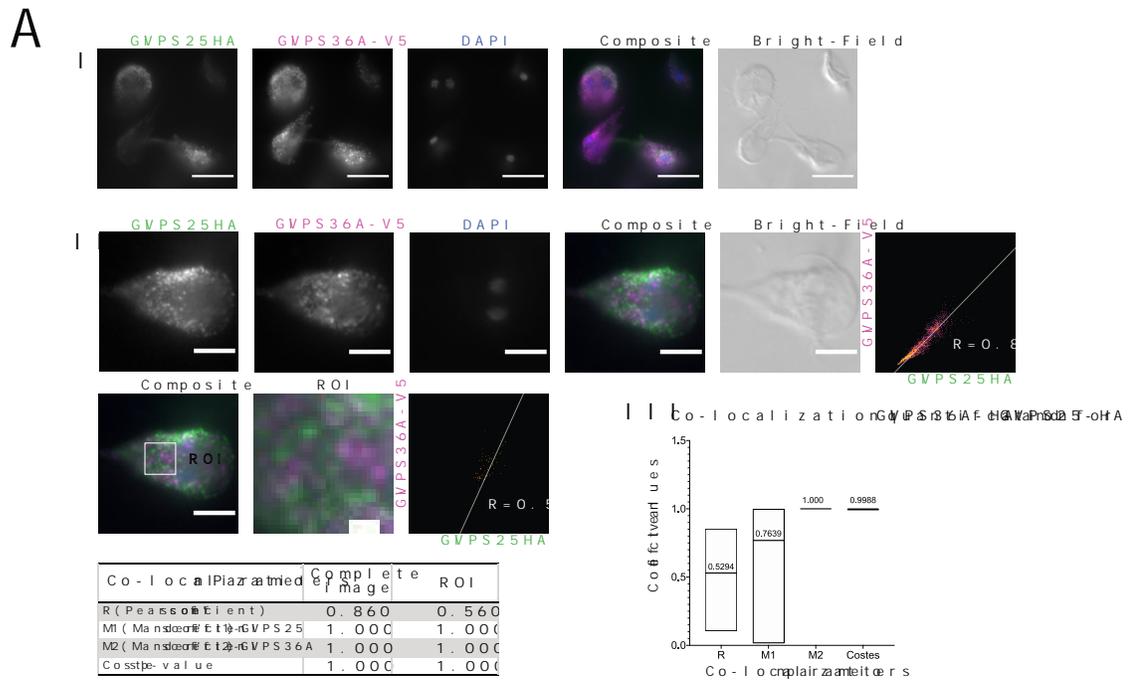


Figure 4.7. Co-expression of epitope-tagged GVPs25 with either GVPs20L or GVPs36. Microscopy analyses of cells co-expressing GVPs25HA (green) with either (A) GVPs36A-V5 (magenta) or (B) GVPs20L (magenta) is depicted. Panels I are representative cell images and percentage of co-labelling. Panels II depict signal overlap analyses in both whole cells and regions of interest (ROI). Scale bars: (I) 10 μ m, (II whole cell) 5 μ m, and (II-ROI) 1 μ m. Panels III are mean values obtained from ≥ 15 analyzed cells for each parameter. All images were obtained using Laser Scanning Confocal Microscopy.

4.4.5 Evolutionary and protein analyses of the newly identified *Giardia* ESCRTIII-CHMP7 reveal unsuspected ancient origins and a novel ER-mitosomal localization

Perhaps the most surprising finding from the comparative genomics analysis was the identification of CHMP7 homologues in multiple Fornicata representatives, despite it being frequently unidentified in many genomes across eukaryotes (Figure 4.2). Fornicate CHMP7 proteins were also highly divergent, missing the C-terminal domain in both *Giardia* and the CLO orthologs.

CHMP7 is currently proposed as being derived from a pre-LECA fusion of two SNF7 domains (Horii et al., 2006). However, the potential homology of the N-terminus to Vps25 has also been suggested (Olmos & Carlton, 2016). In order to first validate our putative CHMP7 candidates as not being divergent in-paralogs of SNF7 or Vps25, we undertook a combined phylogenetic and structural homology approach. HHPRED and iTASSER analyses of the *Giardia* CHMP7 showed a lack of a predicted C-terminal SNF7 domain. Notably, they showed sequence and structural homology of the remainder of this protein (*i.e.*, the N-terminus) to Vps25 (Figure 4.8; Online Appendix Table 4.3). Homology searching analyses with selected CHMP7 N-termini from several representatives of other eukaryotic supergroups confirmed this assessment, retrieving Vps25 as the only homologous protein with any significant e-values (Figure 4.8A; Online Appendix Table 4.3). Notably, SNF7 derived proteins were never retrieved amongst the candidate homologs. Given the exclusive homology to Vps25 indicated by the HHPRED and structural-prediction results, the identity of the fornicate proteins as CHMP7 and not as in-paralogues of Vps25 was also confirmed through our phylogenetic analysis (Online Appendix Figure 4.7). Our collective findings suggest that a duplication event followed by a fusion event between the Vps20/32 SNF7 and Vps25 had occurred prior to the last eukaryotic common ancestor but subsequent to eukaryogenesis from the presumed Asgard archaeal ancestor.

CHMP7 has been demonstrated to have a variety of functions beyond the endocytic pathway in mammalian and yeast model cell systems (Bauer et al., 2015; Gu et al., 2017; Vietri et al., 2020). Therefore, following the identification of this protein in *Giardia*, we aimed to investigate its role in the endomembrane system of this parasite. Based on *Gi*CHMP7's similarity to Vps25 and lack of an SNF7 domain, we expected similar localization patterns as ESCRTII components, specifically at the PVs and the ER. However, our immunofluorescence assay analyses with an N-terminally epitope-tagged *Gi*CHMP7 reporter (HA-*Gi*CHMP7) yielded a distinct localization pattern strongly reminiscent of ER labelling, with no obvious indication of PV association (Figure 4.9A; Online Appendix Figure 4.1-IV; Online Appendix Video 9). As done for *GV*ps25-HA, *GV*ps36A-HA, and HA-*GV*ps20L, HA-*Gi*CHMP7 cells were co-labelled for *Gi*PDI2, and a signal overlap analysis was performed (N \geq 15 cells), showing that HA-*Gi*CHMP7 is partially ER-associated, particularly taking M2 (*i.e.*, the signal overlap between the PDI2 channel and the ESCRT subunit channel) and the Costes' p-values into account (Figure 4.9B -II and III; Online Appendix Video 10; Online Appendix Figure 4.6-Panel VI).

Surprisingly, we repeatedly detected HA-*Gi*CHMP7 signal in compartments consistent with the location of central mitosome complexes (CMC) (Regoes et al., 2005) (Figure 4.9C; Online Appendix Video

11). To test this, we co-labelled HA-*GiCHMP7*-expressing cells with antibodies directed against iron-sulfur cluster assembly component *GiIscU* to detect mitochondria (Rout et al., 2016) (Figure 4.9C-I; Online Appendix Figure 4.6-Panel VII). We measured significant signal overlap limited to the CMC with *GiCHMP7* and *GiIscU*-derived labels, with the low M2 denoting *CHMP7* presence at multiple cellular locales, but very high M2 values indicating substantial overlap with the *IscU* signal (Figure 4.9C -II and III).

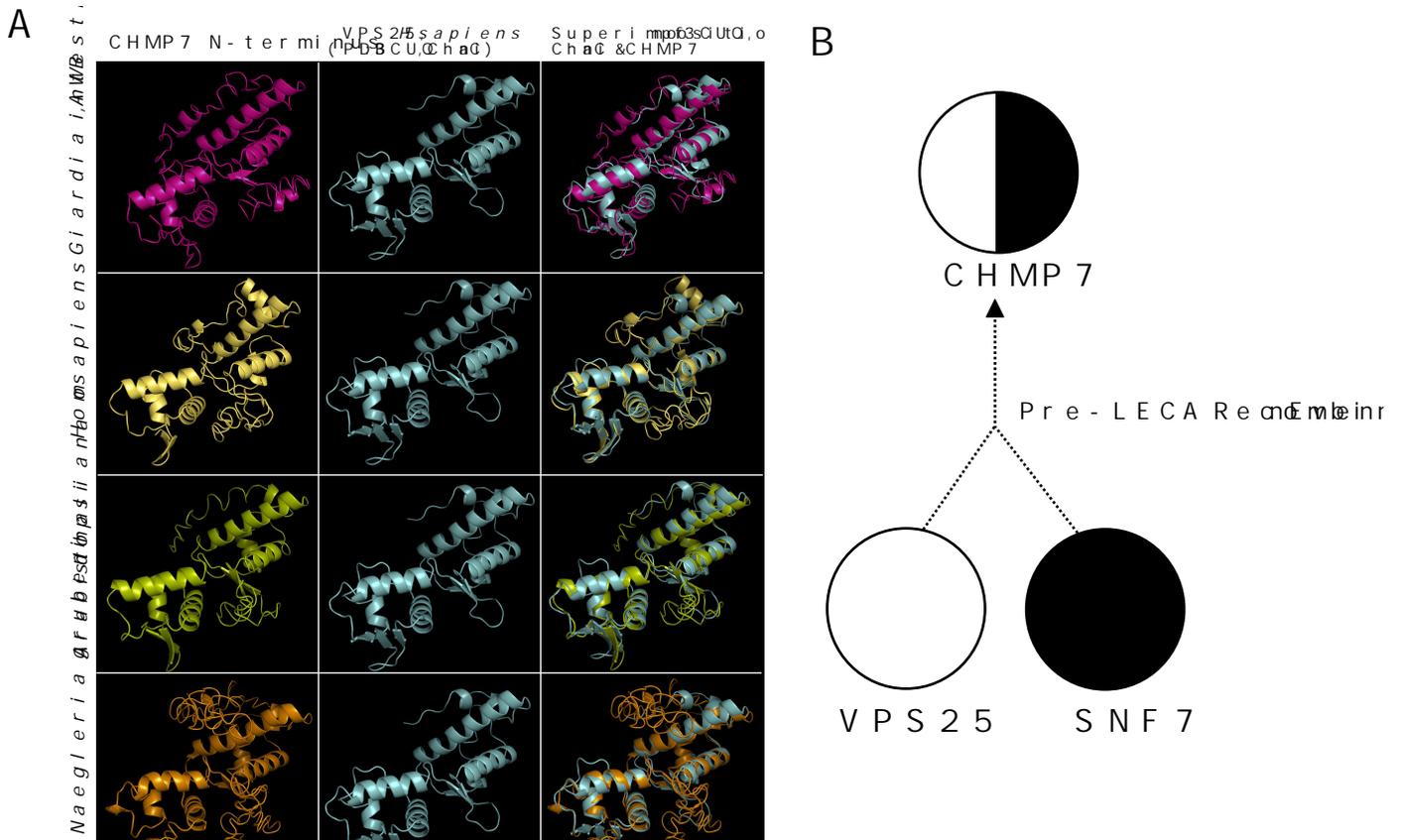
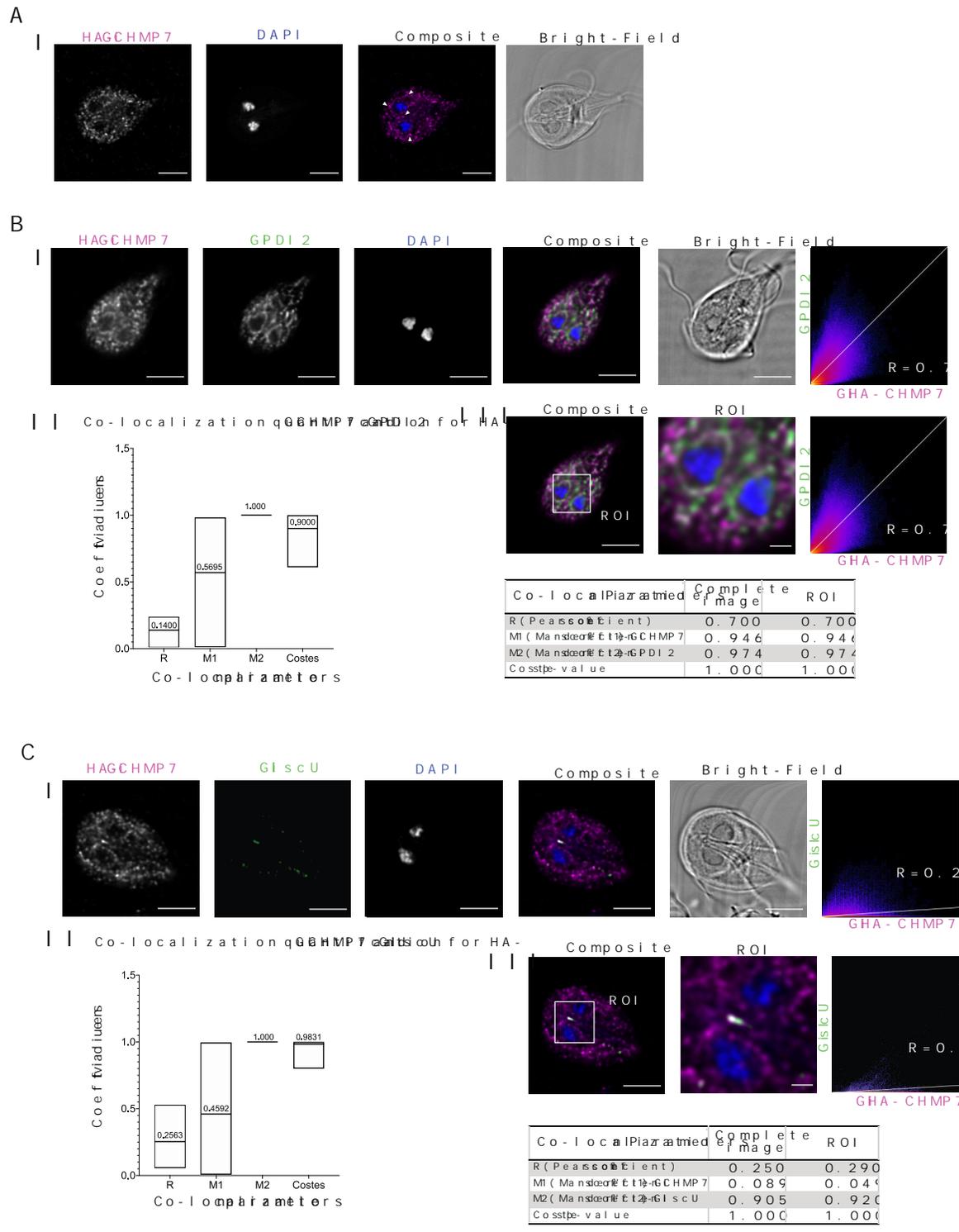


Figure 4.8. *Ab initio* homology-based structural analysis of the CHMP7 N-terminus. (A) Homology-based protein structural analysis of the CHMP7 N-terminus from various pan-eukaryotic representatives was carried out using iTASSER, an *ab initio* structural prediction program, where considerable similarity between the ESCRTII-Vps25 and CHMP7 N-termini was observed. **(B)** Proposed evolution of the pan-eukaryotic CHMP7 protein prior to the last eukaryotic common ancestor, as determined by homology searching, structural analyses, and phylogenetic analysis, which consisted of an evolutionary fusion event between a pre-LECA ESCRTII-Vps25 and ESCRTIII/IIIA-SNF7 progenitor protein (Online Appendix Table 4.3; Online Appendix Figure 4.7)

Figure 4.9. Characterization of GiCHMP7 subcellular location. (A) depicts immunofluorescence microscopy results with HA-GiCHMP7-expressing cells, which yield a diffused punctate pattern with elements of perinuclear ER staining (arrowhead). Scale bars: 5 μm . (B) (I) represents co-labelling of HA-GiCHMP7 (magenta) -expressing cells with GiPDI2 (green). (II) is a boxplot depicting the distribution of the co-localization parameters for HA-GiCHMP7 and GiPDI2 labelling from ≥ 15 analyzed cells. Mean values for each parameter are indicated. (III) depicts signal overlap analyses for all slices of the sample, either using whole-cell or ROI. (C) shows HA-GiCHMP7 to be associated with the *Giardia* mitosomes. (I) depicts co-labelling of HA-GiCHMP7 (magenta) -expressing cells with GilscU (green). (II) is a boxplot distribution of the co-localization parameters for HA-GiCHMP7 and GilscU labelling from ≥ 15 analyzed cells. Mean values for each parameter are indicated. (III) represents signal overlap analyses for all slices of the sample, either using whole-cell or ROI. Scale bar: composite 5 μm and ROI 1 μm . All images were obtained using Laser Scanning Confocal Microscopy.



4.5 Discussion

Giardia intestinalis presents divergent cellular and genomic features and remains an enigma from an evolutionary standpoint. Our work has specifically addressed the reduced endomembrane system observed in *Giardia intestinalis*, focusing on the ESCRT protein machinery from an evolutionary and molecular cell biological perspective. We show that the reduced ESCRT complement is the product of an evolutionary process that spans the shift from free-living to a parasitic state and includes *Giardia* assemblage-specific losses. We also report on previously unidentified ESCRT machinery and sites of ESCRT location in *Giardia*, opening novel avenues for investigation.

4.5.1 Gradual reductive evolution of ESCRTs and MVBs in the Fornicata

Observation of an unusual trait in a prominent parasite can lead to the default assumption that the non-canonical state is due to parasitism. But this correlation need not be causal and can be assessed by more fine-grained taxonomic sampling. Here we have assessed this exact question regarding the ESCRT complement in *Giardia* and its parasitic and non-parasitic relatives. We have found that, a few traits that do seem to correlate with the transition to parasitism in diplomonads, the history of ESCRT system complement modulation is more textured.

Based on the lifestyles of the basally paraphyletic assemblage of CLOs, including *Carpediemonas*, the ancestor of Fornicata was likely a free-living anaerobic flagellate (Leger et al., 2017). In these conditions, membrane trafficking machinery would be expected to play essential roles in phagotrophy, material exchange, osmoregulation, and intracellular homeostasis. From our analysis, this ancestor appears to have possessed a relatively complete complement of ESCRT machinery compared with the deduced complement in the LECA. That said, there were likely some component losses that had already taken place, including the CHMP7 SNF7 C-terminus normally required for association with the ESCRTIII-Vps32 (Figure 4.10). While it is technically possible that “true” orthologs of these proteins may be encoded in the not-yet sequenced genomes of CLOs, given that the pattern remains consistent across 14 different sampling points, it is much more likely for an ancestral loss to have occurred in the ancestor of fornicates, rather than multiple instances of unexpressed protein or independent losses. Loss in the SNF7 domain of CHMP7 may functionally relate to the other deduced loss observed in all free-living fornicates, that of a canonical Vps32 protein.

By contrast, the transition to parasitism appears to have happened by the time of the diplomonad common ancestor. Concurrent with this are losses of Vps28, Vps60, VTA1, and possibly Vps37 (Figures 4.2 and 4.10). These are correlated, though not necessarily causally associated, with this transition. Notably, however, Vps23 is retained in some diplomonads and is characterized by the presence of a UEV domain which is required for interaction with cargo tagged with Ubiquitin for targeted lysosomal degradation. Lineages such as *Tetrahymena*, *Entamoeba*, and *Monocercomonoides* conserving only Vps23 from ESCRTI appear to be capable of forming functional (or at least morphologically identifiable) multivesicular bodies (Cole et al., 2015; Karnkowska et al., 2019; Okada & Nozaki, 2006). In turn, this allows us to predict

that all fornicate lineages possessing ESCRTI-Vps23, including the diplomonad *Trepomonas* sp. PC1 may also possess *bona fide* MVBs.

In the common ancestor of *Giardia* itself, we observed loss in all ubiquitin-binding components and domains. Collectively, these include TOM1-esc, the entire ESCRTI, and Vps36-GRAM and NZF domains. We speculate that the observed lack of canonical MVB morphology in *Giardia intestinalis* specifically corresponds to losses within these components and that the existing repertoire suggests an altered role for ESCRT machinery at the *Giardia*-specific late endo-lysosomal organelle, the PVs. Notably, we also observed variability between the different *Giardia* genomes in their repertoire of ESCRTIII and -IIIA components, indicating inter-strain variability in this membrane-trafficking complement.

These differences are particularly evident between the two human-infecting assemblages A and B and have been noted in other membrane trafficking system proteins such as the ARF GTPase regulatory system proteins (Pipaliya et al., 2021). Previous reports of genome assembly and comparative genomic investigations into virulence gene families such as variant surface proteins also noted similar patterns of inter-assemblage variability. Such differences were particularly within genome sizes (*e.g.*, assemblage B isolates possess a genome that is approximately 11 Mb in size whereas assemblage A isolates are approximately 10 Mb), protein-coding complement, and sequence-level differences in key virulence genes such as variant surface proteins, cathepsins, and cysteine proteases (Allain et al., 2019; Ankarklev et al., 2015; Franzén et al., 2009; Jerlström-Hultqvist et al., 2010). This has also led to the postulation that in humans, Giardiasis may be caused by two different species of *Giardia* with differences in infection potential and underlying disease manifestation. Future clinical studies and *Giardia* assemblage population-level genome and comparative genomics studies should investigate this notion further.

PV vesicle-like contents have been recently observed in *Giardia* (Midlej et al., 2019; Moyano et al., 2019). However, although the parasite may secrete non-exosomal and non-MVB-derived extracellular vesicles, the absence of key ESCRT machinery (*e.g.*, TOM1-esc, ESCRTI, and Vps36-GRAM domain), along with a lack of specific exosomal markers and limited proteomics data keeps the status of PV-associated vesicles in some doubt (Coelho & Singer, 2018; Ma'ayeh et al., 2017). An alternate interpretation is the PVs as a form of reduced and functionally limited MVB-like compartments in a similarly reduced *Giardia* endomembrane system which evolved via a process of merging organelle identity and distribution of endocytic function. Although PVs may not have a direct organellar homologue, it is still meaningful to understand which processes have been distributed to which organelles in this re-organization.

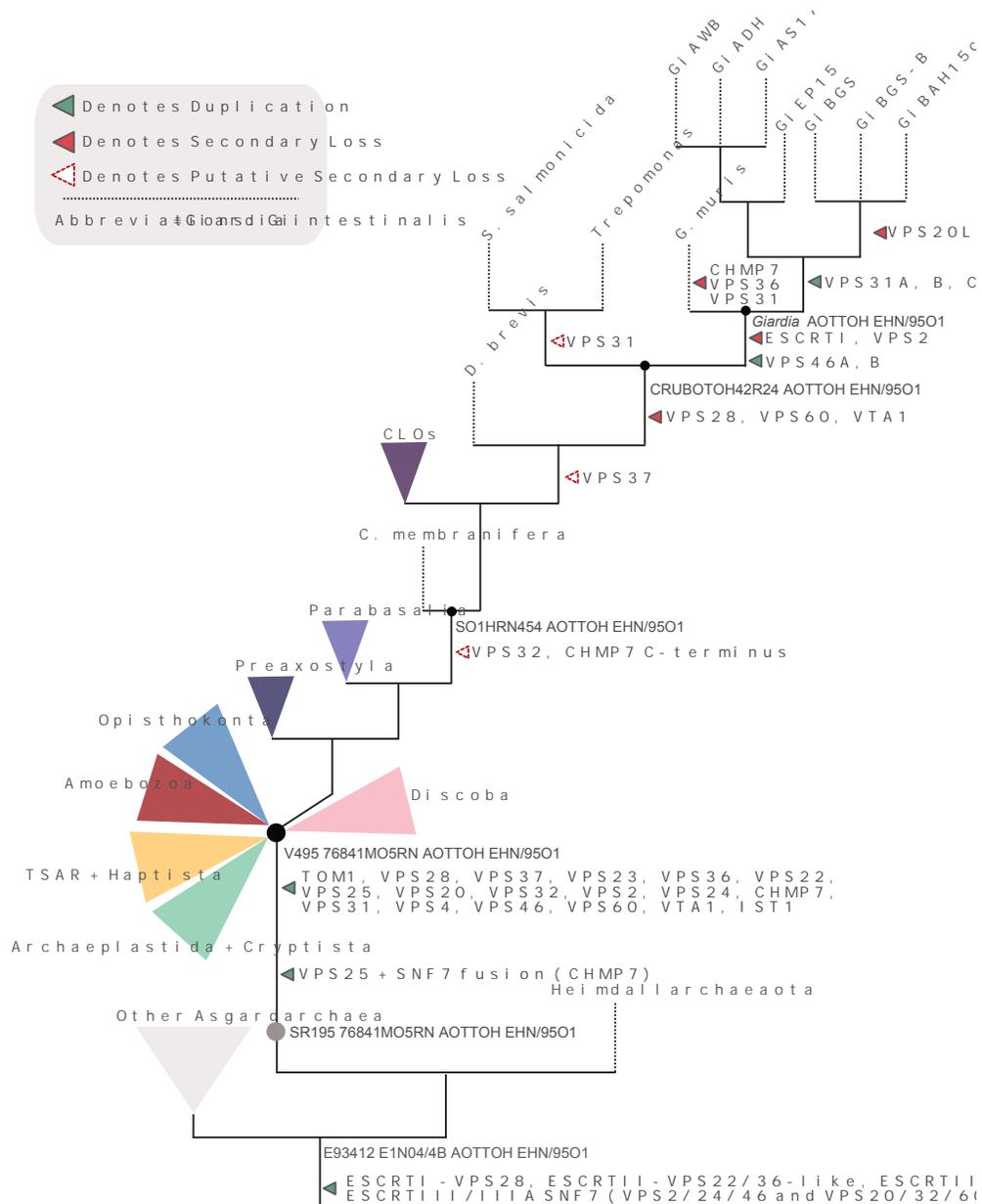


Figure 4.10. Proposed ESCRT evolution in Fornicata. Progenitor ESCRT complexes are present in Asgard archaea, and duplications into the specific subunits are inferred to have occurred between the First Eukaryotic Common Ancestor and the Last Eukaryotic Common Ancestor, which possessed a full complement of the ESCRT subunits. Proposed ESCRT losses in Fornicata inferred previously only using *Giardia intestinalis* are transient, with some losses potentially predating the Last Fornicata Common Ancestor (Leung et al., 2008; Saha et al., 2018). The most prominent of these being loss in the CHMP7 C-terminus SNF7 domain and a canonical Vps32. Examination of diplomonad lineages, specifically genomic data, increases our confidence in additional losses also occurring with progression into parasitism. Most notably, these were within the ESCRT I machinery with its complete loss occurring in the *Giardia* common ancestor and likely to be associated with a lack of canonical MVB morphology. Speculative losses are indicated as unfilled dotted arrows, whereas likely true gene absence is depicted as solid filled arrows.

4.5.2 ESCRT promiscuity at *Giardia* PVs, ER, and mitosomes

Previous investigations of the *Giardia* ESCRTIII components determined a possible role for this complex at the endo-lysosomal peripheral vacuoles (Dutta et al., 2015; Moyano et al., 2019; Saha et al., 2018). While ESCRTIII components Vps4 and Vps46 are universally conserved in all eukaryotes, ESCRTII is not (Leung et al., 2008). Therefore, we aimed to investigate the role of this protein complex that is usually required for bridging an existing ESCRTI and ESCRTIII in the multivesicular body pathway and how *Giardia* may be utilizing it in the absence of ESCRTI.

The imaging data and signal overlap analyses performed with tagged reporters for both *GVps25* and *GVps36* and fluorescent Dextran as a soluble PV lumen marker support a PV association for both ESCRT components. The link between ESCRTs and the endocytic pathway and PVs is further corroborated by cross-referencing previously published co-IP datasets derived from PV-associated endocytic components. This highlights the presence of ESCRT proteins in these PV-centric interactomes (Cernikova et al., 2020; Zumthor et al., 2016). Tagged reporters for α and β subunits of AP-2 collectively immunoprecipitated ESCRT components *GVps36B*, *GVps36A*, *GVps4A*, *GVps4B*, *GiIST1*, *GVps24A*, and the three *GVps31* paralogs (Zumthor et al., 2016). *Giardia*'s first characterized dynamin-related protein (*GiDRP*) pulled down ESCRTIII-Vps46B, Vps31A, and Vps31C. *Giardia* clathrin heavy and putative light chains' interactomes, similar to interactomes for the predicted PH-domain carrying PV-associated *GiNECAP1* protein, include ESCRTIII subfamily components *GVps4A*, *GVps4B*, and the three paralogs of *GVps31* (Cernikova et al., 2020; Zumthor et al., 2016). This wealth of previously-reported targeted proteomics data points to a clear association of *Giardia* ESCRT components to PVs, further strengthening these organelles' status as functionally reduced and non-motile endo-lysosomal compartments. A clear association between ESCRT components and the ER also emerged from our investigations and is in line with reports for ESCRTIII participation in budding vesicles from the ER and CHMP7 deposition at the perinuclear envelope in model animal/fungal systems (Mast et al., 2018; Olmos et al., 2016). In support of our microscopy data, single co-IP experiments with epitope-tagged Vps36, Vps25, and CHMP7 identified the transmembrane ER marker protein disulfide isomerase-2 and other PDIs as interacting proteins by mass-spectrometry (Online Appendix Table 4.5). Co-IP of epitope-tagged Vps20L did not identify any PDI proteins but did show data consistent with interaction with BiP, another ER marker. While these data should not be taken as the basis for quantitative interactomes due to the lack of individual replicates, the fact that all four proteins co-immunoprecipitated ER markers are consistent with our conclusion of ER localization of these ESCRT components. Of note, these datasets also inform on co-immunoprecipitation of several other investigated and annotated ESCRT proteins.

The most surprising association reported here concerns the clear signal-overlap at the CMC of epitope-tagged CHMP7 and the mitochondrial marker IscU, detected with confocal microscopy, from which we infer a role for this ESCRT component at mitosomes. Notably, this inference is corroborated by the presence of *GiCHMP7* and ESCRTIII components, *GVps4B*, *GVps46B*, and *GVps31* in the interactome of mitosome-localized *GiMOMTiP1*, a main interacting partner of *GiTom40* (Rout et al., 2016). Notably, the

single co-IP dataset derived from epitope-tagged *Gi*CHMP7 also reciprocally detected *Gi*Tom40, albeit at low levels, by mass-spectrometry (Online Appendix Table 4.5). This is in line with the above report and our imaging data where mitosome deposition for HA-*Gi*CHMP7 appears limited to the CMC (Rout et al., 2016). Further data from crude sub-fractionation immunoblotting experiments using extracts of HA-*Gi*CHMP7 transgenic cells and non-transgenic control cells also lend support to HA-*Gi*CHMP7 association to endomembranes, consistent with the inferred presence of CHMP7 at mitosomes which contribute to these fractions, albeit also at ER (Jedelský et al., 2011) (Appendix Figure 4.8). Recent reports point to novel links between ESCRTs, mitochondrial membranes, and mitophagy (Anding et al., 2018; Hammerling et al., 2017; Richardson et al., 2014; Zhen et al., 2020). Therefore, although ESCRTs have been associated to mitochondria, to our knowledge, this is the first report to show an association with mitochondria-related organelles, representing a novel facet of MRO biology that should be explored in *Giardia* and other MRO-possessing organisms.

Notably, the co-localization coefficients observed for the various ESCRT components told a consistent, if not entirely straightforward story. In all cases, we observed low coefficients for overall signal overlap and degree of overlap between the ESCRT component and discrete organellar markers, but a high overlap between the organellar markers and the component. However, the overall overlap quantification between ESCRT components, especially Vps25 and Vps20L or Vps36, was higher, indicating that their signals were consistent. Together this tells a story of ESCRT localization at multiple locations, beyond the PV to the ER and even the mitosome in the case of CHMP7.

4.5.3. A comprehensive appreciation of ESCRT evolution and distribution in *Giardia* intestinalis

The definition of *Giardia* ESCRTs subcellular localizations combined with rigorous phylogenetic analyses revealed selective loss of ESCRTI, which mirrors the streamlining and loss of canonical MVB morphology within Fornicata, notably in *Giardia*. In the *Giardia* lineage, we observe duplications in the ESCRTIIIA machinery with paralogs which may compensate for ESCRTI and -III losses while, in combination with remaining ESCRT components, still functioning at PVs (Figure 4.10). We further observe a profound modification in *Giardia*'s ESCRT pathway by ESCRTIII components such as the CHMP7 apparently not associating within the endocytic pathway as first proposed (Horii et al., 2006).

In comparison to ESCRT machinery in characterized model organisms, we observe the localization of ESCRTII together with previously analyzed ESCRTIIIA-Vps46 and Vps4 components in close proximity to PVs and ER, while ESCRTIII-CHMP7 and Vps20L seem to localize almost exclusively in regions overlapping with the ER, with additional unknown roles for ESCRTIII CHMP7 at mitosomes.

Giardia ESCRTIII's association to the ER and mitosomes presents a complex landscape of novel membrane remodelling sites while maintaining PVs as reduced and simplified MVB-like compartments mostly by the action of ESCRTII and ESCRTIIIA subunits. Our collective data sheds light on a potential mode of action for ESCRTII and ESCRTIIIA at the PV membranes. We speculate that these subunits likely associate to the PV outer membrane from a cytosolic pool and perform membrane deformation, as

characteristic of other eukaryotic ESCRT subunits. Contacts sites between ER and PVs have been previously documented and could additionally be mediated by ESCRTs, allowing protein recycling down the endocytic and secretory pathway (Zumthor et al., 2016).

4.6 Conclusions

We have traced the evolutionary trajectories of ESCRTs within the Fornicata, observing a slow streamlining of this machinery across the transition to parasitism, with losses predating, concurrent with, and post-dating. Several groups have recently reported on a broader set of ESCRT functions in the eukaryotic cell than previously understood. In *Giardia*, ESCRTs have been primarily reported at the PVs. Here, we have shown ESCRT association to other membrane locations such as the ER and mitosome surfaces, suggesting this machinery may act more extensively at multiple organelles in *Giardia* than expected. Future functional studies should build on this comprehensive report to better assess the full range of ESCRT functions

4.7 Afterword

The combined findings from this chapter vastly improved our previous understanding of the evolution and molecular locations of ESCRTs in *Giardia*; however, a few limitations still remain. In early 2020, several new *Giardia intestinalis* genomes became available, particularly those belonging to the dog assemblages C and D, which were investigated in Chapter 3 but not in this study (Kooyman et al., 2019). The transcriptome of fornicate-sister lineage, *Barthelona* sp. PAP020, was also published subsequently but also not investigated here (Yazaki et al., 2020). This is because, the bioinformatic data presented here were collected between 2017 until late 2019 prior to these releases, and therefore, sampling of the fornicate genomes and transcriptomes is discrepant compared to those analyzed in the previous chapter. However, these present as important sampling points that should be surveyed in any follow-up analyses. Although not explicitly discussed in this chapter, the long-read genome of *Carpediemonas membranifera* used for comparative genomics and phylogenetics in Chapters 2 and 3 (and provided by co-authors Andrew Roger and Dayana Salas-Leiva) was also used for these analyses to perform homology searching in order to cross-validate ESCRT gene presences and absences determined using the transcriptome. Overall, it was confirmed that the findings laid out in this study held true, and that no additional ESCRTI, II, III, or III-A subunits or paralogues were present in the genome apart from those identified in the publicly available transcriptome. We refrained from a formal discussion of these results as the *Carpediemonas* genome was not publicly released until quite recently (Salas-Leiva et al., 2021).

As already mentioned in the discussion, another limitation of this investigation was the lack of biological or experimental replicates of the co-immunoprecipitation datasets. Replicate analyses, or reverse co-IP investigations, would increase our confidence in the enriched hits as biologically relevant. Although numerous trafficking proteins, including ESCRT components, and several PV, ER, and mitosome-associated proteins were found as interactors of several of baits in these single co-IP proteomics datasets, we were unable to derive any broader conclusions regarding subcomplex-level assemblies or specific

molecular interaction partners due to the absence of these replicates. Additionally, gel-based LC/MS analyses, instead of the filtration-based co-IP and bead-shaved LC/MS approach taken in the next chapter, was opted here, which resulted in a pull-down of large amounts of probable non-specific peptides. Therefore, conclusions other than those pertaining to organellar membrane associations were not made with this data. In future investigations, ESCRT assembly on organellar membranes should also be investigated through super-resolution microscopy (*i.e.*, Stimulated Emission Depletion (STED) Microscopy or Stochastic Optical Reconstruction Microscopy (STORM)) in conjunction with cholera toxin co-labelling of the PV-membranes.

Additionally, several *Giardia* ESCRT proteins described here, and previously, were not functionally investigated due to time and resource constraints. However, future studies should consider follow-up analyses of similar scope with these remaining components, namely *Giardia* ESCRTII-VP22, ESCRTIII-Vps24L, and ESCRTIII-A IST1 and -Vps31. This collective outlook may unveil interesting molecular associations or patterns of protein localization, as was the case with components examined in this analysis. These surveys would also help determine if these individual components have diversified roles within the parasite, if the individual subcomplexes function together but separate from the rest of the machinery, or if the canonical mechanism of sequential subcomplex nucleation for assembly of the entire *Giardia* ESCRT machinery is occurring but at various organelles in an unsuspected manner.

CHAPTER 5

The giardial ARF regulatory system localizes to the peripheral vacuoles and associates with vesicle formation and fusion machinery

5.1 Introduction

The aforementioned vesicle adaptors are conduits of endosomal trafficking between the ER and the Golgi, as well as various organelles that comprise endo-lysosomal pathways. Their timing and nucleation onto endo-membranes are tightly modulated by proteins belonging to the ARF regulatory system that consists of an ARF GTPase (ARF1 or ARF6), an ARF GAP (SMAP, ARFGAP1, ARGAP2/3, ACAP, AGFG, ARFGAP_C2, or ADAP), and a Sec7 domain-containing ARF GEF (BIG, GBF1, or Cytohesin) (Li et al., 2004; Pipaliya et al., 2019; Schlacht et al., 2013; Vargová et al., 2021). The work laid out in Chapters 2 and 3 unveiled the molecular complements of the coat proteins and the ARF regulatory system and pinpointed their evolutionary timepoints of streamlining across fornicates and *Giardia*. One of the principal conclusions conveyed through those chapters was that although many of the vesicle coats and their regulators are dispensable in *Giardia*, conservation in parts of those machinery alludes to some roles within its atypical endomembrane landscape. Cellular functions of some of these giardial proteins have been comprehensively investigated while others not.

Whether dynamic vesicle carriers apart from ESVs exist in this parasite, clathrin-coated or otherwise, remains largely unknown. As discussed in Chapter 2, some coat proteins have been subject to molecular functional investigations to probe their cellular locations and proteins-interaction partners. The collective findings from those studies mainly point to their roles at two *Giardia*-specific compartments, the peripheral vacuoles (PVs) and the encystation-specific vesicles (ESVs). To briefly recap, immunofluorescence and yeast two-hybrid-based (Y2H) co-immunoprecipitation with *Giardia* AP-1 subunits implicated *GiAP-1* to be crucial for the translocation of encystation-specific lysosomal cysteine protease (ESCP) and soluble acid phosphatases to the PVs (Touz et al., 2004). The transport mechanism in the absence of clathrin-coated vesicles is still unclear; however, disruption of AP1 μ subunits resulted in the dispersal of ESCP and AcPH into the ER instead of their final destination (*i.e.*, PVs) (Touz et al., 2004). AP-2, which typically mediates clathrin-mediated endocytosis (CME) at the plasma membrane (PM), localizes to areas corresponding to the giardial plasma membrane and the peripheral vacuole interfaces (Rivero et al., 2010). More recent investigations using a combination of super-resolution fluorescence microscopy, FIB-SEM tomography, and liquid-chromatography mass-spectrometry (LC/MS) based proteomics elucidated clathrin heavy chain to assemble in foci and as steady scaffolds in association with *GiAP-2*, as well as other endocytic proteins (*e.g.*, dynamin and PX domain-containing proteins), at the parasite PV-plasma membrane interfaces (Cernikova et al., 2020; Zumthor et al., 2016). Rather than performing canonical CME, *GiAP-2* and clathrin heavy chain mediate membrane fusion dynamics between the PV-PM,

likely for bulk flow uptake of material (Zumthor et al., 2016). Immunofluorescence microscopy investigation with *Giardia* retromer cargo-selection proteins, Vps26, Vps29, and Vps35, show co-localization with *Giardia* AP-2 subunits in the PV regions and some distribution at the ER (Miras et al., 2013). Subcellular fractionation experiments and Y2H-pulldown using retromer proteins as affinity handles confirm their membrane association and interactions with variant surface proteins and soluble hydrolases for trafficking to the PVs (Miras et al., 2013). Although the canonical role of clathrin in the formation of triskelion cages has diverged in this parasite, AP-1, AP-2, retromer, and clathrin heavy chain at the PVs supports the paradigm for these organelles as steady-state versions of endosomes and lysosomes.

Early secretion in *Giardia* has also diverged, and the corresponding machinery has been rerouted to produce encystation-specific vesicles (ESVs), involving both *Gi*COPI and *Gi*COPII (Marti et al., 2003; Stefanic et al., 2009). As detailed in Chapters 1 and 2, ESV formation is COPII-dependent which recognizes cyst-wall protein 1 (CWP1) followed by its assembly and nucleation on ER-exit sites for membrane budding to sort CWP1-3 into nascent ESVs (Faso et al., 2013; Marti et al., 2003; Stefanic et al., 2009). ESV maturation occurs upon association of the COPI coat as well ARF1 and Rab1 small GTPases (Marti et al., 2003; Stefanic et al., 2009).

ARF GTPases and their regulators modulate the precise dynamics of the above-mentioned vesicle budding and scission processes by recognizing and binding membrane phospholipids, coat proteins, and effector molecules. Apart from identifying losses in the fornicate ARF regulatory system proteins, comparative genomics and phylogenetics investigations in Chapter 3 classified numerous fornicate-specific ARF1 paralogues (*i.e.*, ARF1FA and ARF1FB), aside from canonical ARF1. This survey also identified consistent conservation in three ARF GAPs (*i.e.*, SMAP, ARFGAP1, and AGFG) and two ARF GEFs (*i.e.*, BIG and Cytohesin) across all fornicates. Although the lab strain *Giardia intestinalis* AWB lacks BIG orthologs, it does possess two paralogues of Cytohesin (*Gi*CYTHa and *Gi*CYTHb). The conclusions derived from the previous analysis were that numerous proteins of the ARF regulatory system are present in *Giardia*; however, functional investigations have only delineated the role of canonical *Gi*ARF1 and primarily in the context of encysting cells. As detailed in the previous chapters, *Gi*ARF1 is critical for the neogenesis of ESVs whereby deletion mutants in its cargo encoding gene, *cwp1*, and dominant-negative mutant of ARF1 (Q71L) hampered ESV formation and cyst-wall material trafficking (Ebnetter et al., 2016; Stefanic et al., 2009). Although one of these studies briefly presented an association of *Gi*ARF1 at the PVs of non-encysting trophozoites, no follow-up investigations have been performed since then to assess the specific molecular associations with this small GTPase (Stefanic et al., 2009). Additionally, localization and protein-protein interactions of the other fornicate-specific ARF paralogues and of all *Giardia* ARF GEFs or GAPs remain unelucidated, especially in the context of ESV-lacking trophozoite's endomembrane landscape.

The scope of this chapter is to investigate the localization and protein-protein interactions of *Gi*ARF1, *Gi*ARF1FA, *Gi*ARF1FB1, *Gi*ARFGAP1, and *Gi*CYTHa. ARFGAP1 is well-characterized in other organisms for its roles in hydrolyzing ARF1-bound GTP to regulate COPI vesicle traffic within the ER-Golgi

pathway. In contrast, Cytohesins activate ARF1 and ARF6 for recruitment onto endosomes by facilitating the GDP to GTP exchange (for comprehensive reviews on ARF GAP and ARF GEF-mediated regulation, see Donaldson & Jackson, 2011 and Shiba & Randazzo, 2012). In the absence of a discernable Golgi compartment or endosomal carriers, it was necessary to assess the functions of these typically Golgi and endosome- associating ARF regulatory system proteins in the vegetative trophozoites. Additionally, it would also be interesting to understand whether any of the giardial ARF regulatory system proteins interact with the encoded vesicle coat proteins, as is the case in model systems. The goal of this chapter was to pursue this investigation, particularly to comprehend each component's specific organellar and protein-protein interactions. To do so, a combination of immunofluorescence microscopy, co-immunoprecipitation, and LC/MS-based proteomics was employed to probe episomally expressed epitope-tagged variants of the giardial ARF regulatory system proteins in the trophozoites.

Prior to beginning this molecular survey, the three *Giardia* ARF paralogues identified in the previous chapters were subject to *in silico* sequence comparisons to ensure they conserved the correct biochemical attributes that would confer ARF GTPase enzymatic and effector binding activities. Overall, it was determined that *GiARF1* and *GiARF1FA* bear a high degree of sequence similarity to one another, while *GiARF1FB1* is overall divergent in its amino acid composition in comparison to the other two paralogues (Figure 5.1). Nonetheless, all three still retain correct motifs that are characteristic of ARFs. These included a 17-residue amphipathic myristate switch at their N-terminus, specifically conserving a glycine in position two, and therefore predicted to be correctly N-myristoylated for membrane recruitment (Goldberg, 1999; Liu et al., 2009) (Figure 5.1). All Ras-family GTPases, to which ARFs belong, are characterized by the presence of highly conserved signature motifs that are used for guanine-nucleotide binding. These are termed G-boxes, and generally, five of them are present (G1-G5)(Colicelli, 2004; Pasqualato et al., 2002). The *Giardia* ARF paralogues retain four out of five G-boxes, and therefore, should be capable of binding GTP or GDP. These are G1, also known as P-loop (GxxxxGK(T/S)T), G2 (PT), G3 (WDVGGQ), and G4 (NKxD) (Figure 5.1). The P-loop contains either a threonine or serine residue in position 31 necessary for Mg²⁺ binding to increase GTP-binding affinity (Lee et al., 2009) (indicated with a blue arrow in Figure 5.1). On the other hand, G3 contains glutamine in position 71 and is the catalytic site for GTP hydrolysis (Mishra & Lambright, 2016) (indicated with a blue arrow in Figure 5.1). Both *GiARF1* and *GiARF1FA* contain a glutamine, whereas *GiARF1FB1* encodes a serine, and therefore, regulation of this GTPase may be different compared to *GiARF1* and *GiARF1FA* (Figure 5.1). Ras-GTPases also contain sites for downstream effector binding, termed switch I and switch II regions (and overlap with the G2 and G4 GTP binding sites), which are exposed when the GTP or GDP-bound ARF undergoes a conformational change (Pasqualato et al., 2002). *GiARF1FB1* is divergent in both its switch I/II regions (Figure 5.1). Although it has the requisite residues for GTPase binding activity, effector identity or existing binding capacity compared to ARF1 and ARF1FA could be variable (Figure 5.1). In any case, all three ARF GTPases are predicted to behave canonically and therefore were functionally investigated along with *GiARFGAP1* and *GiCYTHa*.

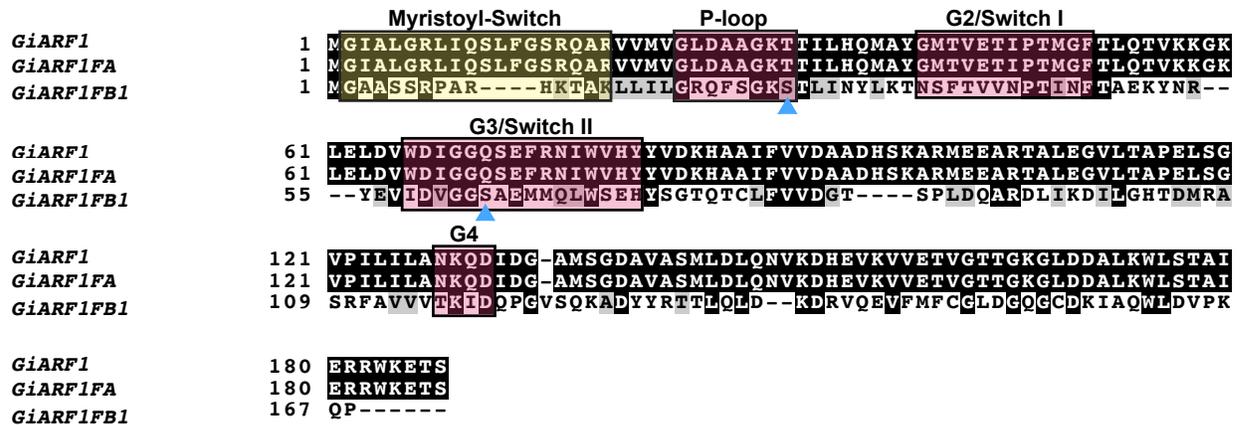


Figure 5.1. Sequence alignment of the three *Giardia intestinalis* AWB ARF1 paralogues for biochemical activity assessment. All three *Giardia* ARF1 sequences were examined to determine whether canonical motifs for myristoylation, Mg²⁺, GTP, GDP, and regulatory protein binding were present. All three ARFs contained the N-terminal myristate switch with a conserved glycine residue in position two necessary for membrane recruitment. The three paralogues were also characterized by the presence of four G-boxes (G1/P-loop, G2, G3, and G4) which coordinate GTP and GDP-binding. The P-loop and G3 also contain residues for Mg²⁺ and GAP binding, T/S31 and Q71, respectively (indicated with blue arrows). Overall, all three ARFs are predicted to retain canonical GTPase activity, but *GiARF1FB1* is most diverged in its sequence composition compared to *GiARF1* and *GiARF1FA*, which share significant similarities (indicated by black shading). Motif predictions were performed according to Pasqualato et al. (2002). Alignment was generated using MUSCLE v. 3.8.31, and the figure was produced using the BoxShade tool (Expasy, Swiss Institute for Bioinformatics).

5.2 Materials and Methods

5.2.1 *Giardia* trophozoite culture and transfection

Cell culture and transfection of *Giardia intestinalis* strain A, isolate WB clone C6 (ATCC 50803) trophozoites were performed as per standard protocol (Cernikova et al., 2020; Morf et al., 2010; Pipaliya, Santos, et al., 2021; Zumthor et al., 2016). Briefly, trophozoites were axenically cultured in Nunc™ polystyrene culture tubes (Thermo Fisher Scientific) containing ca. 11 mL of Diamond's TYI-S-33 medium supplemented with 10% Seraglob bovine serum (Bioswisstec AG), 0.52 mg/mL bovine bile (Sigma B-8381), 10,000 units/mL penicillin-streptomycin (Thermo Fisher), and 22.8 mg/mL ammonium ferric citrate (Fluka 09714; Sigma), and incubated at 37°C until confluent with subculturing performed every 2-4 days (Keister, 1983). 1×10^6 mid-log phase vegetative trophozoites were harvested by placing culture tubes on ice for 30 to 45 minutes, followed by gentle hitting and inversion to dislodge the adherent cells (Keister, 1983). Parasite transgenic lines were prepared by first collecting non-transgenic trophozoites cells by centrifugation at $900 \times g$ at 4°C for 10 minutes, followed by electroporation (350V, 960 μ F, 800 Ω) of 15 μ g of pPacV-Integ-based circular plasmid vectors (episomes) expressing reporter constructs that were prepared in electrocompetent *Escherichia coli* D10HB (Thermo Fisher Scientific). Transgenic parasites were selected and sub-cultured with the addition of 40 μ L Puromycin (50 μ g/ml; InvivoGen) in the culture medium. Parasites were harvested 1.5 to 2 weeks post-transfection and tested for reporter construct expression through immunofluorescence assays and widefield fluorescence microscopy using Leica DM5500 B microscope under 100X oil immersion lens (Leica Microsystems).

5.2.2 Genomic DNA isolation and cloning for episomal vector construct generation

Genomic DNA was isolated from non-transgenic trophozoites by first harvesting confluent trophozoite cultures, as detailed above. Cells were collected as pellets and resuspended in a buffer solution containing 50 nM Tris (pH 8), 10 mM Ethylenediaminetetraacetic acid (EDTA), and 1.5 μ L of 100 mg/mL RNase, and vortexed until a homogenous slurry was observed. 200 mM NaOH and 1% SDS were added to this solution and incubated at room temperature for two minutes, then resuspended in 3M potassium acetate (pH 5.5) followed by centrifugation at $14000 \times g$ for 15 minutes at 4°C. gDNA was precipitated out of solution by adding 900 μ L of 70% cold ethanol, after which supernatant was discarded, and the resulting DNA pellet was dissolved in 20 μ L of ddH₂O. Isolated gDNA yield and purity were measured using a Nanodrop spectrophotometer (Thermo Fisher). Ten nanograms of the isolated gDNA were used for polymerase chain reaction (PCR) amplification of the open reading frames belonging to *Gi*ARF1 (GL50803_7789), *Gi*ARF1FA (GL50803_13930), *Gi*ARF1FB1 (GL50803_75620), *Gi*CYTHa (GL50803_17192), and *Gi*ARFGAP1 (GL50803_2834), along with their native promoters and a C-terminal hemagglutinin tag (HA). As discussed earlier, the C-terminal tagging strategy was employed as the ARF N-terminus is necessary for N-myristoylation and membrane recruitment. PCR amplification was performed using designated oligonucleotide pairs as listed in the Online Appendix Table 5.1, synthesized by

Microsynth (Balgach, Switzerland). Due to the large predicted open reading frame of *GiCYTHa* (GL50803_17192), a cloning strategy using the Gibson Cloning Kit Protocol (Addgene) was employed. For all ORFs, promoter sequences were derived from 150 to 200 base pairs upstream of the predicted start codon. The resulting inserts were cloned into a pPacV-Integ modified vector containing XbaI and PacI restriction enzyme recognition sites, the giardial glutamate dehydrogenase promoter, and a puromycin resistance cassette for constitutive episomal expression of each *Giardia* ARF regulatory system protein and antibiotic selection of transgenic cells (Zumthor et al., 2016). Detailed vector maps for all constructs are provided as Online Appendix Figures 5.1-5.5. Cloning experiments detailed in this section were performed by Dr. Corina Wirdnam and Dr. Carmen Faso at the Institute for Cell Biology (University of Bern) and the Institute for Parasitology (University of Zurich), respectively.

5.2.3 Indirect immunofluorescence and fluid-phase uptake assays

Transgenic lines and non-transgenic control cells were cultured in 1 x Nunc™ T-25 polystyrene flasks (Thermo Fisher) per line until parasites reached confluency. Cells were harvested by cooling and detaching on ice for 30 minutes, followed by centrifugation at 900 x *g* for 10 minutes. The resulting cell pellets were washed with cold PBS and fixed with 3% formaldehyde (Sigma) in phosphate-buffered saline (PBS) for 1 hour, followed by quenching using 0.1M Glycine in PBS for 5 minutes. Fixation and quenching were both performed at room temperature. Cells were permeabilized using 2% bovine serum albumin (BSA)/0.2% Triton X-100 in PBS for 20 minutes at room temperature and blocked in 2% BSA in PBS for two hours. Antibody incubations were performed using a primary rat-derived monoclonal anti-HA high-affinity antibody (dilution 1:250; Roche) and a secondary goat-derived anti-Rat IgG (H+L) conjugated to Alexa Fluor 488 (AF488) (dilution 1:250; Thermo Fisher). Both antibody solutions were prepared in 2% BSA/0.2% Triton X-100 in PBS, and incubations for each was performed for one and a half hour at room temperature on a rotating shaker in the dark. Between each antibody incubation, washes were performed using 1% BSA/0.1% Triton X-100 in PBS, after which samples were fixed in 10 to 40 μ L Vectashield (Reactolab) containing 4'-6-diamidino-2-phenylindole (DAPI) for nuclear staining and cell suspension. Three microlitres of the cell suspension were aliquoted onto microscopy glass slides and covered with 22 mm x 22 mm coverslips which were sealed with nail varnish.

Peripheral vacuolar staining was performed using the fluid-phase marker Dextran (10,000 MW) conjugated to Texas Red (Cat. No. D1863, Thermo Fisher), as previously described (Gaechter et al., 2008). Briefly, confluent T-25 flasks of transgenic and non-transgenic trophozoites were cultured and harvested as described above. Cell pellets were resuspended in ca. 100 μ L of freshly supplemented TYI-S-33 medium to which Dextran-Texas Red (Dextran-TxR) was added to a final concentration of 2 mg/mL. Incubations were performed at 37°C for 30 minutes to allow for the uptake of extracellular Dextran-TxR into the PV lumen, which was then halted by placing the cells on ice for 15 minutes and washing with cold PBS. Cells were chemically fixed with 3% formaldehyde in PBS followed by permeabilization and immunoprobings with primary and secondary antibodies, as per standard immunofluorescence assay protocol described above.

5.2.4 Laser Scanning Confocal Microscopy and image analyses

Visualization of recombinant proteins probed with primary and secondary antibodies was performed using a Leica SP8 x STED super-resolution microscope configured with white light lasers (excitation wavelength between 470 nm to 670 nm), photomultiplier tubes, and HyD detectors, under HC PL APO 100X/1.44 oil immersion objective lens (Leica Microsystems). Appropriate excitation and emission settings were used for the gating and visualization of green (Alexa Fluor 488), red (Texas Red), and blue (DAPI) channels. Brightfield images were acquired using the transmission/differential interference contrast settings (PMT Trans). Laser pinhole opening size was set to Airy 1, and images were acquired with line averaging set to 4 to optimize the signal-to-noise ratio. More than 100 cells per sample were imaged at maximum width in view of both the nuclei and the PV bare zone. Autofluorescence background in the anti-HA AF488 channel was subtracted from reporter construct lines by thresholding the laser settings against control non-transgenic cells that were also subject to immunoprobings with primary and secondary antibodies. Laser settings were kept consistent between all samples.

Single trophozoite images were subject to deconvolution using Huygens Professional (Scientific Volume Imaging) and further analyzed in FIJI/ImageJ (Rueden et al., 2017; Schindelin et al., 2012). Statistical quantification was performed using the coloc2 plugin to determine signal overlap between channels corresponding to Dextran-TxR and Anti-HA AF488. One hundred Costes' iterations and a point-spread function set to three were used to calculate Pearson's, Manders' 1 and 2 coefficients, and Costes' p-values on the whole image or specific regions of interest (Costes et al., 2004; Li et al., 2004; Manders et al., 1993). Two regions of interest were chosen, the bare zone and the peripheral PVs. Co-localization analyses were repeated on \geq ten cells to determine mean and median statistics. Channels corresponding to anti-HA-AF488, Dextran-Texas Red, and DAPI were pseudo-colored in FIJI/ImageJ. Raw signal overlap statistical data for individual cells belonging to each line, as well as expression quantification summaries are provided as Online Appendix Table 5.2. Microscopy was performed using equipment provided by the Microscopy Imaging Centre (University of Bern). Dr. Yury Belyaev at the Theodor Kocher Institute (University of Bern) provided guidance with image deconvolution analyses.

5.2.5 Limited cross-linked co-immunoprecipitation using anti-HA agarose beads

Non-transgenic and transgenic trophozoites expressing reporter constructs were used as baits for crosslinking and protein pulldown via anti-HA agarose beads (Cernikova et al., 2020; Zumthor et al., 2016). Each parasite line was cultured in 1 x T-25 flasks to confluency, followed by harvesting and washing with 20 mL PBS yielding a final volume of ca. 2×10^9 cells. A 500 μ L aliquot was set aside for western blot analyses to use as control cell lysate samples. The remainder were fixed in 2.25% formaldehyde in PBS and incubated for 30 minutes at room temperature while shaking. Cells were quenched in 10 mL 0.1M Glycine in PBS for 15 minutes with gentle shaking and resuspended in 5 mL RIPA-SDS lysis buffer solution with 50 mM Tris (pH 7.4), 150 mM NaCl, 1% IGEPAL, 0.5% sodium deoxycholate, 0.1% SDS, and 10 mM EDTA. 100 μ L 0.1M phenylmethylsulphonyl fluoride (PMSF; Sigma) and 50 μ L protease inhibitor cocktail

(Sigma) was also added, and the resulting suspension was sonicated twice using the following settings: 60 pulses, two output control, 30% duty cycle, and 60 pulses, four output control, 40% duty cycle, respectively, with 10-15 seconds break between each cycle. Sonicated samples were incubated at 4°C on a rotating shaker for 2 hours and centrifuged at 14,000 x g for 10 minutes at 4°C. Supernatants from each sample were syringe filtered through 0.2 µm Acrodisc MS filters (Pall MS-3301) and resuspended in 5 mL RIPA-Triton X-100 solution containing 50mM Tris (pH 7.4), 150mM NaCl, 1% IGEPAL, 0.5% sodium deoxycholate, 1% Triton X-100, and 10 mM EDTA (Sigma). Forty microlitres of the anti-HA agarose beads slurry (Thermo Fisher) was added to the soluble protein fraction and incubated overnight on a rotating shaker at 4°C. Subsequently, beads were washed thrice with 10 mL Tris-Buffered Saline (TBS) with 0.1% Triton X-100. One hundred microlitre aliquot of the beads was saved for immunoblotting, while the remainder of the sample was subject to mass spectrometry analysis as detailed in section 5.2.7. For each transgenic line, co-immunoprecipitation and mass-spectrometry experiments were performed in independent biological replicates of two.

5.2.6 Western Blotting and Coomassie Staining

Proteins from whole cell lysate and bead samples from both transgenic and non-transgenic control were analyzed by SDS-PAGE on a 4%-10% polyacrylamide gel under reducing conditions, as per standard protocol (Smith, 1994). Cross-linking was reversed by resuspending bead and whole cell lysate samples in Laemmli loading buffer and 100 mM Dithiothreitol (DTT) (Sigma), then boiling for 5 minutes. Samples were loaded with PageRuler™ Prestained Protein Ladder (10 to 180 kDa; Cat. No. 26616, Thermo Fisher) and run at 160V for 1 hour. Gels were transferred onto nitrocellulose membranes overnight at 4°C and blocked with 5% dry milk/0.05% TWEEN-20 the following day. Immunoprobings were performed using a rat-derived monoclonal anti-HA antibody (dilution 1:1000; Roche) followed by a secondary anti-rat antibody coupled to horseradish peroxidase (HRP) (dilution 1:5000; Southern Biotech). HRP detection was performed by adding Western chemiluminescent substrate (Thermo Fisher) onto the membrane and visualized using an Amersham Imager 600 under the chemiluminescent settings. Polyacrylamide gels were stained with InstantBlue Coomassie Protein Stain (Sigma) for 30 minutes and de-stained with ddH₂O while shaking.

5.2.7 Liquid Chromatography Mass Spectrometry (LC/MS) for quantitative proteomics

Affinity pulldown beads were suspended in 8M Urea / 50 mM Tris-HCl (pH 8). Proteins were reduced at 37°C with DTT 0.1M / 100 mM Tris-HCl (pH 8) and alkylated at 37°C in the dark with IAA 0.5M / 100 mM Tris-HCl (pH 8) for 30 minutes each. Thereafter, the slurry was diluted with four volumes of 20 mM Tris-HCl (pH 8) / 2 mM CaCl₂ prior to overnight digestion at room temperature with 100 ng sequencing grade trypsin (Promega). Samples were centrifuged in order to extract the peptides in the supernatant. The digests were analyzed by liquid chromatography (LC)-MS/MS using PROXEON coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific) with three injections of five microlitre digests. Peptides were trapped on a µPrecolumn C18 PepMap100 (5 µm, 100 Å, 300 µm× 5 mm, Thermo Fisher Scientific,

Reinach, Switzerland) and separated by backflush on a C18 column (5 µm, 100 Å, 75 µm×15 cm, C18) by applying a 40 minute gradient of 5% acetonitrile to 40% in water, 0.1% formic acid, at a flow rate of 350 nl/min. The Full Scan method was set with resolution at 70,000, an automatic gain control (AGC) target of 1x10⁶, and a maximum ion injection time of 50 ms. The data-dependent method for precursor ion fragmentation was applied with the following settings: resolution 17,500, AGC of 1x10⁵ maximum ion time of 110 milliseconds, mass isolation window 2 *m/z*, collision energy 27, under fill ratio 1%, charge exclusion of unassigned and 1+ ions, and peptide match preferred, respectively.

MS data were interpreted with MaxQuant (v. 1.6.14.0) against GiardiaDB (v. 47) using the default MaxQuant settings, with an allowed mass deviation for precursor ions of 15 ppm in the first search round, the mass deviation for fragments of 20 ppm, a maximum peptide mass of 5500Da, the match between runs activated with a matching time window of 0.7 min, and the use of non-consecutive fractions for the different pulldowns to prevent over-fitting. Other database search settings were a strict trypsin cleavage rule allowing for 3 missed cleavages, fixed carbamidomethylation of cysteines, variable oxidation of methionines, and acetylation of protein N-termini. Resulting MS hits, including intensity-based absolute quantification (iBAQ) and label-free quantitation (LFQ) values, are provided in Online Appendix Table 5.3. Mass-spectrometry experiments detailed in this section were performed by Dr. Sophie Braga and Dr. Manfred Heller at the Proteomics Mass Spectrometry Core Facility (Department for Biomedical Research, University of Bern).

5.2.8 *in silico* evaluation of LC/MS hits

Prior to analyses, contaminants and highly abundant proteins enriched in the non-transgenic control were filtered out from the MS data corresponding to the transgenic lines. Each replicate from a given line was treated independently for initial analyses. MS hits were sorted according to highest to lowest protein abundance as per the intensity-based absolute quantification (iBAQ) values. iBAQ values measured the sum of peak intensities divided by theoretically observed peptides to quantify the relative molar amount of a given protein in the sample (Krey et al., 2014). These were normalized by calculating the relative percent abundance for each protein hit as per the following equation:

$$riBAQ = \frac{iBAQ}{\Sigma iBAQ} * 100$$

Protein hits from replicate datasets belonging to a given line were intersected with one another, wherein only hits enriched in both replicates were kept for investigation. A high stringency and a low stringency threshold of 1% and 0.01% relative (r)iBAQ was applied. All hits ≥ 0.01% threshold were considered significant molar interactions and deemed biologically relevant in the context of the bait. Within these, all identified hypothetical proteins were subject to further analyses using HHPRED analyses (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) and screened for in all previously published *Giardia* co-IP datasets for their functional annotation. Sorted absolute and relative iBAQ hits meeting stringency thresholds are provided in the Online Appendix Table 5.4.

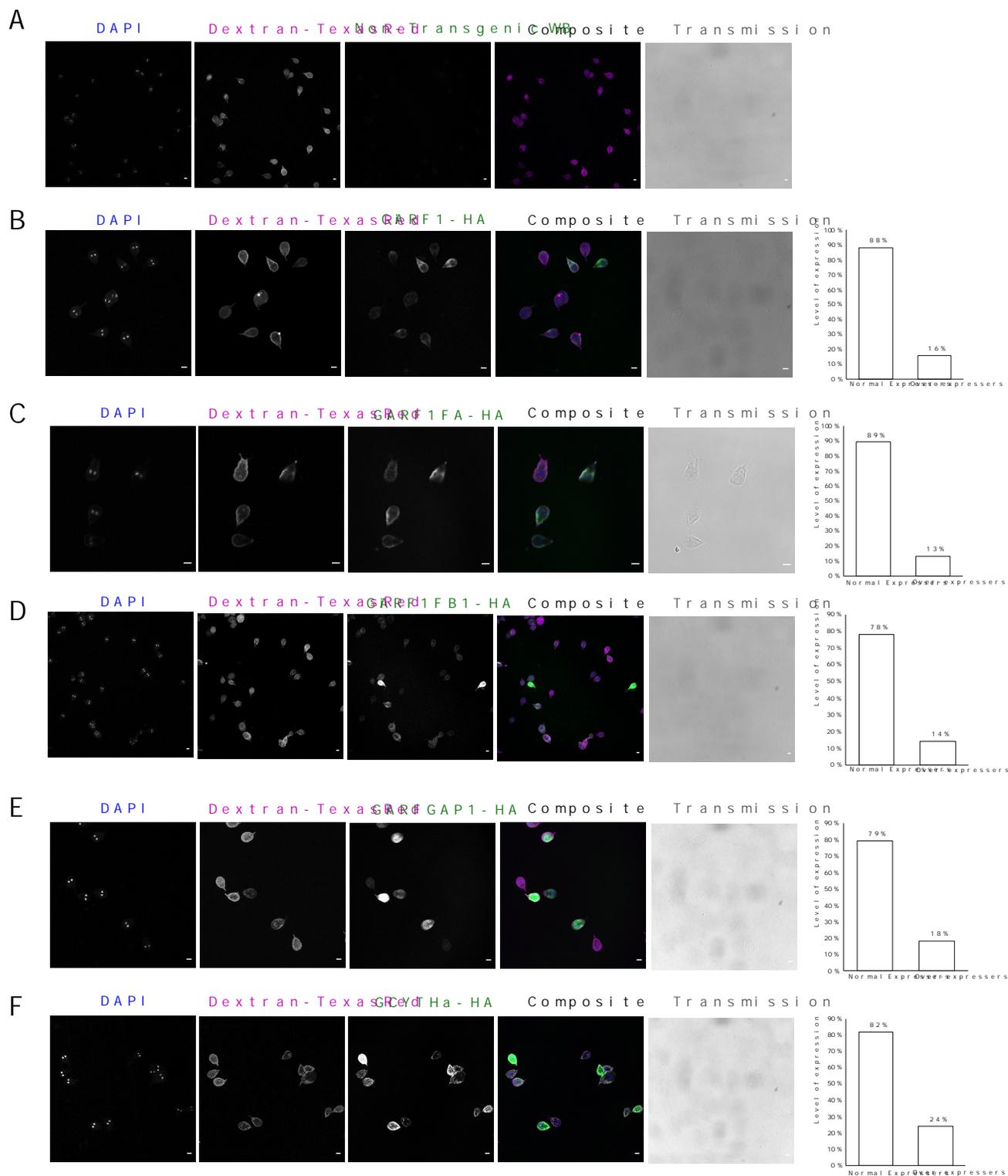
5.3 Results

5.3.1 Canonical ARF1 localizes to the cortical PV populations in vegetative trophozoites

Informatic and phylogenetic investigations determined a complete complement of the ARF regulatory system proteins in *Giardia* and the rest of the Fornicata, as demonstrated in Chapter 3 (Pipaliya et al., 2021). A canonical ARF1 ortholog, which is distributed across the diversity of eukaryotes, is conserved across this phylum and in *Giardia intestinalis*, and so the molecular localization of this protein in *Giardia* trophozoite's cellular landscape was assessed (Vargová et al., 2021). *Gi*ARF1 function has been primarily investigated in the context of encystation, where it is critical in the neogenesis of ESVs for directional transport of cyst-wall material to the cell surface (Ebnetter et al., 2016; Stefanic et al., 2009). *Gi*ARF1 association, among other typically Golgi-localizing small GTPases, to ESVs has prompted *Giardia* biologists to draw parallels with Golgi-like functions. However, general ARF1 roles extend beyond Golgi trafficking and are critical for endo- and exocytic processes that ensue to and from the *trans*-Golgi network (see Donaldson & Jackson, 2011 for a review on this topic). In the absence of these stage-specific ESVs, the role of *Gi*ARF1 was investigated in the vegetative trophozoites.

Transgenic parasites expressing *Gi*ARF1 tagged with C-terminal HA epitope (*Gi*ARF1-HA) were subject to immunofluorescence microscopy. The level of reporter construct expression was assessed in comparison to the non-transgenic control (Figure 5.2A). Approximately 154 cells were counted, where 88% of the population expressed the *Gi*ARF1-HA variant of the protein (Figure 5.2B and G). A sub-population (16%) of these presented as extremely bright cells and were considered to have an overexpression phenotype (Figure 5.2B and G). In order to investigate specific organellar locations of *Gi*ARF1-HA, closer microscopic investigations with isolated trophozoites were performed. Here it was observed that the proteins accumulated in regions corresponding to PV populations in the cortical region of the cell. PV visualization was performed by fluorescently labeling the organellar lumen using Dextran-TxR, a fluid phase marker that indiscriminately stains all PV populations, including those in the bare zone. This was done to inspect signal overlap between the protein and the organellar marker to determine whether *Gi*ARF1-HA is distributed in the regions corresponding to these compartments. Visual inspection of *Gi*ARF1-HA and Dextran-TxR showed considerable signal overlap between green and magenta fluorescence, however, only in regions bearing cortical PVs but not the bare zone (Figure 5.3A and D). Closer regions of interest investigation with the two PV populations confirmed minimal, even complete absence, of *Gi*ARF1-HA at the bare zone (Figure 5.3C).

Figure 5.2. Population-level expression analysis of epitope-tagged ARF regulatory system proteins. (A) Non-transgenic lines (WB) were used as controls and for the subtraction of background anti-HA AF488 signal. (B) *Gi*ARF1-HA was expressed in 88% of the screened cells, of which 16% had an overexpression phenotype. (C) *Gi*ARF1FA-HA was expressed in 89% of the screened cells, of which 13% had an overexpression phenotype. (D) *Gi*ARF1FB1-HA was expressed in 79% of the screened cells, of which 14% had an overexpression phenotype. (E) *Gi*ARFGAP1-HA was expressed in 79% of the screened cells, of which 18% had an overexpression phenotype. (F) *Gi*CYTHa-HA was expressed in 82% of the screened cells, of which 24% had an overexpression phenotype. (G) provides a detailed breakdown of the number of cells that were used for quantification. All scale bars: 5 μ m.



Cell Line	Endogenous/induced expression	Total # of Cells Counted	# of expressed cells	# of overexpressed cells	% expression	% overexpressers
Gi-pE - RFP1-HA	endogenous	154	136	25	88%	16%
Gi-pE - RFP1FA-HA	endogenous	189	162	24	89%	13%
Gi-pE - RFP1FB1-HA	endogenous	471	370	68	79%	14%
Gi-pE - RFP1GAP1-HA	endogenous	461	365	84	79%	18%
Gi-pE CYTH aHA	endogenous	380	313	93	82%	24%

Quantitative assessment of protein localization was performed through signal overlap analyses with at least ten cells to determine whether observed trends were uniformly present across multiple sampling points. Statistical approaches described in Chapter 4 for the investigation of *Gi*ESCRTs were also used here. Briefly, co-localization parameters were calculated using Coloc2 to compute Pearson's coefficients (R), Manders' coefficient (M1 and M2), and Costes' p-value to determine the degree of the signal overlap between *Gi*ARF1-HA and Dextran-TxR across the whole-cell image, bare zone (region of interest 1; ROI1), and cortical PV populations (region of interest; ROI2). Pearson's coefficient values (R) in whole-image analyses ranged between 0.56 to 0.83 (average of 0.75), which suggested a high positive correlation between the two channels (Online Appendix Table 5.2; Figure 5.3A, B and E). This was further corroborated by high M1 and M2 coefficient values, another metric of co-localization that measures the degree of pixel overlap from channel A to channel B and vice versa (Dunn et al., 2011). Average M1 and M2 values were 0.99 and 0.987, respectively (Online Appendix Table 5.2; Figure 5.3B and E). A high degree of confidence in all three coefficients was confirmed by an average Costes' p-value of 1, suggesting that the computed R, M1, and M2 coefficient values are not randomized. A similar positive correlative trend was observed with ROI2 analyses of the cortical PVs with an average R of 0.46, M1 and M2 of 0.99 and 1.00, respectively, and Costes' p-value of 1 (Online Appendix Table 5.2; Figure 5.3D and E). On the other hand, statistical analyses confirmed a lack of co-localization between *Gi*ARF1-HA and Dextran-TxR in the bare zone region (ROI1), as negative or extremely low R values were observed across all sampled trophozoites. Although M1 and M2 coefficients are high, the Costes' p-value is well below the required 0.95 threshold of confidence, suggesting that signal overlap between these two channels is unlikely (Online Appendix Table 5.2; Figure 5.3C and E).

Results from this investigation suggest that within the *Giardia* cell, *Gi*ARF1-HA localizes in regions corresponding to the PV compartments. Although protein distribution is mainly along the cortical PVs, there is a minimal association to the bare zone, hinting at a probable scenario of localization specificity towards different PV populations within the cell.

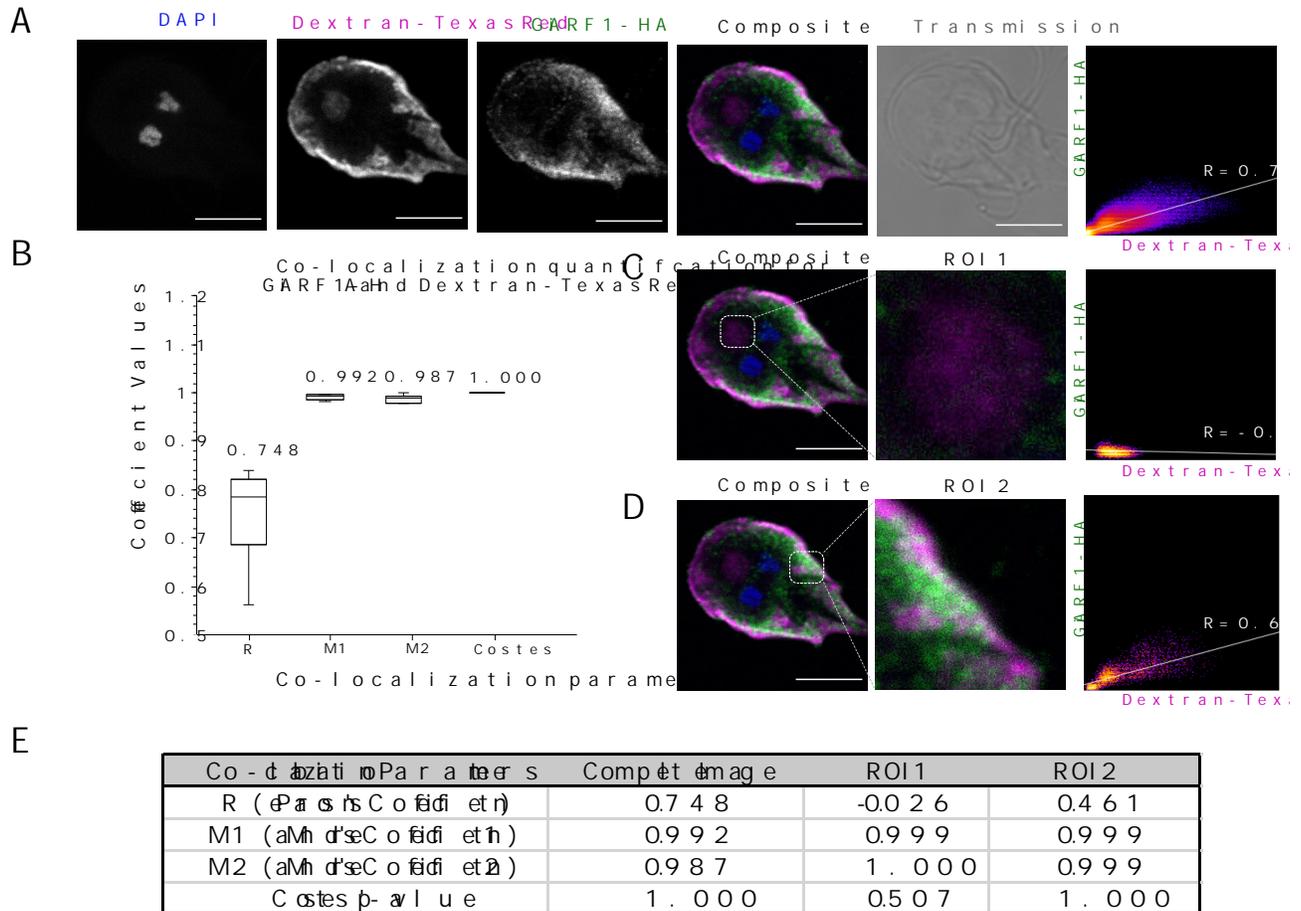


Figure 5.3. Characterization of *GiARF1-HA* subcellular location. (A) depicts results from co-labelling immunoprobings experiments with transgenic *Giardia* WB(C6) trophozoites labelled for epitope-tagged *GiARF1-HA* (green) and Dextran-TxR (magenta). (B) Signal overlap analyses were performed on whole-cell images to determine co-localization between *GiARF1-HA* and Dextran-TxR. Boxplot depiction of the calculated R, M1, M2, and Costes' p-values across ≥ 10 analyzed cells, along with their mean values indicated on top, is also provided. (C) Signal overlap analyses were performed with the bare zone (ROI1) to determine co-localization between *GiARF1-HA* and Dextran-TxR. (D) Signal overlap analyses were performed at the cortical PVs (ROI2) to determine co-localization between *GiARF1-HA* and Dextran-TxR. (E) Summary of average co-localization parameters that were calculated for whole-cell image, ROI1, and ROI2 using ≥ 10 cells. Scale bars: 5 μm . Optical slices were acquired from the middle of the cell for a maximum projection view of the nuclei and the bare zone. All images were acquired using the Leica SP8 x STED laser scanning confocal microscope under the 100x oil immersion objective lens.

5.3.2 ARF1FA and ARF1FB1 localize to the peripheral vacuoles in a pattern similar but not identical to ARF1

Apart from *Gi*ARF1, all assemblages of *Giardia intestinalis* were found to possess numerous additional paralogues of ARF1 that arose at the base of the fornicate lineage, as was previously demonstrated through phylogenetic and comparative genomic investigations detailed in Chapter 3 (Pipaliya et al., 2021). Although there were some differences in the number of individual paralogues, ARF1FA and ARF1FB1 were conserved across all fornicates and in *Giardia intestinalis* AWB. These are currently annotated as ARF3 and ARF2, respectively, on GiardiaDB but will be referred to as per the former nomenclature to reflect shared ancestry with the other fornicate ARF1F proteins. To understand how these proteins localize compared to canonical *Gi*ARF1, similar microscopy investigations as *Gi*ARF1 were performed by generating transgenic lines expressing C-terminally HA-tagged variants of these proteins.

Overall expression levels for *Gi*ARF1FA-HA and *Gi*ARF1FB1-HA were similar to *Gi*ARF1-HA, whereby 89% and 79% of the cells, respectively, had fluorescent signals (Online Appendix Table 5.2; Figure 5.2C, D, and G). Similarly, approximately 14% of both populations had an overexpression phenotype (Figure 5.2C, D, and G). Despite comparable expression trends, closer single-cell examinations showcased slightly different patterns of protein localizations compared to one another and *Gi*ARF1-HA.

Both *Gi*ARF1FA-HA and *Gi*ARF1FB1-HA were localized to the PV regions but in distinct patterns. *Gi*ARF1FA-HA signal was similar to *Gi*ARF1-HA in regions of the cortical PVs but also at the bare zone (Figure 5.4A). Statistical quantification of signal overlap in both regions of interest and the whole image was performed with Dextran-TxR co-labelling (Figure 5.4). Positive Pearson's correlation was observed in complete cell image analyses and ranged between 0.51 to 0.83 (average \approx 0.7; Figure 5.4A, B, and E). Significant overlap between signals in channels corresponding to *Gi*ARF1FA and Dextran-TxR were confirmed as high M1 and M2 coefficients (M1 average \approx 0.99 and M2 average \approx 0.93). All three observations were considered true positives, as the average Costes' p-value equaled to 1. Unlike *Gi*ARF1-HA, some visible bare zone localization (ROI1) was also observed, albeit not as prominently, as positive Pearson's coefficients were obtained across \geq 10 cells (Online Appendix Table 5.2; Figure 5.4C and E). On the other hand, a statistically significant association was observed between *Gi*ARF1FA-HA and Dextran-TxR at the cortical PVs (ROI2), as R fell between 0.4 to 0.77 (average \approx 0.44). Both M1 and M2 ranged between 0.99 and 1 (average \approx 0.99) with an average Costes' p-value of 0.99, also suggesting true signal overlap.

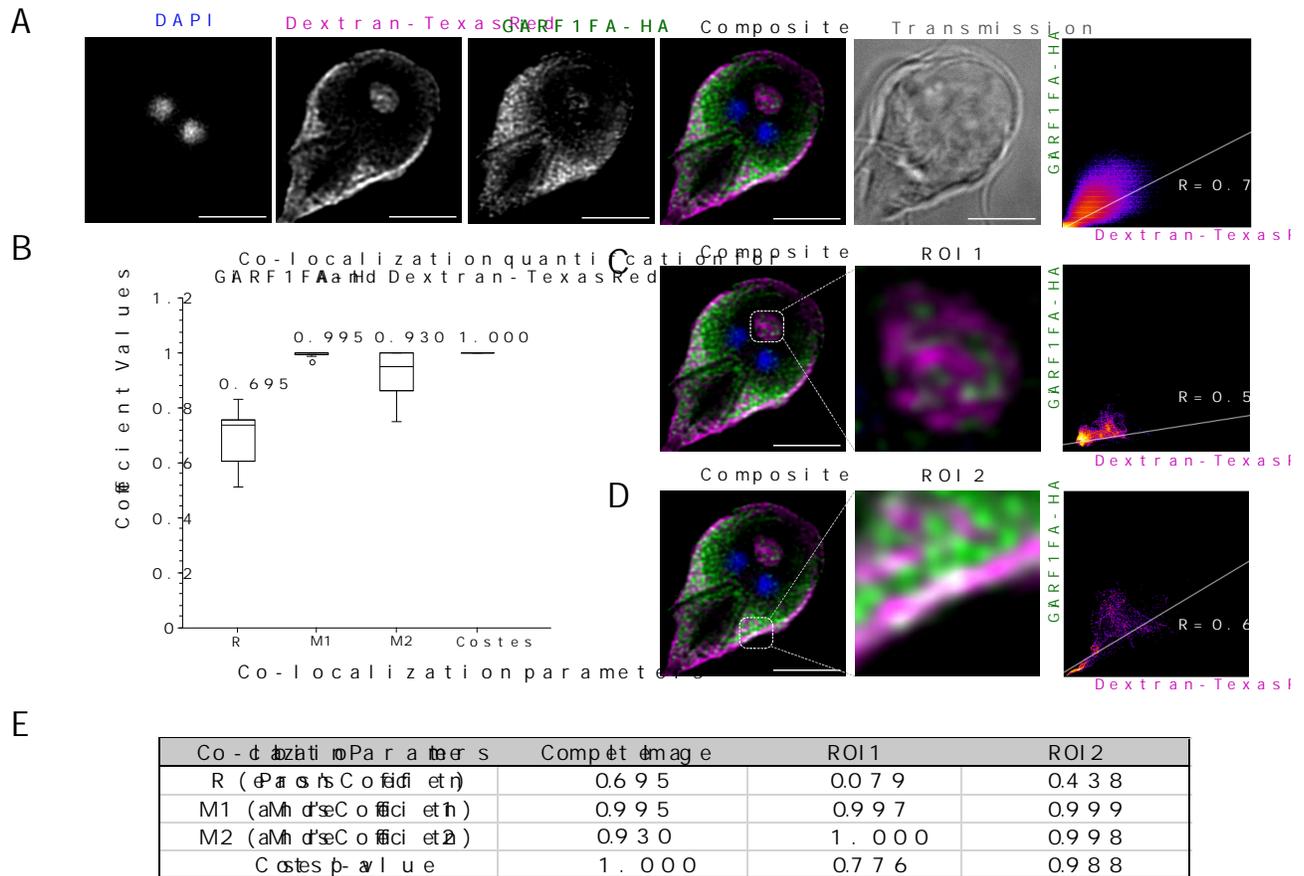


Figure 5.4. Characterization of *GiARF1FA-HA* subcellular location. (A) depicts results from co-labelling immunoprobings experiments with transgenic *Giardia* WB(C6) trophozoites labelled for epitope-tagged *GiARF1FA-HA* (green) and Dextran-TxR (magenta). (B) Signal overlap analyses were performed on complete images to determine co-localization between *GiARF1FA-HA* and Dextran-TxR. Boxplot depiction of the calculated R, M1, M2, and Costes' p-values across ≥ 10 analyzed cells, along with their mean values indicated on top, is provided. (C) Signal overlap analyses were performed at the bare zone (ROI1) to determine co-localization between *GiARF1FA-HA* and Dextran-TxR. (D) Signal overlap analyses were performed at the cortical PVs (ROI2) to determine co-localization between *GiARF1FA-HA* and Dextran-TxR. (E) Summary of average co-localization parameters calculated for the whole image, ROI1, and ROI2 using ≥ 10 cells. Scale bars: 5 μ m. Optical slices were acquired from the middle of the cell for a maximum projection view of the nuclei and the bare zone. All images were acquired using the Leica SP8 x STED laser scanning confocal microscope under the 100x oil immersion objective lens.

Although similar in expression, examination of *Gi*ARF1B1-HA revealed a pattern of localization that was comparatively different to both *Gi*ARF1-HA and *Gi*ARF1FA-HA, whereby a diffuse, punctate signal with some PV overlap was evident (Figure 5.5). Like the previous two proteins, co-localization was performed with Dextran-TxR to assess bare-zone (ROI1) and cortical PV region (ROI2) overlap, along with analyses using whole-cell images (Figure 5.5). Statistical inference from complete image confirms signal overlap between the recombinant ARF1FB1 and Dextran-TxR, as reflected by positive and overall high Pearson's coefficient values (average ≈ 0.75 ; Figure 5.5A, B, and E). Some bare zone-associated signal was also noted, as confirmed by positive R and high M1, M2, and Costes' p-values (average $R \approx 0.1$, $M1 \approx 1.00$, $M2 \approx 1.00$, and Costes' p-value ≈ 0.99 ; Online Appendix Table 5.2; Figure 5.5C and E). Finally, like ARF1FA, a considerable signal overlap was observed at the cortical PVs (Average $R \approx 0.46$, $M1 \approx 0.99$, $M2 \approx 0.99$ and Costes' p-value ≈ 0.96 ; Figure 5.5D and E).

Although the three *Gi*ARF1 paralogues have PVs localizations, there were subtle variations within these distribution patterns. Notably, of the three, *Gi*ARF1FB1 was most different in its cellular localization, which is likely due to a greater overall sequence divergence compared to the *Gi*ARF1 and *Gi*ARF1FA (Figure 5.1). Regulation for organellar recruitment to the PV regions by effectors may be disparate, as the G4 domain from this protein is highly divergent in its sequence. Altogether, these findings argue for a possible scenario where the three paralogues are distinct in their functions. The different *Giardia* ARFs may mark PVs as a heterogeneous population with different endosomal identities based on these variable associations. Expansion within the ARF1 family for functional differentiation is a phenomenon observed in mammalian systems (*i.e.*, duplications to yield class I and class II ARFs), which regulate vesicle formation at distinct endomembrane compartments (Cavenagh et al., 1996). Therefore, in the absence of endosomal ARF6, convergence to replace those molecular functions could have been an evolutionary strategy employed in Fornicata, where further reductive evolution in *Giardia* may be accompanied by neofunctionalization and sub-specialization at its only endo-lysosomal compartment.

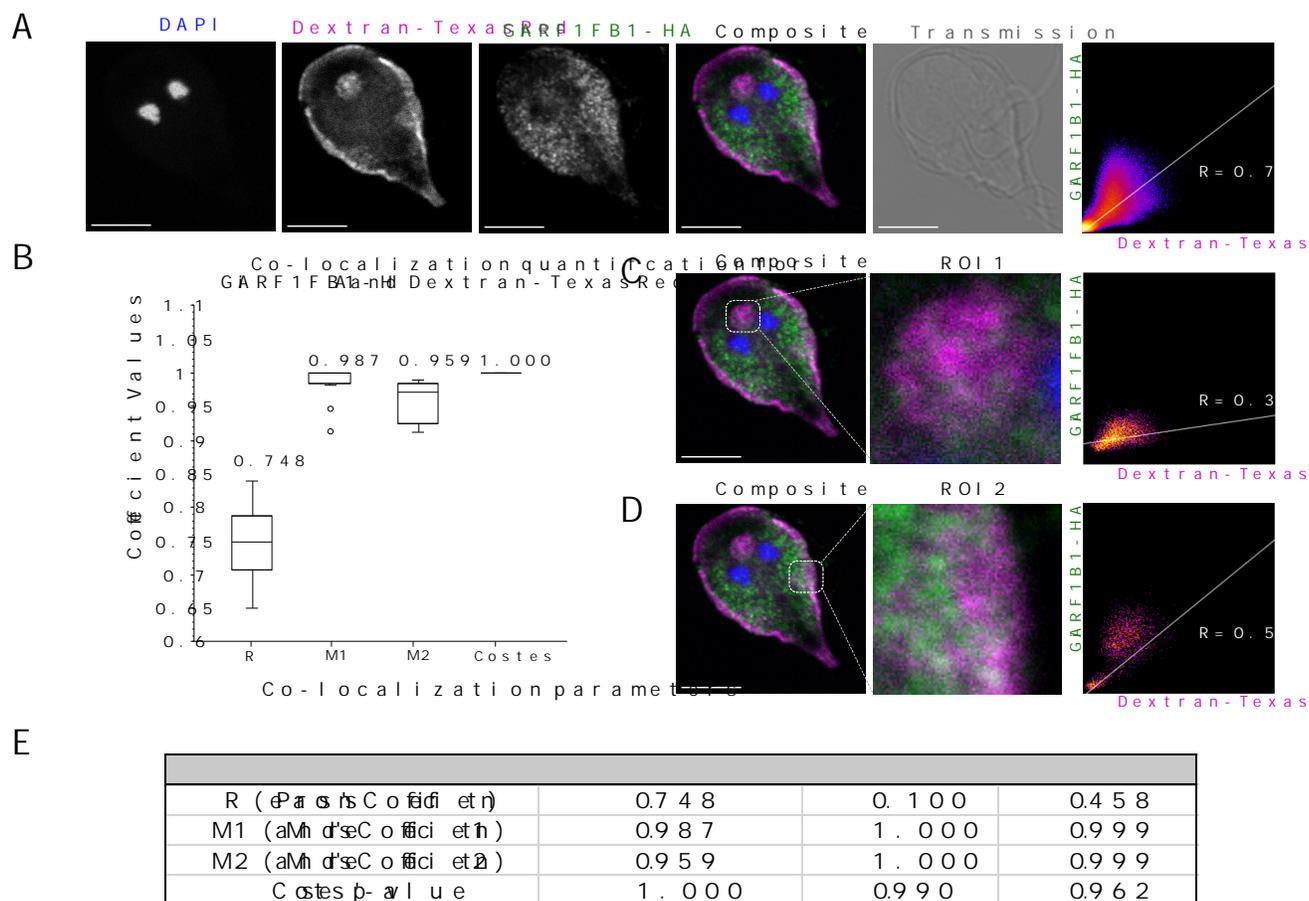


Figure 5.5. Characterization of *GiARF1FB1-HA* subcellular location. (A) depicts results from co-labelling immunoprobings experiments with transgenic *Giardia* WB(C6) trophozoites labelled for epitope-tagged *GiARF1FB1-HA* (green) and Dextran-TxR (magenta). (B) Signal overlap analyses were performed on whole-cell images to determine co-localization between *GiARF1FB1-HA* and Dextran-TxR. Boxplot depiction of the calculated R, M1, M2, and Costes' p-values across ≥ 10 analyzed cells, along with their mean values indicated on top, is provided. (C) Signal overlap analyses were performed at the bare zone (ROI) to determine co-localization between *GiARF1FB1-HA* and Dextran-TexasRed. (D) Signal overlap analyses were performed at the cortical PVs (ROI2) to determine co-localization between *GiARF1FB1-HA* and Dextran-TexasRed. (E) Summary of average co-localization parameters were calculated for the whole image, ROI1, and ROI2 using ≥ 10 cells. Scale bars: 5 μm . Optical slices were acquired from the middle of the cell for a maximum projection view of the nuclei and the bare zone. All images were acquired using the Leica SP8 x STED laser scanning confocal microscope under the 100x oil immersion objective lens.

5.3.3 ARFGAP1 and CYTHa are also PV localized

The GTPase activity of ARFs is regulated by guanine exchange nucleotide factors (ARF GEFs) and GTPase activating proteins (ARF GAPs). As discussed in Chapter 3, *Giardia intestinalis* AWB encodes three ARF GAPs and three ARF GEFs, so for this investigation, one from each family was chosen, namely *GiARFGAP1* and *GiCYTHa*, in order to assess their molecular roles. The rationale for selecting these two specific proteins, compared to the other ARF GAPs and GEFs, lies in their biological relevance to the ARF1 pathway and *Giardia* in general. ARFGAP1 regulates ARF1 for COPI and adaptin-mediated vesicle formation processes at the *cis*-Golgi (Yang et al., 2002). Because *Giardia* possesses both canonical ARF1 and ARF1 paralogues but lacks a stacked Golgi or Golgi-like compartments in the trophozoite stage, the molecular localization of this protein was examined in the absence of those organelles. Unlike ARF1 and ARFGAP1, *Giardia* AWB lacks Golgi-associated ARF GEFs, BIG and GBF1, and instead possesses two paralogs of Cytohesins (termed CYTHa and CYTHb). Of the two, CYTHa was chosen as it has the greatest evidence for biological function due to its expression and pulldown in previously published transcriptomics and proteomics datasets generated from experiments with the trophozoite stage (GiardiaDB).

The Identical cloning strategy as the *GiARF1* paralogs was employed with *GiARFGAP1* and *GiCYTHa*, which were episomally expressed as recombinant proteins with a C-terminal HA-epitope. 461 transgenic trophozoites constitutively expressing *GiARFGAP1*-HA were checked for protein expression, where 79% had fluorescent signals. Like ARFs, 18% of the sub-population had an overexpression phenotype (Online Appendix Table 5.2; Figure 5.2E and G). Similarly, 82% of the examined *GiCYTHa*-HA-expressing trophozoites fluoresced, and roughly 24% had an overexpression phenotype (Figure 5.2F and G). To better pinpoint organellar localizations, representative trophozoites from both transgenic lines were more closely examined.

In *GiARFGAP1*-HA expressing trophozoites, punctate protein distribution occurred in the PV regions along with some dispersal in the cytosol (Figure 5.6). PV localization was evaluated further by co-labeling cells with Dextran-TxR, which were subject to complete-image and specific regions of interest statistical signal overlap analyses. Whole image co-localization quantification yielded positive Pearson's correlation values and was accompanied with high M1 and M2 coefficients and Costes' p-values (average $R \approx 0.51$, $M1 \approx 1.00$, $M2 \approx 1.00$, and Costes' p-value ≈ 1), suggesting likely signal overlap between *GiARFGAP1*-HA and Dextran-TxR (Online Appendix Table 5.2; Figure 5.6A and E). Bare zone (ROI1) and cortical PV (ROI2) associations were assessed through regions of interest analyses. Although strong bare zone association was not observed (average $R \approx -0.03$, $M1 \approx 1.00$, $M2 \approx 1.00$, Costes' p-value ≈ 1), some signal overlap was still present at the cortical PV region (average $R \approx 0.31$, $M1 \approx 1.00$, $M2 \approx 1.00$, Costes' p-value ≈ 0.86) (Online Appendix Table 5.2; Figure 5.6C, D, and E).

Unlike ARFGAP1, *GiCYTH*-HA had an exclusive localization to the PVs. Substantial signal overlap between the recombinant protein and Dextran-TxR was observed both at the bare zone as well as the cortical PV regions (Figure 5.7). Quantitative analyses confirmed these qualitative observations where whole-cell image comparisons yielded high positive Pearson's and Manders' coefficients, and Costes' p-

values across all sampled cells (average $R \approx 0.68$, $M1 \approx 1.00$, $M2 \approx 1.00$, and Costes' $p\text{-value} \approx 1$; Online Appendix Table 5.2; Figure 5.7B and E). Examination of the bare zone (ROI1) also showed statistically significant overlap between the two fluorescent signals (average $R \approx 0.40$, $M1 \approx 1.00$, $M2 \approx 1.00$, and Costes' $p\text{-value} \approx 1$; Online Appendix Table 5.2; Figure 5.7C and E). A similar trend of positive co-localization was also noted at the cortical regions in all sampled trophozoites (average $R \approx 0.56$, $M1 \approx 1.00$, $M2 \approx 1.00$, and Costes' $p\text{-value} \approx 0.99$; Online Appendix Table 5.2, Figure 5.7D and E).

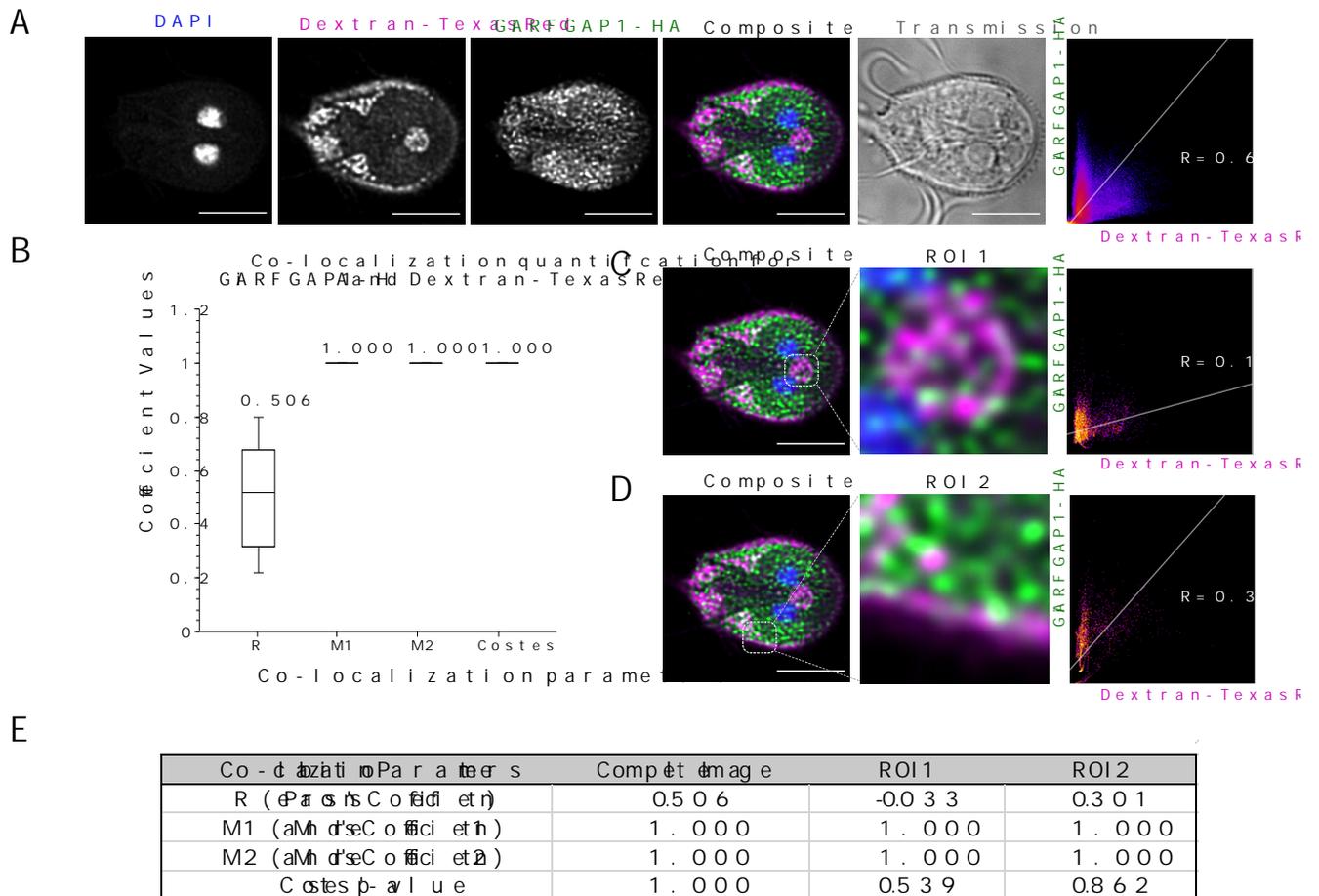


Figure 5.6. Characterization of *GiARFGAP1-HA* subcellular location. (A) depicts results from co-labelling immunoprobings experiments with transgenic *Giardia* WB(C6) trophozoites labelled for epitope-tagged *GiARFGAP1-HA* (green) and Dextran-TxR (magenta). (B) Signal overlap analyses were performed on complete images to determine co-localization between *GiARFGAP1-HA* and Dextran-TxR. Boxplot depiction of the calculated R, M1, M2, and Costes' p-values across ≥ 10 analyzed cells, along with their mean values indicated on top, is provided. (C) Signal overlap analyses were performed with the bare zone (ROI1) to determine co-localization between *GiARFGAP1-HA* and Dextran-TxR. (D) Signal overlap analyses were performed with the cortical PVs (ROI2) to determine co-localization between *GiARFGAP1-HA* and Dextran-TxR. (E) Summary of average co-localization parameters were calculated for the whole image, ROI1, and ROI2 using ≥ 10 cells. Scale bars: 5 μm . Optical slices were acquired from the middle of the cell for a maximum projection view of the nuclei and the bare zone. All images were acquired using the Leica SP8 X STED laser scanning confocal microscope under the 100x oil immersion objective lens.

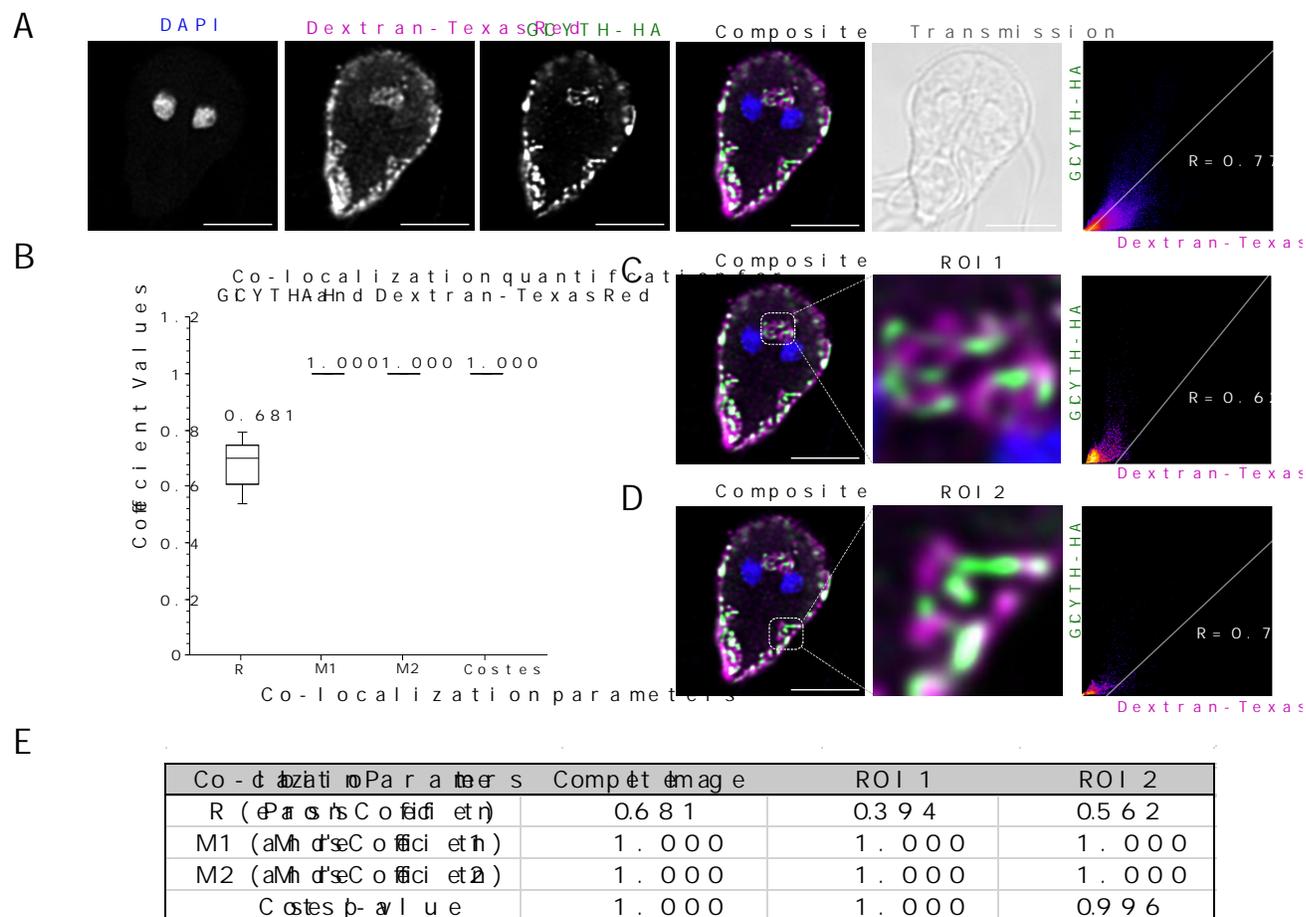


Figure 5.7. Characterization of GiCYTHa-HA subcellular location. (A) depicts results from co-labelling immunoprobings experiments with transgenic *Giardia* WB(C6) trophozoites labelled for epitope-tagged GiCYTHa-HA (green) and Dextran-TxR (magenta). (B) Signal overlap analyses were performed on whole-cell images to determine co-localization between GiCYTHa-HA and Dextran-TxR. Boxplot depiction of the calculated R, M1, M2, and Costes' p-values across ≥ 10 analyzed cells, along with their mean values indicated on top, is provided. (C) Signal overlap analyses were performed with the bare zone (ROI1) to determine co-localization between GiCYTHa-HA and Dextran-TxR. (D) Signal overlap analyses were performed at the cortical PVs (ROI2) to determine co-localization between GiCYTHa-HA and Dextran-TxR. (E) Summary of average co-localization parameters were calculated for the whole image, ROI1, and ROI2 using ≥ 10 cells. Scale bars: 5 μ m. Optical slices were acquired from the middle of the cell for a maximum projection view of the nuclei and the bare zone. All images were acquired using the Leica SP8 X STED laser scanning confocal microscope under the 100x oil immersion objective lens.

The combined microscopy results from the *Giardia* ARF regulatory system proteins suggest their function to be in proximity to the PVs and the plasma membrane interfaces in the Golgi- and endosome-less trophozoites. Note, that although signal overlap for all occurs around the PV regions, perfect co-localization is never observed (*i.e.*, $R=1$) between any of the recombinant proteins and the luminal marker. This means that none of the proteins are internally localized inside the PVs themselves. Instead, they are likely membrane-associated or in the intervening cytosolic spaces between the PVs and the plasma membrane. Collectively, several previously discussed *Giardia* vesicle coat proteins, ESCRT components, and now the ARF regulatory proteins demonstrate a considerable degree of overlap in their patterns of cellular localization. These findings warrant a closer proteomics investigation to define organellar and protein-protein associations better. This investigation was performed while considering two overarching goals. The first was to assess whether proteins of the *Giardia* ARF regulatory system associate with any of the vesicle coat complexes at the PVs, a postulation derived based on the mechanism of ARF-dependent coat protein assembly at endo-membranes in model organisms. The second was to determine whether *GiARFGAP1* or *GiCYTHa* modulate the ARF cycle of any of the three *GiARF1* paralogues.

5.3.4 *ARF1-HA* associates with *Giardia*'s vesicle formation machinery at the PVs

For the proteomics investigation, co-immunoprecipitation (co-IP) experiments were performed using the HA-tagged variants of the *Giardia* ARF regulatory system proteins as affinity handles, which were then cross-linked with their core interaction partners and subject to pulldown using anti-HA agarose beads. Beads were then analyzed using liquid-chromatography mass-spectrometry (LC/MS) to identify all associated proteins. To characterize biologically relevant interactions, co-IP and LC/MS experiments for all proteins were performed in replicates, where only hits identified in both sets meeting the (r)iBAQ threshold of 0.01% were considered as true positives (Online Appendix Tables 5.3 and 5.4). Proteins identified in the non-transgenic control datasets were treated as non-specific enrichments and removed from analyses.

Using this approach, the molecular interactions of *GiARF1-HA* were first investigated. Protein-pulldown was confirmed using immunoblot analyses for both bead and control whole-cell lysate samples. *GiARF1-HA* enrichment was validated by the presence of a 22.8 kDa band, and bait-associating protein complexes were visualized in the Coomassie-stained gel (Figure 5.8A and B). Samples crosslinked to the beads were subject to mass-spectrometry, which generated peptide hits that were classified according to overall relative protein abundance to yield (r)iBAQ values. After applying the above-mentioned thresholds, 195 hits met the 0.01% (r)iBAQ cut-off (Online Appendix Table 5.4). At least 27 of these were well characterized PV-associated proteins, which have been functionally validated through previous molecular functional investigations (Cernikova et al., 2020; Hardin et al., 2017; Miras et al., 2013; Pipaliya, Santos, et al., 2021; Rivero et al., 2010; Saha et al., 2018; Touz et al., 2004; Zumthor et al., 2016) (Online Appendix Table 5.5). Most of these PV-associated proteins belonged to families of vesicle formation and fusion machinery. HTAC components were enriched in both replicates, namely COPI components, γ -COP, δ -COP, α -COP, and β '-Cop, and AP-2 subunits, $\alpha 2$ and $\mu 2$ (Online Appendix Table 5.5; Figure 5.8C). Retromer

and ESCRT subunits, Vps35 and ESCRTIII-Vps31 and ESCRTIII-Vps4, respectively, were also interacting (Online Appendix Table 5.5; Figure 5.8C). Unexpectedly, aside from the vesicle formation machinery, a number of small GTPases that are necessary for catalysis of vesicle fusion dynamics were also identified. Many of these were typically Golgi-associated and endocytic Rab1, Rab2, and Rab11 (Online Appendix Table 5.5; Figure 5.8C). The large trafficking GTPase, Dynamin, which has been implicated in the PV-PM membrane fusion dynamics, was also present in these screens (Gaechter et al., 2008; Zumthor et al., 2016).

Apart from cargo trafficking complexes, PV-membrane markers were also identified. Previous investigations characterized numerous giardial phosphoinositide-binding proteins, which in model systems are adaptors that bind to PI(4,5)P2 plasma membrane phospholipids to recruit clathrin and coat proteins for clathrin-mediated endocytosis (Cernikova et al., 2020). In *Giardia*, these PIP-binding proteins are associated with the PVs in conjunction with clathrin heavy chain (Cernikova et al., 2020). Two of the previously investigated giardial PIP-binding proteins, FYVE domain-containing protein (GL50803_16653) and WD40 domain-containing protein (GL50803_10822), were also enriched with GiARF1-HA (Cernikova et al., 2020) (Figure 5.8C).

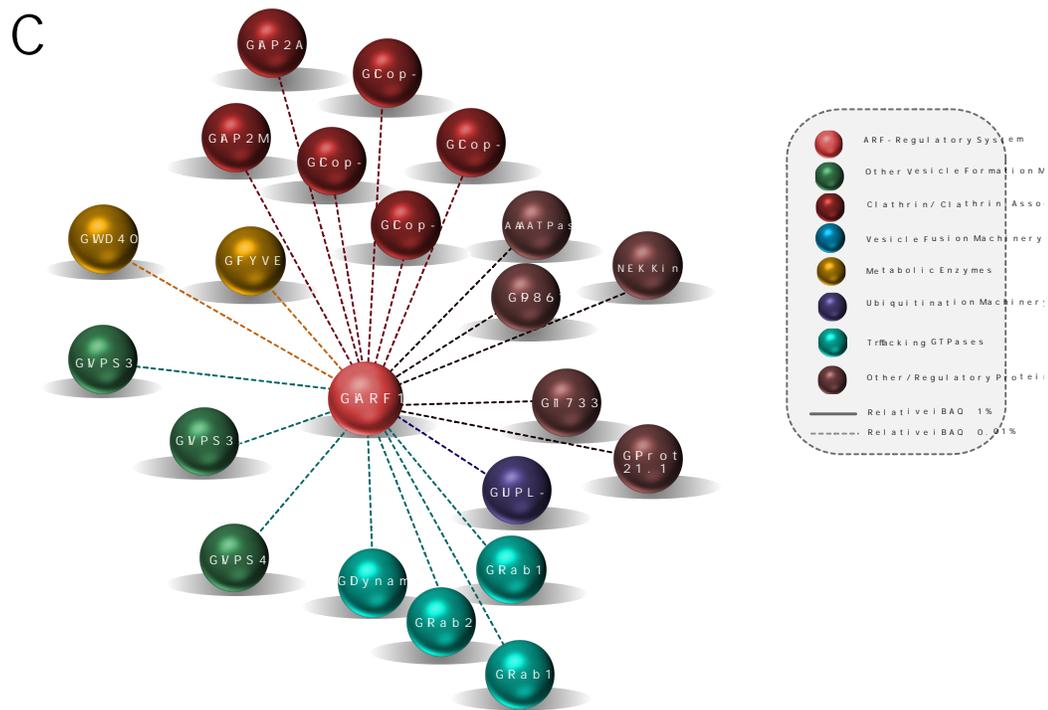
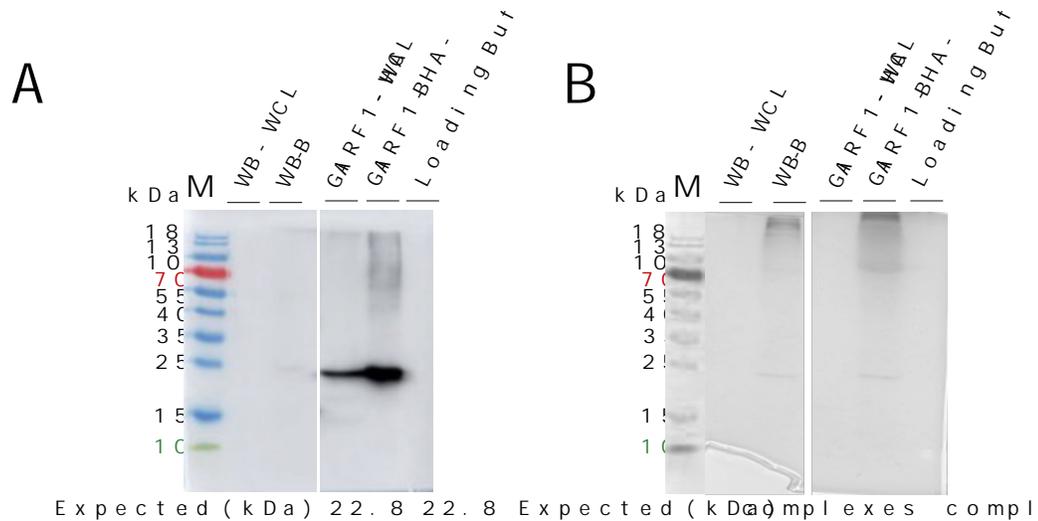


Figure 5.8. Co-immunoprecipitation and mass spectrometry analyses to determine *GiARF1-HA* protein-protein interactions. (A) Western blot confirmation of bait pull-down (*GiARF1-HA*) in bead and cell lysate samples is depicted, compared to the non-transgenic (WB) control samples, which were performed using SDS-PAGE. An expected band size corresponding to the molecular weight of *GiARF1-HA* (22.8 kDa) was observed. (B) Coomassie staining was performed to visualize cross-linked protein complexes in bead samples from non-transgenic and transgenic cells. (C) The interactome of the experimentally validated PV-associating proteins enriched in the *GiARF1-HA* pull-down and identified by LC/MS is depicted. A threshold of 0.01% (r)iBAQ was used to determine biologically relevant interactions and are indicated by the dashed lines. Highly abundant interactions with (r)iBAQ of $\geq 1\%$ are indicated with solid lines. Abbreviations used: M, Thermo Fisher PageRuler pre-stained protein ladder (10 to 180 kDa), WB, non-transgenic control cells, WCL, whole-cell lysate, B, beads.

5.3.5 ARF1FA-HA and ARF1FB1-HA protein interactions are similar to GiARF1 at the peripheral vacuoles

Like *GiARF1*, the two *GiARF1F* paralogues had partial or complete PV localizations. Co-IP and LC/MS experiments were performed with *GiARF1FA-HA* and *GiARF1FB1-HA*, using them as affinity handles for protein pulldown to assess the molecular interactions common or unique to each paralogue.

GiARF1FA-HA enrichment was confirmed by western blotting, where a 21.5 kDa band corresponding to the protein was observed in both bead and pellet samples (Figure 5.9A). Crosslinked protein complexes were visualized through Coomassie staining and analyzed through LC/MS (Figure 5.9B). A total of 132 proteins were enriched, of which 15 have known PV functions (Online Appendix Tables 5.4 and 5.5). Identical to *GiARF1-HA*, several of these associations belonged to HTAC components such as AP-2 μ , β -COP, and retromer-Vps35 (Online Appendix Table 5.5; Figure 5.9C). Other shared interaction partners included Dynamin, Rab1a, and PIP-binding FYVE domain-containing protein. Notably, unique to this interactome was the presence of ubiquitin, which is typically necessary for protein post-translational modification and cargo targeting to endo-lysosomal compartments. Overall, although the interactions were not identical, a demonstrable overlap in the types of machinery present in both *GiARF1* and *GiARF1FA* datasets existed. The presence of PV-associated proteins also corroborated patterns of localization that were observed with fluorescence microscopy. Based on these results, both *GiARF1FA* and *GiARF1* may have canonical ARF1 roles for recruiting coat protein complexes, possibly facilitating membrane budding and scission at the peripheral vacuoles.

GiARF1FB1-HA co-immunoprecipitation was confirmed using immunoblotting where a band size of 19.2 kDa, corresponding to ARF1FB1, was observed in bead and pellet samples (Figure 5.10A). Western blotting and Coomassie staining also confirmed pulldown of cross-linked complexes (Figure 5.10B). Mass-spectrometry of *GiARF1FB1*-derived protein complexes identified 257 different *Giardia* proteins, of which 23 are experimentally validated as PV-associating (Online Appendix Tables 5.4 and 5.5). Once again, several PV-localizing vesicle formation machinery, such as ESCRTIII-Vps31 and retromer-Vps26 and Vps35, associated with *GiARF1FB1* (Online Appendix Table 5.5; Figure 5.10C). However, several important differences within this particular interactome should be noted, compared to the *GiARF1* and *GiARF1FA*. Most notably, none of the HTAC machinery enriched as biologically relevant interactions (Online Appendix Table 5.5; Figure 5.10C). Unlike the other ARFs, Dynamin and Rab11 were among the strongest interacting partners with $\geq 1\%$ (r)iBAQ. Numerous Rabs were also present (*i.e.*, Rab1, Rab32, and Rab2a), but also COPII-SAR1 (Figure 5.10C). Intriguingly, associations with several Q and R SNAREs, namely Syntaxin 1 and Synaptobrevin, which were not identified in the other two ARF datasets, were also enriched with ARF1FB1-HA. These proteins typically assemble within the SNARE complex for vesicle fusion dynamics at the plasma membrane for cargo exocytosis (Jahn, 2004) (Figure 5.10C). As with *GiARF1* and *GiARF1FA*, associations with PIP-binding proteins such as the FYVE domain-containing protein and WD40 domain-containing protein were also observed. However, unique to this repertoire was enrichment of a third PIP-binding protein, PXD3 (Figure 5.10C). Finally, *GiARF1FB1* also associated with

ubiquitination machinery (ubiquitin and UPL-1) as well as other PV-localizing regulatory proteins such as NEK kinases (Figure 5.10C).

Although there are similarities in the interaction patterns of the three ARF paralogues, *Gi*ARF1FB1 did not associate with coat proteins. Additional associations with SNARE proteins and COPII components were uniquely enriched in this interactome. This possibly correlates to the variance in PV localization wherein *Gi*ARF1FB1 differed the most and had a diffused localization compared to the other two paralogues, where the latter were primarily restricted to the cell periphery regions. Differences in interactomes could be incidental to possible divergences in regulatory mechanisms and general ARF-binding functions compared to the other two paralogs, a notion perceptible at the sequence level (Figure 5.1).

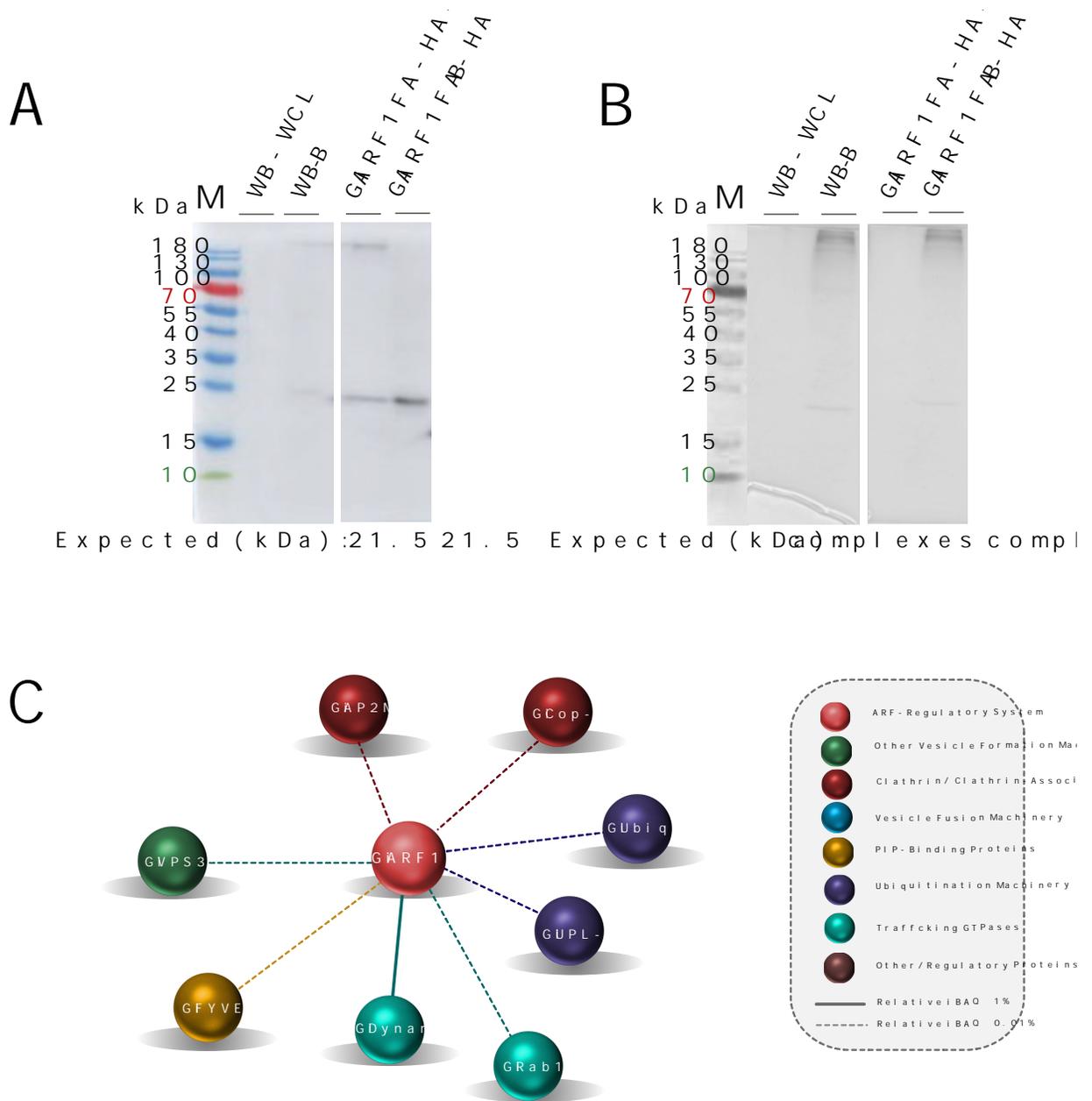


Figure 5.9. Co-immunoprecipitation and mass spectrometry analyses to determine *GiARF1FA-HA* protein-protein interactions. (A) Western blot confirmation of bait pull-down (*GiARF1FA-HA*) in bead and cell lysate samples is depicted, compared to the non-transgenic (WB) control samples, which were performed using SDS-PAGE. An expected band size corresponding to the molecular weight of *GiARF1FA-HA* (21.5 kDa) was observed. (B) Coomassie staining was performed to visualize cross-linked protein complexes in bead samples from non-transgenic and transgenic cells. (C) The interactome of the experimentally validated PV-associating proteins enriched in the *GiARF1FA-HA* pull-down and identified by LC/MS is depicted. A threshold of 0.01% (r)iBAQ was used to determine biologically relevant interactions and are indicated by the dashed lines. Highly abundant interactions with (r)iBAQ of $\geq 1\%$ are indicated with solid lines. Abbreviations used: M, Thermo Fisher PageRuler pre-stained protein ladder (10 to 180 kDa), WB, non-transgenic control cells, WCL, whole-cell lysate, B, beads.

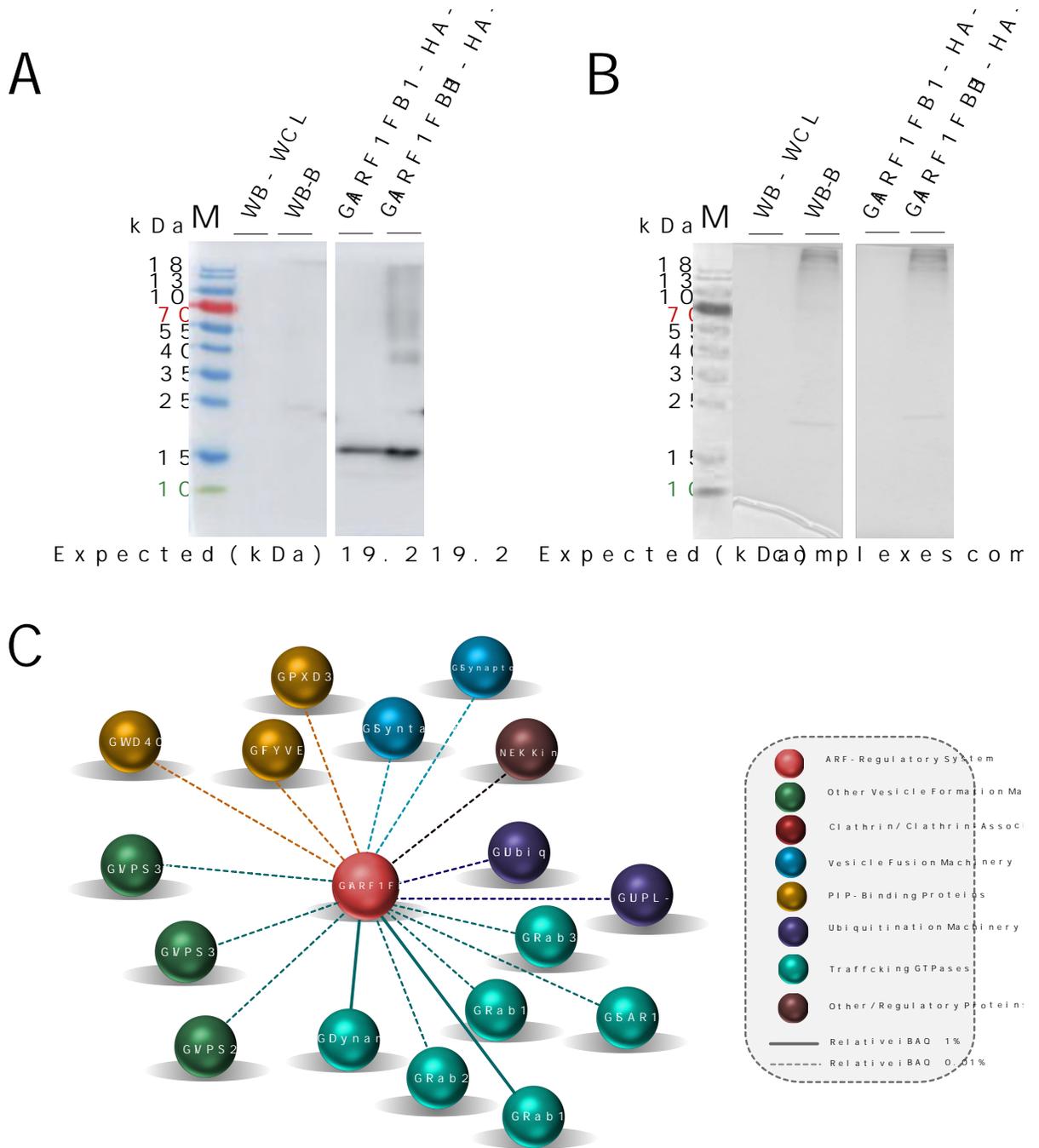


Figure 5.10. Co-immunoprecipitation and mass spectrometry analyses to determine *Gi*ARF1FB1-HA protein-protein interactions. (A) Western blot confirmation of bait pull-down (*Gi*ARF1FB1-HA) in bead and cell lysate samples is depicted, compared to the non-transgenic (WB) control samples, which were performed using SDS-PAGE. An expected band size corresponding to the molecular weight of *Gi*ARF1FB1-HA (19.2 kDa) was observed. (B) Coomassie staining was performed to visualize cross-linked protein complexes in bead samples from non-transgenic and transgenic cells. (C) The interactome of the experimentally validated PV-associating proteins enriched in the *Gi*ARF1FB1-HA pull-down and identified by LC/MS is depicted. A threshold of 0.01% (r)iBAQ was used to determine biologically relevant interactions and are indicated by the dashed lines. Highly abundant interactions with (r)iBAQ of $\geq 0.1\%$ are indicated with solid lines. Abbreviations used: M, Thermo Fisher PageRuler pre-stained protein ladder (10 to 180 kDa), WB, non-transgenic control cells, WCL, whole-cell lysate, B, beads.

5.3.6 ARFGAP1 and CYTHa also associate with the PV machinery

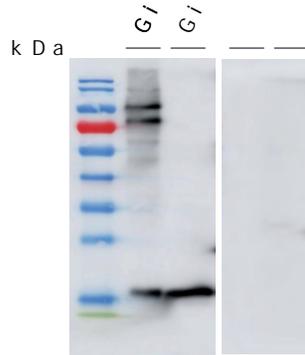
Apart from the GTPases, *Gi*ARFGAP1-HA and *Gi*CYTHa-HA also had patterns of localization at the PVs. These GEFs and GAPs typically regulate ARF recruitment to organellar membranes and cycling back into a cytosolic pool by promoting GDP to GTP exchange and hydrolysis of the GTP phosphate group in an 'on' and 'off' switch-like mechanism. Since both *Gi*ARFGAP1-HA and *Gi*CYTHa-HA localized to similar PV regions as the ARF paralogues, the possibility of ARF cycle regulation by these proteins at these subcellular interfaces was also probed through proteomics.

*Gi*ARFGAP1-HA recombinant protein served as a bait for co-immunoprecipitation and crosslinking of complexes for pulldown using anti-HA agarose beads. Immunoblotting and Coomassie staining confirmed pulldown of the *Gi*ARFGAP1-HA bait (20 kDa) and complexed proteins (large, smeared band) (Figure 5.11A and B). LC/MS identified an overlapping protein interaction pattern with *Gi*ARF1, *Gi*ARF1FA, and *Gi*ARF1FB1, where vesicle coat components, α -COP, β '-COP, γ -COP, and AP-2 μ , interacted (Online Appendix Tables 5.3-5.5; Figure 5.11C). ESCRTIII-Vps4, three paralogues of ESCRTIII-Vps31, and retromer-Vps35 were also enriched. Of the GTPases, Rab11, SAR1, and Dynamin were present along with three PIP-binding proteins (PXD3, WD40, and FYVE domain-containing proteins).

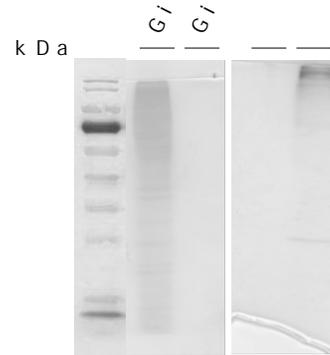
Interestingly, proteomics analyses with *Gi*ARFGAP1 identified ARF1FA and another ARF GAP (*i.e.*, AGFG) to be associating (Figure 5.11C). Although ARFGAP1 was not identified in the ARF1FA interactome, it is present in this reverse co-IP, hinting at a scenario that involves ARFGAP1-mediated regulation of ARF1FA. In addition to this, an association with AGFG suggests this protein may have effector functions for the recruitment of other trafficking GTPases, such as Rabs, in a cascade-like mechanism. Nonetheless, interaction with the PV-associating proteins corroborates the localization results at the peripheral vacuoles.

*Gi*CYTHa was chosen as the representative Sec7 domain-containing ARF GEF for investigation. Unlike *Gi*ARFs and *Gi*ARFGAP1, the predicted molecular weight of this protein is much larger (*i.e.*, 266 kDa). Therefore, *Gi*CYTHa-HA bait separation was performed on a 7.5% polyacrylamide gel compared to the standard 10% gel. Instead of observing a single band with a molecular weight of 266 kDa, numerous bands of different sizes were observed (Figure 5.12A and B). This may be due to several possible alternate start codons in the predicted open reading frame or are products of protein degradation. Nonetheless, LC/MS confirmed peptides corresponding to *Gi*CYTHa-HA to be highest in enrichment (Online Appendix Tables 5.3-5.5). 128 common interactors were shared between the replicate datasets, where a large fraction of these corresponded to ribosomal proteins (Online Appendix Table 5.4). However, others were mostly vesicle trafficking proteins, namely, Rab11, Rab2a, Synaptobrevin, Syntaxin 16, and an unclassified Qa-SNARE. Notably, strong associations with PDX3 and AP-2 γ were also identified. Other PV-associating hypothetical proteins and transporters were also enriched. Although no ARF or ARF GAP interactions were identified in this screen, proteomics with *Gi*CYTHa still supports microscopy observations and its localization to the PVs.

A



B



C

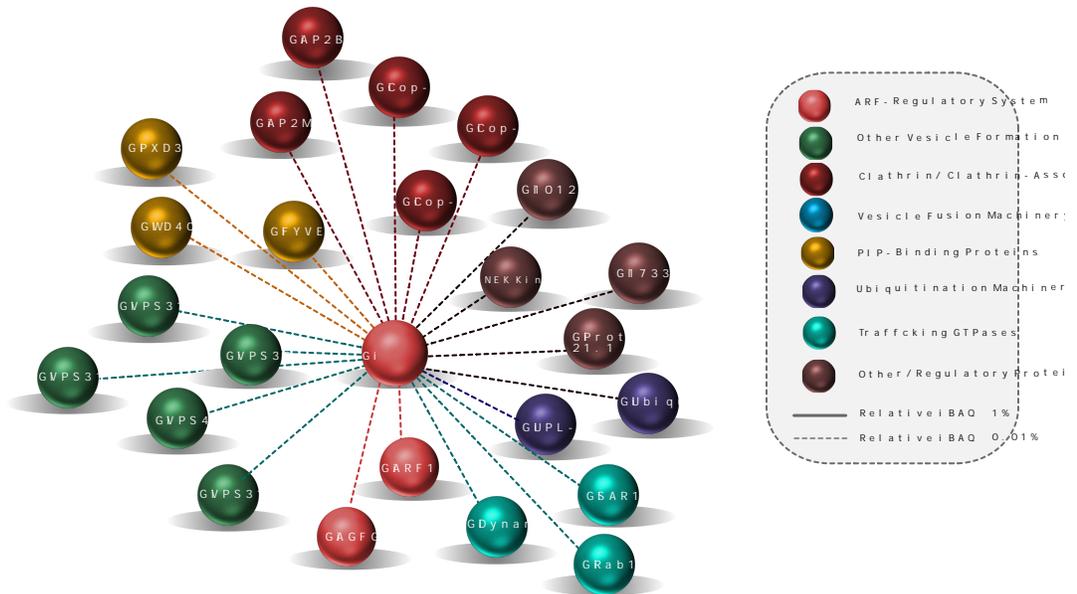


Figure 5.11. Co-immunoprecipitation and mass spectrometry analyses to determine *GiARFGAP1-HA* protein-protein interactions. (A) Western blot confirmation of bait pull-down (*GiARFGAP1-HA*) in bead and whole-cell lysate samples is depicted, compared to the non-transgenic (WB) control samples, which were performed using SDS-PAGE. An expected band size corresponding to the molecular weight of *GiARFGAP1-HA* (19.8 kDa) was observed. (B) Coomassie staining was performed to visualize cross-linked protein complexes in bead samples from non-transgenic and transgenic cells. (C) The interactome of the experimentally validated PV-associated proteins enriched in the *GiARFGAP1-HA* pull-down and identified by LC/MS is depicted. A threshold of 0.01% (r)iBAQ was used to determine biologically relevant interactions and are indicated by the dashed lines. Highly abundant interactions with (r)iBAQ of $\geq 1\%$ are indicated with the solid lines. Abbreviations used: M, Thermo Fisher PageRuler pre-stained protein ladder (10 to 180 kDa), WB, non-transgenic control cells, WCL, whole-cell lysate, B, beads.

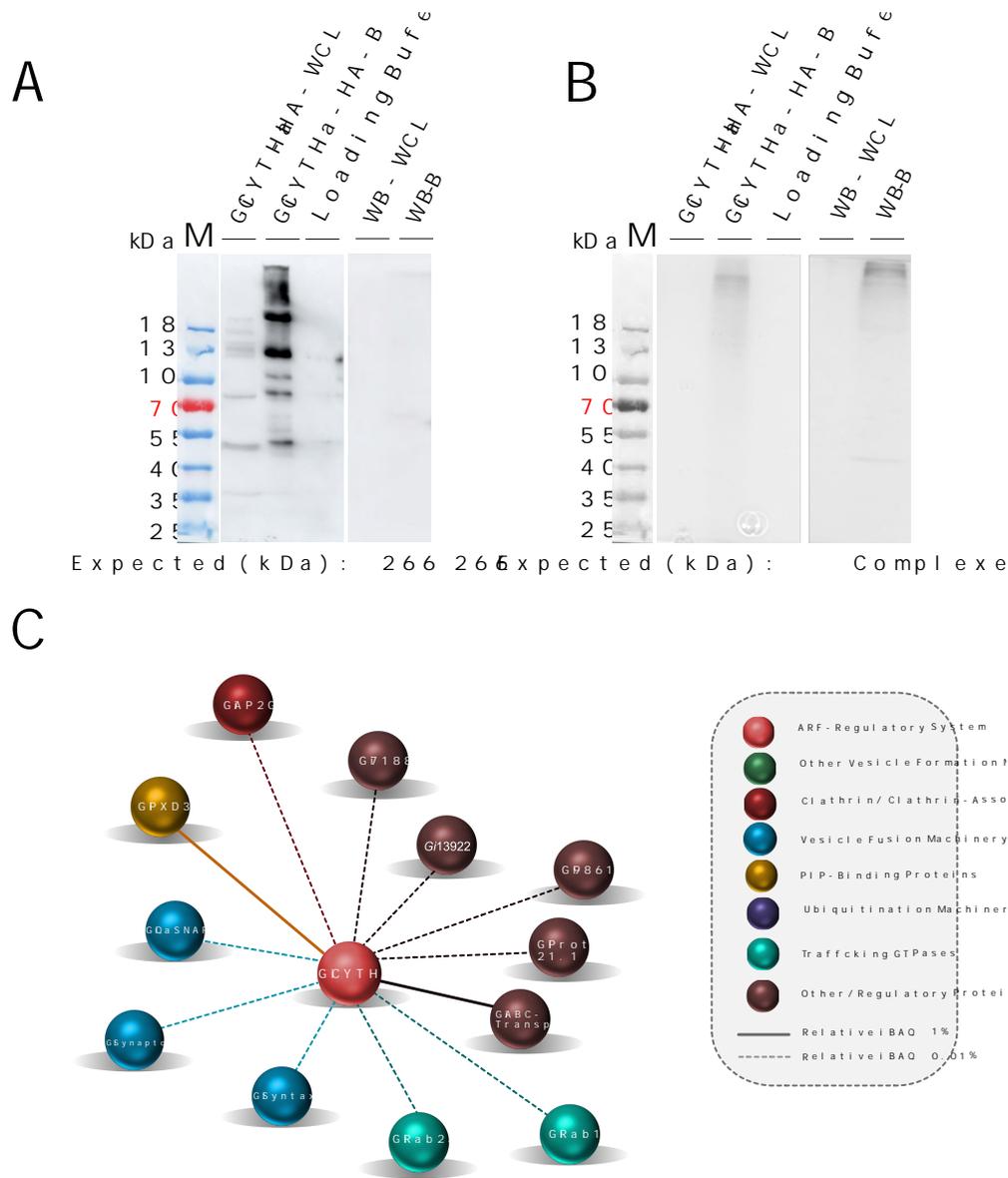


Figure 5.12. Co-immunoprecipitation and mass spectrometry analyses to determine *GiCYTHa-HA* protein-protein interactions. (A) Western blot confirmation of bait pull-down (*GiCYTHa-HA*) in bead and whole-cell lysate samples is depicted, compared to the non-transgenic (WB) control samples, which were performed using SDS-PAGE. An expected band size corresponding to the molecular weight of *GiCYTHa-HA* (226 kDa) was not observed. Instead, numerous bands of variable sizes in both the bead and cell lysate samples were present in the bait-associated lanes. (B) Coomassie staining was performed to visualize cross-linked protein complexes in non-transgenic and transgenic bead samples. (C) The interactome of the core PV-associating proteins enriched in the *GiCYTHa-HA* pull-down and identified by LC/MS is depicted. A threshold of 0.01% (r)iBAQ was used to determine biologically relevant interactions and are indicated by the dashed lines. Highly abundant interactions with (r)iBAQ of $\geq 1\%$ are indicated with solid lines. Abbreviations used: M, Thermo Fisher PageRuler pre-stained protein ladder (10 to 180 kDa), WB, non-transgenic control cells, WCL, whole-cell lysate, B, beads.

5.3.7 Intersection of binding partners confirm the association of ARF-regulatory system proteins at the PVs

Individual proteomics investigations with each of *Giardia*'s ARF-regulatory system component provides evidence for their associations at the peripheral vacuoles. Common trends in the types of interaction partners were present between one or more components. Therefore, in order to consolidate these findings and determine overlapping networks between these PV associations, all five proteomics datasets were intersected (Figure 5.13). Interactions are mainly with trafficking proteins that typically mark early and late endosomal pathways. These were AP-2 and COPI components, ubiquitination machinery, ESCRTs, and retromer. Vesicle fusion proteins such as SNAREs and Rabs were also identified. Overall, these findings confirm two attributes of the ARF pathway that are conserved in other systems. The first is that the giardial ARF regulatory system likely modulates the assembly dynamics of vesicle coat complexes around the PV and PM membranes to mark PVs as the singular endo-lysosomal compartment. Second is that apart from canonical aspects, these analyses yield exciting insights into the role of Rabs as mediators of crosstalk between the two small GTPase pathways, as is the case in mammalian endocytic systems.

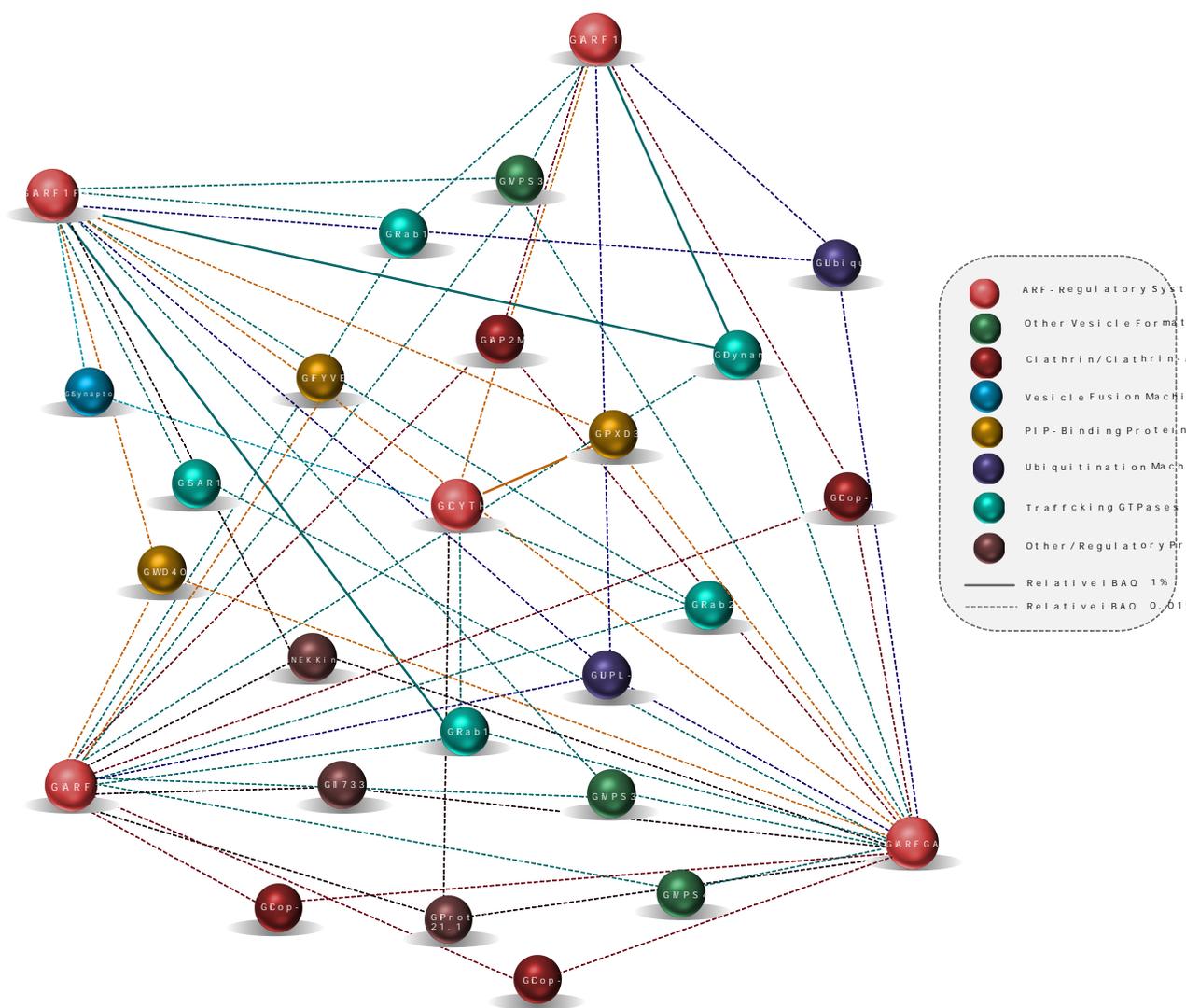


Figure 5.13. Network analysis of the PV-associated interacting partners. LC/MS hits enriched in all five *Giardia* ARF regulatory system proteomic datasets with known PV localizations and functions were intersected to identify common trends in the molecular machinery involved in vesicle formation and fusion pathways, as well as those with other endo-lysosomal activities. Highly abundant interactions with (r)iBAQ of $\geq 1\%$ are indicated with solid lines.

5.4 Discussion

The work performed in this chapter is the first comprehensive investigation into the entire ARF regulatory system and its molecular roles in *Giardia*. Previous studies have been limited to the characterization of *Gi*ARF1 function during encystation. However, this study provides a detailed view into the roles of all three paralogues and their regulatory proteins during the trophozoite stages, which participate in active intra- and extracellular material exchange at the PVs. Using a combination of fluorescent microscopy and proteomics, the molecular localization and interaction partners for each aspect of the *Giardia* ARF regulatory protein at the specialized peripheral vacuoles (PVs) were determined. Although PVs are highly distinguishable compartments, their exact homology and the precise dynamics underpinning organellar function and neogenesis are still an active area of investigation. This study identifies an additional layer of mechanistic complexity and identities of the trafficking machinery that may be participating in PV function regulation. Identification of the associating molecular players has also cast light onto notions of crosstalk between different GTPase pathways and the involvement of the vesicle fusion machinery. Especially in this latter case, *Giardia* could therefore serve as a model for a stripped-down system to probe these questions that would otherwise be more challenging to address in systems with greater cellular complexities.

5.4.1 Paralogous expansions within fornicate ARF regulatory system likely accommodates existing trafficking complexity

Studies investigating ARF GTPases, ARF GAPs, and ARF GEFs in metazoan systems have illuminated functional expansions within this family, which are diagnostic for trafficking processes at unique intracellular destinations. These are within the context of intricate cellular complexities that arose with a transition to multicellularity and specialization in different cell types. (Li et al., 2004; Pipaliya et al., 2019; Schlacht et al., 2013; Yüzbaşıoğlu et al., 2017). In a reversed scenario, investigations into more simplified systems, such as yeast, have recapitulated the versatility and the generalist nature of ARF1 and ARF1-associated machinery for both exocytic and endocytic processes (Yahara et al., 2001). Comparative genomic assessments have identified expansions and reductive evolutionary events in the ARF and its regulators across the diversity of eukaryotes, including parasitic protists such as *Giardia*, many of which have highly modified trafficking landscapes (Li et al., 2004; Pipaliya et al., 2019; Schlacht et al., 2013; Vargová et al., 2021). However, ARF biology in many of these lineages is largely unelucidated. This investigation lends a newfound appreciation for how ARFs facilitate unconventional trafficking processes in the absence of a canonical Golgi-dependent secretory pathway. It also provides a new outlook on how different ARF paralogues can adapt to secondary organellar losses paired with lineage-specific organellogenesis.

Chapter 3 detailed duplications within the ARF1 GTPase in fornicate lineages to yield lineage-specific paralogues. This occurred in conjunction with loss of ARF6 and streamlining within the ARF GEFs and the ARF GAPs, which implies tight modulation and specificity to accommodate distinct trafficking

processes in a diversified endomembrane system in the basal free-living fornicates. A similar evolutionary trend of expansion and loss exists across the tree of eukaryotes. For example, in Holozoa, new families of ARFs (*i.e.*, class I, II, and III ARFs), ARF GAPs (*i.e.*, AGAP, ARAP, and GIT), and ARF GEFs (*i.e.*, FBX8, EFA6, and Cytohesins 1-4) arose in parallel to cellular specialization with distinct functions that depend on these different ARF, ARF GAP, or ARF GEF activities (Li et al., 2004; Pipaliya et al., 2019; Schlacht et al., 2013). Examples include induction of filopodia, cholesterol trafficking, cell migration, platelet-dense granule secretion, neurite growth, and epidermal growth factor receptor (EGFR) trafficking, to name a few (Cavet et al., 2008; Engel et al., 2004; Heun et al., 2020; Kobayashi & Fukuda, 2012; Lucanic & Cheng, 2008; Sztul et al., 2019). Plants exemplify lineage-specific patterns of gain accompanied by loss within the ARF regulatory system proteins. Embryophytes do not encode any Cytohesin or Cytohesin-derived proteins but possess numerous paralogues of Golgi-associated BIG that facilitate the transport of the plant-hormone auxin within the TGN-early endosomal pathway (Geldner et al., 2003; Kitakura et al., 2017; Pipaliya et al., 2019).

In fornicates, secondary reductive evolution followed complexity within the organization of the endomembrane system. Ultrastructural studies with the free-living *Carpodiemonas membranifera* show the presence of both a stacked Golgi and numerous endosomal-like compartments (Simpson & Patterson, 1999). It is likely that one or more ARF1 and ARF1F proteins mediate trafficking processes between the ER and Golgi and in post-Golgi endo- and exocytic processes, similar to mammalian class I, II, and III ARFs (Cavenagh et al., 1996; Chun et al., 2008; Gu & Gruenberg, 2000; Nakai et al., 2013). The presence of distinct Golgi-associated and endosomal ARF GAPs and GEFs is congruent with a scenario for the regulatory activity of one or more of these towards the three lineage-specific ARFs to mediate trafficking between the Golgi and endosomal compartments. Although gradual simplification in trafficking organelles persists across all fornicates (*i.e.*, loss of stacked Golgi), the presence of endosome-like compartments in most lineages may be indicative of latent ARF-dependent vesicle formation. In *Giardia*, the indispensability of its three ARFs, *GiARFGAP1*, and *GiCYTHa* at *Giardia*-specific PVs conforms to the notion of plasticity within the ARF regulatory system.

5.4.2 Conventional roles of the *Giardia* ARF regulatory system to functionally diversify the PVs

Although all investigated giardial ARFs and their regulators localize to the PVs, they do so in variable patterns supporting a plausible notion wherein different PVs have specialized and distinct functions. Similar to the early and late endosomal stages, a mechanism of organelle maturation may exist within the static-state PVs. Populations closer to the ER may have *trans*-Golgi network-like functions for the extracellular secretion of the newly synthesized material. Those closer to the plasma membrane, including the bare zone, may be involved in selective uptake and sorting of host macromolecules for retrograde trafficking for downstream metabolic breakdown. Support for such a model is provided by ongoing super-resolution optical microscopy (STORM) and focused ion beam scanning electron microscopy (FIB-SEM) investigations with the PVs (Santos et al., in preparation). The results from these

high-resolution microscopy studies and three-dimensional reconstruction of the PV populations classify PVs as morphologically distinct heterogeneous populations of PVs that vary in size and shape (Santos et al., in preparation). To validate this selective association model of *Gi*ARFs, *Gi*ARFGAP1, and *Gi*CYTHa, additional investigations with higher resolution microscopy, such as Stochastic Optical Reconstruction Microscopy (STORM) and correlated light and electron microscopy (CLEM), for nanoscale visualization of the fluorescent protein in combination with the ultrastructures of the PV membranes is necessary (de Boer et al., 2015; Vicidomini et al., 2018). Other viable routes of investigation that should be pursued are co-localization analyses using dual-reporter constructs and co-labeling of PV-membranes with fluorescently labeled cholera toxin (Zumthor et al., 2016). Finally, investigations of similar scope with remaining *Giardia* ARF GAPs (AGFG and SMAP) and ARF GEFs (CYTHb) may also lend further support to the notion of non-overlapping localization of different ARF regulatory proteins with different PV populations.

Mechanisms of PV biogenesis and the donor membranes which give rise to them remain a mystery. It is unclear whether these endo-lysosomal organelles originate from the plasma membrane, the endoplasmic reticulum, or both. The ARF regulatory system may be integral to vesicle-budding and maturation-like processes for PV neogenesis from either of these compartments. The different *Gi*ARF1 paralogues and the regulators may coordinate coat protein assembly and nucleation (e.g., COPII, COPI, adaptins, and retromer), followed by tethering and fusion in a SNARE-dependent manner (i.e., Syntaxins and Synaptobrevins) for synergistic vesicle scission and fusion to yield different PV-populations. For example, Syntaxin 1 and Synaptobrevin, identified in these proteomics screens, are implicated in maintaining PV morphology through fusion dynamics with the plasma membrane (Cernikova, Hehl, and Faso, in preparation). Association with the giardial COPII machinery (e.g., SAR1) and *Gi*Syntaxin 1 and *Gi*Syntaxin16 is a scenario reminiscent of mammalian ARF1-dependent formation of COPII-coated endosomes that arise at ER/ERGIC interfaces for fusion with the *cis*-Golgi in coordination with SNARE-helical bundles (Honda et al., 2005). *Giardia* trophozoites that lack stacked-Golgi or ESVs may have stages of the PVs, which are functionally analogous to COPII-vesicles or the *cis*-Golgi. Fluorescent microscopy with *Gi*ARF1FB1 supports this notion as the localization of this protein extends close to regions spanning the *Giardia* ER. Trafficking and possibly the organellogenesis of the PV network that extends into the cytosol may be harmonized with the *Giardia* ER, which is highly tubular and spans extensively throughout the cytosol.

*Gi*ARF1 and *Gi*ARF1FA interact with post-Golgi endosomal machinery, which suggests a link between cortical PV populations and early/late endosomes. Both proteins are associated with components of COPI, AP-2, retromer, and the ESCRT subcomplexes, which typically facilitate the formation and maturation of early/recycling endosomes or multivesicular bodies. Although the role of *Giardia* clathrin has diverged away from clathrin-coated vesicle formation and the parasite is devoid of clathrin uncoating factors, vesicle coats are identified at the cortical and bare zone PV populations near the plasma membrane (Marti et al., 2003; Miras et al., 2013; Pipaliya, Santos, et al., 2021; Zumthor et al., 2016). Using methods employed in this analysis, investigations of similar scope have shown PV populations to be marked and

modulated by PIP-binding proteins that contain standard FYVE, PX, and NECAP1 domains. These proteins that are typically associated with early and late endosomes for the recruitment of clathrin and coat-protein complexes likely interact with adaptin and ARF-regulatory components in *Giardia*, as suggested from the findings from this and previous analyses (Cernikova et al., 2020).

Several components of the *Giardia* ARF regulatory system are also associated with these previously identified PIP-binding domains and kinases. Investigations by Cernikova and colleagues demonstrated phenotypic defects and aberrations in PV morphology upon depletion of canonical endosomal phospholipids (*i.e.*, PI(3)P, PI(3,4)P2, and PI(3,4,5)P3) which impaired fluid-phase uptake dynamics (Cernikova et al., 2020). In *Dictyostelium*, arrestins with FYVE-domain architecture were implicated in the selective recruitment of lineage-specific ArfA for early endosomal transport of arrestins via recognition and binding of PI(3)P at the plasma membrane (Guetta et al., 2010). Similarly, in mammalian systems, ARF6 is essential for the activation of PtdIns4P 5'-kinase (PIP5K) and to control cargo sorting of PX domain-containing phospholipase D (PLD) through AP-2 mediated clathrin-dependent endocytosis at the plasma membrane (for a detailed review on the role of PIPs and membrane trafficking, see Krauß & Haucke, 2007). The findings from this study hint at *Gi*ARF1 and *Gi*ARF1A to have roles in endosome-like trafficking processes that occur in synchrony with AP-2, PIP-binding proteins, and kinases.

Finally, all three *Gi*ARFs as well as *Gi*ARFGAP1 interact with the parasite's retromer components (*i.e.*, *Gi*Vps35 and *Gi*Vps26), reflecting another conventional paradigm of ARF-mediated endosomal trafficking to exist in this parasite. In mice fibroblast and HeLa cells, ARF6 regulates recruitment of retromer-mediated trafficking of mannose-6-receptors, where ARF6 deletion mutants exhibit mislocalization of retromer components and cargo away from the TGN pathway, which in turn results in an aberrant tubular morphology of the endosomes (Marquer et al., 2016).

Of note, it is crucial to recognize that for a robust consolidation of these endo-lysosomal associations as direct interactions, reverse co-IP with most or all of the PV-associated enrichments is the critical next step, especially ones that momentarily lack any existing co-IP validation. This will also provide an expanded view into the protein networks that interact upstream or downstream of the ARF regulatory proteins at the PVs. Additionally, only proteins with known PV interactions validated through previous experimental investigations were discussed here. Therefore, it is likely that many more proteins with PV functions that currently remain unelucidated were enriched in these interactomes. Organellar proteomics with PVs would also be a fruitful approach to capture the diversity of endo-lysosomal proteins associated with this compartment. Finally, numerous hypothetical proteins, whose functions currently remain unknown, were also identified in all analyzed proteomes (Online Appendix Table 5.4). *In silico* functional orthology assignment of these hits was attempted using HHPRED, however to little success. Hypothetical proteins in *Giardia* have the potential for novel discoveries and serve as an avenue to illuminate unique parasite biology that may be veiled. Successful examples wherein investigations with hypothetical proteins that have led to the discovery of exciting trafficking processes in *Giardia* is the characterization of a clathrin light chain analog that functions in conjunction with clathrin heavy chain at the peripheral vacuoles to mediate PV-PM

fusion dynamics (Zumthor et al., 2016; Santos et al., in preparation). Therefore, functions of select hypothetical proteins enriched in the interactomes of the giardial ARF regulatory system proteins should also be explored through molecular investigations of similar scope.

5.4.3 Possibility of ARF regulation in a cascade-like mechanism in crosstalk with the giardial Rab GTPases

One of the intriguing findings from this investigation was the association of *Gi*ARFGAP1 with *Gi*ARF1FA and AGFG, where *Gi*ARFGAP1 may potentiate GDP to GTP exchange on this ARF in part with AGFG effector functions that is otherwise proposed to lack an intrinsic ARF GAP activity (Schlacht et al., 2013). Unexpectedly, our proteomics investigation also identified *Giardia* ARF and regulators to interact with trafficking-associated Rabs consistently. This suggests possible signaling-cascade recruitment of Rabs (or vice versa) and crosstalk between the two pathways to ensure the fidelity of these trafficking processes, which has been the case in mammalian systems. For example, mammalian ARF4 requires associations with GTPase activating protein, ASAP1, and Rab11 at the *trans*-Golgi network, which results in cascade recruitment of other Rabs, and ARF- and Rab-interacting effectors (e.g., Rab8 and FIP3) to provide specificity for ciliary trafficking of rhodopsin (Deretic, 2013). Emerging lines of evidence in parasite endomembrane systems have implicated a role of ARF-Rab cross talk as well. In *Plasmodium*, late endosomal Rab5 and Golgi-associated Rab1 regulates the trafficking of N- myristoylated adenylate kinase 2 (*Pf*AK2) and transmembrane protein Rifin to the parasite-induced parasitophorous vacuole. This process relies on sorting these cargoes by *Pf*ARF1 at the parasite ER-exit sites, followed by recruitment of Rab5 and Rab1 (Taku et al., 2021). Other glimpses of crosstalk between these pathways are evident throughout the eukaryotic tree of life in the form of domain fusions. The most prominent example is the discovery of alveolate and haptophyte-specific TBC-Sec7 (TBS) ARF GEF, which contains a RabGAP TBC-N domain (Mouratou et al., 2005; Pipaliya et al., 2019).

Finally, ARF GAPs and their roles as effectors within the ARF-Rab cascade are also not unprecedented. For example, ASAP1 is suggested to have indirect effector functions towards GTP-bound ARF6 for downstream GAP activity towards mammalian ARF1 and ARF5 for trafficking of paxillin to invadopodia produced by breast tumor cells for their migration and systemic invasion (Sabe et al., 2006). Another example includes effector activity towards Rabs, such as Rab35, which recruits ACAP2 at recycling endosomes and during phagocytosis to inactivate ARF6 for phagosome closure. Numerous studies continue to highlight the effector functions of ARF GAPs towards both ARFs and Rabs (for a comprehensive review on this topic, see the following reviews: Kjos et al., 2018; Spang et al., 2010; Tanna et al., 2019). Therefore, it is not unlikely that similar regulatory mechanisms exist within the giardial ARF/Rab pathways, and results from this investigation are likely providing a glimpse into a much more convoluted cargo trafficking regulation that may be at interplay within this parasite and eukaryotes in general.

5.5 Conclusions

Overall, this investigation shed new light on the roles of ARF regulatory system proteins at *Giardia* trophozoites' endomembrane compartments. A combination of proteomics and microscopy analyses elucidated molecular localizations and meaningful protein interactions at the PVs while also providing glimpses into the cascade mechanisms that may exist within this pathway. While the abovementioned scenarios still require detailed biochemical and gene ablation experiments, which are either out of scope for this thesis or currently not viable in *Giardia*, the results from this investigation are still informative as they confirm many of the notions pertaining to general ARF biology and their conservation in this parasite. Most importantly, these findings robustly allow for functional inference of the peripheral vacuoles as endo-lysosomal compartments and directly build on findings with vesicle coat investigations previously put forward by us and others in this field.

CHAPTER 6

Population-level survey maps differences in the encoded complement of vesicle formation machinery between the two human-infecting *Giardia intestinalis* assemblages

6.1 Introduction

Although morphologically identical, multi-locus sequence genotyping has sub-divided *Giardia intestinalis* into eight distinct assemblages (*i.e.*, genotypes), A through H, with broad animal host tropism, including humans (Heyworth, 2016). Of these, A and B have the greatest host range and are the predominant assemblages that infect humans to cause human giardiasis. Despite similarities in their ability to cause human disease, phylogenetic analyses place assemblage A closely related to the cattle-infecting assemblage E and cat assemblage F (see Figure 1.3 from Chapter 1). Meanwhile, assemblage B is closer in relation to assemblage G, which infects murine hosts such as mice and rats (Cacciò et al., 2008). This suggests that zoonotic transmission into humans has occurred twice and through independent convergence.

Advancements in genome sequencing technologies have resulted in an abundance of new genomic data from numerous human and animal-infecting *Giardia* isolates for comparative analyses. Of the human-infecting isolates, assemblage A, isolate WB was the first available genome and lent tremendous insights into molecular-level losses and sequence divergence in this parasite compared to other eukaryotes (Morrison et al., 2007). Since then, other assemblage A isolates, such as AS175 and AS98, as well as assemblage B isolates, GS, GS_B, and BAH15c1, have also been sequenced and made available (Adam et al., 2013; Franzén et al., 2009; Wielinga et al., 2015; Xu et al., 2020). Recent advances in long-read sequencing technologies have also allowed cost-efficient improvements of these previous assemblies (Pollo et al., 2020; Xu et al., 2020). Other animal-infecting isolates from dog assemblages, C and D, and cattle assemblage E have also been sequenced and assembled, which have shed light on the molecular differences between the various *Giardia* assemblages (Jerlström-Hultqvist et al., 2010; Kooyman et al., 2019). Investigating these additional animal isolates is necessary due to their implications in the potential for anthroozoonotic transmission between humans and domesticated animals (Fantinatti et al., 2016).

Fundamental biological and genetic differences have been enumerated between the two human-infecting strains. From an *in vitro* standpoint, compared to assemblage A, assemblage B isolates are slow-growing in axenic cultures and have lower transfection efficiency with episomally or stably integrated plasmids (Singer et al., 1998). On the other hand, assemblage A isolates are challenging to study in mice models where infection cannot be established or is cleared immediately compared to assemblage B

isolates, which can readily infect and cause disease (Byrd et al., 1994). Although structurally otherwise identical, various cytogenetic differences are also present, such as the number of chromosomes per nuclei and their sizes (Adam, 1992; Tůmová et al., 2007). Although generally lower compared to other polyploid eukaryotes, genetic allelic heterozygosity (ASH) comparisons between assemblages A and B have shown variances in the levels between the two. Assemblage A isolate WB is markedly lower in overall ASH (<0.01%) compared to assemblage B isolate, GS, which has a much higher genomic allelic divergence (0.5%) (Ankarklev et al., 2010).

From a clinical standpoint, differences in human Giardiasis outcomes are also variable. Several cross-sectional clinical studies have been conducted in numerous countries (*i.e.*, United Kingdom, Scotland, Sweden, US, Saudi Arabia, Egypt, and more) to associate whether specific strain can be attributed to a higher probability of developing symptomatic or chronic disease outcomes (Cacciò & Ryan, 2008; Ferguson et al., 2020; Minetti et al., 2015a; Minetti et al., 2015b; El Basha et al., 2016). These surveys predominantly associated *ca.* 60-65% of the infections were due to assemblage B, while the remaining 30-35% of the cases seem to be caused by assemblage A isolates. Additionally, the latter was correlated to asymptomatic or milder disease outcomes, whereas infections with assemblage B isolates had a greater likelihood of causing a prolonged or refractory disease characterized by severe diarrheal and other gastroenteritis symptoms (Lebbad et al., 2011). Patients infected with assemblage B were also less likely to respond to first-line treatments, metronidazole and albendazole (Lalle & Hanevik, 2018; Lecová et al., 2018).

Molecular-level variances underpin disease and biological differences. Availability of whole-genome sequencing data from numerous assemblage A and B isolates enabled large-scale and multi-family comparative genomic investigation, where growing evidence suggests considerable inter-assemblage variability at the genetic level. These are within multiple *Giardia*-specific virulence gene families, such as the cysteine-rich variant-specific surface proteins (VSPs) that are expressed by the trophozoites during a gut infection for parasite antigenic variation in order to modulate and evade the host immune cells (Prucca & Lujan, 2009). For example, isolates belonging to assemblage B encode almost 700 different VSPs compared to their assemblage A counterparts, which have considerably lower VSP repertoires (*i.e.*, 190 in WB isolate and 250 in the DH isolate) (Ankarklev et al., 2010; Jerlström-Hultqvist et al., 2010).

Other differences in the encoded protein machinery, particularly those belonging to the membrane trafficking system pathways, are also evident. In the preceding chapters, comparative genomics and phylogenetic investigations were performed to determine the complement and evolution of the ESCRT proteins, vesicle coats, and the ARF regulatory system proteins. Within each of these systems, consistent patterns of inter-assemblage variabilities were identified that would have substantial implications on the molecular intricacies underpinning the processes occurring at the parasite endomembrane organelles. Considering these findings, a fine-tuned investigation at a population level was necessary to elucidate and corroborate the degree of variability that exists between these systems, especially between the two human-

infecting assemblages. Large-scale genomic data from assemblage A and B isolates were essential to evaluate these observed trends, no matter how subtle or prominent. Recently, the British Columbia Centre for Disease Control Public Health Laboratory (BCCDC PHL) undertook large-scale Illumina MiSeq genome sequencing with 89 *Giardia intestinalis* isolates belonging to either assemblage A or B for subsequent genome assembly using SPAdes (Tsui et al., 2018). Archived isolates were previously collected from fecal samples, surface waters, and beavers during periods of sporadic and regional Giardiasis outbreaks that occurred in British Columbia (Canada) between the years 1989 and 1995 (Prystajecy et al., 2015). A combination of PCR and whole-genome sequencing classified these outbreak-associated *Giardia* isolates within assemblage AI, AII, and B (Prystajecy et al., 2015; Tsui et al., 2018). Although raw paired-end short reads from this study are published, fully assembled genomes are not yet publicly available. Nonetheless, these sequencing reads still serve as a valuable resource for exploratory comparative analyses within these isolates, but unfortunately, cannot be used standalone.

Therefore, for the scope of this chapter, *de novo* genome assembly was first performed using the BCCDC PHL paired-end short-read data from each isolate to yield full-length genomes. These were then used to conduct a population-level survey of the giardial vesicle formation machinery for a telescopic view into notions pertaining to assemblage A and B variabilities identified in the previous chapters. Additionally, these analyses also aimed to determine whether there are any intra-assemblage isolate-level variations. Findings from the literature and this thesis with the pan-global isolates have provided a glimpse into the possible existence of both of these postulations, and hence subject to further testing in this chapter.

In this chapter, Maryland Super-Read Celera Assembler (MaSuRCA) was opted to assemble the tetraploid *Giardia* genomes. Compared to other short-read genome assemblers, MaSuRCA is a ploidy-aware software that combines the efficiency of classical *de Bruijn* graphing and the accuracy of overlap consensus methods to generate super-reads in order to produce genome assemblies that are comparatively more contiguous than those produced by other short-read assembly methods (Zimin et al., 2013). A *de novo* genome assembly approach was specifically chosen as opposed to reference-mapping in order to eliminate reference-associated bias because the goal of this investigation was to uncover subtle inter-assemblage differences within the vesicle formation machinery (Günther & Nettelblad, 2019). Comparative analyses were performed using these newly assembled genomes to survey the complement and differences in the giardial vesicle formation machinery of the two human-infecting *Giardia* assemblages. Although outside of the scope for this thesis, the increased availability of genomes belonging to isolates from both assemblages will be valuable resources to other *Giardia* biologists to investigate their systems of interest and better address other genetic variabilities (*i.e.*, single nucleotide polymorphisms and allelic heterozygosity). These concerted efforts will lead to a better understanding of whether *Giardia intestinalis* is a lineage composed of multiple species.

6.2 Materials and Methods

6.2.1 Paired-end read information and data retrieval

De novo genome assembly analyses with the BCCDC PHL isolates of *Giardia intestinalis* assemblage A and B were performed using the previously sequenced and archived paired-end inserts generated using the Illumina MiSeq next-generation sequencing technology. Cysts belonging to each isolate were previously archived by the BCCDC PHL that were initially obtained from surface water, beaver, and human fecal samples (Prystajecy et al., 2015; Tsui et al., 2018). Detailed geographical sources for isolate retrieval and assemblage classification were provided through Table 1 and Supplementary Table S1 in the previously published study (Tsui et al., 2018). Tsui and colleagues performed genome sequencing on each isolate by first purifying the cysts for subsequent *in vitro* and *in vivo* trophozoite excystation in separate TYI-S-33 supplemented media and gerbils, respectively (Tsui et al., 2018). The newly excysted trophozoites were further cultured and sub-cultured as separate parasite lines for extraction of genomic DNA (gDNA) using a QIAmp DNA Minikit (Qiagen, Mississauga, ON, Canada). Using the purified gDNA, paired-end DNA libraries were prepared using the Nextera XT DNA kit (Illumina, San Diego, CA, USA). Six to ten samples were pooled per run in order to generate 250 bp paired-end inserts.

Raw sequencing reads from a single Illumina MiSeq run belonging to each biosample (*i.e.*, isolate) were deposited at the National Centre for Biotechnology Information (NCBI) under the BioProject ID PRJNA280606. For this investigation, libraries corresponding to this project accession were batch retrieved through the European Nucleotide Archive (The European Molecular Biology Laboratory, Heidelberg, Germany). Metadata detailing sequencing run statistics and biosample accessions for the 89 isolates is provided through the Online Appendix Table 6.1 as well as in the previous study (Tsui et al., 2018).

6.2.2 Read quality assessment, taxonomic classification, and de-contamination using Kraken2

Prior to assembly, the overall quality of the paired-end inserts was assessed using FastQC (Babraham Bioinformatics) to generate HTML reports detailing sequence and base quality (*i.e.*, Phred scores), sequence length distribution, percent GC content, adaptor content, and levels of overrepresented and de-duplicated sequences (Andrews, 2010). Although contaminating Nextera transposase adapter sequences and poor-quality base positions (*i.e.*, reads with Phred score <20) were present across all datasets, trimming was not performed prior to assembly, as per instructions detailed within the MaSuRCA assembler workflow (Zimin et al., 2013). Instead, quality control and read error correction were performed as part of the assembly pipeline by the in-built tool, Quorum (Marçais et al., 2015; Zimin et al., 2013).

Although reads were not trimmed or error-corrected, all datasets were still inspected for non-*Giardia* contaminating sequences. To do so, paired-end reads belonging to each biosample were subject to taxonomic classification using the metagenomic classification software, Kraken2, and against a custom-built database termed “kraken2-microbial-soup-with-giardia” for k-mer matching against query sequences (Wood et al., 2019). The database structure consisting of three human-readable directories (hash.k2d, opts.k2d, and taxo.k2d) were built by retrieving taxonomic information and full-length genomes

corresponding to archaeal, bacterial, viral, human, fungal, protozoa, and vector contaminant (UniVec_Core) datasets from the NCBI Reference Sequence Database (RefSeq). In addition, genomes belonging to *Giardia muris*, *Giardia intestinalis* EP15, *G. intestinalis* BAH15c1, *G. intestinalis* BGS, *G. intestinalis* BGS_B, *G. intestinalis* AWB, *G. intestinalis* AS175, *G. intestinalis* ADH, *G. intestinalis* assemblage C pooled cysts, and *G. intestinalis* assemblage D pooled cysts were also added to the library using the kraken2-build option. K-mers from each read dataset were classified against these genomes belonging to different taxonomic groups in order to generate hierarchical output reports detailing sequence classification (Wood et al., 2019). Sequences that remained unclassified or were assigned to taxa other than the *Giardia* genus (NCBI taxid ID: 5741) were considered contaminants. Datasets were flagged if 20% or higher number of the total reads were not characterized as *Giardia*. These biosamples were removed from analyses and not subject to downstream assembly process or comparative genomics. For the remainder datasets with low contamination levels, decontamination was performed using the Python script, `extract_kraken_reads.py`, available through KrakenTools (<https://github.com/jenniferlu717/KrakenTools>; Wood et al., 2019). Post-decontamination Kraken2 output was visualized using KronaTools (Ondov et al., 2011).

Kraken2 database building and read taxonomic classification was performed on Compute Canada high-performance computing clusters using the pre-installed Kraken2 v. 2.0.8 AVX2 module (www.computeCanada.ca). Database building and read taxonomic classification for all 89 isolates were parallel processed by specifying parameters for both steps, as detailed within the Kraken2 manual, using a job array script on Compute Canada's Simple Linux Utility for Research Management (SLURM) scheduling system (https://slurm.schedmd.com/job_array.html). Kraken2 taxonomic classification output of all paired-end read datasets is available through Online Appendix Table 6.2. Commands for Kraken2 custom database building and taxonomic classification have been made available through Online Appendix File 6.1.

6.2.3 MaSuRCA de novo genome assembly

Kraken2 analyses flagged three datasets to be highly contaminated with bacterial sequences and three with a mixed-assemblage taxonomic assignment. These were not subject to genome assembly. The remaining 83 paired-end read datasets were kept for *de novo* genome assembly after taxonomy-based decontamination. To do so, MaSuRCA v. 3.3.0, which uses a modified version of the Celera Assembler with the Best Overlap Graph (CABOG) assembler, was used to generate 'super-reads' (also termed k-unitigs) from the initial paired-end inserts by *de Bruijn* graphing (Miller et al., 2008; Zimin et al., 2013). The program then used an overlap consensus layout algorithm to combine the super-reads into longer contig sequences (Miller et al., 2008; Zimin et al., 2013). In general, the assembly pipeline consisted of the following steps: 1) read error-correction and pre-processing using QuorUM, 2) creation of k-unitigs, 3) super-read generation from paired-end reads, 4) contigging and scaffolding using CABOG and, 4) scaffold gap-filling (Zimin et al., 2013) (Figure 6.1).

Assembly parameters were specified through a configuration shell script. Briefly, read type specification was set to paired-end, 250 bp in size, and a standard deviation of 38 bp. The optimal k-mer size was set to auto for *de Bruijn* graphing, which estimated $K > 127$ bp for all datasets based on the input read and GC content. MaSuRCA also uses Jellyfish v. 3.2.4 to generate a hash table for efficient k-mer counting to estimate genome size (Marçais & Kingsford, 2011). Jellyfish hash size was required to be set to 20 times the predicted genome size (Zimin et al., 2013). Because both assemblage A and B genomes were expected to be between 10 to 13 Mbp, this parameter was set to 200,000,000. Since *de novo* genome assembly is computationally resource-intensive, additional parameters detailing multiple CPU utility (*i.e.*, n-threads=16) on a SLURM grid engine were also specified. Parallel genome assembly for all isolates was performed in batch by providing configuration parameters on a job array script and performed on Compute Canada clusters using the pre-installed MaSuRCA v. 3.3.0 module. Final nucleotide contig and scaffold sequences in a FASTA output format were generated for each genome and made available through the Figshare repository (<https://doi.org/10.6084/m9.figshare.12668813>). MaSuRCA configuration script has been made available as Online Appendix File 6.2.

6.2.4 Genome assembly completeness and contiguity evaluations

The resulting nucleotide assemblies were subject to post-assembly quality assessments by computing genome contiguity and completeness statistics. Both of these are standard measures of determining if the *de novo* assembly process was minimally error-prone and has yielded assemblies that are biologically meaningful in size (Del Angel et al., 2018). Contiguity was assessed using the N50 metric to evaluate the degree of genome fragmentation (Earl et al., 2011; Vezzi et al., 2012). In general, the N50 is a weighted median statistic of the shortest contig sequence length that is necessary to cover 50% of the genome. Theoretically, a larger N50 score suggests a more contiguous assembly, while lower scores indicate fragmentation. N50 was determined for assembled contigs belonging to all 83 isolates by setting various sequence length thresholds (*i.e.*, 0 bp, 1000 bp, 5000 bp, 10,000 bp, 25,000 bp, and 50,000 bp). These were generated using the Quality Assessment Tool for Genome Assemblies (QUAST) v. 5.0.2, which also determined other contig statistics such as total length in base pairs and the final percent GC for each assembly (Gurevich et al., 2013). This detailed report is made available through Online Appendix Table 6.3.

In addition to genome contiguity, to assess gene set completeness and report on the number of evolutionarily conserved near-universal single-copy orthologs, Benchmarking Universal Single-Copy Orthologs (BUSCO) v. 4. 1.1 was used as a tool for a translated Hidden Markov Model-based searching against the OrthoDB v.10 (Kriventseva et al., 2019; Simão et al., 2015). Run parameters for BUSCO analyses were specified through a configuration shell script. Briefly, the lineage dataset selection was set to eukaryote_odb10, analysis mode set to 'genome,' and TBLASTN e-value cut-off set to the default 1×10^{-3} . All other BUSCO parameters, including those for the in-built AUGUSTUS v. 3.2.3 gene prediction software, were kept to default. BUSCO output scores for each assembly were summarized in the following

notation: C: % complete [D: % duplicated], F: % fragmented, M: % missing, N: number of genes used for analyses (Simão et al., 2015). All BUSCO output scores were plotted and visualized using the ggplot2 library available through the R Tidyverse package (Wickham et al., 2019).

QUAST v. 5.0.2. was installed through the conda package manager (Bioconda) and run locally. BUSCO v. 4.1.1 and OrthoDB v.10 were installed in a VirtualEnv python/3.7.4 virtual environment on Compute Canada. Necessary BUSCO dependencies such as DendroPy and SEP were also installed. Ggplot2 library was accessed by installing the entire Tidyverse package locally (Wickham, 2016; Wickham et al., 2019). Data visualization specifications were provided through an R script (R Core Team, 2020).

6.2.5 Reference- mapped gene predictions and functional annotation using Liftoff and TBLASTN

The nucleotide contigs from all isolates were used to predict gene features and annotations to screen for protein orthologs. To do so, the newly available Liftoff v.1.5.1 was used for accurate annotation mapping from closely related reference genomes (Shumate & Salzberg, 2021). Liftoff was preferred for two reasons instead of opting for other popular *ab initio* programs such as AUGUSTUS. First, AUGUSTUS requires a test dataset to re-train the software from its default settings currently optimized for model eukaryotes. This is a challenging and a computationally resource-intensive process, with the potential for the end output to have many inaccuracies. The second reason is that the core gene conservation between *Giardia intestinalis* species' is predicted to be approximately 91%. Because the aim was to screen subtle molecular differences, the trade-off between an accurate and efficient estimation of protein-coding genes was higher with Liftoff than with AUGUSTUS.

Liftoff performs gene mapping through pair-wise alignment of nucleotide coding sequences using the Minimap2 aligner and aims to preserve exon, transcript, and overall gene sequence structure (Li, 2018; Shumate & Salzberg, 2021). Of the currently available assemblage A and B genomes, those belonging to *Giardia intestinalis* assemblage A, isolate WB (UU_20 assembly) and assemblage B, isolate BGS (GL50801 assembly) have been richly annotated (Adam et al., 2013; Xu et al., 2020). Additionally, gene annotations for both are available through general feature format (GFF) files that detail the coordinates of the predicted genes with their specific intron and exon boundaries, confidence scores for each prediction, specific strand features, and orthology assignments. Therefore, UU_20 and GL50801 assemblies were used as references for gene prediction in the BCCDC PHL A and B genomes, respectively. Input and reference genomes were first indexed and aligned using Minimap2 v. 2.17 using the following parameters: -a -eqx -end-bonus 5 -N 50 -p 0.5 (Li, 2018). Genes were successfully aligned if the coverage along the coding sequence (CDS) met a threshold of 50% or higher in sequence similarity. They were used to generate output GFF files detailing coordinates, accessions of the mapped ORFs, and accompanying gene features. Unmapped gene accessions were parsed out in a separate text file. Liftoff v. 1.5.1 was installed via Python Package Index (PyPI), meanwhile, Minimap2 v. 2.17 was directly cloned from the project's GitHub repository using git (<https://github.com/lh3/minimap2>). Both tools were run locally, and program usage parameters were specified as per Liftoff instructions. GFF and unmapped features text files for all

assembled BCCDC PHL genomes have been made available through the Figshare repository (<https://doi.org/10.6084/m9.figshare.13567655.v1>).

These resulting GFF files were then used for presence/absence analyses of the vesicle formation machinery, previously curated through comprehensive comparative genomics and phylogenetics in the preceding chapters. Specifically, WB and GS accessions corresponding to subunits of the giardial ESCRT machinery, adaptins, retromer, COPII, COPI, clathrin, and the ARF regulatory system proteins, were searched into the newly generated GFF files. To consolidate these preliminary hits, assign orthology to any unmapped genes, and cross-identify and validate assemblage-specific orthologs/paralogs, TBLASTN searches, with an e-value threshold set to 0.01, were also performed. Orthologs of the identified vesicle formation machinery from *Giardia intestinalis* ADH, AWB, BGS, and BGS_B were used as queries for homology searching into the contigs and scaffolds of all newly assembled genomes. Results of this survey are summarized as tile-plots, but specific scaffold locations from the TBLASTN hits have been made available through Online Appendix Table 6.4.

6.3 Results

6.3.1 Genome contiguity and completeness analyses yield uniform results across all isolates

Previously, Tsui and colleagues performed genome sequencing and assembly with 89 *Giardia intestinalis* assemblage A and B isolates archived at the BCCDC PHL to evaluate specific disease transmission dynamics underpinning the 1980s Giardiasis outbreak in British Columbia (Tsui et al., 2018). Although the sequenced short paired-end reads from this analysis are publicly available, the assembled genomes are not. Therefore, it was necessary to perform *de novo* genome assembly prior to comparative genomic analyses with the vesicle formation machinery, as per the summarized workflow (Figure 6.1). To do so, read taxonomic classification was performed first to assess contamination and remove sequences that did not correspond to *Giardia* (Figure 6.1). Of the 89 initially retrieved datasets, six were removed from final genome analyses, either due to significant contamination or if the samples were characterized as ‘mixed’ assemblages. In the latter case, according to Tsui et al., a mixed assemblage assignment was denoted if the sequencing reads mapped equally to both *Giardia* assemblage A and B. This does not imply a genetically hybrid strain but is instead a by-product of gDNA contamination or its isolation from mixed cultures containing trophozoites belonging to both strains. Therefore, these data were treated as metagenomic. Because this investigation aimed to elucidate nuanced molecular-level differences between the two assemblages, samples with ‘mixed’ assemblage classification were also removed from analyses. This resulted in a final 83 datasets used for downstream analyses, 42 of which were classified to assemblage A and 41 to assemblage B (Online Appendix Table 6.5).

First, it was necessary to ensure that the genomes under investigation were biologically relevant to increase confidence in gene presence and absence conclusions. Therefore, to technically validate the assembly process, completeness and contiguity of the output contigs were first evaluated (Figure 6.1). Genome contiguity was determined by generating N50 scores for the individual assemblies, then used to

calculate the average and median for all assemblies. Because the N50 is a metric of genome contiguity, a lower score corresponded to a more fragmented assembly containing a greater number of short contigs, and therefore, the two were inversely reciprocal. Genome fragmentation can result from contamination, polyploidy, genetic polymorphism, and allelic heterogeneity. In the case of *Giardia*, the parasite's genomic tetraploidy has historically hampered efforts to yield a contiguous genome assembly, mainly because ploidy-aware assembly tools or long-read sequencing technology were not yet available. In this study, the way to circumvent this issue was by performing genome assembly using a ploidy-aware assembler, MaSuRCA (Zimin et al., 2013).

Both N50 scores and contig sizes for all assemblies were determined using QUAST. The results of this analysis calculated an average N50 of 70,543 bp and a median of 74,269 bp across all assemblies (Online appendix Table 6.5). Genomes were also assembled in an average of 382 contigs that were \geq 1kb in size (Online Appendix Table 6.5). In addition to genome contiguity, completeness was also assessed by determining the number of evolutionarily conserved single-copy orthologs of eukaryotic proteins. Genome completeness was evaluated by generating BUSCO scores, which reported the percent of conserved eukaryotic genes out of the 255 proteins curated within the eukaryotic OrthoDB v. 10 database (eukaryote_odb10) (Figure 6.1). Overall, the percent BUSCO of complete plus fragmented genes was approximately 25 to 30% across all assemblies (Online Appendix Figure 6.1). While these are much lower than what is typically expected of a eukaryotic genome from animal, plant, or fungal lineages (*i.e.*, 85-95%), low BUSCO scores in *Giardia* are a drawback of limited inclusion of diverse protist lineages within the eukaryote_odb10 database. Additionally, in general, metamonad lineages are highly divergent in their sequences and ancestrally lack many of the proteins present in model eukaryotes, and therefore all suffer from poor BUSCO scores (Karnkowska et al., 2019; Salas-Leiva et al., 2021; Tanifuji et al., 2018). Instead of using these as absolute measures, BUSCO scores were evaluated against those published for other *Giardia* assemblies. The short-read reference genome of assemblage A, isolate WB was previously determined to have a BUSCO score of C:23.9%[S:0%, D:23.9%], F:5.9%, M:70.2%, n=255, while the short-read reference genome from assemblage B, isolate GS has a BUSCO score of C:23.1%[S:22.7%,D:0.4%], F:7.1%, M:69.8%, n:255 (<https://www.uniprot.org/proteomes/UP000001548>; <https://www.uniprot.org/proteomes/UP000002488>; Franzén et al., 2009; Morrison et al., 2007). Therefore, the BCCDC PHL assemblies are comparable, if not slightly better, in their BUSCO scores to those of the previously published short-read reference genomes belonging to A and B assemblages.

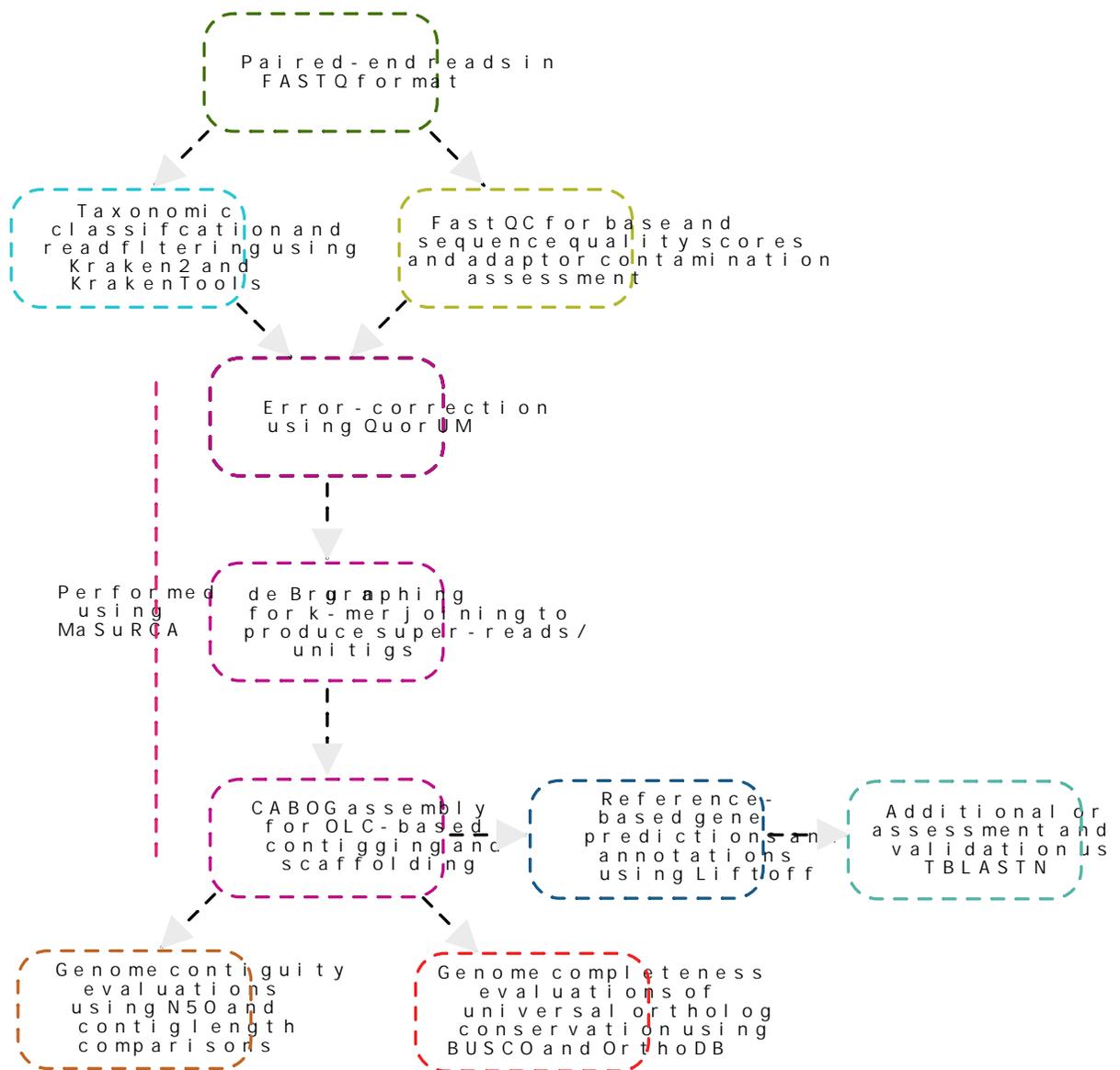


Figure 6.1. Workflow for *de novo* assembly and reference-based gene predictions. Steps included read-quality assessment and filtering using Kraken2 and FastQC, followed by *de novo* genome assembly using MaSuRCA. Assembled contigs and scaffolds were used to determine contiguity and completeness using metrics such as N50, the total number of contigs, and BUSCO scores. The resulting contigs were used for reference-based genome annotation using LiftOff for orthology mapping of genes and presence/absence analyses of the vesicle formation proteins. False negatives and existing orthology assignments were cross-validated using TBLASTN.

6.3.2 Overall GC content and genome sizes are comparable to other isolates of assemblages A and B

Aside from genome completeness and contiguity, the percent GC content (%GC) and genome size metrics for the new *Giardia* assemblies were also determined.

As described earlier, of the 83 isolates assembled and kept for analyses, 42 belonged to assemblage AI or All, and 41 to assemblage B. Previous genomic investigations with assemblage AI and All isolates (*i.e.*, DH, WB, and AS175) approximated these to be 48% GC rich (Adam et al., 2013; Ankarklev et al., 2015; Morrison et al., 2007). A similar trend was observed across the 42 BCCDC PHL *G. intestinalis* assemblage A isolates, where the %GC across all isolates ranged between 47.8 and 48.9 (Figure 6.2B; Online Appendix Tables 6.3 and 6.5). A few outliers had slightly higher scores (*i.e.*, 49 to 49.96%), but overall, mean and median were determined to be 48.5% and 48.36%, respectively (Figure 6.2C; Online Appendix Table 6.5). Similar to assemblage A, previously investigated assemblage B isolates also had %GC ranging between 47 and 49 (*i.e.*, BGS, BGS_B, and BAH15c1) (Adam et al., 2013; Franzén et al., 2009; Wielinga et al., 2015). The results from this study are identical to those values, wherein the 41 BCCDC B isolates had 47 to 49 %GC, with a mean of 48.7% and a median of 48.9% (Figure 6.2A and C; Online Appendix Table 6.5).

Genome sizes were also evaluated and compared against the previously published *Giardia* genomes. In the literature, inter-assemblage variances within the overall genome sizes of assemblage A and assemblage B isolates have been noted, wherein assemblage AI and All isolates (*i.e.*, WB, DH, AS175, AS98, ISS17, and ZX15) generally ranged between 10.2 Mbp and 11.7 Mbp (Adam et al., 2013; Ankarklev et al., 2015; Franzén et al., 2009; Morrison et al., 2007). In contrast, assemblage B isolates (*i.e.*, GS, GS_B, and GS/M clone H7) were comparatively larger, ranging between *ca.* 11 Mbp to 13 Mbp (Adam et al., 2013; Ankarklev et al., 2015; Franzén et al., 2009; Morrison et al., 2007). This trend was also consistent in this investigation. Apart from one outlier, the genome size range for all assemblage A isolates ranged between *ca.* 10.6 Mbp and 11.9 Mbp, with an average of 11.0 Mbp and a median of 10.9 Mbp (Figure 6.3B and Figure 6.3C). Assemblage B isolates, on the other, were comparatively larger and ranged between 10.8 Mbp and 13.7 Mbp, with an average of 12 Mbp and a median of 11.9 Mbp (Figure 6.3A and C). The reference isolate GS has a genome size of 12 Mbp, whereas GS_B is 13 Mbp in size. Therefore, assemblies from this study are similar in size compared to the previously published isolates from the same assemblage (Figure 6.3A and C).

The GC content and genome size's collective examination yielded identical values as the previously published pan-global isolates. These findings additionally validate assembly correctness and corroborate the previously observed trends at a greater population level. Although the two assemblages are similar in their overall %GC content, isolates belonging to assemblage B are consistently larger than A.

Figure 6.2. Percent GC content comparisons between assemblage A and B BCCDC PHL genome assemblies. (A) depicts %GC for the 41 isolates belonging to assemblage B. These ranged between a minimum of 47.7% and a maximum of 49.3%. (B) depicts the %GC content for assemblage A isolates, which, like assemblage B, ranged between 47.8 and 49.7. (C) is a box-plot depiction of the percent ranges and calculated means for GC content for assemblage A and B isolates, where a similar overall %GC was present for both (*i.e.*, ca. 48%).

Figure 6.3. Genome size comparisons between assemblage A and B BCCDC PHL genome assemblies. (A) depicts genome sizes for the 41 isolates belonging to assemblage B. Genome sizes are relatively uniform across all isolates and range between 10.9 Mbp and 13.7Mbp. (B) represents genome sizes for the 42 isolates belonging to assemblage A which range between 10.6 Mbp and 12.3 Mbp. (C) is a box-plot depiction of the range and calculated mean of the genome sizes for all isolates belonging to each assemblage. It is evident that assemblage A genomes are comparatively smaller in size than assemblage B genomes.

6.3.3 Gene predictions and functional annotations suggests close similarity to reference genomes

To comprehend the trends previously observed in the repertoire of the endo-lysosomal vesicle formation machinery, a fine-grained population-level survey of these proteins was undertaken with these 83 isolates. To do this in a computationally and time-efficient manner, the newly available Liftoff was used, which accurately maps gene features and annotations from closely related species of organisms. This was done by providing the assemblage A, isolate WB (UU_20 assembly) and assemblage B, isolate GS (GL50801 assembly) genomes as references for mapping. Many of the annotations have been curated based on experimental evidence (*i.e.*, molecular protein characterizations, transcriptomics, and proteomics) and are regularly updated by the *Giardia* research community members. To supplement and cross-validate Liftoff annotations, TBLASTN analyses were also performed for genes of interest to eliminate false-negative absence artifacts due to improper or missed gene mapping.

As per recent estimations, the newly updated genome of *Giardia intestinalis* AWB (UU_20 assembly) is predicted to encode 4,963 protein-coding genes and 85 non-coding RNA (ncRNAs) genes, yielding a total of 5048 genes (Xu et al., 2020). From this total, Liftoff mapped an average of 4688 genes onto the BCCDC PHL assemblage A genomes, with approximately 72 remaining unmapped (Online Appendix Table 6.5). Similarly, albeit slightly lower, the total number of protein-coding and ncRNAs in the assemblage B reference GS isolate (GL50801 assembly) is estimated at 4470 and 92, respectively, resulting in a total of 4562 predicted genes. An average of 4481 genes were mapped across the BCCDC PHL assemblage B genomes, with 62 remaining unmapped (Online Appendix Table 6.5). Unmapped genes may have resulted from sequence divergence or fragmented orthologs that did not share enough similarity with the target for a Minimap2 cut-off of 50%. The other possibility is that a subset of these unmapped genes are protein repertoires exclusive to the reference WB and GS genomes. Nonetheless, these findings suggest that for both assemblages A and B, approximately 98.2% of the genes and annotations from AWB and BGS genomes were successfully mapped, and therefore, encoded in the genomes of the newly assembled isolates.

6.3.4 ESCRT repertoire varies between assemblage A and B isolates

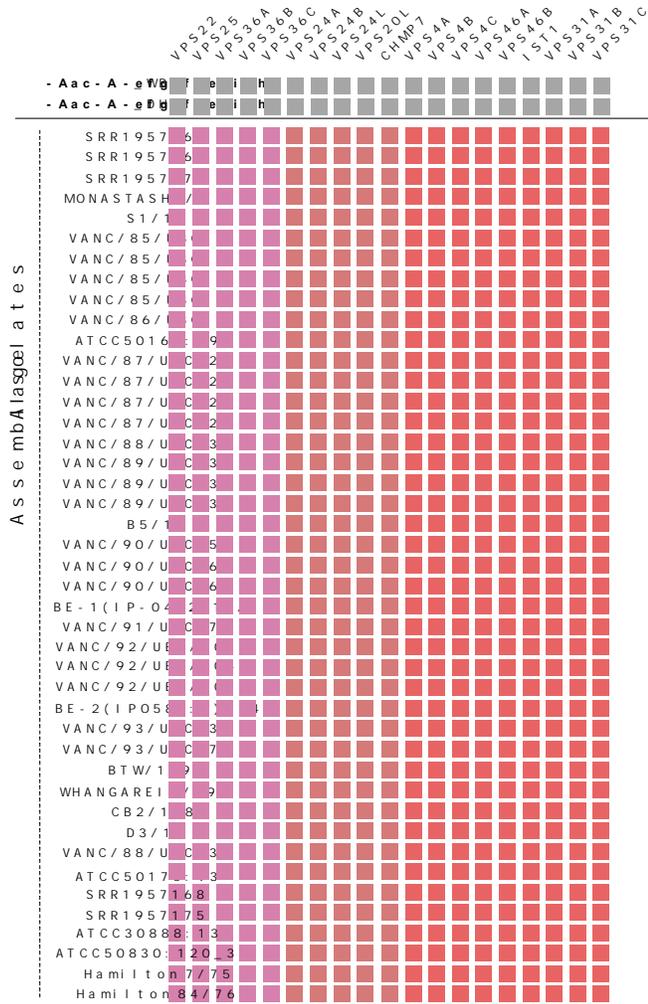
In Chapter 4, comparative genomic and phylogenetic analyses with the late-endosomal ESCRT subcomplexes were performed to trace their evolution across fornicates, including several human-infecting isolates of *Giardia*. Here patterns of universal streamlining within ESCRTs were observed across the entire *Giardia* genus, such as the complete loss of the ESCRTI subcomplex (Pipaliya, Santos, et al., 2021). However, aside from absences, *Giardia*-specific duplication events also occurred to yield several paralogs in the following subunits: ESCRTII-Vps36 (*i.e.*, Vps36A, B, and C), ESCRTIII-Vps24 (*i.e.*, Vps24A and Vps24B), ESCRTIIIA-Vps4 (*i.e.*, Vps4A, B, and C), ESCRTIIIA-Vps46 (*i.e.*, Vps46A and B), and ESCRTIIIA-Vps31 (*i.e.*, Vps31A, B, and C) (Pipaliya, Santos, et al., 2021). Comparisons between various pan-global isolates revealed distinct differences within the repertoire of the individual subunits and the number of paralogs that comprise the giardial ESCRTII, ESCRTIII, and ESCRTIIIA. To better elucidate these inter-

assemblage differences within ESCRTs and to assess if these trends exist at a population level, the repertoire of the giardial ESCRT machinery in the newly assembled BCCDC PHL *Giardia* genomes was reconstructed. This was done by surveying the Liftoff GFF annotations combined with TBLASTN searches into the nucleotide contigs and scaffolds. Overall, it was evident that a pattern of ESCRT machinery retention and loss is identical to the one observed in Chapter 4.

All 42 assemblage A isolates are conserved in their *Giardia* repertoire of the ESCRT complexes without any variabilities at the individual isolate level (Figure 6.4A). Contrary to this, previously noted streamlining exists in several ESCRT machinery components in assemblage B isolates (Figure 6.4B). The most prominent of these was the loss of Vps20L. The survey of 41 new assemblage B isolates confirms that this crucial ESCRTIII component is universally absent across all sampled *Giardia* assemblage B genomes (Figure 6.4B). TBLASTN searches using Vps20L orthologs identified in pan-global assemblage A isolates (*i.e.*, DH, WB, and AS175) were used as queries to rule out false-negative artifacts. Despite this approach, no protein hits resembling Vps20L orthology were identified. Similar trends were observed within one of the three Vps4 paralogs (*i.e.*, Vps4C), which also remained unidentified in the pan-global isolates of assemblage B. Using both Liftoff and TBLASTN, only Vps4A and Vps4B were retrieved across all 41 isolates (Figure 6.4B). Therefore, the absence of Vps20L and Vps4 points towards a global loss of these components in this assemblage.

Altogether, these findings validate previous observations of molecular differences within the endolysosomal ESCRT subunits between the two assemblages and allow for an extension of those conclusions at a greater population level.

A



B

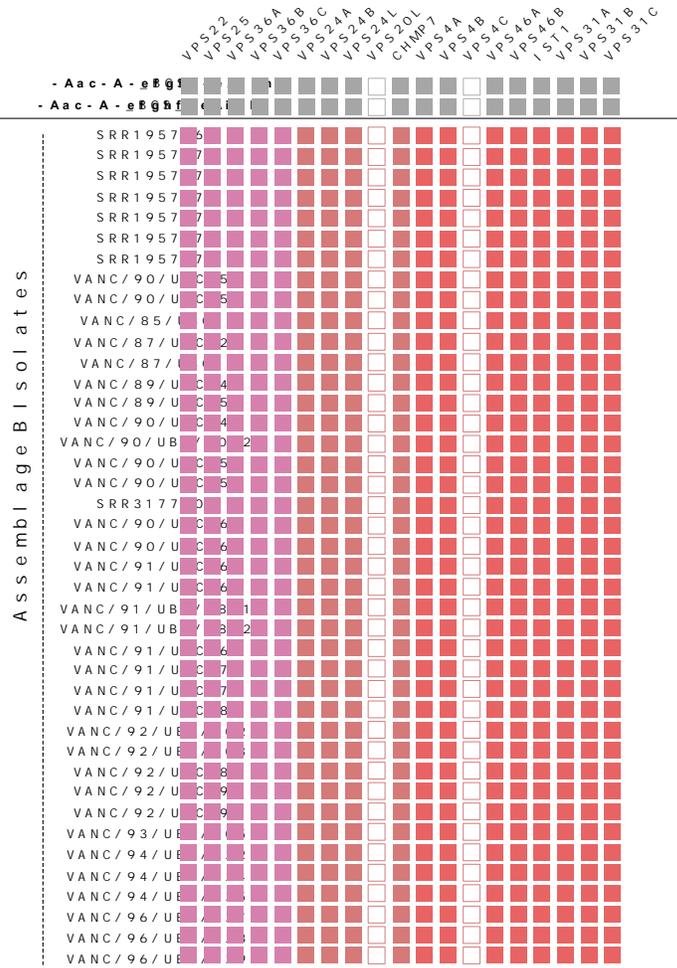


Figure 6.4. Tile-plot depictions of the ESCRT repertoire in the BCCDC PHL isolates. (A) depicts the giardial ESCRT repertoire distribution in the newly assembled BCCDC PHL isolates compared to the two pan-global reference assemblage A isolates, WB (AI) and DH (AII). No absences were identified within any assemblage A isolates. **(B)** shows the distribution of the giardial ESCRT repertoire in the newly assembled BCCDC PHL assemblage B isolates compared to the two pan-global reference assemblage B isolates, BGS and BGS_B. Assemblage-specific losses were identified within the ESCRTIII-Vps20L and ESCRTIII-Vps4C, and are consistent with the findings in the pan-global isolates, indicated in grey.

6.3.5 Vesicle coat complexes follow a similar pattern of inter-assembly molecular differences

Identical to the analyses performed with the ESCRTs, the evolution and the molecular complement of the heterotetrameric complexes (adaptins and COPI), COPII, retromer, and clathrin were recapitulated across fornicates and *Giardia* in Chapter 2. That investigation determined differences to exist within the repertoire of the fornicate-specific Sec24FII paralogs between assemblage A and assemblage B. Specifically, phylogenetics traced secondary loss within one of the two Sec24FII (*i.e.*, Sec24C) in assemblage B. To consolidate those findings as well as to identify any other possible isolate-level differences, the *Giardia* complement of HTACs, COPI, COPII, retromer, and clathrin in the 83 BCCDC PHL A and B isolates were determined.

All 41 assemblage A isolates encoded all three paralogs of the fornicate-specific COPII-Sec24F (*i.e.*, Sec24A, B, and C; Figure 6.5A). On the other hand, indeed, COPII-Sec24C was absent from all BCCDC PHL assemblage B isolates (Figure 6.5B). Once again, like the ESCRTs, these results were corroborated by TBLASTN analyses using Sec24C sequences from assemblage A as queries for homology searching into the genomes of the assemblage B isolates which yielded no additional hits other ones belonging to Sec24A and B (Figure 6.5B). Apart from this, the giardial repertoire of adaptins, COPI, retromer, and clathrin heavy chain were conserved from an inter-assembly standpoint, as was the case previously (Figure 6.5). However, isolate-level variabilities were present within the adaptins.

Unexpectedly, two instances of gene absences within components of the AP-1 subcomplex in isolates of assemblage B were noted. First was the lack of AP-1 μ in the SRR1957167 assembly and the second absence was in AP-1 σ in the VANC/96/UBC/129 assembly (Figure 6.6B). TBLASTN searching using AP-1 μ and AP-1 σ orthologs from other assemblage B isolates could not identify μ 1 or σ 1 subunits in these genomes other than those belonging to the AP-2 sub-complex. This is an unprecedented finding as AP-1 components have so far been reported to be universally conserved, even in highly reduced parasites. The only exception was in the kinetoplastid endosymbiont *Perkinsela* sp. (Herman and Dacks, personal communication).

Altogether, the population-level survey confirms the previously reported secondary loss of COPII-Sec24C in assemblage B. At the same time, other vesicle coats such as adaptins, retromer, COPI, and clathrin heavy chain remain universally conserved without any inter-assembly variations. The absence of AP-1 μ and AP-1 σ , although surprising, is highly unlikely and should be ascertained as a technical assembly or sequencing-related issue. AP-1 μ is absent from the SRR1957167 belonging to assemblage B, which is an outlier as it has a comparatively smaller genome size (10.9 Mbp) and lower N50 (44795 bp) with its counterparts. Both attributes are likely a result of low starting sequencing depth compared to the other isolates and hence a cause for gene absence artifacts. In the case of AP-1 σ , which is missing from the VANC/96/UBC/129 assembly, may have been due to low coverage k-mers in this region and therefore discarded during the assembly process. Both of these represent more likely scenarios than instances of actual loss, as the remaining 39 isolates from assemblage B encode both of these adaptin subunits.

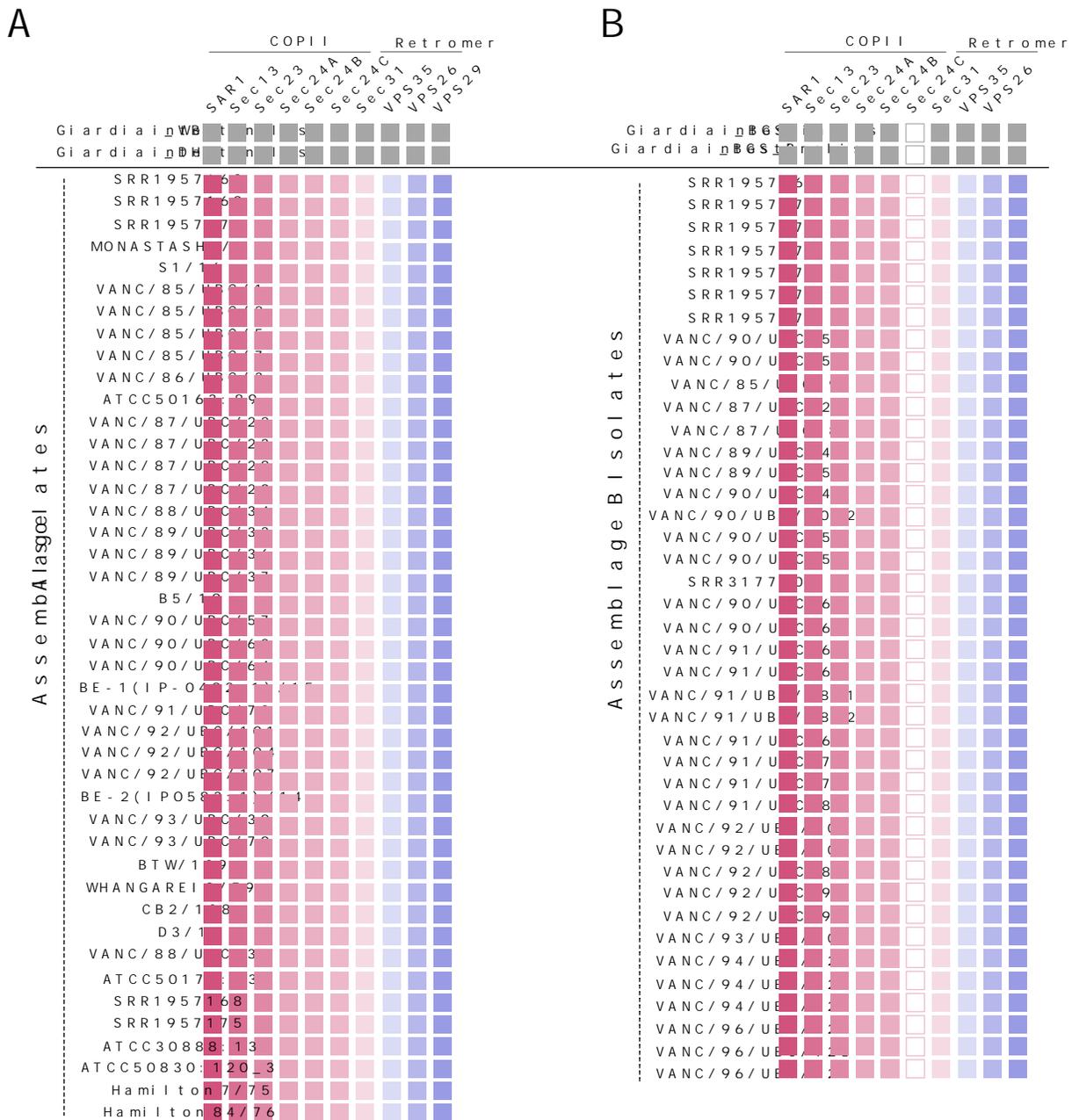


Figure 6.5. Tile-plot depictions of COPII and retromer repertoire in the BCCDC PHL isolates. (A) depicts the distribution of the giardial COPII and retromer components in the newly assembled genomes of the BCCDC PHL isolates, compared to the two pan-global reference assemblage A isolates, WB (AI) and DH (AII). No absences were identified within any of the assemblage A genomes. **(B)** depicts the distribution of the giardial COPII and retromer repertoire in the newly assembled BCCDC PHL assemblage B genomes and compared with the two pan-global reference assemblage B isolates, BGS and BGS_B. Assemblage B-specific losses are evident within one of the fornicate-specific paralogs of COPII-Sec24FII (Sec24C) and remain consistent with the findings reported in Chapter 2.

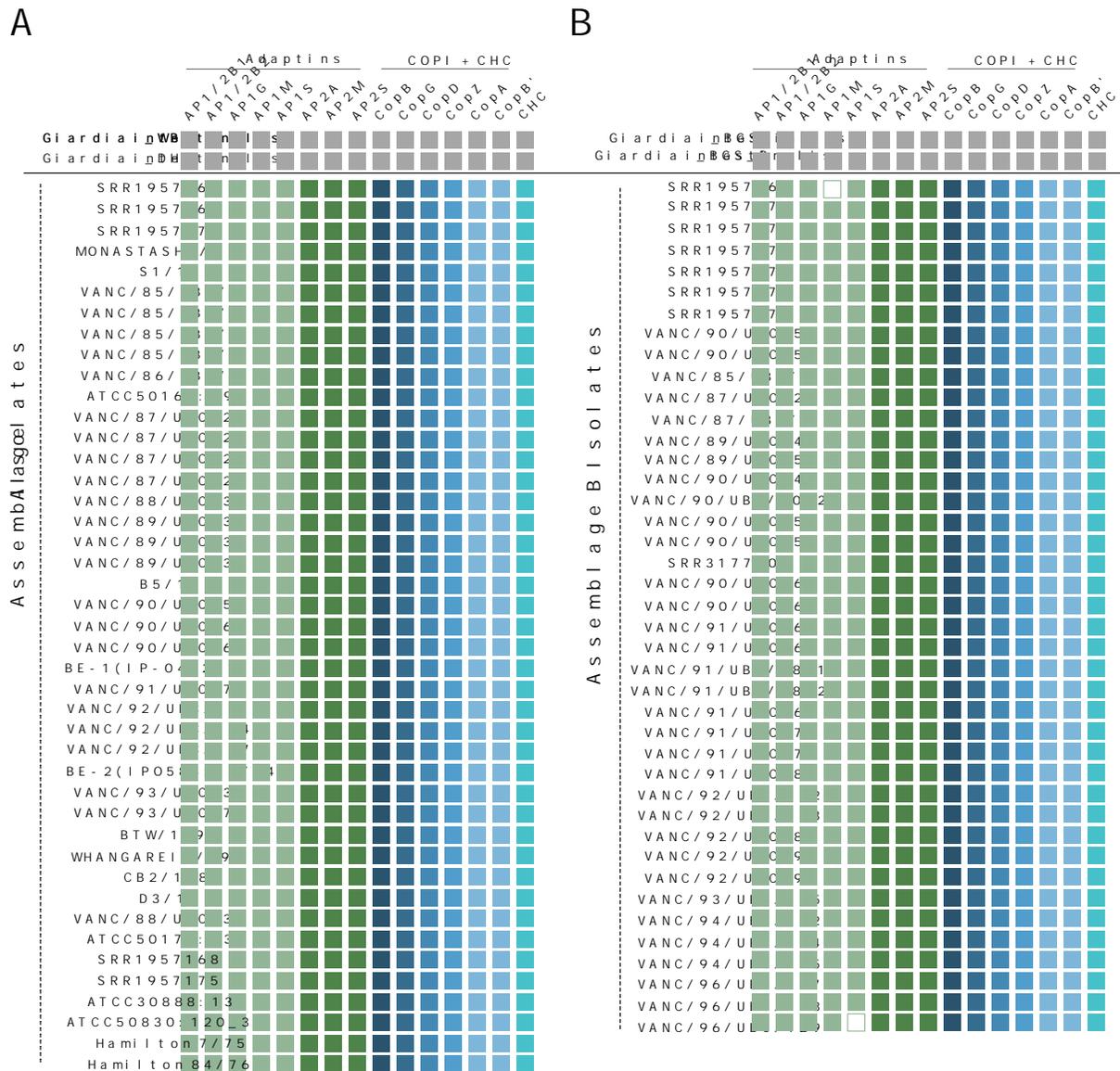


Figure 6.6. Tile-plot depictions of heterotetrameric adaptor complexes and clathrin repertoire in the BCCDC PHL isolates. (A) depicts the distribution of the previously identified HTACs and clathrin components in the newly assembled BCCDC PHL isolates and compared with the two pan-global reference assemblage A isolates, WB (AI) and DH (AII). No absences were identified in any assemblage A genomes. (B) depicts the distribution of adaptin, COPI, and clathrin components in the newly assembled BCCDC PHL assemblage B genomes in comparison to the two pan-global reference assemblage B isolates, BGS and BGS_B. Although no large assemblage-wide losses are evident, instances of individual isolate-specific absences within AP-1 subunits (*i.e.*, AP-1 μ and AP-1 σ) were present.

6.3.6. Paralogues of the ARF regulatory system proteins continue to differ between the two assemblages

Evolutionary bioinformatics analyses with proteins belonging to the ARF regulatory system were performed in Chapter 3. Although ancestral losses shaped the fornicate ARF regulatory system, surprisingly tight conservations within the overall complement of ARF1, ARF GAPs, and ARF GEFs exists in *Giardia* and its free-living relatives (Pipaliya et al., 2021). Despite this, there still were critical differences in the encoded proteins between the two assemblages. Namely, paralogs of the fornicate-specific ARF1F proteins and Sec7 domain-containing ARF GEFs differed (Pipaliya et al., 2021).

In this previous survey, of the three fornicate-specific ARF paralogues traced in *Giardia intestinalis*, a secondary loss within one of the two ARF1FB paralogs (*i.e.*, ARF1FB2) was identified across all examined pan-global isolates belonging to assemblage B (Pipaliya et al., 2021). On the other hand, assemblage A possessed all three ARF1 variants (*i.e.*, ARF1FA, ARF1FB1, and ARF1B2). This trend was confirmed in this survey, where all new assemblage A isolates possessed these three paralogs, while ARF1FB1 was universally absent across all 41 BCCDC PHL assemblage B isolates (Figure 6.7B). Unlike the previously discussed systems where absences were strictly observed in assemblage B, that was not the case with the ARF regulatory system proteins. In the previous investigation, inter-assemblage differences were identified in the ARF GEFs, namely within paralogues of Cytohesin and BIG. Assemblage A lacked one of the two BIGL proteins, while assemblage B did not possess one of the two Cytohesin identified in assemblages A and E (Pipaliya et al., 2021). Once again, the population-level survey confirms the absence of one of the two BIGL proteins from assemblage A and CYTHL from assemblage B (Figure 6.7).

Finally, no isolate-level variations were observed in either assemblage A or B in any of the proteins, and the fornicate ARF GAP complement, as well as canonical ARF1, was also conserved across all 83 isolates. This suggests that within *Giardia*, plasticity is greatest in ARF1Fs and the ARF GEFs. Overall, like the ESCRTs and the vesicle coats, previous trends of *Giardia*-level modulation within the molecular complement of the ARF regulatory system exists at the population level.

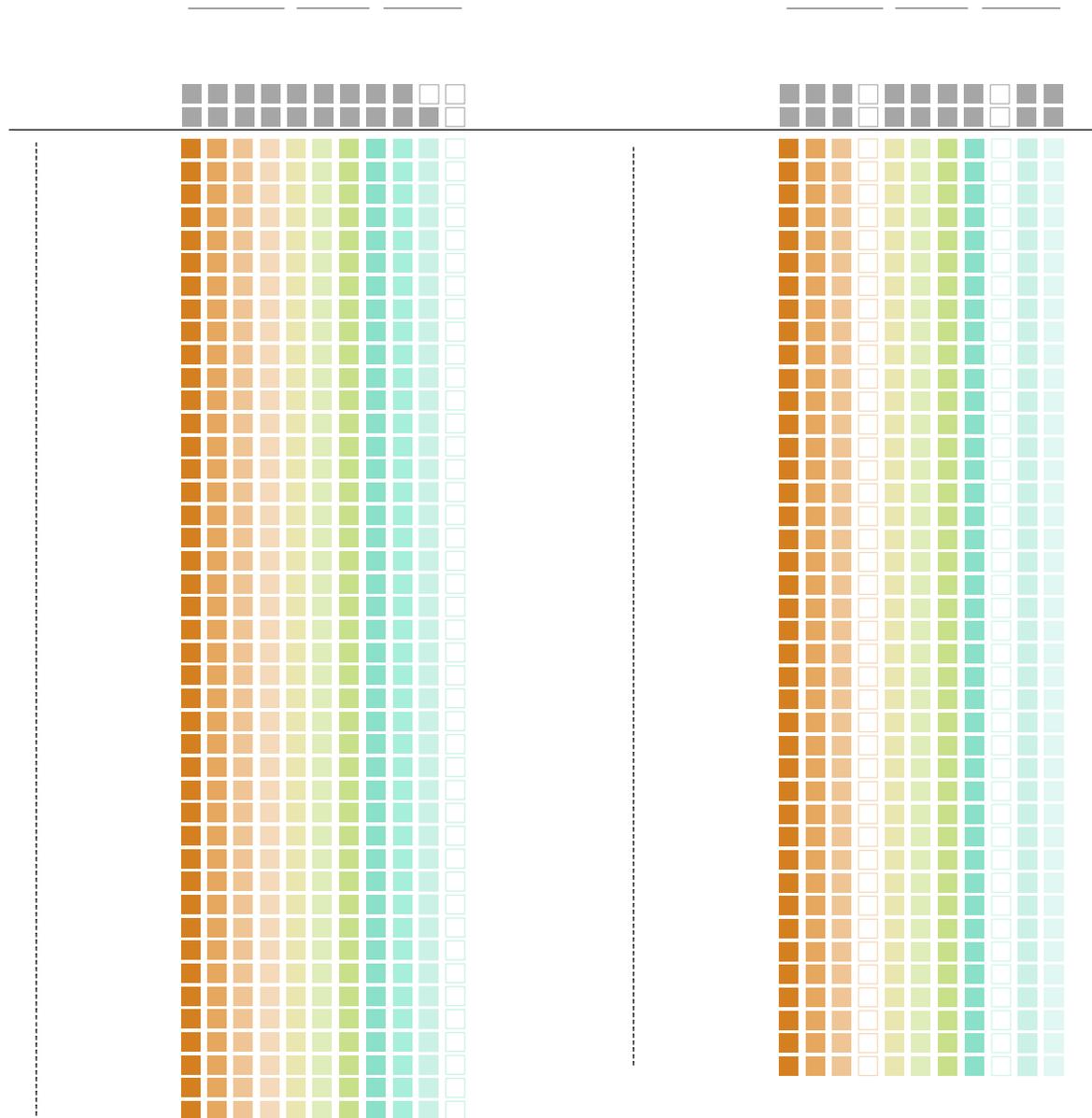


Figure 6.7. Tile-plot depictions of the giardial ARF regulatory system protein repertoire in the BCCDC PHL isolates. (A) depicts the distribution of the previously identified giardial ARF, ARF GEF, and ARF GAP repertoire in newly assembled BCCDC PHL *Giardia* genomes compared to the two pan-global reference assemblage A isolates, WB (AI) and DH (AII). In the previous survey, isolate WB was shown to lack both paralogues of BIG. However, loss within only one of the two BIGL paralogues is apparent in the new BCCDC assemblage A isolates. (B) depicts the distribution of the giardial ARF regulatory system proteins in the newly assembled BCCDC PHL assemblage B isolates in comparison to the two pan-global reference assemblage B isolates, BGS and BGS_B. Assemblage-wide losses within ARF1FB2 and one of the two CYTHL paralogs are also evident and consistent with the previous results.

6.4 Discussion

Comparative genomic investigations of the vesicle formation machinery in the preceding chapters noted distinct differences within subunits of the trafficking complexes encoded by the two human-infecting *Giardia intestinalis* assemblages. Because the previous sampling was limited to few genomes, a population-scale investigation to better define the heterogeneity within these protein complements was necessary. The assembly of 83 full-length genomes belonging to BCCDC PHL isolates of *Giardia intestinalis* assemblage A and B permitted a large-scale comparative analysis of the vesicle formation machinery. The findings from this study align with the previously published literature that highlights differences in the genomic features and extend results presented in the previous chapters of this thesis pertaining to inter-assemblage variances within the complement of the vesicle formation machinery to occur at the population level. While isolate-level differences were not ubiquitous, some instances were identified, which should be considered a technical assembly or sequencing-related issue.

6.4.1 Genome size differences and implications for parasite cell biology

Because the underlying aim of this investigation was to uncover species-level molecular differences, a *de novo* approach for genome assembly was pursued instead of reference-aligned assembly to limit reference bias and account for possible structural and sequence variants. This was because the working hypothesis of this investigation presumed the assemblages in question to be genetically distinct strains. Therefore, an impartial examination into the differences between the trafficking complement and other genomic attributes (*i.e.*, genome size and GC content) could have only been made appropriately through a *de novo* approach.

Using MaSuRCA, a roughly equal number of genomes belonging to both BCCDC PHL assemblage A and B isolates (*i.e.*, 42 assemblage A and 41 assemblage B genomes) were assembled. These were used to compare the %GC and genome sizes against one another and with the previously published pan-global isolates. Currently, there exist a total of 12 publicly available assemblies for both assemblages. Upon further breakdown, approximately half belong to AI and AII, which consist of the following isolates: WB, DH, AS175, AS98, ISS17, and ZX15 (Adam et al., 2013; McArthur et al., 2000; Morrison et al., 2007; Weisz et al., 2019; Xu et al., 2020). When these isolates are examined for their %GC, a range of 48.5 to 49.5% is obtained, apart from AS98, which is markedly lower in its %GC (45.50). In comparison, %GC of the BCCDC PHL assemblage A genomes yielded an average of 48.5%, which is akin to the GC content of the previously published AI and AII genomes. Within assemblage B, previously published short-read assemblies correspond to GS, GS_B, GS/M clone H7, and BAH15c1 isolates (Adam et al., 2013; Franzén et al., 2009; Weisz et al., 2019; Wielinga et al., 2015). Similar to assemblage A, the %GC for these genomes ranged between 47.0 to 49.2%. The average %GC for BCCDC PHL assemblage B genomes from this investigation comparable and yielded a value of 48.7%. Combining these results, *Giardia intestinalis* assemblage A and B isolates are similar in their overall GC content to the values reported in the literature, and little variability exists at the inter-assemblage level.

In contrast, the findings from these investigations, in combination with the previous studies, indicate that the two assemblages vary in their overall genome sizes. The BCCDC PHL assemblage A isolates ranged between 10.6 to 12.3 Mbp, similar to WB, DH, AS175, AS98, ISS17, and ZX15, which also ranged between 10.3 to 12.08 Mbp (Adam et al., 2013; Morrison et al., 2007; Weisz et al., 2019). In comparison, assemblage B genomes were larger (*i.e.*, 11 to 13.7 Mbp) and similar to genome sizes belonging to GS, GS_B, GS/M clone H7, and BAH15c1 that ranged between 10.4 to 13.6 Mbp. These inter-assemblage differences in genome sizes have implications on the encoded protein-coding repertoire as well. Although the BUSCO scores suggest conservation in 27% of the eukaryotic orthologs in all BCCDC PHL isolates, previous comparative genomic differences have identified several considerable cytogenetic differences between assemblages A and B isolates. These are in the structural organization of chromosomes, protein-coding syntenic regions, encoded repertoires of gene families, and inter-assemblage sequence divergences within the conserved orthologs.

Pulsed-field gel electrophoresis for separation and comparative assessment of molecular weights of the five *Giardia* chromosomes (Chr) belonging to isolate WB and GS revealed distinct differences. WB has chromosomes that are 1.5, 2, 3, and 3.5 Mbp in size. On the other hand, GS has a slightly larger chr1 sized 1.8 Mbp due to a recombination event with Chr2 (Upcroft et al., 1996, 2010). Optical mapping between these five chromosomes identified considerable structural rearrangements, inversions, and translocations (Adam et al., 2013). Notably, even within the syntenic regions, only 70% sequence similarity was present between GS and WB. Comparisons between the number of protein-coding genes were also dissimilar, where ortholog overlap analyses identified 2962 heterologous open reading frames (ORFs) in assemblage B isolate GS_B. In contrast, assemblage A isolates, DH and WB, were reduced in this number by half, where only 1935 and 1067 unique ORFs were identified, respectively (Adam et al., 2013). Within these ORFs, the number of genes corresponding to major families of virulence genes, especially in the cysteine-rich variant surface proteins (VSPs), differed significantly. In assemblage B isolate GS_B, approximately 6.7% of all predicted open reading frames corresponded to VSPs, whereas assemblage A isolates, DH and WB, were comparably lower in their proportion of the encoded VSPs, ranging at *ca.* 3-4% (Adam et al., 2013). Additional sequence level assessments of the shared VSP profiles only yielded a maximum of 55% sequence similarity between the two assemblages (Ankarklev et al., 2015). Identical differences likely exist between isolates of assemblage A and B examined here.

6.4.2 Potential impacts of missing vesicle formation machinery on secretory processes within parasites of the two assemblages

While the previous studies have focused on elucidating differences within the virulence genes and various other genomic aspects, this investigation supplements those findings by shedding light on the heterozygosity within the encoded trafficking machinery. The population-level survey consolidated the findings from Chapters 2, 3, and 4 and extended them at a broader scale. More specifically, a combination of the results from Chapters 2, 3, and 4, and now this survey, identified assemblage-specific differences in

subunits of the ESCRT, COPII, and ARF regulatory system machinery, which may confer disparities in the role of those subcomplexes and assembly dynamics at the PVs and ESVs where they primarily associate. As has been well established in the preceding chapters, the role of the vesicle formation machinery within *Giardia* is critical to the parasite's ability to facilitate secretory and uptake processes in trophozoites and encysting cells. Therefore, divergences in its molecular complement could implicate crucial functional variabilities within giardial trafficking processes at these organism-specific compartments.

Within ESCRTs, variations in ESCRTIII-Vps20L and ESCRTIIIA-Vps4 were identified. Cellular localization investigations in Chapter 4 and by previous others in the field have lent unparalleled insights into the promiscuity of ESCRTs at different giardial endomembrane compartments. To briefly remind the reader, *G*Vps20L was primarily localized to the ER in conjunction with numerous ESCRT subunits (Pipaliya, Santos, et al., 2021). While the specific dynamics of inter-ESCRT association by sequential recruitment onto various organellar membranes are currently unknown in *Giardia*, some conserved aspects such as cargo recognition and membrane remodelling must still exist at those different compartments. Therefore, the absence of critical subunits such as Vps20L in assemblage B isolates implies functional compensation mechanisms to fulfill those same roles or divergence in the underlying pathway wherein Vps20L is simply not necessary in this assemblage. While the former scenario may very well be the case for the missing paralogs such as Vps4C, which may be having redundant cellular roles as other Vps4 proteins at the peripheral vacuoles in assemblage A isolates, the latter scenario could signify crucial biological differences. Differences in paralog functions and molecular association of the giardial trafficking machinery, including the ESCRTs (*i.e.*, Vps4 and Vps46), have been hinted at previously and therefore is a plausible scenario, but one that still requires robust molecular and biochemical testing (Saha et al., 2018). Nonetheless, these results posit a potentially differentiated underlying mechanism by which the existing ESCRT subunits function within the *Giardia* trophozoites of the two assemblages.

Similar to the ESCRTs, although most vesicle coats (*i.e.*, adaptins, COPI, and retromer) are conserved in their complement between the two assemblages, differences were noted within the COPII components specifically in the number of lineage-specific paralogs of Sec24. As discussed previously, the giardial COPII is fundamental to ESV biogenesis and maturation to ensure the transport of cyst-wall material to the parasite surface (Faso et al., 2013; Stefanic et al., 2009). Sec24, along with Sec23, forms pre-budding complexes upon cyst-wall protein recognition for a COPII-coat assembly at the ER-exit sites (Faso et al., 2013; Zamponi et al., 2017). In the absence of one of the three Sec24 paralogues, it could be that the remaining two Sec24 proteins may be fulfilling this role in assemblage B isolates and that plasticity within this system exists to accommodate losses such as this one. A contrary scenario may also be possible, wherein absence in Sec24 paralogs could indicate differences in COPII-assembly dynamics. All fornicates lack Sec16, which typically stabilizes Sec23/Sec24 and Sec13/Sec31 coats during the vesicle budding process. Therefore, ancestral neofunctionalization of these lineage-specific paralogs may have occurred to compensate for missing Sec16 roles. A secondary loss of this protein in *Giardia* assemblage B could then have implications on the molecular dynamics of how the COPII-coat is stabilized in those parasites, which

in turn could translate to altered ESV-biogenesis mechanisms or cyst-wall material trafficking. *In vitro* encystation is currently not possible in assemblage B, and therefore, these postulations remain untestable conjectures (Barash et al., 2017).

Finally, differences in the ARF regulatory system proteins, especially in the ARF1F GTPases, again may indicate functional redundancy of these paralogues in assemblage B at the PVs. Differences within the different GEF profiles may indicate variable regulation of the giardial ARFs. While testing these scenarios through fluorescent microscopy and proteomics experiments in the trophozoite stages of assemblage B isolates would be interesting (as showcased with lab strain of assemblage A in Chapter 5), it is challenging to generate and establish parasite cultures of transgenic variants of assemblage B isolates, as they are slow growing and not amenable to stable or episomal integration of plasmids for recombination protein expression. Once technical advancements are made in this area of *Giardia* biology, these experiments should be pursued as they will open avenues to elucidate novel ARF biology and GTPase regulation mechanisms in *Giardia* and eukaryotes in general.

6.4.3 General implications on giardial parasitism and species complex

The abovementioned scenarios are postulations that would require further testing, some of which are out of scope for this thesis and *Giardia* in general. Nevertheless, the results from this chapter are sufficiently concomitant with previous studies and substantiate notions of divergences in the protein machinery which underpin molecular intricacies in the two-human infecting *Giardia* strains.

Perhaps, the previously noted differences in the number and biochemical profiles of VSPs and the now-reported variation in trafficking system proteins may be directly related to one another. Recognition and secretion of virulence factors in parasites are directly dependent on vesicle-mediated transport. For example, the closely related *Trichomonas vaginalis* releases exosome-like vesicles to secrete parasite-specific tetraspanins (Tsp1) and small RNAs to immunomodulate inflammatory cytokine-mediated responses within the genitourinary tract (Rai & Johnson, 2019; Twu et al., 2013). Other examples include *Trypanosoma brucei*, which also produce and secrete EV-like structures containing parasite-specific variant surface glycoproteins for antigenic variation and immune avoidance (Cross, 1975; Szempruch et al., 2016). Growing evidence suggests that this strategy of delivery and transport of virulence factors in membrane-bound vesicles is employed by parasites spanning the breadth of eukaryotic diversity and is fundamental to their pathogenesis (Ofir-Birin & Regev-Rudzki, 2019). In *Giardia*, EV-like structures have been documented, although their origins and specific mechanisms for biogenesis are still highly contested (Midlej et al., 2019). While the specific biochemical profiles of these EVs and the cargoes they carry remain unelucidated, it is known that *Giardia* secretes metabolic enzymes, such as arginine deiminase, ornithine carbamoyltransferase, and enolase that alter host pathophysiology and have immunomodulatory functions (Ma'ayeh et al., 2017; Stadelmann et al., 2012). Previous and ongoing microscopy and proteomics investigations have determined that many of these proteins accumulate within the peripheral vacuoles and

are in association with trafficking complexes from Chapters 4 and 5 and those previously examined (Zumthor et al., 2016; Balmer and Faso, personal communication).

The transport of these cargoes to the PVs and into the extracellular environment is likely modulated by the giardial cargo trafficking system. Direct evidence of this is the regulation of VSP translocation from the ER to the PVs for cell-surface expression. Their secretion from the ER to the PVs is implicated to occur in an ARF1 and coatmer-dependent manner in the trophozoites (Luján et al., 1995). The absence of different ARF paralogs in *Giardia intestinalis* assemblage B may have direct or indirect implications on the expression of specific VSP profiles. These biological differences, in turn, may underpin variable clinical outcomes of Giardiasis. A scenario wherein differential functional modulation of the virulence proteins via distinct trafficking pathways in assemblage A and B may very well exist. Divergences in these biological mechanisms between the two assemblages, even if subtle, will have implications for the trophozoite-mediated disease establishment and subsequent parasite propagation.

Finally, although this study does not immediately resolve whether the *Giardia intestinalis* assemblages are separate species, it supports previously illuminated genomic and molecular complement-level differences at the population level. The definition of what constitutes 'species' remains a hotly debated philosophical discussion within the fields of ecology, evolution, and medicine. These lines are especially blurred in microbiology. Historically, the biological species concept has defined two organisms to belong to the same species if, through sexual reproduction, they can produce viable progeny (De Queiroz, 2005). There is limited evidence for sexual reproduction or inter-assemblage genetic exchange in *Giardia*. Therefore, their species status under the strict umbrella of the biological species concept remains contested. On the other hand, the phylogenetic species concept defines species as a group of organisms with shared and unique evolutionary histories that possess a defined set of traits (De Queiroz, 2007). Phylogenetic placement using multi-locus sequence genotyping places *Giardia intestinalis* assemblages A and B as paraphyletic lineages nested within other animal-infecting assemblages (Cacciò et al., 2008; Huey et al., 2013; Lee et al., 2020). This implies that zoonoses for human pathogenesis occurred convergently. Combined with these phylogenetic classifications, differences in genome attributes and encoded protein repertoires strengthen the notion that *Giardia intestinalis* assemblage A and B (and all other assemblages) can be defined as different species as per the phylogenetic-species concept. Additionally, previous whole-genome comparisons estimate only 77% nucleotide-level sequence conservation and 78% amino-acid similarity in the protein-coding regions between assemblages A and B isolates (Franzén et al., 2009). Genome recombination analyses performed on assemblage A, B, and E isolates further supported this notion (Xu et al., 2012). Therefore, from evolutionary, molecular, and genomic standpoints, the findings from this thesis, as well as those from the literature, present a strong case for *Giardia intestinalis* assemblages to be considered as separate species. Ultimately, the interpretation of this paragraph and its derived conclusions of whether this stance should be taken will also depend on the reader's biases.

Nevertheless, the contribution this study makes to the *Giardia* community is substantial. The 83 genome assemblies will provide a wealth of new genomic data for biologists to probe relevant questions

for a holistic view into this notion and to form a consensus on whether *Giardia intestinalis* is a species complex or not.

6.5 Conclusions

Findings from this study confirm genomic differences between *Giardia intestinalis* assemblages A and B to exist at the population level. Although this investigation aimed not to derive broader conclusions regarding whether these two constitute different species, the variabilities in the vesicle formation machinery and genomic attributes corroborate the previously identified trends and extend that view.

CHAPTER 7

General Discussion

7.1 Overview

Cargo trafficking in eukaryotes is essential, especially during host-parasite interactions where pathogens rely on these cellular processes to establish disease and ensure propagation between host and environment. This thesis investigated the vesicle formation machinery in *Giardia intestinalis* to discern the evolution and mechanistic frameworks of transport processes within this parasite's atypical endomembrane landscape. Both at the morphological and genomic level, *Giardia* has highly modified its cellular organization and protein-coding complement compared to other eukaryotes. Therefore, a comparative genomic approach was taken against its free-living and parasitic fornicate relatives to discover the time-points at which these divergences occurred. This approach identified patterns of lineage-wide, parasitism-associated, and *Giardia*-specific duplications and losses within the surveyed families of vesicle formation protein complexes. The advancements made through this lineage-wide evolutionary approach, compared to previous *in silico* reconstructions of this machinery where *Giardia* was used as a sole representative, highlighted the notion that reductive pressures were at interplay prior to parasitic states and likely shaped the transition to parasitism that is observed within fornicates. Aside from broad evolutionary patterns, subtle yet important differences were also identified in the molecular complement between the different *Giardia* assemblages and were especially prominent between isolates of the human-infecting assemblages A and B. A fine-grained view into these patterns was taken by performing genome assembly and comparative genomics with the Canadian isolates sequenced by the BCCDC PHL. The comparative genomic results from the pan-global isolates were corroborated and existed at a greater population level. Moreover, molecular functional investigations with select trafficking proteins were also pursued in the lab strain of *Giardia intestinalis* WB, C6. Immunofluorescence microscopy and co-immunoprecipitation-derived proteomics elucidated intracellular localizations and molecular interactions in the context of trophozoite-specific endomembrane compartments. Altogether, this multidisciplinary approach provided a broadened outlook into the mode and tempo by which cargo trafficking processes were subject to evolutionary modulation and the level of functional plasticity that exists in the absence of canonical secretory and endocytic pathways.

7.2 Cellular implications of conservations and dispensability in the fornicate vesicle formation machinery

Although previous studies have determined the molecular complement of several vesicle formation proteins in *Giardia intestinalis* assemblage A, isolate WB, as an isolated sampling point, this thesis undertook a comprehensive assessment through a broader evolutionary looking glass (Dutta et al., 2015; Leung et al., 2008; Marti et al., 2003; Miras et al., 2013; Pipaliya et al., 2019; Rivero et al., 2010; Saha et

al., 2018). The purpose of this was two-fold. First, *Giardia* belongs to the sub-phylum Fornicata, which comprises organisms with diverse trophic strategies (*i.e.*, free-living, commensal, and parasitic). Understanding the accompanying evolutionary pressures by which cellular systems are shaped in these different environments is fundamental to deducing the mode by which parasitism arises. The second rationale for employing a comparative strategy was to use the identified protein complement from the close relatives for iterative homology searching to characterize lineage-specific or highly divergent repertoire of proteins that have so far remained unelucidated. Overall, these evolutionary bioinformatic approaches successfully uncovered the precise road to parasitism and allowed for the characterization of the lineage-specific novelties within the vesicle formation machinery, absent elsewhere within the tree of eukaryotes.

Often, a parasitic lifestyle is marked by certain types of cellular endpoints that are generally accompanied by reductive molecular evolution and repurposing of machinery for parasite-specific functions. A combination of these have been exemplified in apicomplexans from the TSAR supergroup, trypanosomatids within Discoba, and now *Giardia* within Fornicata, indicating this to be a frequently occurring convergent phenomenon within the eukaryotic tree of life (Ebenezer et al., 2019; Rueckert et al., 2019; Woo et al., 2015). Work performed in this thesis builds a robust case for evolutionary adaptations that are synchronous with functional redirection of trafficking machinery towards *Giardia*-specific organelles. Conservation as well as plasticity within the evolution and function of ESCRTs, vesicle coats, and the ARF regulatory system proteins exist across fornicates and the *Giardia* assemblages. These results allow for an integrated view into what factors can be considered critical versus dispensable for vesicle formation processes in this lineage. The overarching trend seems to be that the late endo-lysosomal machinery is amenable to many losses, whereas those canonically associated with the early endo- and exocytic pathways are less likely to be discarded. These findings are important because examination of *Giardia* alone would not have yielded insight into why it lacks proteins typically associated to the early and late endosomal compartments but kept Golgi-associated machinery, despite the presence of endo-lysosomal peripheral vacuoles and absence of a stacked Golgi. A surface-level view would suggest that *Giardia* has entirely restructured its endomembrane organization in comparison to its relatives. However, these detailed investigations help comprehend the underlying processes of the manner and timing by which this took place.

Evolution within the late-endo-lysosomal trafficking pathway was assessed through the evolution of the ESCRT complexes that are required for the biogenesis of multivesicular bodies. The existence of MVB-like endosomes can be discerned in the previously published electron micrographs detailing the ultrastructures of several fornicate lineages (*i.e.*, *Carpodemonas* and *Kipferlia*), but which are absent in *Giardia* (Morecki & Parker, 1967; Simpson & Patterson, 1999; Yubuki et al., 2013). It was determined that a pattern of loss coinciding with a lack of MVB-morphology could likely be attributed to the absence of ESCRTI in *Giardia* (Figure 4.1). Apart from *Giardia*-specific modifications, streamlining prior to the transition to parasitism may be indicative of dispensability within various subcomplexes and their subunits within this pathway as a whole (Figure 4.1). Apart from ESCRTI, components from ESCRTIII (*i.e.*, Vps32) were the

first to be lost. At the same time, conservations and even expansions within ESCRTII and III-A were observed, indicating redundancy or possibility for compensatory mechanisms by other subunits for ubiquitin-binding, membrane recruitment and assembly, or scission. Pre-duplicates of these conserved ESCRT subunits, specifically the SNF7 domain-containing progenitors and Vps22/Vps25 in most of the Asgard archaeal ancestors highlight, that although not all ESCRTs may be necessary, ESCRTII and ESCRTIII components likely make up the core functional units of this machinery for membrane recruitment and deformation (Spang et al., 2015; Zaremba-Niedzwiedzka et al., 2017). These trends have also been observed in other pan-eukaryotic investigations of the ESCRT machinery in other parasites such as *Plasmodium*, *Trypanosoma*, and *Entamoeba*, which also mainly conserve ESCRTII, III, and IIIA (Leung et al., 2008).

Similar observations were noted in Chapter 2, where the molecular complement and evolution of the vesicle coat complexes were determined and which of the pathways are inessential versus critically conserved within the fornicates. The evolutionary trend within these proteins follows that observed in other eukaryotes, where TSET and AP-5 are lost before the Last Fornicata Common Ancestor (Figures 2.2 and 2.5). This may indicate that functions of these two HTAC subcomplexes may be compensated by other mechanisms or not typically required within early or late endosomal pathways in many free-living organisms and parasites (Hirst et al., 2011, 2014, 2018). Interesting to note, however, was the pre-fornicate loss of AP-4 that regulates early endosome to *trans*-Golgi network (TGN) trafficking (Burgos et al., 2010). While *Carpodemonas membranifera* is the only exception in its presentation of a stacked Golgi, all other CLOs, retortamonads, and diplomonads lack readily identifiable Golgi compartments. Therefore, the absence of AP-4 may correlate to a state of ‘phasing’ out to discard the TGN-associated machinery from its molecular repertoire, while simultaneously restructuring the morphological states of those trafficking pathways. AP-3 loss in *Giardia*, which is characterized for its involvement in TGN to late endosomal cargo transport, represents an additional level of endomembrane organellar reorganization from canonical endosomes, which are present in its parasitic relative *Spironucleus*, to static-state PVs (Santos et al., in preparation).

In comparison, patterns of conservation within COPI, COPII, AP-1, AP-2, and the cargo-selective retromer complexes (*i.e.*, Vps26, Vps29, and Vps35) resembles similar to that observed in other eukaryotes and suggests their indispensability within the fornicate endosomal systems (Hirst et al., 2011; Koumandou et al., 2011; Schlacht & Dacks, 2015). COPI and COPII mediate ER-Golgi trafficking for early secretion, whereas AP-1, AP-2, and retromer are critical for early endosome biogenesis. This suggests that although streamlining within numerous late endo-lysosomal pathways is present, ER-associated trafficking persists irrespective of the absence of a canonically stacked Golgi. Furthermore, any existing endo-lysosomal functions and organelles in fornicates are likely sculpted by a repertoire of what may constitute a core set of coat complexes in fornicates, which in this case would likely be AP-1, AP-2, and retromer. The existence of a cryptic Golgi morphology (*i.e.*, punctate or tubular) in this lineage should not be ruled out, as has been demonstrated in many other protists such as *Naegleria* and *Mastigamoeba* (Barlow et al., 2018; Herman et

al., 2018). A cryptic Golgi could be a hub for cargo sorting between the ER and the endo-lysosomal organelles, however, until the discovery of any such compartment, the current best scenario is the one described above.

Among the unexpected findings was the universal absence of clathrin light chain across all fornicates (Figures 2.2 and 2.5). The absence of a light chain would hamper canonical clathrin triskelion assembly mechanisms and altogether disbar clathrin-mediated endocytosis (CME). Therefore, the existence of clathrin-independent endocytosis (e.g., Rac1/Cdc42/actin-dependent) could likely be at play, especially in endosome-bearing fornicates (Huotari & Helenius, 2011; Parton & Simons, 2007). Alternatively, if CME is still utilized, other lineage-specific proteins could have compensatory functions to structurally replace CLC within clathrin triskelion assemblies. The discovery of these altered mechanisms for endosomal biogenesis is worthy of future exploration in fornicate lineages that are now amenable to genetic manipulation (e.g., *Spiroucleus salmonicida* and *Carpodidomonas membranifera*).

The pattern of loss within the coat proteins but retention of Golgi-associated trafficking is also observed within the ARF regulatory pathway, the family of proteins that regulate adaptor coat assembly dynamics at membranes. Losses were mainly within the endosomal ARF6 and numerous endo-lysosomal ARF GAPs, while conservations remained in the Golgi-associated BIG and ARF1 (and even expansions in this case) (Pipaliya et al., 2021). These trends highlighted that dispensability within the endo-lysosomal pathway is tolerated, whereas conservation in the Golgi-associated and early secretion machinery is prone to expansions and possible cellular repurposing.

Altogether, these results conclude that the giardial endomembrane organizational remodeling is a corollary to the combination of gradual molecular modifications that predate and coincide with parasitism. These findings also demonstrate that the endo-lysosomal vesicle formation processes within these lineages are constrained to only what is most essential and can be mediated by a limited yet core complement. The absence of dynamic endosomal carriers and a discernible Golgi, paired with the organellogenesis of *Giardia*-specific compartments (e.g., peripheral vacuoles and encystation-specific vesicles), implies a scenario where the distinct endo-lysosomal organelles are functionally fused. These represent simplified static-state compartments that continue to perform the same secretory and endocytic functions within a minimal organellar landscape.

Findings from each of these chapters also identified differences within the repertoire of *Giardia*'s trafficking system proteins and that strain-specific variances exist within the molecular complement of the different *Giardia intestinalis* assemblages, especially between the two human-infecting assemblages, A and B. Chapter 6 undertook a population-level survey of isolates belonging to these assemblages to assess better whether inter-assemblage differences in the trafficking complement noted in the preceding chapters were a by-product of limited sampling or a reflection of genuine biological differences conforming to the notion of a species-complex within this parasite. This survey determined that, indeed, *Giardia intestinalis* assemblages A and B are distinct in their repertoire of trafficking proteins. These findings have profound implications on the mechanistic intricacies of cargo transport within each of these *Giardia* lineages. The

genomes generated in Chapter 6 will serve as valuable resources for the *Giardia* scientific community to probe and discern variabilities in other systems such as virulence genes and molecular pathways that could underpin disease establishment and pathophysiology. The clinical and translational outcomes of pursuing such studies are discussed further in Section 7.5.

7.3 An integrated model for peripheral vacuoles as the singular yet multidimensional hub for endo-lysosomal processes

In the absence of canonical endosomes and lysosomes, the other goal of this thesis was to determine where and how the identified vesicle formation machinery functions within the trophozoite stages, which actively participate in material exchange. Findings from Chapters 4 and 5 were fundamental to postulate the homology of *Giardia*-specific peripheral vacuoles by shedding light on the types of cargo transport proteins associated with these organelles. Specifically, molecular localization and protein-protein interactions with several of the trafficking proteins investigated in this thesis, in conjunction with previously characterized mechanisms, further confirm the *Giardia* field's working hypothesis that the static-state PVs are likely a fused and a multifunctional endo-lysosomal system serving as a singular 'in' and 'out' destination for cargo uptake and secretion. By synthesizing the various findings of the roles of the giardial vesicle formation machinery, an integrated model for uptake, secretion, and cargo sorting for anterograde and retrograde material transport via these compartments can also be postulated. Aspects of this model are testable and should be pursued in future investigations.

Giardia lacks both defined endosomes and lysosomes, so the PVs in the vegetative trophozoites are organelles that likely perform functions for both. Investigations in Chapter 4 revealed ESCRTII and III components to be primarily associated with the PVs and directly or indirectly interact with several ESCRTIII and III-A components (Pipaliya, Santos et al., 2021; Saha et al., 2018). In comparison to the characterized functions of the ESCRT machinery in model organisms, promiscuity in the locations of giardial subunits was observed. These were at the PVs, the ER, and the mitosomes. Molecular localization and proteomics from this analysis hint at ESCRTII and ESCRTIII assembly at these organellar membranes, most likely from a cytosolic pool. Contacts sites between the *Giardia* ER and PVs have been previously documented and could be mediated by ESCRTs for possible unconventional anterograde or retrograde trafficking (Zumthor et al., 2016). ESCRTIII-CHMP7 specifically could also have roles in ER-synchronized mitosomal release from the central mitosomal complex, which are hypothesized to be the source for peripheral mitosomes (Rout et al., 2016). Altogether, localizations of the ESCRT machinery at the PVs imply these compartments have late-endosomal characteristics.

Early and late endosomal identities of the PVs are additionally confirmed by the previously examined roles of vesicle adaptor complexes, which also participate in the trafficking of material between the PVs and the ER. This is evident in previous investigations that have elucidated the role of the giardial retromer machinery, which localizes to both the parasite's PV and ER membranes (Miras et al., 2013). The retromer machinery was also colocalized with soluble hydrolase receptors (AcPh), and is the proposed

cargo recycled between the PVs and the ER. This implies that retromer function in *Giardia* is somewhat conserved, namely in retrograde recycling of cargo between the endosomes and the TGN. *Giardia* retromer function between these two compartments posits a functional model of PVs behaving as early/recycling endosomes, while parts of the ER may have sorting functions like the TGN. Previous biochemical and microscopic investigations have identified the association of retromer subunits with AP-1 and have postulated a role for this adaptor protein complex and its involvement in the trafficking of AcPh and cysteine proteases from the ER-exit sites to the peripheral vacuoles (Miras et al., 2013; Rivero et al., 2012). Like retromer, AP-1 is typically necessary for the movement of cargo between early/recycling endosomes and the TGN, and so anterograde AP-1 mediated trafficking of lysosomal proteins to the PVs from the ER-exit sites support the notion that PVs are the singular endosomal destination within the cell with functionalities of both early and recycling endosomes. These findings also theorize that the *Giardia* ER serves as a site for cargo sorting and transport in the absence of a Golgi. Both *Giardia* COPI and COPII assemble at the perinuclear ER and traffic VSPs to the PVs in non-encysting cells and mediate cyst-wall material loading into ESVs during encystation (Faso et al., 2013; Luján et al., 1995). Strikingly, both processes are sensitive to the antifungal brefeldin-A, which is routinely used to hamper Golgi-trafficking (Faso et al., 2013; Luján et al., 1995; Touz & Zamponi, 2017). This hints at a scenario where specialized sites within the *Giardia* ER may have Golgi-like characteristics for regulated (encystation) and constitutive (VSP trafficking) secretory processes.

Support for early endocytic PV functions at the plasma membrane interfaces is also rooted in their role in bulk flow uptake of cargo material from the extracellular environment (Zumthor et al., 2016). Instead of clathrin and adaptin assembly at the plasma membranes to temporally regulate neogenesis of dynamic coated vesicles and early-stage endosomes, PVs represent steady-state compartments that identically perform these same functions but in an altered manner. In the absence of a light chain, the clathrin heavy chain (CHC), in conjunction with the giardial Dynamin and AP-2, likely scaffold the PV-PM junctions to facilitate membrane fusion dynamics for pore formation and material uptake (Zumthor et al., 2016). Recognition of AP-2 and CHC at the PVs also occurs through interactions with the plasma membrane-associated PX-domain (PXD) - containing proteins (PXD1 and 6), as well as with PXD2 and PXD4 on the peripheral vacuoles (Cernikova et al., 2020; Zumthor et al., 2016). This is identical to canonical adaptor recruitment onto endosomes in model organisms, whereby AP-2 and Dynamin recruitment to the plasma-membrane interfaces also occurs by PX-domain proteins and binding to PI(4,5)P2 to scaffold, mark, and modulate clathrin-mediated endocytosis of extracellular material into the cell (Daumke et al., 2014; Praefcke et al., 2004; Robinson, 2004). Association of numerous PXDs to both PVs and the plasma membranes, and with AP-2 and clathrin, supports the PV functional model of early endosomes for cargo uptake.

While PVs currently represent the only endo-lysosomal compartments in this parasite, a scenario wherein these organelles are a functionally heterogeneous population with distinct functional capabilities that range from uptake, sorting, and anterograde trafficking is plausible. Previous and ongoing microscopic investigations for nanoscale ultrastructural visualization and 3D reconstructions have shed light on the

morphological differences between the PVs, where both tubular and vesicular structures of these compartments have been observed (Zumthor et al., 2016; Santos et al., in preparation). These studies have also illuminated PVs to span large amounts of cytosolic space and contact with the ER. Therefore, differences in the biochemical properties and endosomal characteristics between the populations closer to the ER versus those near the cell periphery are possible. This notion of parasite-specific organelles with diversified morphophysiological characteristics is not unique to *Giardia*. Emerging new evidence collected using high-resolution fluorescent microscopy with *Toxoplasma gondii*'s endo-lysosomal Rab GTPases (*i.e.*, Rab5 paralogs) has elucidated the parasite micronemes, which are secretory organelles part of the apical complex located at the parasite cell periphery (similar to the *Giardia* PVs), to be a collection of functionally and biochemically distinct organelles (Kremer et al., 2013). Investigations of similar scope by first identifying distinct protein markers with differential and non-overlapping patterns of PV localizations would be necessary. The various paralogs of ARF GTPases in *Giardia* and the results described in Chapter 5 suggest these may be crucial markers that hold the key to biochemically distinguish PV populations as different in their trafficking and cargo sorting functions. While the results of that chapter serve as an important starting point and hint at differences in their PV localizations and molecular interactions, additional experiments are necessary to consolidate this proposed model. A direct extension of that work and experiments that should be employed next is generating and expressing dual-construct reporters with different combinations of the *Gi*ARF1 paralogues and visualizing them using high-resolution fluorescent microscopy (similar to the strategy undertaken for ESCRTII-Vps25 and 36 in Chapter 4) for co-localization analyses. This will determine whether or not signal overlap exists between the different ARFs and which specific PV populations they mark. Nonetheless, the combination of results from this thesis and past investigations on their own conjecture PVs as the singular yet dynamic junction for exo- and endocytosis between host and parasite in the absence of a typical endomembrane complexity.

7.4 Limitations and future perspectives

While this thesis takes a comprehensive approach to studying the evolution and functions of the vesicle formation machinery using improved methodologies, there still remain many technical challenges to fully appreciate this parasite's biology and evolution. Four years ago, when I first began the bioinformatic investigations, a limited number of genomes and transcriptomes from the free-living fornicates and *Giardia* spp. were available. Since then, high-quality sequence data from new fornicate lineages, new isolates from assemblages A and B, and other *Giardia* assemblages and species have vastly proliferated (Kooyman et al., 2019; Pollo et al., 2020; Xu et al., 2020; Yazaki et al., 2020). It is crucial to include these new lineages as additional sampling points in future studies of similar scope to accurately reconstruct the evolutionary histories and assess inter-assemblage cellular and molecular diversity. *In vitro* cultures of *Giardia* assemblages or other *Giardia* species other than assemblage A isolates have so far been challenging or unattainable, which has hampered the ability to perform classical genome sequencing or transcriptomics from axenic parasite monocultures. In recent years, the advent of single-cell genomics and

transcriptomics has helped overcome the need for establishing *in vitro* cultures, and instead, FACS-sorted trophozoites and cysts can be acquired and sequenced directly from fecal samples (Kooyman et al., 2019). With the availability of long-read sequencing technologies, assembling contiguous genomes from organisms with polyploid lifestyle stages has also become easier (Pollo et al., 2020; Xu et al., 2020). Therefore, in combination with long-read sequencing and single-celled genomics, attempts to increase the overall availability of *Giardia* and fornicate genomes through these approaches should also be made. Using these data, trafficking machinery investigated in this thesis, as well as others, should be studied as direct extensions of this work for a comprehensive outlook into the evolution of this system and path to parasitism in this lineage, as has been done with apicomplexan and apicomplexan-like parasites (Mathur et al., 2019, 2021). Population genomics with different strains of *Giardia intestinalis* should especially be pursued for a holistic view into the conceptions of species-level biological differences and what implications they have on parasite zoonoses and the manifestation of clinical Giardiasis.

Although I focused my investigations on the vesicle formation processes within membrane trafficking pathways, results from molecular studies performed in this parasite and those in other organisms have clearly revealed that they stand alone provides an incomplete picture. These pathways intertwine with other trafficking machinery, especially the vesicle fusion and tethering complexes. This was especially evident in the proteomics investigations with the *Giardia* ARF regulatory system proteins in Chapter 5, pointing to cross-talk between Rabs and SNAREs with the ARF regulatory proteins and coat complexes. For an exhaustive perspective into the intricacies of the giardial transport mechanisms, which was out of scope for this thesis, the next critical step and direct extension of this work requires assessment of the vesicle fusion machinery. The principal players that should be studied are the remainder of the Ras-small GTPases (Rabs and Arls) and their regulators, SNAREs, and the multi-subunit tethering complexes (e.g., HOPS/CORVET, TRAPPs, exocyst, GARP, COG, and more). Analyses of these families will permit a thorough reconstruction of the precise cascades and networks by which the trafficking proteins regulate cargo transport processes within this parasite. *In silico* characterization of these proteins will also open exciting avenues for *in vivo* molecular probing to discover any novel functions of these trafficking proteins in *Giardia* and in the context of general eukaryotic endomembrane systems, both of which have been successfully demonstrated in this thesis.

The molecular functional approach taken in this thesis to elucidate protein mechanisms in *Giardia* was performed through microscopy and proteomics. However, gene deletion remains the gold-standard approach to correlate protein function with derived phenotypes in a specific pathway or organelle. Perhaps the most critical limitation of performing molecular functional studies in *Giardia* currently is the inability to achieve complete and efficient gene knockouts through otherwise highly selective genome editing approaches such as the type II CRISPR/Cas9 system or homologous recombination. Although CRISPR, as well as other strategies such as RNAi and morpholinos, have been successful in gene knockdown for silencing effects, full-knockout has only been achieved in one instance (Carpenter & Cande, 2009; Horáčková et al., 2021; Lin et al., 2019; Marcial-Quino et al., 2017; McNally et al., 2019). CRISPR

ineffectiveness in *Giardia* is a multivariate problem. First is the lack of non-homologous end-joining machinery (NHEJ) and a sexual cycle that precludes Cas9 nuclease-mediated break and repair of the double-stranded DNA (Jinek et al., 2012; Morrison et al., 2007; Poxleitner et al., 2008). The second is that problems exist with appropriate translocation of the Cas9 nuclease to both nuclei using standard nuclear localization signal (Ebnetter et al., 2016). Finally, the parasite's tetraploidy hampers efficient targeting of all four copies of a gene of interest by the guide RNA (Ebnetter et al., 2016). Instead, the Cre/loxP-mediated homologous recombination represents the only successful method that has achieved complete ablation of cyst-wall protein-1 through sequential deletion of all four encoded gene copies (Ebnetter et al., 2016; Wampfler et al., 2014). This method is, however, tedious and unsuitable for routine usage. A recent report has detailed efficient and successful gene knockout of several *Giardia*-specific genes using an improved 'self-propagating' CRISPR/Cas9 system (Horáčková et al., 2021). Thus, future investigations should further test this new technology and implement it to generate trafficking gene knockouts to determine different phenotypic effects on organellar morphologies and how they hamper cargo transport.

Finally, improved approaches to assess molecular localization and protein interactions should also be considered. Although laser scanning confocal microscopy in these investigations allowed for inference of approximate organellar localization, specific membrane assessments were not possible at the current resolution, especially at the PVs, which are small in diameter (50 to 150 nm), and the intervening cytosolic space between the plasma membrane and the PVs. Advancements in high-resolution microscopy techniques have allowed the mapping of precise protein localizations at specific organelles. Electron and ion-based microscopy, such as FIB-SEM, have successfully reconstructed the structural dynamics of intracellular compartments through their three-dimensional rendition (Kizilyaprak et al., 2014). These, combined with super-resolution immunofluorescence approaches such as gSTED, PALM/STORM, or correlative light and electron microscopy (CLEM), should be explored as viable routes for nanometer-resolution imaging especially for proteins investigated in this thesis that are likely membrane-associating (*i.e.*, ARF regulatory proteins and the ESCRTs) (Betzig et al., 2006; de Boer et al., 2015; Hell & Wichmann, 1994). In Chapters 4 and 5, the approach for capturing protein-protein interactions was made through *in vivo* chemical cross-linking of complexes. One of the disadvantages of this technique is that during cell disruption and membrane solubilizing steps, weak protein-protein interactions or protein-membrane interactions may be lost. In the case of the ARF GTPases, the kinetics of the GEF and GAP-mediated cycling of GTP and GDP is rapid and transient. Therefore, the co-IP approach may not have been sufficient to cross-link these interactions for enrichment at sufficient thresholds. This issue is notoriously difficult within the field of ARF biology in general. Therefore, alternative approaches such as the proximity-dependent biotinylation method (BioID) using an *E. coli* derived BirA ligase have been pursued to reconstruct interactome networks of the ARF regulatory system in model organisms (Chan et al., 2019; Gillingham et al., 2019). This BioID approach of determining protein-protein interactions in *Giardia* has also been attempted previously; however it has been difficult to reproduce (Martincová et al., 2015). In order to circumvent some of these issues, mass-spectrometry and density-based approaches such as hyper-LOPIT

should be considered as an alternative strategy for steady-state spatial proteomics and mapping of cellular pathways without requiring aggressive cellular disruption or mutational strategies (Mulvey et al., 2017). This method has proven to be highly successful in parasites such as *Toxoplasma* to accurately recapitulate high-density networks of protein-interaction profiles at individual organelles, including the parasite-specific apicoplast, rhoptries, and micronemes (Barylyuk et al., 2020). Similar experiments should be pursued to elucidate protein expression, functions, and trafficking networks at the *Giardia* ER, ESVs, and PVs in vegetative trophozoites and encysting cells.

Historically, technical limitations have impeded our understanding of the biology and evolution of this parasite. However, with rapid advancements in molecular technologies in recent years, new and exciting toolkits have become available to study organisms that are not tractable through traditional cell biological approaches. These methods should be employed in *Giardia* to test the various hypotheses laid out in this chapter, as the significance of the findings from these investigations would lie beyond illuminating this parasite's basic cell biology. Nevertheless, the results from this thesis have on their own advanced our understanding of the evolution and unconventional dynamics of key trafficking proteins to shape a modified endomembrane system in this pathogen. Overall, this thesis also highlights a continual need to pursue studies of similar scope in *Giardia*, as biology ultimately underpins pathogenesis and clinical disease outcomes.

7.5 Clinical and public health translational significance

It is unequivocal that *Giardia*'s trafficking organelles are at a critical interface between host and parasite for material exchange and secretion of parasite-specific factors to establish virulence and for environmental transmission. Therefore, advancing our understanding of the biology of these organelles and the underlying cellular processes is crucial because basic cell biological discoveries are a cornerstone in translational research and public health outcomes.

Currently, there are no approved vaccines or pharmacotherapeutics that target *Giardia*-specific factors for treating the disease in humans. Identifying the molecular machinery that largely remains uncharacterized is the principal first step in this process. Through comparative genomics, the work presented here has identified numerous lineage-specific membrane trafficking markers that are absent elsewhere in the eukaryotic tree and this parasite's hosts (*i.e.*, animals and humans). By determining how these proteins behave and their essentiality within the giardial cellular landscape, novel pathways for drug targeting can be uncovered. This approach of targeting *Giardia*-specific proteins has been used to develop the first Giardiasis vaccine for veterinary use. The vaccination strategy entailed inoculation of parasites in companion animals where *Giardia*'s RNAi machinery was repressed, resulting in the surface expression of a complete variant-surface protein repertoire in the trophozoites. This in turn, elicited a strong humoral immune response and monoclonal antibody production in the host (Serradell et al., 2016). Immunization in dogs also prevented recurrent transmission and alleviated chronic Giardiasis, and impressively reduced the zoonotic transmission and infections in children residing in endemic regions (Serradell et al., 2016).

Giardiasis causes an immense burden in regions that face healthcare inequities, and as a result, the disease has impacts that lie beyond the health of the community and manifest into long-term socio-economic stagnancy and regression. Increasing rates of global antimicrobial resistance in clinical settings have exacerbated the overall endemicity and recurrence of protist infections in general. This is also the case for Giardiasis, where increased treatment failures against the first-line therapeutics (*e.g.*, metronidazole and other nitroimidazoles) are observed. Additionally, a paucity of new treatments has made it difficult to reduce the global incidence of disease (Escobedo et al., 2010; Delfino et al., 2016; Lalle & Hanevik, 2018; Leitsch, 2015). Breakthroughs in understanding parasite cell biology are significant for the development of new drugs that are urgently needed. In the case of *Giardia*, promising avenues for pharmaco-targeting can be exploited by disrupting protein pathways that ensure proper functioning or biogenesis of *Giardia*-specific compartments such as ESVs or PVs. This would severely impede the transition between the trophozoite and the cyst stages.

Understanding genetic variabilities and shedding light on biological dissimilarities between the *Giardia intestinalis* assemblages is also essential from a public health standpoint. This would allow for an improved assessment of host tropism and pathogenicity to develop and implement strategic measures for epidemiological surveillance and disease prevention. Improved modeling of which assemblages are outbreak-associated and have the greatest anthrozoönotic potential becomes critical, especially in endemic regions where parasite transmission occurs readily between humans and animals. Identifying variations in virulence potentials would also permit improved clinical diagnostics for early intervention to prevent severe disease outcomes in patients. Overall, neglected diseases such as Giardiasis benefit immensely from research advances in molecular, genomic, and evolutionary aspects of parasite biology. Ultimately, this is pertinent to limiting the spread and endemicity of these preventable infectious diseases to improve the lives and livelihoods of millions of people affected globally.

7.6 Conclusions

A systems approach investigation was taken to determine the molecular evolution and roles of the vesicle formation machinery belonging to the gut parasite *Giardia intestinalis*. This diarrhea-causing pathogen is enigmatic from evolutionary and cellular standpoints due to its reduced overall biology, such as its membrane trafficking system. The work presented here, combined with the previous body of knowledge, substantially improves our understanding of the evolutionary path taken to yield lineage-specific compartments and the versatility of protein machinery associated with them. Based on the results from this work, the mode of eukaryotic parasitism, in general, expands to include notions of adaptive modulation in free-living ancestors to prime and permit the transition to a host-obligate lifestyle. The simple and enigmatic endomembrane system in this parasite also has the utility of being an interesting cell biological model to investigate conventional and atypical aspects of general eukaryotic trafficking. Overall, the work compiled in this thesis provides many leaps into our understanding of the evolutionary and molecular aspects that

underpin existing cellular intricacies within *Giardia*, a parasite that is of paramount biomedical and clinical importance.

REFERENCES

- Abodeely, M., DuBois, K. N., Hehl, A., Stefanic, S., Sajid, M., deSouza, W., Attias, M., Engel, J. C., Hsieh, I., Fetter, R. D., & McKerrow, J. H. (2009). A Contiguous Compartment Functions as Endoplasmic Reticulum and Endosome/Lysosome in *Giardia lamblia*. *Eukaryotic Cell*, 8(11), 1665–1676. <https://doi.org/10.1128/EC.00123-09>
- Achkar, J. M., & Fries, B. C. (2010). *Candida* infections of the genitourinary tract. *Clinical Microbiology Reviews*, 23(2), 253–273. <https://doi.org/10.1128/CMR.00076-09>
- Adam, R. D. (1992). Chromosome-size variation in *Giardia lamblia*: The role of rDNA repeats. *Nucleic Acids Research*, 20(12), 3057–3061. <https://doi.org/10.1093/nar/20.12.3057>
- Adam, R. D., Dahlstrom, E. W., Martens, C. A., Bruno, D. P., Barbian, K. D., Ricklefs, S. M., Hernandez, M. M., Narla, N. P., Patel, R. B., Porcella, S. F., & Nash, T. E. (2013). Genome Sequencing of *Giardia lamblia* Genotypes A2 and B Isolates (DH and GS) and Comparative Analysis with the Genomes of Genotypes A1 and E (WB and Pig). *Genome Biology and Evolution*, 5(12), 2498–2511. <https://doi.org/10.1093/gbe/evt197>
- Adam, R. D., Nigam, A., Seshadri, V., Martens, C. A., Farneth, G. A., Morrison, H. G., Nash, T. E., Porcella, S. F., & Patel, R. (2010). The *Giardia lamblia* vsp gene repertoire: Characteristics, genomic organization, and evolution. *BMC Genomics*, 11, 424.
- Adam, R. D. (2001). Biology of *Giardia lamblia*. *Clinical Microbiology Reviews*, 14(3), 447–475. <https://doi.org/10.1128/CMR.14.3.447-475.2001>
- Adams, E. (2014). Mammalian COPII Coat Component Sec24C Is Required for Embryonic Development in Mice. *Journal of Biological Chemistry*, 289(30), P20858-20870
- Adl, S. M., Bass, D., Lane, C. E., Lukeš, J., Schoch, C. L., Smirnov, A., Agatha, S., Berney, C., Brown, M. W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., ... Zhang, Q. (2019). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology*, 66(1), 4–119. <https://doi.org/10.1111/jeu.12691>
- Aebi, M. (2013). N-linked protein glycosylation in the ER. *Functional and Structural Diversity of the Endoplasmic Reticulum*, 1833(11), 2430–2437. <https://doi.org/10.1016/j.bbamcr.2013.04.001>
- Allain, T., & Buret, A. G. (2020). Chapter Five—Pathogenesis and post-infectious complications in giardiasis. In M. G. Ortega-Pierres (Ed.), *Advances in Parasitology* (Vol. 107, pp. 173–199). Academic Press. <https://doi.org/10.1016/bs.apar.2019.12.001>
- Allain, T., Fekete, E., & Buret, A. G. (2019). *Giardia* Cysteine Proteases: The Teeth behind the Smile. *Trends in Parasitology*, 35(8), 636–648. <https://doi.org/10.1016/j.pt.2019.06.003>
- Allan, B. B., Moyer, B. D., & Balch, W. E. (2000). Rab1 Recruitment of p115 into a cis-SNARE Complex: Programming Budding COPII Vesicles for Fusion. *Science*, 289(5478), 444. <https://doi.org/10.1126/science.289.5478.444>

- Al-Mohammed, H. I. (2011). Genotypes of *Giardia intestinalis* clinical isolates of gastrointestinal symptomatic and asymptomatic Saudi children. *Parasitology Research*, 108(6), 1375–1381. <https://doi.org/10.1007/s00436-010-2033-5>
- Alsford, S., Eckert, S., Baker, N., Glover, L., Sanchez-Flores, A., Leung, K. F., Turner, D. J., Field, M. C., Berriman, M., & Horn, D. (2012). High-throughput decoding of antitrypanosomal drug efficacy and resistance. *Nature*, 482(7384), 232–236. <https://doi.org/10.1038/nature10771>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). *Basic Local Alignment Search Tool*. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Anding, A. L., Wang, C., Chang, T.-K., Sliter, D. A., Powers, C. M., Hofmann, K., Youle, R. J., & Baehrecke, E. H. (2018). Vps13D Encodes a Ubiquitin-Binding Protein that Is Required for the Regulation of Mitochondrial Size and Clearance. *Current Biology*, 28(2), 287–295.e6. <https://doi.org/10.1016/j.cub.2017.11.064>
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Ankarklev, J., Franzén, O., Peirasmaki, D., Jerlström-Hultqvist, J., Lebbad, M., Andersson, J., Andersson, B., & Svärd, S. G. (2015). Comparative genomic analyses of freshly isolated *Giardia intestinalis* assemblage A isolates. *BMC Genomics*, 16(1), 697. <https://doi.org/10.1186/s12864-015-1893-6>
- Ankarklev, J., Jerlström-Hultqvist, J., Ringqvist, E., Troell, K., & Svärd, S. G. (2010). Behind the smile: Cell biology and disease mechanisms of *Giardia* species. *Nature Reviews Microbiology*, 8(6), 413–422. <https://doi.org/10.1038/nrmicro2317>
- Archibald, J. M. (2015). Endosymbiosis and Eukaryotic Cell Evolution. *Current Biology*, 25(19), R911–R921. <https://doi.org/10.1016/j.cub.2015.07.055>
- Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A., & Sebastian Seung, H. (2017). Trainable Weka Segmentation: A machine learning tool for microscopy pixel classification. *Bioinformatics*, 33(15), 2424–2426. <https://doi.org/10.1093/bioinformatics/btx180>
- Arseculeratne, S. N. (2005). Rhinosporidiosis: What is the cause? *Current Opinion in Infectious Diseases*, 18(2).
- Balch, W. E., Dunphy, W. G., Braell, W. A., & Rothman, J. E. (1984). Reconstitution of the transport of protein between successive compartments of the golgi measured by the coupled incorporation of N-acetylglucosamine. *Cell*, 39(2), 405–416. [https://doi.org/10.1016/0092-8674\(84\)90019-9](https://doi.org/10.1016/0092-8674(84)90019-9)
- Balmer, E. A., & Faso, C. (2021). The Road Less Traveled? Unconventional Protein Secretion at Parasite-Host Interfaces. *Frontiers in Cell and Developmental Biology*, 9, 662711–662711. <https://doi.org/10.3389/fcell.2021.662711>
- Banik, S., Renner Viveros, P., Seeber, F., Klotz, C., Ignatius, R., & Aebischer, T. (2013). *Giardia duodenalis* Arginine Deiminase Modulates the Phenotype and Cytokine Secretion of Human

- Dendritic Cells by Depletion of Arginine and Formation of Ammonia. *Infection and Immunity*, 81(7), 2309–2317. <https://doi.org/10.1128/IAI.00004-13>
- Bannykh, S. I., Rowe, T., & Balch, W. E. (1996). The organization of endoplasmic reticulum export complexes. *The Journal of Cell Biology*, 135(1), 19–35. <https://doi.org/10.1083/jcb.135.1.19>
- Barash, N. R., Nosala, C., Pham, J. K., McNally, S. G., Gourguechon, S., McCarthy-Sinclair, B., & Dawson, S. C. (2017). *Giardia* Colonizes and Encysts in High-Density Foci in the Murine Small Intestine. *MSphere*, 2(3), e00343-16. <https://doi.org/10.1128/mSphere.00343-16>
- Barlow, L. D., Dacks, J. B., & Wideman, J. G. (2014). From all to (nearly) none: Tracing adaptin evolution in Fungi. *Cellular Logistics*, 4(1), e28114–e28114. <https://doi.org/10.4161/cl.28114>
- Barlow, L. D., Nývltová, E., Aguilar, M., Tachezy, J., & Dacks, J. B. (2018). A sophisticated, differentiated Golgi in the ancestor of eukaryotes. *BMC Biology*, 16(1), 27. <https://doi.org/10.1186/s12915-018-0492-9>
- Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama, N., Rexach, M. F., Ravazzola, M., Amherdt, M., & Schekman, R. (1994). COPII: A membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell*, 77(6), 895–907. [https://doi.org/10.1016/0092-8674\(94\)90138-4](https://doi.org/10.1016/0092-8674(94)90138-4)
- Barois, N., & Bakke, O. (2005). The adaptor protein AP-4 as a component of the clathrin coat machinery: A morphological study. *The Biochemical Journal*, 385(Pt 2), 503–510. PubMed. <https://doi.org/10.1042/BJ20041010>
- Barylyuk, K., Koreny, L., Ke, H., Butterworth, S., Crook, O. M., Lassadi, I., Gupta, V., Tromer, E., Mourier, T., Stevens, T. J., Breckels, L. M., Pain, A., Lilley, K. S., & Waller, R. F. (2020). A Comprehensive Subcellular Atlas of the Toxoplasma Proteome via hyperLOPIT Provides Spatial Context for Protein Functions. *Cell Host & Microbe*, 28(5), 752-766.e9. <https://doi.org/10.1016/j.chom.2020.09.011>
- Bauer, I., Brune, T., Preiss, R., & Kölling, R. (2015). Evidence for a Nonendosomal Function of the *Saccharomyces cerevisiae* ESCRT-III-Like Protein Chm7. *Genetics*, 201(4), 1439–1452. <https://doi.org/10.1534/genetics.115.178939>
- Beck, R., Ravet, M., Wieland, F. T., & Cassel, D. (2009). The COPI system: Molecular mechanisms and function. *FEBS Letters*, 583(17), 2701–2709. <https://doi.org/10.1016/j.febslet.2009.07.032>
- Benchimol, M., & De Souza, W. (2011). The Ultrastructure of *Giardia* During Growth and Differentiation. In H. D. Luján & S. Svärd (Eds.), *Giardia: A Model Organism* (pp. 141–160). Springer Vienna. https://doi.org/10.1007/978-3-7091-0198-8_9
- Benchimol, M., Piva, B., Campanati, L., & de Souza, W. (2004). Visualization of the funis of *Giardia lamblia* by high-resolution field emission scanning electron microscopy—New insights. *Journal of Structural Biology*, 147(2), 102–115. <https://doi.org/10.1016/j.jsb.2004.01.017>

- Bernander, R., Palm, J. E. D., & Svärd, S. G. (2001). Genome ploidy in different stages of the *Giardia lamblia* life cycle. *Cellular Microbiology*, 3(1), 55–62. <https://doi.org/10.1046/j.1462-5822.2001.00094.x>
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Böhme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C. M., Arrowsmith, C., ... El-Sayed, N. M. (2005). The Genome of the African Trypanosome *Trypanosoma brucei*. *Science*, 309(5733), 416. <https://doi.org/10.1126/science.1112642>
- Bertrand, I., Albertini, L., & Schwartzbrod, J. (2005). Comparison of Two Target Genes for Detection and Genotyping of *Giardia lamblia* in Human Feces by PCR and PCR-Restriction Fragment Length Polymorphism. *Journal of Clinical Microbiology*, 43(12), 5940–5944. <https://doi.org/10.1128/JCM.43.12.5940-5944.2005>
- Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., & Hess, H. F. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, 313(5793), 1642. <https://doi.org/10.1126/science.1127344>
- Beznoussenko, G. V., Dolgikh, V. V., Seliverstova, E. V., Semenov, P. B., Tokarev, Y. S., Trucco, A., Micaroni, M., Di Giandomenico, D., Auinger, P., Senderskiy, I. V., Skarlato, S. O., Snigirevskaya, E. S., Komissarchik, Y. Yu., Pavelka, M., De Matteis, M. A., Luini, A., Sokolova, Y. Ya., & Mironov, A. A. (2007). Analogs of the Golgi complex in Microsporidia: Structure and vesicular mechanisms of function. *Journal of Cell Science*, 120(7), 1288–1298. <https://doi.org/10.1242/jcs.03402>
- Bigay, J., Casella, J.-F., Drin, G., Mesmin, B., & Antonny, B. (2005). ArfGAP1 responds to membrane curvature through the folding of a lipid packing sensor motif. *The EMBO Journal*, 24(13), 2244–2253. <https://doi.org/10.1038/sj.emboj.7600714>
- Blader, I. J., Coleman, B. I., Chen, C.-T., & Gubbels, M.-J. (2015). Lytic Cycle of *Toxoplasma gondii*: 15 Years Later. *Annual Review of Microbiology*, 69(1), 463–485. <https://doi.org/10.1146/annurev-micro-091014-104100>
- Blobel, G. (1980). Intracellular protein topogenesis. *Proceedings of the National Academy of Sciences*, 77(3), 1496. <https://doi.org/10.1073/pnas.77.3.1496>
- Boehm, M., Aguilar, R. C., & Bonifacino, J. S. (2001). Functional and physical interactions of the adaptor protein complex AP-4 with ADP-ribosylation factors (ARFs). *The EMBO Journal*, 20(22), 6265–6276. <https://doi.org/10.1093/emboj/20.22.6265>
- Boettner, D. R., Huston, C. D., Linford, A. S., Buss, S. N., Houpt, E., Sherman, N. E., & Petri, W. A., Jr. (2008). *Entamoeba histolytica* Phagocytosis of Human Erythrocytes Involves PATMK, a Member of the Transmembrane Kinase Family. *PLOS Pathogens*, 4(1), e8. <https://doi.org/10.1371/journal.ppat.0040008>

- Bonifacino, J. S., & Glick, B. S. (2004). The Mechanisms of Vesicle Budding and Fusion. *Cell*, 116(2), 153–166. [https://doi.org/10.1016/S0092-8674\(03\)01079-1](https://doi.org/10.1016/S0092-8674(03)01079-1)
- Bonifacino, J. S., & Lippincott-Schwartz, J. (2003). Coat proteins: Shaping membrane transport. *Nature Reviews Molecular Cell Biology*, 4(5), 409–414. <https://doi.org/10.1038/nrm1099>
- Boucher, S. E., & Gillin, F. D. (1990). Excystation of in vitro-derived *Giardia lamblia* cysts. *Infection and Immunity*, 58(11), 3516–3522. <https://doi.org/10.1128/iai.58.11.3516-3522.1990>
- Bowman, S. L., Bi-Karchin, J., Le, L., & Marks, M. S. (2019). The road to lysosome-related organelles: Insights from Hermansky-Pudlak syndrome and other rare diseases. *Traffic*, 20(6), 404–435. <https://doi.org/10.1111/tra.12646>
- Braakman, I., & Hebert, D. N. (2013). Protein folding in the endoplasmic reticulum. *Cold Spring Harbor Perspectives in Biology*, 5(5), a013201–a013201. <https://doi.org/10.1101/cshperspect.a013201>
- Bredeston, L. M., Caffaro, C. E., Samuelson, J., & Hirschberg, C. B. (2005). Golgi and Endoplasmic Reticulum Functions Take Place in Different Subcellular Compartments of *Entamoeba histolytica*. *Journal of Biological Chemistry*, 280(37), 32168–32176. <https://doi.org/10.1074/jbc.M507035200>
- Bröcker, C., Engelbrecht-Vandré, S., & Ungermann, C. (2010). Multisubunit Tethering Complexes and Their Role in Membrane Fusion. *Current Biology*, 20(21), R943–R952. <https://doi.org/10.1016/j.cub.2010.09.015>
- Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., & White, T. C. (2012). Hidden Killers: Human Fungal Infections. *Science Translational Medicine*, 4(165), 165rv13. <https://doi.org/10.1126/scitranslmed.3004404>
- Burd, C., & Cullen, P. J. (2014). Retromer: A master conductor of endosome sorting. *Cold Spring Harbor Perspectives in Biology*, 6(2), a016774. <https://doi.org/10.1101/cshperspect.a016774>
- Buret, A. G., & Cotton, J. (2011). Pathophysiological Processes and Clinical Manifestations of Giardiasis. In H. D. Luján & S. Svärd (Eds.), *Giardia: A Model Organism* (pp. 301–318). Springer Vienna. https://doi.org/10.1007/978-3-7091-0198-8_19
- Burgess, J., Jauregui, M., Tan, J., Rollins, J., Lallet, S., Leventis, P. A., Boulianne, G. L., Chang, H. C., Le Borgne, R., Krämer, H., & Brill, J. A. (2011). AP-1 and clathrin are essential for secretory granule biogenesis in *Drosophila*. *Molecular Biology of the Cell*, 22(12), 2094–2105. <https://doi.org/10.1091/mbc.E11-01-0054>
- Burgos, P. V., Mardones, G. A., Rojas, A. L., daSilva, L. L. P., Prabhu, Y., Hurley, J. H., & Bonifacino, J. S. (2010). Sorting of the Alzheimer's disease amyloid precursor protein mediated by the AP-4 complex. *Developmental Cell*, 18(3), 425–436. <https://doi.org/10.1016/j.devcel.2010.01.015>
- Burki, F., Roger, A. J., Brown, M. W., & Simpson, A. G. B. (2020). The New Tree of Eukaryotes. *Trends in Ecology & Evolution*, 35(1), 43–55. <https://doi.org/10.1016/j.tree.2019.08.008>
- Byrd, L. G., Conrad, J. T., & Nash, T. E. (1994). *Giardia lamblia* infections in adult mice. *Infection and Immunity*, 62(8), 3583–3585. <https://doi.org/10.1128/iai.62.8.3583-3585.1994>

- Cacciò, S. M., Beck, R., Lalle, M., Marinculic, A., & Pozio, E. (2008). Multilocus genotyping of *Giardia duodenalis* reveals striking differences between assemblages A and B. *International Journal for Parasitology*, 38(13), 1523–1531. <https://doi.org/10.1016/j.ijpara.2008.04.008>
- Cacciò, S. M., & Ryan, U. (2008). Molecular epidemiology of giardiasis. *Molecular and Biochemical Parasitology*, 160(2), 75–80. <https://doi.org/10.1016/j.molbiopara.2008.04.006>
- Cai, Y., Chin, H. F., Lazarova, D., Menon, S., Fu, C., Cai, H., Sclafani, A., Rodgers, D. W., De La Cruz, E. M., Ferro-Novick, S., & Reinisch, K. M. (2008). The Structural Basis for Activation of the Rab Ypt1p by the TRAPP Membrane-Tethering Complexes. *Cell*, 133(7), 1202–1213. <https://doi.org/10.1016/j.cell.2008.04.049>
- Camacho, E., González-de la Fuente, S., Rastrojo, A., Peiró-Pastor, R., Solana, J. C., Tabera, L., Gamarro, F., Carrasco-Ramiro, F., Requena, J. M., & Aguado, B. (2019). Complete assembly of the *Leishmania donovani* (HU3 strain) genome and transcriptome annotation. *Scientific Reports*, 9(1), 6127. <https://doi.org/10.1038/s41598-019-42511-4>
- Capella-Gutiérrez, S., Marcet-Houben, M., & Gabaldón, T. (2012). Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC Biology*, 10(1), 47. <https://doi.org/10.1186/1741-7007-10-47>
- Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C. M., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., ... Johnson, P. J. (2007). Draft Genome Sequence of the Sexually Transmitted Pathogen *Trichomonas vaginalis*. *Science*, 315(5809), 207–212. <https://doi.org/10.1126/science.1132894>
- Carpenter, M. L., & Cande, W. Z. (2009). Using Morpholinos for Gene Knockdown in *Giardia intestinalis*. *Eukaryotic Cell*, 8(6), 916–919. <https://doi.org/10.1128/EC.00041-09>
- Carruthers, V. B. (2002). Host cell invasion by the opportunistic pathogen *Toxoplasma gondii*. *Acta Tropica*, 81(2), 111–122. [https://doi.org/10.1016/S0001-706X\(01\)00201-7](https://doi.org/10.1016/S0001-706X(01)00201-7)
- Carter, R. F. (1970). Description of a *Naegleria* sp. Isolated from two cases of primary amoebic meningo-encephalitis, and of the experimental pathological changes induced by it. *The Journal of Pathology*, 100(4), 217–244. <https://doi.org/10.1002/path.1711000402>
- Castillo-Romero, A., Leon-Avila, G., Wang, C. C., Perez Rangel, A., Camacho Nuez, M., Garcia Tovar, C., Ayala-Summano, J. T., Luna-Arias, J. P., & Hernandez, J. M. (2010). Rab11 and Actin Cytoskeleton Participate in *Giardia lamblia* Encystation, Guiding the Specific Vesicles to the Cyst-wall. *PLoS Neglected Tropical Diseases*, 4(6), e697. <https://doi.org/10.1371/journal.pntd.0000697>
- Cavalier-Smith, T. (1989). Archaeobacteria and Archezoa. *Nature*, 339(6220), 100–101. <https://doi.org/10.1038/339100a0>
- Cavalier-Smith, T. (1993). Kingdom protozoa and its 18 phyla. *Microbiological Reviews*, 57(4), 953–994. PubMed. <https://doi.org/10.1128/mr.57.4.953-994.1993>

- Cavalier-Smith, T., Fiore-Donno, A. M., Chao, E., Kudryavtsev, A., Berney, C., Snell, E. A., & Lewis, R. (2015). Multigene phylogeny resolves deep branching of Amoebozoa. *Molecular Phylogenetics and Evolution*, 83, 293–304. <https://doi.org/10.1016/j.ympev.2014.08.011>
- Cavenagh, M. M., Whitney, J. A., Carroll, K., Zhang, C., Boman, A. L., Rosenwald, A. G., Mellman, I., & Kahn, R. A. (1996). Intracellular Distribution of Arf Proteins in Mammalian Cells. *Journal of Biological Chemistry*, 271(36), 21767–21774. <https://doi.org/10.1074/jbc.271.36.21767>
- Cavet, M. E., Pang, J., Yin, G., & Berk, B. C. (2008). An epidermal growth factor (EGF) -dependent interaction between GIT1 and sorting nexin 6 promotes degradation of the EGF receptor. *The FASEB Journal*, 22(10), 3607–3616. <https://doi.org/10.1096/fj.07-094086>
- Cepicka, I., Hampl, V., & Kulda, J. (2010). Critical Taxonomic Revision of Parabasalids with Description of one New Genus and three New Species. *Protist*, 161(3), 400–433. <https://doi.org/10.1016/j.protis.2009.11.005>
- Cernikova, L., Faso, C., & Hehl, A. B. (2020). Phosphoinositide-binding proteins mark, shape and functionally modulate highly-diverged endocytic compartments in the parasitic protist *Giardia lamblia*. *PLOS Pathogens*, 16(2), e1008317. <https://doi.org/10.1371/journal.ppat.1008317>
- Chan, C. J., Le, R., Burns, K., Ahmed, K., Coyaud, E., Laurent, E. M. N., Raught, B., & Melançon, P. (2019). BioID Performed on Golgi Enriched Fractions Identify C10orf76 as a GBF1 Binding Protein Essential for Golgi Maintenance and Secretion. *Molecular & Cellular Proteomics: MCP*, 18(11), 2285–2297. <https://doi.org/10.1074/mcp.RA119.001645>
- Chatterjee, A., Carpentieri, A., Ratner, D. M., Bullitt, E., Costello, C. E., Robbins, P. W., & Samuelson, J. (2010). *Giardia* Cyst-wall Protein 1 Is a Lectin That Binds to Curled Fibrils of the GalNAc Homopolymer. *PLOS Pathogens*, 6(8), e1001059. <https://doi.org/10.1371/journal.ppat.1001059>
- Chen, K.-E., Healy, M. D., & Collins, B. M. (2019). Towards a molecular understanding of endosomal trafficking by Retromer and Retriever. *Traffic*, 20(7), 465–478. <https://doi.org/10.1111/tra.12649>
- Chen, P.-W., Randazzo, P. A., & Parent, C. A. (2010). ACAP-A/B Are ArfGAP Homologs in *Dictyostelium* Involved in Sporulation but Not in Chemotaxis. *PLOS ONE*, 5(1), e8624. <https://doi.org/10.1371/journal.pone.0008624>
- Chinappi, M., Via, A., Marcatili, P., & Tramontano, A. (2010). On the Mechanism of Chloroquine Resistance in *Plasmodium falciparum*. *PLOS One*, 5(11), e14064. <https://doi.org/10.1371/journal.pone.0014064>
- Chun, J., Shapovalova, Z., Dejgaard, S. Y., Presley, J. F., & Melançon, P. (2008). Characterization of Class I and II ADP-Ribosylation Factors (Arfs) in Live Cells: GDP-bound Class II Arfs Associate with the ER-Golgi Intermediate Compartment Independently of GBF1. *Molecular Biology of the Cell*, 19(8), 3488–3500. <https://doi.org/10.1091/mbc.e08-04-0373>
- Clark, C. G., Kaffashian, F., Tawari, B., Windsor, J. J., Twigg-Flesner, A., Davies-Morel, M. C. G., Blessmann, J., Ebert, F., Peschel, B., Van, A. L., Jackson, C. J., Macfarlane, L., & Tannich, E. (2006). New insights into the phylogeny of *Entamoeba* species provided by analysis of four new

- small-subunit rRNA genes. *International Journal of Systematic and Evolutionary Microbiology*, 56(9), 2235–2239. <https://doi.org/10.1099/ijs.0.64208-0>
- Coelho, C. H., & Singer, S. M. (2018). Recent advances in the *Giardia*–host relationship reveal danger lurking behind the smile. *PLOS Neglected Tropical Diseases*, 12(9), e0006625. <https://doi.org/10.1371/journal.pntd.0006625>
- Cogliati, M. (2013). Global Molecular Epidemiology of *Cryptococcus neoformans* and *Cryptococcus gattii*: An Atlas of the Molecular Types. *Scientifica*, 2013, 675213–675213. PubMed. <https://doi.org/10.1155/2013/675213>
- Cole, E. S., Giddings, T. H., Ozzello, C., Winey, M., O'Toole, E., Orias, J., Hamilton E., Guerrier, S., Ballard, A., & Aronstein, T. (2015). Membrane Dynamics at the Nuclear Exchange Junction during Early Mating (One to Four Hours) in the Ciliate *Tetrahymena thermophila*. *Eukaryotic Cell*, 14(2), 116–127. <https://doi.org/10.1128/EC.00164-14>
- Colicelli, J. (2004). Human RAS superfamily proteins and related GTPases. *Science's STKE : Signal Transduction Knowledge Environment*, 2004(250), RE13–RE13. PubMed. <https://doi.org/10.1126/stke.2502004re13>
- Conner, S. D., & Schmid, S. L. (2003). Differential requirements for AP-2 in clathrin-mediated endocytosis. *Journal of Cell Biology*, 162(5), 773–780. <https://doi.org/10.1083/jcb.200304069>
- Cook, W. J., Smith, C. D., Senkovich, O., Holder, A. A., & Chattopadhyay, D. (2010). Structure of *Plasmodium falciparum* ADP-ribosylation factor 1. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, 66(11), 1426–1431. <https://doi.org/10.1107/S1744309110036997>
- Corvera, S., D'Arrigo, A., & Stenmark, H. (1999). Phosphoinositides in membrane traffic. *Current Opinion in Cell Biology*, 11(4), 460–465. [https://doi.org/10.1016/S0955-0674\(99\)80066-0](https://doi.org/10.1016/S0955-0674(99)80066-0)
- Costes, S. V., Daelemans, D., Cho, E. H., Dobbin, Z., Pavlakis, G., & Lockett, S. (2004). Automatic and Quantitative Measurement of Protein-Protein Co-localization in Live Cells. *Biophysical Journal*, 86(6), 3993–4003. <https://doi.org/10.1529/biophysj.103.038422>
- Cotton, J. A., Beatty, J. K., & Buret, A. G. (2011). Host parasite interactions and pathophysiology in *Giardia* infections. *International Journal for Parasitology*, 41(9), 925–933. <https://doi.org/10.1016/j.ijpara.2011.05.002>
- Cross, G. A. M. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. *Parasitology*, 71(3), 393–417. Cambridge Core. <https://doi.org/10.1017/S003118200004717X>
- Cullen, P. J., & Steinberg, F. (2018). To degrade or not to degrade: Mechanisms and significance of endocytic recycling. *Nature Reviews Molecular Cell Biology*, 19(11), 679–696. <https://doi.org/10.1038/s41580-018-0053-7>
- Current, W. L., & Garcia, L. S. (1991). Cryptosporidiosis. *Clinical Microbiology Reviews*, 4(3), 325–358. <https://doi.org/10.1128/CMR.4.3.325>

- Curtis, B. A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M. C., Ball, S. G., Gile, G. H., Hirakawa, Y., Hopkins, J. F., Kuo, A., Rensing, S. A., Schmutz, J., Symeonidi, A., Elias, M., Eveleigh, R. J. M., Herman, E. K., Klute, M. J., ... Archibald, J. M. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature*, *492*(7427), 59–65.
<https://doi.org/10.1038/nature11681>
- Dacks, J. B., & Field, M. C. (2007). Evolution of the eukaryotic membrane-trafficking system: Origin, tempo and mode. *Journal of Cell Science*, *120*(17), 2977–2985.
<https://doi.org/10.1242/jcs.013250>
- Dacks, J. B., & Field, M. C. (2018). Evolutionary origins and specialisation of membrane transport. *Membrane Trafficking*, *53*, 70–76. <https://doi.org/10.1016/j.ceb.2018.06.001>
- Dacks, J. B., Field, M. C., Buick, R., Eme, L., Gribaldo, S., Roger, A. J., Brochier-Armanet, C., & Devos, D. P. (2016). The changing view of eukaryogenesis – fossils, cells, lineages and how they all come together. *Journal of Cell Science*, *129*(20), 3695–3703. <https://doi.org/10.1242/jcs.178566>
- Dacks, J. B., Poon, P. P., & Field, M. C. (2008). Phylogeny of endocytic components yields insight into the process of nonendosymbiotic organelle evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(2), 588–593.
<https://doi.org/10.1073/pnas.0707318105>
- Dacks, J. B., & Robinson, M. S. (2017). Outerwear through the ages: Evolutionary cell biology of vesicle coats. *Cell Organelles*, *47*, 108–116. <https://doi.org/10.1016/j.ceb.2017.04.001>
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics*, *27*(8), 1164–1165.
<https://doi.org/10.1093/bioinformatics/btr088>
- Daumke, O., Roux, A., & Haucke, V. (2014). BAR Domain Scaffolds in Dynamin-Mediated Membrane Fission. *Cell*, *156*(5), 882–892. <https://doi.org/10.1016/j.cell.2014.02.017>
- Davids, B. J., Reiner, D. S., Birkeland, S. R., Preheim, S. P., Cipriano, M. J., McArthur, A. G., & Gillin, F. D. (2006). A New Family of Giardial Cysteine-Rich Non-VSP Protein Genes and a Novel Cyst Protein. *PLOS One*, *1*(1), e44. <https://doi.org/10.1371/journal.pone.0000044>
- Dawson, S. C., & House, S. A. (2010). Life with eight flagella: Flagellar assembly and division in *Giardia*. *Host–Microbe Interactions: Fungi/Parasites/Viruses*, *13*(4), 480–490.
<https://doi.org/10.1016/j.mib.2010.05.014>
- de Boer, P., Hoogenboom, J. P., & Giepmans, B. N. G. (2015). Correlated light and electron microscopy: Ultrastructure lights up! *Nature Methods*, *12*(6), 503–513. <https://doi.org/10.1038/nmeth.3400>
- De Cádiz, A. E., Jeelani, G., Nakada-Tsukui, K., Caler, E., & Nozaki, T. (2013). Transcriptome Analysis of Encystation in *Entamoeba invadens*. *PLOS One*, *8*(9), e74840.
<https://doi.org/10.1371/journal.pone.0074840>

- De Jonckheere, J. F. (2011). Origin and evolution of the worldwide distributed pathogenic amoeboflagellate *Naegleria fowleri*. *Infection, Genetics and Evolution*, 11(7), 1520–1528. <https://doi.org/10.1016/j.meegid.2011.07.023>
- De Jonckheere, J. F. (2014). What do we know by now about the genus *Naegleria*? *SI: FLAM_2013*, 145, S2–S9. <https://doi.org/10.1016/j.exppara.2014.07.011>
- de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, 102, 6600. <https://doi.org/10.1073/pnas.0502030102>
- De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, 56(6), 879–886. <https://doi.org/10.1080/10635150701701083>
- Delfino, B. M., Campos, R. G., Pereira, T. M., Mantovani, S. A. S., Oliart-Guzmán, H., Martins, A. C., Braña, A. M., Branco, F. L. C. C., Filgueira-Júnior, J. A., Santos, A. P., Araújo, T. S., Oliveira, C. S. M., Ramalho, A. A., Muniz, P. T., Codeço, C. T., & da Silva-Nunes, M. (2016). Evolution of Socioeconomic Conditions and Its Relation to Spatial–Temporal Changes of Giardiasis and Helminthiasis in Amazonian Children. *EcoHealth*, 13(4), 743–760. <https://doi.org/10.1007/s10393-016-1167-z>
- Delprato, A., & Lambright, D. G. (2007). Structural basis for Rab GTPase activation by VPS9 domain exchange factors. *Nature Structural & Molecular Biology*, 14(5), 406–412. <https://doi.org/10.1038/nsmb1232>
- Deretic, D. (2013). Crosstalk of Arf and Rab GTPases en route to cilia. *Small GTPases*, 4(2), 70–77. <https://doi.org/10.4161/sgtp.24396>
- Dergai, M., Iershov, A., Novokhatska, O., Pankivskiy, S., & Rynditch, A. (2016). Evolutionary Changes on the Way to Clathrin-Mediated Endocytosis in Animals. *Genome Biology and Evolution*, 8(3), 588–606. <https://doi.org/10.1093/gbe/evw028>
- Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B. T., Sali, A., & Rout, M. P. (2004). Components of Coated Vesicles and Nuclear Pore Complexes Share a Common Molecular Architecture. *PLOS Biology*, 2(12), e380. <https://doi.org/10.1371/journal.pbio.0020380>
- Didier, E. S. (2005). Microsporidiosis: An emerging and opportunistic infection in humans and animals. *Acta Tropica*, 94(1), 61–76. <https://doi.org/10.1016/j.actatropica.2005.01.010>
- Didier, E. S., & Weiss, L. M. (2006). Microsporidiosis: Current status. *Current Opinion in Infectious Diseases*, 19(5), 485–492. <https://doi.org/10.1097/01.qco.0000244055.46382.23>
- Dobell, C. (1920). The Discovery of the Intestinal Protozoa of Man. *Proceedings of the Royal Society of Medicine*, 13(Sect Hist Med), 1–15.
- Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Amselem, J., Bouri, L., Bocs, S., Klopp, C., Gibrat, J.-F., Vlasova, A., Leskosek, B. L., Soler, L., Binzer-Panchal, M., & Lantz, H. (2018). Ten steps to get started in Genome Assembly and Annotation. *F1000Research*, 7, 148. <https://doi.org/10.12688/f1000research.13598.1>

- Donaldson, J. G., & Jackson, C. L. (2011). ARF family G proteins and their regulators: Roles in membrane transport, development and disease. *Nature Reviews Molecular Cell Biology*, 12(6), 362–375. <https://doi.org/10.1038/nrm3117>
- Drouin, E., Piloquet, P., & Péréon, Y. (2015). The first illustrations of neurons by Camillo Golgi. *The Lancet Neurology*, 14(6), 567. [https://doi.org/10.1016/S1474-4422\(15\)00051-4](https://doi.org/10.1016/S1474-4422(15)00051-4)
- Dubey, J. P., & Jones, J. L. (2008). *Toxoplasma gondii* infection in humans and animals in the United States. *Cover Image © Copyright of Frank Balthis (Http://Www.Digitalrailroad.Net/FrankBalthis)*, 38(11), 1257–1278. <https://doi.org/10.1016/j.ijpara.2008.03.007>
- Dubey, J. P., Lindsay, D. S., & Speer, C. A. (1998). Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. *Clinical Microbiology Reviews*, 11(2), 267–299. <https://doi.org/10.1128/CMR.11.2.267>
- Dunn, K. W., Kamocka, M. M., & McDonald, J. H. (2011). A practical guide to evaluating co-localization in biological microscopy. *American Journal of Physiology. Cell Physiology*, 300(4), C723–C742. <https://doi.org/10.1152/ajpcell.00462.2010>
- Dutta, S., Saha, N., Ray, A., & Sarkar, S. (2015). Significantly Diverged Did2/Vps46 Orthologues from the Protozoan Parasite *Giardia lamblia*. *Current Microbiology*, 71(3), 333–340. <https://doi.org/10.1007/s00284-015-0844-4>
- Earl, D., Bradnam, K., St John, J., Darling, A., Lin, D., Fass, J., Yu, H. O. K., Buffalo, V., Zerbino, D. R., Diekhans, M., Nguyen, N., Ariyaratne, P. N., Sung, W.-K., Ning, Z., Haimel, M., Simpson, J. T., Fonseca, N. A., Birol, I., Docking, T. R., ... Paten, B. (2011). Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Research*, 21(12), 2224–2241. <https://doi.org/10.1101/gr.126599.111>
- Ebenezer, T. E., Zoltner, M., Burrell, A., Nenarokova, A., Novák Vanclová, A. M. G., Prasad, B., Soukal, P., Santana-Molina, C., O'Neill, E., Nankissoor, N. N., Vadakedath, N., Daiker, V., Obado, S., Silva-Pereira, S., Jackson, A. P., Devos, D. P., Lukeš, J., Lebert, M., Vaughan, S., ... Field, M. C. (2019). Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biology*, 17(1), 11. <https://doi.org/10.1186/s12915-019-0626-8>
- Ebneter, J. A., Heusser, S. D., Schraner, E. M., Hehl, A. B., & Faso, C. (2016). Cyst-Wall-Protein-1 is fundamental for Golgi-like organelle neogenesis and cyst-wall biosynthesis in *Giardia lamblia*. *Nature Communications*, 7(1), 13859. <https://doi.org/10.1038/ncomms13859>
- Eckmann, L., Laurent, F., Langford, T. D., Hetsko, M. L., Smith, J. R., Kagnoff, M. F., & Gillin, F. D. (2000). Nitric Oxide Production by Human Intestinal Epithelial Cells and Competition for Arginine as Potential Determinants of Host Defense Against the Lumen-Dwelling Pathogen *Giardia lamblia*. *The Journal of Immunology*, 164(3), 1478–1487. <https://doi.org/10.4049/jimmunol.164.3.1478>
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9), 755–763. <https://doi.org/10.1093/bioinformatics/14.9.755>

- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5(1), 113. <https://doi.org/10.1186/1471-2105-5-113>
- Eichinger, L., Pachebat, J. A., Glöckner, G., Rajandream, M.-A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., Tunggal, B., Kummerfeld, S., Madera, M., Konfortov, B. A., Rivero, F., Bankier, A. T., Lehmann, R., Hamlin, N., Davies, R., ... Kuspa, A. (2005). The genome of the social amoeba *Dictyostelium discoideum*. *Nature*, 435(7038), 43–57. <https://doi.org/10.1038/nature03481>
- Einarsson, E., Ástvaldsson, Á., Hultenby, K., Andersson, J. O., Svärd, S. G., & Jerlström-Hultqvist, J. (2016). Comparative Cell Biology and Evolution of Annexins in Diplomonads. *MSphere*, 1(2), mSphere.00032-15, e00032-15. <https://doi.org/10.1128/mSphere.00032-15>
- Einarsson, E., & Svärd, S. G. (2015). Encystation of *Giardia intestinalis*—A Journey from the Duodenum to the Colon. *Current Tropical Medicine Reports*, 2(3), 101–109. <https://doi.org/10.1007/s40475-015-0048-9>
- Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M., Tallon, L. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Farzad, M., Carlton, J. M., Smith, R. K., Jr., ... Orias, E. (2006). Macronuclear Genome Sequence of the Ciliate *Tetrahymena thermophila*, a Model Eukaryote. *PLOS Biology*, 4(9), e286. <https://doi.org/10.1371/journal.pbio.0040286>
- Elias, M., Brighouse, A., Gabernet-Castello, C., Field, M. C., & Dacks, J. B. (2012). Sculpting the endomembrane system in deep time: High resolution phylogenetics of Rab GTPases. *Journal of Cell Science*, 125(10), 2500–2508. <https://doi.org/10.1242/jcs.101378>
- Elkin, S. R., Lakoduk, A. M., & Schmid, S. L. (2016). Endocytic pathways and endosomal trafficking: A primer. *Wiener Medizinische Wochenschrift*, 166(7), 196–204. <https://doi.org/10.1007/s10354-016-0432-7>
- Ellis, R. J., Ellis, R. J., Laskey, R. A., & Lorimer, G. H. (1993). The general concept of molecular chaperones. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 339(1289), 257–261. <https://doi.org/10.1098/rstb.1993.0023>
- El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A.-N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Åslund, L., Attipoe, P., ... Andersson, B. (2005). The Genome Sequence of *Trypanosoma cruzi*, Etiologic Agent of Chagas Disease. *Science*, 309(5733), 409. <https://doi.org/10.1126/science.1112631>
- Eme, L., Spang, A., Lombard, J., Stairs, C. W., & Ettema, T. J. G. (2017). Archaea and the origin of eukaryotes. *Nature Reviews Microbiology*, 15(12), 711–723. <https://doi.org/10.1038/nrmicro.2017.133>

- Emery, S. J., Lacey, E., & Haynes, P. A. (2016). Quantitative proteomics in *Giardia duodenalis*—Achievements and challenges. *Molecular and Biochemical Parasitology*, 208(2), 96–112. <https://doi.org/10.1016/j.molbiopara.2016.07.002>
- Engel, T., Lueken, A., Bode, G., Hobohm, U., Lorkowski, S., Schlueter, B., Rust, S., Cullen, P., Pech, M., Assmann, G., & Seedorf, U. (2004). ADP-ribosylation factor (ARF)-like 7 (ARL7) is induced by cholesterol loading and participates in apolipoprotein AI-dependent cholesterol export. *FEBS Letters*, 566(1–3), 241–246. <https://doi.org/10.1016/j.febslet.2004.04.048>
- Erlandsen, S. L., Macechko, P. T., Keulen, H. V., & Jarroll, E. L. (1996). Formation of the *Giardia* Cyst-wall: Studies on Extracellular Assembly Using Immunogold Labeling and High Resolution Field Emission SEM. *Journal of Eukaryotic Microbiology*, 43(5), 416–430. <https://doi.org/10.1111/j.1550-7408.1996.tb05053.x>
- Escobedo, A., Almirall, P., J. Robertson, L., M.B. Franco, R., Hanevik, K., Morch, K., & Cimerman, S. (2010). Giardiasis: The Ever-Present Threat of a Neglected Disease. *Infectious Disorders - Drug Targets*, 10(5), 329–348. <https://doi.org/10.2174/187152610793180821>
- Esch, K. J., & Petersen, C. A. (2013). Transmission and Epidemiology of Zoonotic Protozoal Diseases of Companion Animals. *Clinical Microbiology Reviews*, 26(1), 58–85. <https://doi.org/10.1128/CMR.00067-12>
- Fagone, P., & Jackowski, S. (2009). Membrane phospholipid synthesis and endoplasmic reticulum function. *Journal of Lipid Research*, 50, S311–S316. <https://doi.org/10.1194/jlr.R800049-JLR200>
- Fantinatti, M., Bello, A. R., Fernandes, O., & Da-Cruz, A. M. (2016). Identification of *Giardia lamblia* Assemblage E in Humans Points to a New Anthrozoönotic Cycle. *The Journal of Infectious Diseases*, 214(8), 1256–1259. <https://doi.org/10.1093/infdis/jiw361>
- Faso, C., & Hehl, A. B. (2011). Membrane trafficking and organelle biogenesis in *Giardia lamblia*: Use it or lose it. *International Journal for Parasitology*, 41(5), 471–480. <https://doi.org/10.1016/j.ijpara.2010.12.014>
- Faso, C., & Hehl, A. B. (2019). A cytonaut's guide to protein trafficking in *Giardia lamblia*. In *Advances in Parasitology* (Vol. 106, pp. 105–127). Elsevier. <https://doi.org/10.1016/bs.apar.2019.08.001>
- Faso, C., Konrad, C., Schraner, E. M., & Hehl, A. B. (2013). Export of cyst-wall material and Golgi organelle neogenesis in *Giardia lamblia* depend on endoplasmic reticulum exit sites: ER exit sites in *Giardia lamblia*. *Cellular Microbiology*, 15(4), 537–553. <https://doi.org/10.1111/cmi.12054>
- Fasshauer, D., Sutton, R. B., Brunger, A. T., & Jahn, R. (1998). Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proceedings of the National Academy of Sciences*, 95(26), 15781. <https://doi.org/10.1073/pnas.95.26.15781>
- Faubert, G. (2000). Immune Response to *Giardia duodenalis*. *Clinical Microbiology Reviews*, 13(1), 35–54. <https://doi.org/10.1128/CMR.13.1.35>

- Feely, D. E., & Dyer, J. K. (1987). Localization of Acid Phosphatase Activity in *Giardia lamblia* and *Giardia muris* Trophozoites. *The Journal of Protozoology*, *34*(1), 80–83. <https://doi.org/10.1111/j.1550-7408.1987.tb03137.x>
- Felsenstein, J. (1985). Confidence Limits on Phylogenies: An approach using the Bootstrap. *Evolution*, *39*(4), 783–791. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x>
- Felsenstein, J. (1989). *PHYLIP-Phylogeny Inference Package (Ver. 3.2)*. 5, 164–166.
- Feng, Y., & Xiao, L. (2011). Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis. *Clinical Microbiology Reviews*, *24*(1), 110–140. <https://doi.org/10.1128/CMR.00033-10>
- Ferguson, L. C., Smith-Palmer, A., & Alexander, C. L. (2020). An update on the incidence of human giardiasis in Scotland, 2011–2018. *Parasites & Vectors*, *13*(1), 291. <https://doi.org/10.1186/s13071-020-04160-9>
- Field, M. C., Sali, A., & Rout, M. P. (2011). Evolution: On a bender—BARs, ESCRTs, COPs, and finally getting your coat. *The Journal of Cell Biology*, *193*(6), 963–972. <https://doi.org/10.1083/jcb.201102042>
- Fisk, D. G., Ball, C. A., Dolinski, K., Engel, S. R., Hong, E. L., Issel-Tarver, L., Schwartz, K., Sethuraman, A., Botstein, D., & Michael Cherry, J. (2006). *Saccharomyces cerevisiae* S288C genome annotation: A working hypothesis. *Yeast*, *23*(12), 857–865. <https://doi.org/10.1002/yea.1400>
- Flanagan, P. A. (1992). *Giardia*—Diagnosis, clinical course and epidemiology. A review. *Epidemiology and Infection*, *109*(1), 1–22.
- Franzén, O., Jerlström-Hultqvist, J., Castro, E., Sherwood, E., Ankarklev, J., Reiner, D. S., Palm, D., Andersson, J. O., Andersson, B., & Svärd, S. G. (2009). Draft Genome Sequencing of *Giardia intestinalis* Assemblage B Isolate GS: Is Human Giardiasis Caused by Two Different Species? *PLoS Pathogens*, *5*(8), e1000560. <https://doi.org/10.1371/journal.ppat.1000560>
- Friend, D. S. (1966). The fine structure of *Giardia muris*. *The Journal of Cell Biology*, *29*(2), 317–332. <https://doi.org/10.1083/jcb.29.2.317>
- Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J. B., Carpenter, M. L., Field, M. C., Kuo, A., Paredes, A., Chapman, J., Pham, J., Shu, S., Neupane, R., Cipriano, M., Mancuso, J., Tu, H., Salamov, A., Lindquist, E., Shapiro, H., Lucas, S., ... Dawson, S. C. (2010). The Genome of *Naegleria gruberi* Illuminates Early Eukaryotic Versatility. *Cell*, *140*(5), 631–642. <https://doi.org/10.1016/j.cell.2010.01.032>
- Fuji, K., Shirakawa, M., Shimono, Y., Kunieda, T., Fukao, Y., Koumoto, Y., Takahashi, H., Hara-Nishimura, I., & Shimada, T. (2016). The Adaptor Complex AP-4 Regulates Vacuolar Protein Sorting at the trans-Golgi Network by Interacting with Vacuolar Protein Sorting. *Plant Physiology*, *170*(1), 211–219. JSTOR.
- Gabernet-Castello, C., O'Reilly, A. J., Dacks, J. B., & Field, M. C. (2013). Evolution of Tre-2/Bub2/Cdc16 (TBC) Rab GTPase-activating proteins. *Molecular Biology of the Cell*, *24*(10), 1574–1583. <https://doi.org/10.1091/mbc.e12-07-0557>

- Gabler, F., Nam, S.-Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A. N., & Alva, V. (2020). Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics*, 72(1), e108. <https://doi.org/10.1002/cpbi.108>
- Gadeyne, A., Sánchez-Rodríguez, C., Vanneste, S., Di Rubbo, S., Zauber, H., Vanneste, K., Van Leene, J., De Winne, N., Eeckhout, D., Persiau, G., Van De Slijke, E., Cannoot, B., Vercruyse, L., Mayers, J. R., Adamowski, M., Kania, U., Ehrlich, M., Schweighofer, A., Ketelaar, T., ... Van Damme, D. (2014). The TPLATE Adaptor Complex Drives Clathrin-Mediated Endocytosis in Plants. *Cell*, 156(4), 691–704. <https://doi.org/10.1016/j.cell.2014.01.039>
- Gaechter, V., Schraner, E., Wild, P., & Hehl, A. B. (2008). The Single Dynamin Family Protein in the Primitive Protozoan *Giardia lamblia* Is Essential for Stage Conversion and Endocytic Transport. *Traffic*, 9(1), 57–71. <https://doi.org/10.1111/j.1600-0854.2007.00657.x>
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., ... Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906), 498–511. <https://doi.org/10.1038/nature01097>
- Gardner, T. B. & Hill, D. R. (2001). Treatment of Giardiasis. *Clinical Microbiology Reviews*, 14(1), 114–128. <https://doi.org/10.1128/CMR.14.1.114-128.2001>
- Gargantini, P. R., Serradell, M. del C., Ríos, D. N., Tenaglia, A. H., & Luján, H. D. (2016). Antigenic variation in the intestinal parasite *Giardia lamblia*. *Host-Microbe Interactions: Parasites/Fungi/Viruses*, 32, 52–58. <https://doi.org/10.1016/j.mib.2016.04.017>
- Geldner, N., Anders, N., Wolters, H., Keicher, J., Kornberger, W., Muller, P., Delbarre, A., Ueda, T., Nakano, A., & Jürgens, G. (2003). The *Arabidopsis* GNOM ARF-GEF Mediates Endosomal Recycling, Auxin Transport, and Auxin-Dependent Plant Growth. *Cell*, 112(2), 219–230. [https://doi.org/10.1016/S0092-8674\(03\)00003-5](https://doi.org/10.1016/S0092-8674(03)00003-5)
- Gill, D. J., Teo, H., Sun, J., Perisic, O., Veprintsev, D. B., Emr, S. D., & Williams, R. L. (2007). Structural insight into the ESCRT-I/II link and its role in MVB trafficking. *The EMBO Journal*, 26(2), 600–612. <https://doi.org/10.1038/sj.emboj.7601501>
- Gill, E. E., & Fast, N. M. (2006). Assessing the microsporidia-fungi relationship: Combined phylogenetic analysis of eight genes. *Gene*, 375, 103–109. <https://doi.org/10.1016/j.gene.2006.02.023>
- Gillingham, A. K., Bertram, J., Begum, F., & Munro, S. (2019). *In vivo* identification of GTPase interactors by mitochondrial relocalization and proximity biotinylation. *ELife*, 8, e45916. <https://doi.org/10.7554/eLife.45916>
- Githeko, A. K., Lindsay, S. W., Confalonieri, U. E., & Patz, J. A. (2000). Climate change and vector-borne diseases: A regional analysis. *Bulletin of the World Health Organization*, 78(9), 1136–1147.
- Glick, B. S. (2014). Integrated self-organization of transitional ER and early Golgi compartments. *BioEssays*, 36(2), 129–133. <https://doi.org/10.1002/bies.201300131>

- Goldberg, J. (1999). Structural and Functional Analysis of the ARF1–ARFGAP Complex Reveals a Role for Coatamer in GTP Hydrolysis. *Cell*, 96(6), 893–902. [https://doi.org/10.1016/S0092-8674\(00\)80598-X](https://doi.org/10.1016/S0092-8674(00)80598-X)
- Gray, M. W. (1992). The Endosymbiont Hypothesis Revisited. In D. R. Wolstenholme & K. W. Jeon (Eds.), *International Review of Cytology* (Vol. 141, pp. 233–357). Academic Press. [https://doi.org/10.1016/S0074-7696\(08\)62068-9](https://doi.org/10.1016/S0074-7696(08)62068-9)
- Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harbor Perspectives in Biology*, 4(9), a011403–a011403. <https://doi.org/10.1101/cshperspect.a011403>
- Gu, F., & Gruenberg, J. (2000). ARF1 Regulates pH-dependent COP Functions in the Early Endocytic Pathway. *Journal of Biological Chemistry*, 275(11), 8154–8160. <https://doi.org/10.1074/jbc.275.11.8154>
- Gu, M., LaJoie, D., Chen, O. S., von Appen, A., Ladinsky, M. S., Redd, M. J., Nikolova, L., Bjorkman, P. J., Sundquist, W. I., Ullman, K. S., & Frost, A. (2017). LEM2 recruits CHMP7 for ESCRT-mediated nuclear envelope closure in fission yeast and human cells. *Proceedings of the National Academy of Sciences*, 114(11), E2166. <https://doi.org/10.1073/pnas.1613916114>
- Guetta, D., Langou, K., Grunwald, D., Klein, G., & Aubry, L. (2010). FYVE-Dependent Endosomal Targeting of an Arrestin-Related Protein in Amoeba. *PLOS One*, 5(12), e15249. <https://doi.org/10.1371/journal.pone.0015249>
- Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLOS Genetics*, 15(7), e1008302. <https://doi.org/10.1371/journal.pgen.1008302>
- Guo, Y., Sirkis, D. W., & Schekman, R. (2014). Protein Sorting at the trans-Golgi Network. *Annual Review of Cell and Developmental Biology*, 30(1), 169–206. <https://doi.org/10.1146/annurev-cellbio-100913-013012>
- Gupta, A., & Rawlings, J. B. (2014). Comparison of Parameter Estimation Methods in Stochastic Chemical Kinetic Models: Examples in Systems Biology. *AIChE Journal. American Institute of Chemical Engineers*, 60(4), 1253–1268. <https://doi.org/10.1002/aic.14409>
- Gupta, S. N., Kloster, M. M., Rodionov, D. G., & Bakke, O. (2006). Re-routing of the invariant chain to the direct sorting pathway by introduction of an AP3-binding motif from LIMP II. *European Journal of Cell Biology*, 85(6), 457–467. <https://doi.org/10.1016/j.ejcb.2006.02.001>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, 29(8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Hallée, S., Thériault, C., Gagnon, D., Kehrer, J., Frischknecht, F., Mair, G. R., & Richard, D. (2018). Identification of a Golgi apparatus protein complex important for the asexual erythrocytic cycle of the malaria parasite *Plasmodium falciparum*. *Cellular Microbiology*, 20(8), e12843. <https://doi.org/10.1111/cmi.12843>

- Halliez, M. C. M., & Buret, A. G. (2013). Extra-intestinal and long term consequences of *Giardia duodenalis* infections. *World Journal of Gastroenterology*, *19*(47), 8974–8985. PubMed. <https://doi.org/10.3748/wjg.v19.i47.8974>
- Hammerling, B. C., Najor, R. H., Cortez, M. Q., Shires, S. E., Leon, L. J., Gonzalez, E. R., Boassa, D., Phan, S., Thor, A., Jimenez, R. E., Li, H., Kitsis, R. N., Dorn II, G. W., Sadoshima, J., Ellisman, M. H., & Gustafsson, Å. B. (2017). A Rab5 endosomal pathway mediates Parkin-dependent mitochondrial clearance. *Nature Communications*, *8*(1), 14050. <https://doi.org/10.1038/ncomms14050>
- Hardin, W. R., Li, R., Xu, J., Shelton, A. M., Alas, G. C. M., Minin, V. N., & Paredez, A. R. (2017). Myosin-independent cytokinesis in *Giardia* utilizes flagella to coordinate force generation and direct membrane trafficking. *Proceedings of the National Academy of Sciences*, *114*(29), E5854. <https://doi.org/10.1073/pnas.1705096114>
- Health Canada. (2019). *Guidelines for Canadian drinking water quality: Guideline technical document: enteric protozoa: Giardia and Cryptosporidium*. http://epe.lac-bac.gc.ca/100/201/301/weekly_acquisitions_list-ef/2019/19-16/publications.gc.ca/collections/collection_2019/sc-hc/H144-13-10-2018-eng.pdf
- Hegedűs, K., Takáts, S., Boda, A., Jipa, A., Nagy, P., Varga, K., Kovács, A. L., & Juhász, G. (2016). The Ccz1-Mon1-Rab7 module and Rab5 control distinct steps of autophagy. *Molecular Biology of the Cell*, *27*(20), 3132–3142. <https://doi.org/10.1091/mbc.e16-03-0205>
- Hehl, A. B., Marti, M., & Köhler, P. (2000). Stage-Specific Expression and Targeting of Cyst-wall Protein-Green Fluorescent Protein Chimeras in *Giardia*. *Molecular Biology of the Cell*, *11*(5), 1789–1800. <https://doi.org/10.1091/mbc.11.5.1789>
- Hell, S. W., & Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, *19*(11), 780–782. <https://doi.org/10.1364/OL.19.000780>
- Hellard, M. E., Sinclair, M. I., Hogg, G. G., & Fairley, C. K. (2000). Prevalence of enteric pathogens among community based asymptomatic individuals. *Journal of Gastroenterology and Hepatology*, *15*(3), 290–293. <https://doi.org/10.1046/j.1440-1746.2000.02089.x>
- Helms, J. B., & Rothman, J. E. (1992). Inhibition by brefeldin A of a Golgi membrane enzyme that catalyses exchange of guanine nucleotide bound to ARF. *Nature*, *360*(6402), 352–354. <https://doi.org/10.1038/360352a0>
- Herman, E. K. (2018). *Understanding the evolution of the membrane trafficking system in diverse eukaryotes through comparative genomics and transcriptomics*. 432.
- Herman, E. K., Ali, M., Field, M. C., & Dacks, J. B. (2018). Regulation of early endosomes across eukaryotes: Evolution and functional homology of Vps9 proteins. *Traffic*, *19*(7), 546–563. <https://doi.org/10.1111/tra.12570>

- Herman, E. K., Greninger, A., van der Giezen, M., Ginger, M. L., Ramirez-Macias, I., Miller, H. C., Morgan, M. J., Tsaousis, A. D., Velle, K., Vargová, R., Záhonová, K., Najle, S. R., MacIntyre, G., Muller, N., Wittwer, M., Zysset-Burri, D. C., Eliáš, M., Slamovits, C. H., Weirauch, M. T., ... Dacks, J. B. (2021). Genomics and transcriptomics yields a system-level view of the biology of the pathogen *Naegleria fowleri*. *BMC Biology*, *19*(1), 142. <https://doi.org/10.1186/s12915-021-01078-1>
- Herman, E. K., Yiangou, L., Cantoni, D. M., Miller, C. N., Marciano-Cabral, F., Anthonyrajah, E., Dacks, J. B., & Tsaousis, A. D. (2018). Identification and characterisation of a cryptic Golgi complex in *Naegleria gruberi*. *Journal of Cell Science*, *131*(7), jcs213306. <https://doi.org/10.1242/jcs.213306>
- Herman, E., Siegesmund, M. A., Bottery, M. J., van Aerle, R., Shather, M. M., Caler, E., Dacks, J. B., & van der Giezen, M. (2017). Membrane Trafficking Modulation during *Entamoeba* Encystation. *Scientific Reports*, *7*(1), 12854. <https://doi.org/10.1038/s41598-017-12875-6>
- Herr, R. A., Ajello, L., Taylor, J. W., Arseculeratne, S. N., & Mendoza, L. (1999). Phylogenetic analysis of *Rhinosporidium seeberi*'s 18S small-subunit ribosomal DNA groups this pathogen among members of the protoctistan Mesomycetozoa clade. *Journal of Clinical Microbiology*, *37*(9), 2750–2754. PubMed. <https://doi.org/10.1128/JCM.37.9.2750-2754.1999>
- Heun, Y., Gräff, P., Lagara, A., Schelhorn, R., Mettler, R., Pohl, U., & Mannell, H. (2020). The GEF Cytohesin-2/ARNO Mediates Resistin induced Phenotypic Switching in Vascular Smooth Muscle Cells. *Scientific Reports*, *10*(1), 3672. <https://doi.org/10.1038/s41598-020-60446-z>
- Heuser, J. E., & Reese, T. S. (1973). Evidence For Recycling of Synaptic Vesicle Membrane During Transmitter Release at the Frog Neuromuscular Junction. *Journal of Cell Biology*, *57*(2), 315–344. <https://doi.org/10.1083/jcb.57.2.315>
- Heyworth M F. (1986). Antibody response to *Giardia muris* trophozoites in mouse intestine. *Infection and Immunity*, *52*(2), 568–571. <https://doi.org/10.1128/iai.52.2.568-571.1986>
- Heyworth, M. F. (2016). *Giardia duodenalis* genetic assemblages and hosts. *Parasite (Paris, France)*, *23*, 13–13. <https://doi.org/10.1051/parasite/2016013>
- Hirst, J., Borner, G. H. H., Edgar, J., Hein, M. Y., Mann, M., Buchholz, F., Antrobus, R., & Robinson, M. S. (2013). Interaction between AP-5 and the hereditary spastic paraplegia proteins SPG11 and SPG15. *Molecular Biology of the Cell*, *24*(16), 2558–2569. <https://doi.org/10.1091/mbc.E13-03-0170>
- Hirst, J., D. Barlow, L., Francisco, G. C., Sahlender, D. A., Seaman, M. N. J., Dacks, J. B., & Robinson, M. S. (2011). The Fifth Adaptor Protein Complex. *PLOS Biology*, *9*(10), e1001170. <https://doi.org/10.1371/journal.pbio.1001170>
- Hirst, J., Irving, C., & Borner, G. H. H. (2013). Adaptor Protein Complexes AP-4 and AP-5: New Players in Endosomal Trafficking and Progressive Spastic Paraplegia. *Traffic*, *14*(2), 153–164. <https://doi.org/10.1111/tra.12028>

- Hirst, J., Itzhak, D. N., Antrobus, R., Borner, G. H. H., & Robinson, M. S. (2018). Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. *PLOS Biology*, *16*(1), e2004411. <https://doi.org/10.1371/journal.pbio.2004411>
- Hirst, J., Schlacht, A., Norcott, J. P., Traynor, D., Bloomfield, G., Antrobus, R., Kay, R. R., Dacks, J. B., & Robinson, M. S. (2014). Characterization of TSET, an ancient and widespread membrane trafficking complex. *eLife*, *3*, e02866. <https://doi.org/10.7554/eLife.02866>
- Hirt, R. P., Noel, C. J., Sicheritz-Ponten, T., Tachezy, J., & Fiori, P.-L. (2007). *Trichomonas vaginalis* surface proteins: A view from the genome. *Trends in Parasitology*, *23*(11), 540–547. <https://doi.org/10.1016/j.pt.2007.08.020>
- Honda, A., Al-Awar, O. S., Hay, J. C., & Donaldson, J. G. (2005). Targeting of Arf-1 to the early Golgi by membrin, an ER-Golgi SNARE. *Journal of Cell Biology*, *168*(7), 1039–1051. <https://doi.org/10.1083/jcb.200409138>
- Hoppins, S., Collins, S. R., Cassidy-Stone, A., Hummel, E., Devay, R. M., Lackner, L. L., Westermann, B., Schuldiner, M., Weissman, J. S., & Nunnari, J. (2011). A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria. *The Journal of Cell Biology*, *195*(2), 323–340. PubMed. <https://doi.org/10.1083/jcb.201107053>
- Horáčková, V., Voleman, L., Hagen, K. D., Petrů, M., Vinopalová, M., Weisz, F., Janowicz, N., Marková, L., Motyčková, A., Tůmová, P., Dawson, S. C., & Doležal, P. (2021). Cas9-mediated gene disruption in tetraploid *Giardia intestinalis*. *BioRxiv*, 2021.04.21.440745. <https://doi.org/10.1101/2021.04.21.440745>
- Horii, M., Shibata, H., Kobayashi, R., Katoh, K., Yorikawa, C., Yasuda, J., & Maki, M. (2006). CHMP7, a novel ESCRT-III-related protein, associates with CHMP4b and functions in the endosomal sorting pathway. *Biochemical Journal*, *400*(1), 23–32. <https://doi.org/10.1042/BJ20060897>
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- Huey, C. S., Mahdy, M. A. K., Al-Mekhlafi, H. M., Nasr, N. A., Lim, Y. A. L., Mahmud, R., & Surin, J. (2013). Multilocus genotyping of *Giardia duodenalis* in Malaysia. *Infection, Genetics and Evolution*, *17*, 269–276. <https://doi.org/10.1016/j.meegid.2013.04.013>
- Huotari, J., & Helenius, A. (2011). Endosome maturation. *The EMBO Journal*, *30*(17), 3481–3500. <https://doi.org/10.1038/emboj.2011.286>
- Hurley, J. H. (2015). ESCRTs are everywhere. *The EMBO Journal*, *34*(19), 2398–2407. <https://doi.org/10.15252/embj.201592484>
- Jackson, A. P., Otto, T. D., Aslett, M., Armstrong, S. D., Bringaud, F., Schlacht, A., Hartley, C., Sanders, M., Wastling, J. M., Dacks, J. B., Acosta-Serrano, A., Field, M. C., Ginger, M. L., & Berriman, M. (2016). Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the

- Origins of Parasitism. *Current Biology: CB*, 26(2), 161–172. PubMed.
<https://doi.org/10.1016/j.cub.2015.11.055>
- Jackson, C. L., & Bouvet, S. (2014). Arfs at a Glance. *Journal of Cell Science*, 127(19), 4103–4109.
<https://doi.org/10.1242/jcs.144899>
- Jahn, R. (2004). Principles of Exocytosis and Membrane Fusion. *Annals of the New York Academy of Sciences*, 1014(1), 170–178. <https://doi.org/10.1196/annals.1294.018>
- Jahn, R., & Scheller, R. H. (2006). SNAREs—Engines for membrane fusion. *Nature Reviews Molecular Cell Biology*, 7(9), 631–643. <https://doi.org/10.1038/nrm2002>
- Jahn, R., & Südhof, T. C. (1999). Membrane Fusion and Exocytosis. *Annual Review of Biochemistry*, 68(1), 863–911. <https://doi.org/10.1146/annurev.biochem.68.1.863>
- James, T. Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., & Stajich, J. E. (2013). Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia. *Current Biology*, 23(16), 1548–1553. <https://doi.org/10.1016/j.cub.2013.06.057>
- Jedelský, P. L., Doležal, P., Rada, P., Pyrih, J., Šmíd, O., Hrdý, I., Šedinová, M., Marcinčíková, M., Voleman, L., Perry, A. J., Beltrán, N. C., Lithgow, T., & Tachezy, J. (2011). The Minimal Proteome in the Reduced Mitochondrion of the Parasitic Protist *Giardia intestinalis*. *PLOS ONE*, 6(2), e17285. <https://doi.org/10.1371/journal.pone.0017285>
- Jerlström-Hultqvist, J., Ankarklev, J., & Svärd, S. G. (2010). Is human giardiasis caused by two different *Giardia* species? *Gut Microbes*, 1(6), 379–382. <https://doi.org/10.4161/gmic.1.6.13608>
- Jerlström-Hultqvist, J., Franzén, O., Ankarklev, J., Xu, F., Nohýnková, E., Andersson, J. O., Svärd, S. G., & Andersson, B. (2010). Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. *BMC Genomics*, 11(1), 543. <https://doi.org/10.1186/1471-2164-11-543>
- Jiménez-González, A., Andersson, J. O., & Lutz, H. L. (2020). Metabolic Reconstruction Elucidates the Lifestyle of the Last Diplomonadida Common Ancestor. *MSystems*, 5(6), e00774-20. <https://doi.org/10.1128/mSystems.00774-20>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), 816. <https://doi.org/10.1126/science.1225829>
- Johnson, A., Dahhan, D. A., Gnyliukh, N., Kaufmann, W. A., Zheden, V., Costanzo, T., Mahou, P., Hrtyan, M., Wang, J., Aguilera-Servin, J., van Damme, D., Beaurepaire, E., Loose, M., Bednarek, S. Y., & Friml, J. (2021). The TPLATE complex mediates membrane bending during plant clathrin-mediated endocytosis. *BioRxiv*, 2021.04.26.441441. <https://doi.org/10.1101/2021.04.26.441441>
- Joshi, A. S., Nebenfuehr, B., Choudhary, V., Satpute-Krishnan, P., Levine, T. P., Golden, A., & Prinz, W. A. (2018). Lipid droplet and peroxisome biogenesis occur at the same ER subdomains. *Nature Communications*, 9(1), 2940. <https://doi.org/10.1038/s41467-018-05277-3>
- Kahn, R. A. (2009). Toward a model for Arf GTPases as regulators of traffic at the Golgi. *FEBS Letters*, 583(23), 3872–3879. <https://doi.org/10.1016/j.febslet.2009.10.066>

- Kahn, R. A., Cherfils, J., Elias, M., Lovering, R. C., Munro, S., & Schurmann, A. (2006). Nomenclature for the human Arf family of GTP-binding proteins: ARF, ARL, and SAR proteins. *Journal of Cell Biology*, *172*(5), 645–650. <https://doi.org/10.1083/jcb.200512057>
- Kaksonen, M., & Roux, A. (2018). Mechanisms of clathrin-mediated endocytosis. *Nature Reviews Molecular Cell Biology*, *19*(5), 313–326. <https://doi.org/10.1038/nrm.2017.132>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kano, R. (2020). Emergence of Fungal-Like Organisms: Prototheca. *Mycopathologia*, *185*(5), 747–754. <https://doi.org/10.1007/s11046-019-00365-4>
- Karnkowska, A., Treitli, S. C., Brzoň, O., Novák, L., Vacek, V., Soukal, P., Barlow, L. D., Herman, E. K., Pipaliya, S. V., Pánek, T., Žihala, D., Petrželková, R., Butenko, A., Eme, L., Stairs, C. W., Roger, A. J., Eliáš, M., Dacks, J. B., & Hampl, V. (2019). The Oxymonad Genome Displays Canonical Eukaryotic Complexity in the Absence of a Mitochondrion. *Molecular Biology and Evolution*, *36*(10), 2292–2312. <https://doi.org/10.1093/molbev/msz147>
- Keeling, P. J. (1998). A kingdom's progress: Archezoa and the origin of eukaryotes. *BioEssays*, *20*(1), 87–95. [https://doi.org/10.1002/\(SICI\)1521-1878\(199801\)20:1<87::AID-BIES12>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1521-1878(199801)20:1<87::AID-BIES12>3.0.CO;2-4)
- Keeling, P. J., Burger, G., Durnford, D. G., Lang, B. F., Lee, R. W., Pearlman, R. E., Roger, A. J., & Gray, M. W. (2005). The tree of eukaryotes. *Trends in Ecology & Evolution*, *20*(12), 670–676. <https://doi.org/10.1016/j.tree.2005.09.005>
- Keeling, P. J., & Doolittle, W. F. (1997). Widespread and ancient distribution of a noncanonical genetic code in diplomonads. *Molecular Biology and Evolution*, *14*(9), 895–901. <https://doi.org/10.1093/oxfordjournals.molbev.a025832>
- Keeling, P. J., & Fast, N. M. (2002). Microsporidia: Biology and Evolution of Highly Reduced Intracellular Parasites. *Annual Review of Microbiology*, *56*(1), 93–116. <https://doi.org/10.1146/annurev.micro.56.012302.160854>
- Keeling, P. J., Fast, N. M., Law, J. S., Williams, B. A. P., & Slamovits, C. H. (2005). Comparative genomics of microsporidia. *Folia Parasitologica (Praha)*, *52*(1–2), 8–14. <https://doi.org/10.14411/fp.2005.002>
- Keeling, P. J., & Slamovits, C. H. (2005). Causes and effects of nuclear genome reduction. *Genomes and Evolution*, *15*(6), 601–608. <https://doi.org/10.1016/j.gde.2005.09.003>
- Keister, D. B. (1983). Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *77*(4), 487–488. [https://doi.org/10.1016/0035-9203\(83\)90120-7](https://doi.org/10.1016/0035-9203(83)90120-7)
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. <https://doi.org/10.1038/nprot.2015.053>

- Kim, J.-Y. (2016). Human fungal pathogens: Why should we learn? *Journal of Microbiology*, 54(3), 145–148. <https://doi.org/10.1007/s12275-016-0647-8>
- Kirchhausen, T. (1999). Adaptors for Clathrin-Mediated Traffic. *Annual Review of Cell and Developmental Biology*, 15(1), 705–732. <https://doi.org/10.1146/annurev.cellbio.15.1.705>
- Kirchhausen, T. (2000). Three ways to make a vesicle. *Nature Reviews Molecular Cell Biology*, 1(3), 187–198. <https://doi.org/10.1038/35043117>
- Kirk, M. D., Pires, S. M., Black, R. E., Caipo, M., Crump, J. A., Devleeschauwer, B., Döpfer, D., Fazil, A., Fischer-Walker, C. L., Hald, T., Hall, A. J., Keddy, K. H., Lake, R. J., Lanata, C. F., Torgerson, P. R., Havelaar, A. H., & Angulo, F. J. (2015). World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLOS Medicine*, 12(12), e1001921. <https://doi.org/10.1371/journal.pmed.1001921>
- Kissinger, P. (2015). *Trichomonas vaginalis*: A review of epidemiologic, clinical and treatment issues. *BMC Infectious Diseases*, 15(1), 307. <https://doi.org/10.1186/s12879-015-1055-0>
- Kitakura, S., Adamowski, M., Matsuura, Y., Santuari, L., Kouno, H., Arima, K., Hardtke, C. S., Friml, J., Kakimoto, T., & Tanaka, H. (2017). BEN3/BIG2 ARF GEF is Involved in Brefeldin A-Sensitive Trafficking at the trans-Golgi Network/Early Endosome in *Arabidopsis thaliana*. *Plant and Cell Physiology*, 58(10), 1801–1811. <https://doi.org/10.1093/pcp/pcx118>
- Kizilyaprak, C., Bittermann, A. G., Daraspe, J., & Humbel, B. M. (2014). FIB-SEM Tomography in Biology. In J. Kuo (Ed.), *Electron Microscopy: Methods and Protocols* (pp. 541–558). Humana Press. https://doi.org/10.1007/978-1-62703-776-1_24
- Kjaerulf, O., Verstreken, P., & Bellen, H. J. (2002). Synaptic vesicle retrieval: Still time for a kiss. *Nature Cell Biology*, 4(11), E245–E248. <https://doi.org/10.1038/ncb1102-e245>
- Kjos, I., Vestre, K., Guadagno, N. A., Borg Distefano, M., & Progida, C. (2018). Rab and Arf proteins at the crossroad between membrane transport and cytoskeleton dynamics. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1865(10), 1397–1409. <https://doi.org/10.1016/j.bbamcr.2018.07.009>
- Klinger, C. M., Klute, M. J., & Dacks, J. B. (2013). Comparative Genomic Analysis of Multi-Subunit Tethering Complexes Demonstrates an Ancient Pan-Eukaryotic Complement and Sculpting in Apicomplexa. *PLOS One*, 8(9), e76278. <https://doi.org/10.1371/journal.pone.0076278>
- Klug, D. M., Gelb, M. H., & Pollastri, M. P. (2016). Repurposing strategies for tropical disease drug discovery. *Bioorganic & Medicinal Chemistry Letters*, 26(11), 2569–2576. <https://doi.org/10.1016/j.bmcl.2016.03.103>
- Kobayashi, H., & Fukuda, M. (2012). Rab35 regulates Arf6 activity through centaurin β /ACAP2 during neurite outgrowth. *Journal of Cell Science*, jcs.098657. <https://doi.org/10.1242/jcs.098657>
- Kolisko, M., Cepicka, I., Hampl, V., Leigh, J., Roger, A. J., Kulda, J., Simpson, A. G., & Flegr, J. (2008). Molecular phylogeny of diplomonads and enteromonads based on SSU rRNA, alpha-tubulin and

- HSP90 genes: Implications for the evolutionary history of the double karyomastigont of diplomonads. *BMC Evolutionary Biology*, 8(1), 205. <https://doi.org/10.1186/1471-2148-8-205>
- Kolisko, M., Silberman, J. D., Cepicka, I., Yubuki, N., Takishita, K., Yabuki, A., Leander, B. S., Inouye, I., Inagaki, Y., Roger, A. J., & Simpson, A. G. B. (2010). A wide diversity of previously undetected free-living relatives of diplomonads isolated from marine/saline habitats. *Environmental Microbiology*, 12(10), 2700–2710. <https://doi.org/10.1111/j.1462-2920.2010.02239.x>
- Konrad, C., Spycher, C., & Hehl, A. B. (2010). Selective Condensation Drives Partitioning and Sequential Secretion of Cyst-wall Proteins in Differentiating *Giardia lamblia*. *PLOS Pathogens*, 6(4), e1000835. <https://doi.org/10.1371/journal.ppat.1000835>
- Kooyman, F. N. J., Wagenaar, J. A., & Zomer, A. (2019). Whole-genome sequencing of dog-specific assemblages C and D of *Giardia duodenalis* from single and pooled cysts indicates host-associated genes. *Microbial Genomics*, 5(12), e000302. <https://doi.org/10.1099/mgen.0.000302>
- Koumandou, V. L., Klute, M. J., Herman, E. K., Nunez-Miguel, R., Dacks, J. B., & Field, M. C. (2011). Evolutionary reconstruction of the retromer complex and its function in *Trypanosoma brucei*. *Journal of Cell Science*, 124(Pt 9), 1496–1509. <https://doi.org/10.1242/jcs.081596>
- Krauß, M., & Haucke, V. (2007). Phosphoinositide-metabolizing enzymes at the interface between membrane traffic and cell signalling. *EMBO Reports*, 8(3), 241–246. <https://doi.org/10.1038/sj.embor.7400919>
- Kremer, K., Kamin, D., Rittweger, E., Wilkes, J., Flammer, H., Mahler, S., Heng, J., Tonkin, C. J., Langsley, G., Hell, S. W., Carruthers, V. B., Ferguson, D. J. P., & Meissner, M. (2013). An Overexpression Screen of *Toxoplasma gondii* Rab-GTPases Reveals Distinct Transport Routes to the Micronemes. *PLOS Pathogens*, 9(3), e1003213. <https://doi.org/10.1371/journal.ppat.1003213>
- Krey, J. F., Wilmarth, P. A., Shin, J.-B., Klimek, J., Sherman, N. E., Jeffery, E. D., Choi, D., David, L. L., & Barr-Gillespie, P. G. (2014). Accurate label-free protein quantitation with high- and low-resolution mass spectrometers. *Journal of Proteome Research*, 13(2), 1034–1044. <https://doi.org/10.1021/pr401017h>
- Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, 47(D1), D807–D811. <https://doi.org/10.1093/nar/gky1053>
- Kulda, J., Nohýnková, E., & Čepička, I. (2017). Retortamonadida (with Notes on *Carpediemonas*-Like Organisms and Caviomonadidae). In J. M. Archibald, A. G. B. Simpson, C. H. Slamovits, L. Margulis, M. Melkonian, D. J. Chapman, & J. O. Corliss (Eds.), *Handbook of the Protists* (pp. 1–32). Springer International Publishing. https://doi.org/10.1007/978-3-319-32669-6_3-1
- Kung, L. F., Pagant, S., Futai, E., D’Arcangelo, J. G., Buchanan, R., Dittmar, J. C., Reid, R. J. D., Rothstein, R., Hamamoto, S., Snapp, E. L., Schekman, R., & Miller, E. A. (2012). Sec24p and

- Sec16p cooperate to regulate the GTP cycle of the COPII coat. *The EMBO Journal*, 31(4), 1014–1027. <https://doi.org/10.1038/emboj.2011.444>
- Kural, C., Tacheva-Grigorova, S. K., Boulant, S., Cocucci, E., Baust, T., Duarte, D., & Kirchhausen, T. (2012). Dynamics of Intracellular Clathrin/AP1- and Clathrin/AP3-Containing Carriers. *Cell Reports*, 2(5), 1111–1119. <https://doi.org/10.1016/j.celrep.2012.09.025>
- Lalle, M., & Hanevik, K. (2018). Treatment-refractory giardiasis: Challenges and solutions. *Infection and Drug Resistance*, 11, 1921–1933. PubMed. <https://doi.org/10.2147/IDR.S141468>
- Lanata, C. F., Fischer-Walker, C. L., Olascoaga, A. C., Torres, C. X., Aryee, M. J., Black, R. E., & for the Child Health Epidemiology Reference Group of the World Health Organization and UNICEF. (2013). Global Causes of Diarrheal Disease Mortality in Children <5 Years of Age: A Systematic Review. *PLOS One*, 8(9), e72788. <https://doi.org/10.1371/journal.pone.0072788>
- Larson, R. T., Dacks, J. B., & Barlow, L. D. (2019). Recent gene duplications dominate evolutionary dynamics of adaptor protein complex subunits in embryophytes. *Traffic*, 20(12), 961–973. <https://doi.org/10.1111/tra.12698>
- Lass-Flörl, C., & Mayr, A. (2007). Human Protothecosis. *Clinical Microbiology Reviews*, 20(2), 230–242. <https://doi.org/10.1128/CMR.00032-06>
- Lauwaet, T., Davids, B. J., Reiner, D. S., & Gillin, F. D. (2007). Encystation of *Giardia lamblia*: A model for other parasites. *Growth and Development*, 10(6), 554–559. <https://doi.org/10.1016/j.mib.2007.09.011>
- Lebbad, M., Petersson, I., Karlsson, L., Botero-Kleiven, S., Andersson, J. O., Svenungsson, B., & Svärd, S. G. (2011). Multilocus Genotyping of Human *Giardia* Isolates Suggests Limited Zoonotic Transmission and Association between Assemblage B and Flatulence in Children. *PLOS Neglected Tropical Diseases*, 5(8), e1262. <https://doi.org/10.1371/journal.pntd.0001262>
- Lecová, L., Weisz, F., Tůmová, P., Tolarová, V., & Nohýnková, E. (2018). The first multilocus genotype analysis of *Giardia intestinalis* in humans in the Czech Republic. *Parasitology*, 145(12), 1577–1587. Cambridge Core. <https://doi.org/10.1017/S0031182018000409>
- Lee, H., Jung, B., Lim, J.-S., Seo, M.-G., Lee, S.-H., Choi, K.-H., Hwang, M.-H., Kim, T.-H., Kwon, O.-D., & Kwak, D. (2020). Multilocus genotyping of *Giardia duodenalis* from pigs in Korea. *Parasitology International*, 78, 102154. <https://doi.org/10.1016/j.parint.2020.102154>
- Leger, M. M., Kolisko, M., Kamikawa, R., Stairs, C. W., Kume, K., Čepička, I., Silberman, J. D., Andersson, J. O., Xu, F., Yabuki, A., Eme, L., Zhang, Q., Takishita, K., Inagaki, Y., Simpson, A. G. B., Hashimoto, T., & Roger, A. J. (2017). Organelles that illuminate the origins of *Trichomonas* hydrogenosomes and *Giardia* mitosomes. *Nature Ecology & Evolution*, 1(4), 0092. <https://doi.org/10.1038/s41559-017-0092>
- Leitsch, D. (2015). Drug Resistance in the Microaerophilic Parasite *Giardia lamblia*. *Current Tropical Medicine Reports*, 2(3), 128–135. <https://doi.org/10.1007/s40475-015-0051-1>

- Leung, K. F., Dacks, J. B., & Field, M. C. (2008). Evolution of the Multivesicular Body ESCRT Machinery; Retention Across the Eukaryotic Lineage. *Traffic*, 9(10), 1698–1716.
<https://doi.org/10.1111/j.1600-0854.2008.00797.x>
- Lewin, D. A., & Mellman, I. (1998). Sorting out adaptors. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1401(2), 129–145. [https://doi.org/10.1016/S0167-4889\(97\)00130-4](https://doi.org/10.1016/S0167-4889(97)00130-4)
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, Q., Lau, A., Morris, T. J., Guo, L., Fordyce, C. B., & Stanley, E. F. (2004). A Syntaxin 1, G α , and N-Type Calcium Channel Complex at a Presynaptic Nerve Terminal: Analysis by Quantitative Immunoco-localization. *The Journal of Neuroscience*, 24(16), 4070.
<https://doi.org/10.1523/JNEUROSCI.0346-04.2004>
- Li, Y., Kelly, W. G., Logsdon, J. M., Schurko, A. M., Harfe, B. D., Hill-Harfe, K. L., & Kahn, R. A. (2004). Functional genomic analysis of the ADP-ribosylation factor family of GTPases: Phylogeny among diverse eukaryotes and function in *C. elegans*. *The FASEB Journal*, 18(15), 1834–1850.
<https://doi.org/10.1096/fj.04-2273com>
- Liechti, N., Schürch, N., Bruggmann, R., & Wittwer, M. (2019). Nanopore sequencing improves the draft genome of the human pathogenic amoeba *Naegleria fowleri*. *Scientific Reports*, 9(1), 16040.
<https://doi.org/10.1038/s41598-019-52572-0>
- Liendo, A., Stedman, T. T., Ngô, H. M., Chaturvedi, S., Hoppe, H. C., & Joiner, K. A. (2001). *Toxoplasma gondii* ADP-ribosylation Factor 1 Mediates Enhanced Release of Constitutively Secreted Dense Granule Proteins. *Journal of Biological Chemistry*, 276(21), 18272–18281.
<https://doi.org/10.1074/jbc.M008352200>
- Liew, F. Y., Vickerman, K., & Nash, T. E. (1997). Antigenic variation in *Giardia lamblia* and the host's immune response. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1359), 1369–1375. <https://doi.org/10.1098/rstb.1997.0122>
- Lin, Z.-Q., Gan, S.-W., Tung, S.-Y., Ho, C.-C., Su, L.-H., & Sun, C.-H. (2019). Development of CRISPR/Cas9-mediated gene disruption systems in *Giardia lamblia*. *PLOS ONE*, 14(3), e0213594. <https://doi.org/10.1371/journal.pone.0213594>
- Lipoldová, M. (2014). *Giardia* and Vilém Dušan Lambl. *PLOS Neglected Tropical Diseases*, 8(5), e2686.
<https://doi.org/10.1371/journal.pntd.0002686>
- Liu, J., Ma'ayeh, S., Peirasmaki, D., Lundström-Stadelmann, B., Hellman, L., & Svärd, S. G. (2018). Secreted *Giardia intestinalis* cysteine proteases disrupt intestinal epithelial cell junctional complexes and degrade chemokines. *Virulence*, 9(1), 879–894.
<https://doi.org/10.1080/21505594.2018.1451284>
- Liu, Y., Kahn, R. A., & Prestegard, J. H. (2009). Structure and membrane interaction of myristoylated ARF1. *Structure (London, England): 1993*, 17(1), 79–87.
<https://doi.org/10.1016/j.str.2008.10.020>

- Ljungström I, Holmgren J, Svennerholm A M, & Ferrante A. (1985). Changes in intestinal fluid and mucosal immune responses to cholera toxin in *Giardia muris* infection and binding of cholera toxin to *Giardia muris* trophozoites. *Infection and Immunity*, 50(1), 243–249. <https://doi.org/10.1128/iai.50.1.243-249.1985>
- Llor, C., & Bjerrum, L. (2014). Antimicrobial resistance: Risk associated with antibiotic overuse and initiatives to reduce the problem. *Therapeutic Advances in Drug Safety*, 5(6), 229–241. <https://doi.org/10.1177/2042098614554919>
- Loftus, B., Anderson, I., Davies, R., Alsmark, U. C. M., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R. P., Mann, B. J., Nozaki, T., Suh, B., Pop, M., Duchene, M., Ackers, J., Tannich, E., Leippe, M., Hofer, M., Bruchhaus, I., ... Hall, N. (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature*, 433(7028), 865–868. <https://doi.org/10.1038/nature03291>
- Lorenzi, H., Khan, A., Behnke, M. S., Namasivayam, S., Swapna, L. S., Hadjithomas, M., Karamycheva, S., Pinney, D., Brunk, B. P., Ajioka, J. W., Ajzenberg, D., Boothroyd, J. C., Boyle, J. P., Dardé, M. L., Diaz-Miranda, M. A., Dubey, J. P., Fritz, H. M., Gennari, S. M., Gregory, B. D., ... Sibley, L. D. (2016). Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. *Nature Communications*, 7(1), 10147. <https://doi.org/10.1038/ncomms10147>
- Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, 48(D1), D265–D268. <https://doi.org/10.1093/nar/gkz991>
- Lucanic, M., & Cheng, H.-J. (2008). A RAC/CDC-42–Independent GIT/PIX/PAK Signaling Pathway Mediates Cell Migration in *C. elegans*. *PLOS Genetics*, 4(11), e1000269. <https://doi.org/10.1371/journal.pgen.1000269>
- Luján, H. D., Marotta, A., Mowatt, M. R., Sciaky, N., Lippincott-Schwartz, J., & Nash, T. E. (1995). Developmental Induction of Golgi Structure and Function in the Primitive Eukaryote *Giardia lamblia*. *Journal of Biological Chemistry*, 270(9), 4612–4618. <https://doi.org/10.1074/jbc.270.9.4612>
- Lukeš, J., Lys, G. D., Votýpka, J., Zíková, A., Benne, R., & Englund, P. T. (2002). Kinetoplast DNA Network: Evolution of an Improbable Structure. *Eukaryotic Cell*, 1(4), 495–502. <https://doi.org/10.1128/EC.1.4.495-502.2002>
- Lundquist, P. K., Shivaiah, K.-K., & Espinoza-Corral, R. (2020). Lipid droplets throughout the evolutionary tree. *Progress in Lipid Research*, 78, 101029. <https://doi.org/10.1016/j.plipres.2020.101029>
- Lynes, E. M., & Simmen, T. (2011). Urban planning of the endoplasmic reticulum (ER): How diverse mechanisms segregate the many functions of the ER. *Biochimica et Biophysica Acta*, 1813(10), 1893–1905. <https://doi.org/10.1016/j.bbamcr.2011.06.011>

- Ma'ayeh, S. Y., Liu, J., Peirasmaki, D., Hörnaeus, K., Bergström Lind, S., Grabherr, M., Bergquist, J., & Svärd, S. G. (2017). Characterization of the *Giardia intestinalis* secretome during interaction with human intestinal epithelial cells: The impact on host cells. *PLOS Neglected Tropical Diseases*, *11*(12), e0006120. <https://doi.org/10.1371/journal.pntd.0006120>
- Maddison, W., & Maddison, D. (2019). *Mesquite: A modular system for evolutionary analysis*. Version 3.40. <http://www.mesquiteproject.org>
- Malik, S.-B., Brochu, C. D., Bilic, I., Yuan, J., Hess, M., Logsdon, J. M., Jr, & Carlton, J. M. (2011). Phylogeny of Parasitic Parabasalia and Free-Living Relatives Inferred from Conventional Markers vs. Rpb1, a Single-Copy Gene. *PLOS ONE*, *6*(6), e20774. <https://doi.org/10.1371/journal.pone.0020774>
- Malik, S.-B., Ramesh, M. A., Hulstrand, A. M., & Logsdon, J. M., Jr. (2007). Protist Homologs of the Meiotic Spo11 Gene and Topoisomerase VI reveal an Evolutionary History of Gene Duplication and Lineage-Specific Loss. *Molecular Biology and Evolution*, *24*(12), 2827–2841. <https://doi.org/10.1093/molbev/msm217>
- Manders, E. M. M., Verbeek, F. J., & Aten, J. A. (1993). Measurement of co-localization of objects in dual-colour confocal images. *Journal of Microscopy*, *169*(3), 375–382. <https://doi.org/10.1111/j.1365-2818.1993.tb03313.x>
- Mani, S., & Thattai, M. (2016). Wine glasses and hourglasses: Non-adaptive complexity of vesicle traffic in microbial eukaryotes. *Diversity, Specialisation and Conservation within the Cellular Blueprints of Parasitic Protists*, *209*(1), 58–63. <https://doi.org/10.1016/j.molbiopara.2016.03.006>
- Manna, P. T., Kelly, S., & Field, M. C. (2013). Adaptin evolution in kinetoplastids and emergence of the variant surface glycoprotein coat in African trypanosomatids. *Molecular Phylogenetics and Evolution*, *67*(1), 123–128. <https://doi.org/10.1016/j.ympev.2013.01.002>
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
- Marçais, G., Yorke, J. A., & Zimin, A. (2015). QuorUM: An Error Corrector for Illumina Reads. *PLOS One*, *10*(6), e0130821–e0130821. PubMed. <https://doi.org/10.1371/journal.pone.0130821>
- Marcial-Quino, J., Gómez-Manzo, S., Fierro, F., Rufino-González, Y., Ortega-Cuellar, D., Sierra-Palacios, E., Vanoye-Carlo, A., González-Valdez, A., Torres-Arroyo, A., Oria-Hernández, J., & Reyes-Vivas, H. (2017). RNAi-Mediated Specific Gene Silencing as a Tool for the Discovery of New Drug Targets in *Giardia lamblia*; Evaluation Using the NADH Oxidase Gene. *Genes*, *8*(11), 303. PubMed. <https://doi.org/10.3390/genes8110303>
- Marciano-Cabral, F. (1988). Biology of *Naegleria* spp. *Microbiological Reviews*, *52*(1), 114–133. <https://doi.org/10.1128/mr.52.1.114-133.1988>

- Marcusson, E. G., Horazdovsky, B. F., Cereghino, J. L., Gharakhanian, E., & Emr, S. D. (1994). The sorting receptor for yeast vacuolar carboxypeptidase Y is encoded by the VPS10 gene. *Cell*, 77(4), 579–586. [https://doi.org/10.1016/0092-8674\(94\)90219-4](https://doi.org/10.1016/0092-8674(94)90219-4)
- Marquer, C., Tian, H., Yi, J., Bastien, J., Dall'Armi, C., Yang-Klingler, Y., Zhou, B., Chan, R. B., & Di Paolo, G. (2016). Arf6 controls retromer traffic and intracellular cholesterol distribution via a phosphoinositide-based mechanism. *Nature Communications*, 7(1), 11919. <https://doi.org/10.1038/ncomms11919>
- Marti, M., Regös, A., Li, Y., Schraner, E. M., Wild, P., Müller, N., Knopf, L. G., & Hehl, A. B. (2003). An Ancestral Secretory Apparatus in the Protozoan Parasite *Giardia intestinalis*. *Journal of Biological Chemistry*, 278(27), 24837–24848. <https://doi.org/10.1074/jbc.M302082200>
- Martin, W., Brinkmann, H., Savonna, C., & Cerff, R. (1993). Evidence for a chimeric nature of nuclear genomes: Eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *Proceedings of the National Academy of Sciences*, 90(18), 8692. <https://doi.org/10.1073/pnas.90.18.8692>
- Martincová, E., Voleman, L., Pyrih, J., Žárský, V., Vondráčková, P., Kolísko, M., Tachezy, J., & Doležal, P. (2015). Probing the Biology of *Giardia intestinalis* Mitosomes Using In Vivo Enzymatic Tagging. *Molecular and Cellular Biology*, 35(16), 2864–2874. <https://doi.org/10.1128/MCB.00448-15>
- Mast, F. D., Herricks, T., Strehler, K. M., Miller, L. R., Saleem, R. A., Rachubinski, R. A., & Aitchison, J. D. (2018). ESCRT-III is required for scissioning new peroxisomes from the endoplasmic reticulum. *The Journal of Cell Biology*, 217(6), 2087–2102. <https://doi.org/10.1083/jcb.201706044>
- Mathur, V., Kolísko, M., Hehenberger, E., Irwin, N. A. T., Leander, B. S., Kristmundsson, Á., Freeman, M. A., & Keeling, P. J. (2019). Multiple Independent Origins of Apicomplexan-Like Parasites. *Current Biology*, 29(17), 2936–2941.e5. <https://doi.org/10.1016/j.cub.2019.07.019>
- Mathur, V., Kwong, W. K., Husnik, F., Irwin, N. A. T., Kristmundsson, Á., Gestal, C., Freeman, M., & Keeling, P. J. (2021). Phylogenomics Identifies a New Major Subgroup of Apicomplexans, Marosporida class nov., with Extreme Apicoplast Genome Reduction. *Genome Biology and Evolution*, 13(2). <https://doi.org/10.1093/gbe/evaa244>
- Mattera, R., Park, S. Y., De Pace, R., Guardia, C. M., & Bonifacino, J. S. (2017). AP-4 mediates export of ATG9A from the *trans*-Golgi network to promote autophagosome formation. *Proceedings of the National Academy of Sciences*, 114(50), E10697. <https://doi.org/10.1073/pnas.1717327114>
- Maziarz, E. K., & Perfect, J. R. (2016). Cryptococcosis. *Infectious Disease Clinics of North America*, 30(1), 179–206. <https://doi.org/10.1016/j.idc.2015.10.006>
- McArthur, A. G., Knodler, L. A., Silberman, J. D., Davids, B. J., Gillin, F. D., & Sogin, M. L. (2001). The Evolutionary Origins of Eukaryotic Protein Disulfide Isomerase Domains: New Evidence from the Amitochondriate Protist *Giardia lamblia*. *Molecular Biology and Evolution*, 18(8), 1455–1463. <https://doi.org/10.1093/oxfordjournals.molbev.a003931>

- McArthur, A. G., Morrison, H. G., Nixon, J. E. J., Passamaneck, N. Q. E., Kim, U., Hinkle, G., Crocker, M. K., Holder, M. E., Farr, R., Reich, C. I., Olsen, G. E., Aley, S. B., Adam, R. D., Gillin, F. D., & Sogin, M. L. (2000). The *Giardia* genome project database. *FEMS Microbiology Letters*, *189*(2), 271–273. <https://doi.org/10.1111/j.1574-6968.2000.tb09242.x>
- McAuley, J. B. (2014). Congenital Toxoplasmosis. *Journal of the Pediatric Infectious Diseases Society*, *3*, S30–S35. <https://doi.org/10.1093/jpids/piu077>
- Mccaffery, J. M., & Gillin, F. D. (1994). *Giardia lamblia*: Ultrastructural Basis of Protein Transport during Growth and Encystation. *Experimental Parasitology*, *79*(3), 220–235. <https://doi.org/10.1006/expr.1994.1086>
- McCardell, B. A., Madden, J. M., Stanfield, J. T., Tall, B. D., & Stephens, M. J. (1987). Binding of cholera toxin to *Giardia lamblia*. *Journal of Clinical Microbiology*, *25*(9), 1786–1788. <https://doi.org/10.1128/jcm.25.9.1786-1788.1987>
- McConville, M. J., Mullin, K. A., Ilgoutz, S. C., & Teasdale, R. D. (2002). Secretory Pathway of Trypanosomatid Parasites. *Microbiology and Molecular Biology Reviews*, *66*(1), 122–154. <https://doi.org/10.1128/MMBR.66.1.122-154.2002>
- McInally, S. G., Hagen, K. D., Nosala, C., Williams, J., Nguyen, K., Booker, J., Jones, K., & Dawson, S. C. (2019). Robust and stable transcriptional repression in *Giardia* using CRISPRi. *Molecular Biology of the Cell*, *30*(1), 119–130. PubMed. <https://doi.org/10.1091/mbc.E18-09-0605>
- McNally, K. E., Faulkner, R., Steinberg, F., Gallon, M., Ghai, R., Pim, D., Langton, P., Pearson, N., Danson, C. M., Nägele, H., Morris, L. L., Singla, A., Overlee, B. L., Heesom, K. J., Sessions, R., Banks, L., Collins, B. M., Berger, I., Billadeau, D. D., ... Cullen, P. J. (2017). Retriever is a multiprotein complex for retromer-independent endosomal cargo recycling. *Nature Cell Biology*, *19*(10), 1214–1225. <https://doi.org/10.1038/ncb3610>
- Mercer, F., & Johnson, P. J. (2018). *Trichomonas vaginalis*: Pathogenesis, Symbiont Interactions, and Host Cell Immune Responses. *Trends in Parasitology*, *34*(8), 683–693. <https://doi.org/10.1016/j.pt.2018.05.006>
- Midlej, V., de Souza, W., & Benchimol, M. (2019). The peripheral vesicles gather multivesicular bodies with different behavior during the *Giardia intestinalis* life cycle. *Journal of Structural Biology*, *207*(3), 301–311. <https://doi.org/10.1016/j.jsb.2019.07.002>
- Miller, J. R., Delcher, A. L., Koren, S., Venter, E., Walenz, B. P., Brownley, A., Johnson, J., Li, K., Mobarry, C., & Sutton, G. (2008). Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics*, *24*(24), 2818–2824. <https://doi.org/10.1093/bioinformatics/btn548>
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1–8. <https://doi.org/10.1109/GCE.2010.5676129>

- Minetti, C., Lamden, K., Durband, C., Cheesbrough, J., Fox, A., & Wastling, J. M. (2015). Determination of *Giardia duodenalis* assemblages and multi-locus genotypes in patients with sporadic giardiasis from England. *Parasites & Vectors*, 8, 444–444. <https://doi.org/10.1186/s13071-015-1059-z>
- Minetti, C., Lamden, K., Durband, C., Cheesbrough, J., Platt, K., Charlett, A., O'Brien, S. J., Fox, A., & Wastling, J. M. (2015). Case-Control Study of Risk Factors for Sporadic Giardiasis and Parasite Assemblages in North West England. *Journal of Clinical Microbiology*, 53(10), 3133–3140. <https://doi.org/10.1128/JCM.00715-15>
- Miras, S. L., Merino, M. C., Gottig, N., Rópolo, A. S., & Touz, M. C. (2013). The giardial VPS35 retromer subunit is necessary for multimeric complex assembly and interaction with the vacuolar protein sorting receptor. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833(12), 2628–2638. <https://doi.org/10.1016/j.bbamcr.2013.06.015>
- Montagnac, G., Sibarita, J.-B., Loubéry, S., Daviet, L., Romao, M., Raposo, G., & Chavier, P. (2009). ARF6 Interacts with JIP4 to Control a Motor Switch Mechanism Regulating Endosome Traffic in Cytokinesis. *Current Biology*, 19(3), 184–195. <https://doi.org/10.1016/j.cub.2008.12.043>
- Moore, R. B., Oborník, M., Janouškovec, J., Chrudimský, T., Vancová, M., Green, D. H., Wright, S. W., Davies, N. W., Bolch, C. J. S., Heimann, K., Šlapeta, J., Hoegh-Guldberg, O., Logsdon, J. M., & Carter, D. A. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. *Nature*, 451(7181), 959–963. <https://doi.org/10.1038/nature06635>
- Moran, G. P., Coleman, D. C., & Sullivan, D. J. (2011). Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. *Eukaryotic Cell*, 10(1), 34–42. <https://doi.org/10.1128/EC.00242-10>
- Morecki, R., & Parker, J. G. (1967). Ultrastructural Studies of the Human *Giardia lamblia* and Subjacent Jejunal Mucosa in a Subject with Steatorrhea. *Gastroenterology*, 52(2), 151–164. [https://doi.org/10.1016/S0016-5085\(67\)80002-7](https://doi.org/10.1016/S0016-5085(67)80002-7)
- Morf, L., Spycher, C., Rehrauer, H., Fournier, C. A., Morrison, H. G., & Hehl, A. B. (2010). The Transcriptional Response to Encystation Stimuli in *Giardia lamblia* Is Restricted to a Small Set of Genes. *Eukaryotic Cell*, 9(10), 1566–1576. <https://doi.org/10.1128/EC.00100-10>
- Morré, D. J., & Mollenhauer, Hilton. H. (1974). The endomembrane concept: A functional integration of endoplasmic reticulum and Golgi apparatus. In *Dynamic Aspects of Plant infrastructure* (pp. 84–137). McGraw-Hill.
- Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., ... Sogin, M. L. (2007). Genomic Minimalism in the Early Diverging Intestinal Parasite *Giardia lamblia*. *Science*, 317(5846), 1921. <https://doi.org/10.1126/science.1143837>

- Mouratou, B., Biou, V., Joubert, A., Cohen, J., Shields, D. J., Geldner, N., Jürgens, G., Melançon, P., & Cherfils, J. (2005). The domain architecture of large guanine nucleotide exchange factors for the small GTP-binding protein Arf. *BMC Genomics*, *6*(1), 20. <https://doi.org/10.1186/1471-2164-6-20>
- Moyano, S., Musso, J., Feliziani, C., Zamponi, N., Frontera, L. S., Ropolo, A. S., Lanfredi-Rangel, A., Lalle, M., & Touz, M. (2019). Exosome Biogenesis in the Protozoa Parasite *Giardia lamblia*: A Model of Reduced Interorganellar Crosstalk. *Cells*, *8*(12), 1600. <https://doi.org/10.3390/cells8121600>
- Mulvey, C. M., Breckels, L. M., Geladaki, A., Britovšek, N. K., Nightingale, D. J. H., Christoforou, A., Elzek, M., Deery, M. J., Gatto, L., & Lilley, K. S. (2017). Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. *Nature Protocols*, *12*(6), 1110–1135. <https://doi.org/10.1038/nprot.2017.026>
- Muñoz-Gómez, S. A., Hess, S., Burger, G., Lang, B. F., Susko, E., Slamovits, C. H., & Roger, A. J. (2019). An updated phylogeny of the Alphaproteobacteria reveals that the parasitic Rickettsiales and Holosporales have independent origins. *eLife*, *8*, e42535. <https://doi.org/10.7554/eLife.42535>
- Nakai, W., Kondo, Y., Saitoh, A., Naito, T., Nakayama, K., & Shin, H.-W. (2013). ARF1 and ARF4 regulate recycling endosomal morphology and retrograde transport from endosomes to the Golgi apparatus. *Molecular Biology of the Cell*, *24*(16), 2570–2581. <https://doi.org/10.1091/mbc.e13-04-0197>
- Neuman, S. D., Terry, E. L., Selegue, J. E., Cavanagh, A. T., & Bashirullah, A. (2021). Mistargeting of secretory cargo in retromer-deficient cells. *Disease Models & Mechanisms*, *14*(1), dmm046417. <https://doi.org/10.1242/dmm.046417>
- Nevin, W. D., & Dacks, J. B. (2009). Repeated secondary loss of adaptin complex genes in the Apicomplexa. *Parasitology International*, *58*(1), 86–94. <https://doi.org/10.1016/j.parint.2008.12.002>
- Ngô, H. M., Yang, M., Paprotka, K., Pypaert, M., Hoppe, H., & Joiner, K. A. (2003). AP-1 in *Toxoplasma gondii* Mediates Biogenesis of the Rhoptry Secretory Organelle from a Post-Golgi Compartment. *Journal of Biological Chemistry*, *278*(7), 5343–5352. <https://doi.org/10.1074/jbc.M208291200>
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Niu, T.-K., Pfeifer, A. C., Lippincott-Schwartz, J., & Jackson, C. L. (2005). Dynamics of GBF1, a Brefeldin A-Sensitive Arf1 Exchange Factor at the Golgi. *Molecular Biology of the Cell*, *16*, 10.
- Nobile, C. J., & Johnson, A. D. (2015). *Candida albicans* Biofilms and Human Disease. *Annual Review of Microbiology*, *69*, 71–92. <https://doi.org/10.1146/annurev-micro-091014-104330>
- Noda, S., Mantini, C., Meloni, D., Inoue, J.-I., Kitade, O., Viscogliosi, E., & Ohkuma, M. (2012). Molecular Phylogeny and Evolution of Parabasalia with Improved Taxon Sampling and New Protein

- Markers of Actin and Elongation Factor-1 α . *PLOS ONE*, 7(1), e29938.
<https://doi.org/10.1371/journal.pone.0029938>
- Noussa, R. E. B., Mayssa, M. Z., Omayma, M. H., Mohamed, K. R., & Dalia, O. (2016). *Giardia* Assemblages A and B in Diarrheic Patients: A Comparative Study in Egyptian Children and Adults. *Journal of Parasitology*, 102(1), 69–74. <https://doi.org/10.1645/14-676>
- Novick, P., Field, C., & Schekman, R. (1980). Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. *Cell*, 21(1), 205–215.
[https://doi.org/10.1016/0092-8674\(80\)90128-2](https://doi.org/10.1016/0092-8674(80)90128-2)
- Oborník, M., Vancová, M., Lai, D.-H., Janouškovec, J., Keeling, P. J., & Lukeš, J. (2011). Morphology and Ultrastructure of Multiple Life Cycle Stages of the Photosynthetic Relative of Apicomplexa, *Chromera velia*. *Protist*, 162(1), 115–130. <https://doi.org/10.1016/j.protis.2010.02.004>
- Odorizzi, G., Cowles, C. R., & Emr, S. D. (1998). The AP-3 complex: A coat of many colours. *Trends in Cell Biology*, 8(7), 282–288. [https://doi.org/10.1016/S0962-8924\(98\)01295-1](https://doi.org/10.1016/S0962-8924(98)01295-1)
- Ofir-Birin, Y., & Regev-Rudzki, N. (2019). Extracellular vesicles in parasite survival. *Science*, 363(6429), 817. <https://doi.org/10.1126/science.aau4666>
- Ohkuma, M., Iida, T., Ohtoko, K., Yuzawa, H., Noda, S., Viscogliosi, E., & Kudo, T. (2005). Molecular phylogeny of parabasalids inferred from small subunit rRNA sequences, with emphasis on the Hypermastigae. *Molecular Phylogenetics and Evolution*, 35(3), 646–655.
<https://doi.org/10.1016/j.ympev.2005.02.013>
- Ohkuma, S., & Poole, B. (1978). Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents. *Proceedings of the National Academy of Sciences*, 75(7), 3327. <https://doi.org/10.1073/pnas.75.7.3327>
- Ohsaki, Y., Soltysik, K., & Fujimoto, T. (2017). The Lipid Droplet and the Endoplasmic Reticulum. In M. Tagaya & T. Simmen (Eds.), *Organelle Contact Sites: From Molecular Mechanism to Disease* (pp. 111–120). Springer Singapore. https://doi.org/10.1007/978-981-10-4567-7_8
- Okada, M., & Nozaki, T. (2006). New Insights into Molecular Mechanisms of Phagocytosis in *Entamoeba histolytica* by Proteomic Analysis. *Amebiasis*, 37(2), 244–251.
<https://doi.org/10.1016/j.arcmed.2005.10.003>
- Oliver, J. M., Berlin, R. D., & Davis, B. H. (1984). Use of horseradish peroxidase and fluorescent dextrans to study fluid pinocytosis in leukocytes. In G. Di Sabato, J. J. Langone, & H. Van Vunakis (Eds.), *Methods in Enzymology* (Vol. 108, pp. 336–347). Academic Press. [https://doi.org/10.1016/S0076-6879\(84\)08100-3](https://doi.org/10.1016/S0076-6879(84)08100-3)
- Oliviusson, P., Heinzerling, O., Hillmer, S., Hinz, G., Tse, Y. C., Jiang, L., & Robinson, D. G. (2006). Plant retromer, localized to the prevacuolar compartment and microvesicles in *Arabidopsis*, may interact with vacuolar sorting receptors. *The Plant Cell*, 18(5), 1239–1252.
<https://doi.org/10.1105/tpc.105.035907>

- Olmos, Y., & Carlton, J. (2016). The ESCRT machinery: New roles at new holes. *Cell Architecture*, 38, 1–11. <https://doi.org/10.1016/j.ceb.2015.12.001>
- Olmos, Y., Perdrix-Rosell, A., & Carlton, J. G. (2016). Membrane Binding by CHMP7 Coordinates ESCRT-III-Dependent Nuclear Envelope Reformation. *Current Biology*, 26(19), 2635–2641. <https://doi.org/10.1016/j.cub.2016.07.039>
- Olzmann, J. A., & Carvalho, P. (2019). Dynamics and functions of lipid droplets. *Nature Reviews. Molecular Cell Biology*, 20(3), 137–155. <https://doi.org/10.1038/s41580-018-0085-z>
- Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, 12(1), 385. <https://doi.org/10.1186/1471-2105-12-385>
- Ooi, C. E., Dell'Angelica, E. C., & Bonifacino, J. S. (1998). ADP-Ribosylation factor 1 (ARF1) regulates recruitment of the AP-3 adaptor complex to membranes. *The Journal of Cell Biology*, 142(2), 391–402. <https://doi.org/10.1083/jcb.142.2.391>
- Opperdoes, F. R., Butenko, A., Flegontov, P., Yurchenko, V., & Lukeš, J. (2016). Comparative Metabolism of Free-living *Bodo saltans* and Parasitic Trypanosomatids. *Journal of Eukaryotic Microbiology*, 63(5), 657–678. <https://doi.org/10.1111/jeu.12315>
- Orci, L., Glick, B. S., & Rothman, J. E. (1986). A new type of coated vesicular carrier that appears not to contain clathrin: Its possible role in protein transport within the Golgi stack. *Cell*, 46(2), 171–184. [https://doi.org/10.1016/0092-8674\(86\)90734-8](https://doi.org/10.1016/0092-8674(86)90734-8)
- Owen, D. J., Vallis, Y., Pearse, B. M., McMahon, H. T., & Evans, P. R. (2000). The structure and function of the beta 2-adaptin appendage domain. *The EMBO Journal*, 19(16), 4216–4227. PubMed. <https://doi.org/10.1093/emboj/19.16.4216>
- Painter, J. E., Gargano, J. W., Collier, S. A., & Yoder, J. S. (2015). Giardiasis surveillance—United States, 2011–2012. *MMWR Supplements*, 64(3), 15–25.
- Palade, G. (1975). Intracellular Aspects of the Process of Protein Synthesis. *Science*, 189(4206), 867. <https://doi.org/10.1126/science.189.4206.867-b>
- Paleotti, O., Macia, E., Luton, F., Klein, S., Partisani, M., Chardin, P., Kirchhausen, T., & Franco, M. (2005). The Small G-protein Arf6GTP Recruits the AP-2 Adaptor Complex to Membranes. *Journal of Biological Chemistry*, 280(22), 21661–21666. <https://doi.org/10.1074/jbc.M503099200>
- Pan, X., Eathiraj, S., Munson, M., & Lambright, D. G. (2006). TBC-domain GAPs for Rab GTPases accelerate GTP hydrolysis by a dual-finger mechanism. *Nature*, 442(7100), 303–306. <https://doi.org/10.1038/nature04847>
- Papanikou, E., & Glick, B. S. (2014). Golgi compartmentation and identity. *Cell Organelles*, 29, 74–81. <https://doi.org/10.1016/j.ceb.2014.04.010>
- Papkou, A., Guzella, T., Yang, W., Koepper, S., Pees, B., Schalkowski, R., Barg, M.-C., Rosenstiel, P. C., Teotónio, H., & Schulenburg, H. (2019). The genomic basis of Red Queen dynamics during rapid reciprocal host–pathogen coevolution. *Proceedings of the National Academy of Sciences*, 116(3), 923. <https://doi.org/10.1073/pnas.1810402116>

- Park, J. S., Kolisko, M., & Simpson, A. G. B. (2010). Cell Morphology and Formal Description of *Ergobibamus cyprinoides* n. g., n. sp., Another *Carpediemonas*-Like Relative of Diplomonads: *Ergobibamus* n. g., A New Diplomonad Relative. *Journal of Eukaryotic Microbiology*, 57(6), 520–528. <https://doi.org/10.1111/j.1550-7408.2010.00506.x>
- Parton, R. G., & Simons, K. (2007). The multiple faces of caveolae. *Nature Reviews Molecular Cell Biology*, 8(3), 185–194. <https://doi.org/10.1038/nrm2122>
- Pasqualato, S., Renault, L., & Cherfils, J. (2002). Arf, Arl, Arp and Sar proteins: A family of GTP-binding proteins with a structural device for “front-back” communication. *EMBO Reports*, 3(11), 1035–1041. <https://doi.org/10.1093/embo-reports/kvf221>
- Peirasmaki, D., Ma’ayeh, S. Y., Xu, F., Ferella, M., Campos, S., Liu, J., & Svärd, S. G. (2020). High Cysteine Membrane Proteins (HCMPs) Are Up-Regulated During *Giardia*-Host Cell Interactions. *Frontiers in Genetics*, 11, 913. <https://doi.org/10.3389/fgene.2020.00913>
- Pereira-Leal, J. B., Hume, A. N., & Seabra, M. C. (2001). Prenylation of Rab GTPases: Molecular mechanisms and involvement in genetic disease. *FEBS Letters*, 498(2–3), 197–200. [https://doi.org/10.1016/S0014-5793\(01\)02483-8](https://doi.org/10.1016/S0014-5793(01)02483-8)
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, Ş., Tiwary, S., Cox, J., Audain, E., Walzer, M., ... Vizcaino, J. A. (2019). The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Research*, 47(D1), D442–D450. <https://doi.org/10.1093/nar/gky1106>
- Phillips, M. A., Burrows, J. N., Manyando, C., van Huijsduijnen, R. H., Van Voorhis, W. C., & Wells, T. N. C. (2017). Malaria. *Nature Reviews Disease Primers*, 3(1), 17050. <https://doi.org/10.1038/nrdp.2017.50>
- Pickart, C. M. (2001). Mechanisms Underlying Ubiquitination. *Annual Review of Biochemistry*, 70(1), 503–533. <https://doi.org/10.1146/annurev.biochem.70.1.503>
- Pipaliya, S. V., Santos, R., Salas-Leiva, D., Balmer, E. A., Wirdnam, C. D., Roger, A. J., Hehl, A. B., Faso, C., & Dacks, J. B. (2021). Unexpected organellar locations of ESCRT machinery in *Giardia intestinalis* and complex evolutionary dynamics spanning the transition to parasitism in the lineage Fornicata. *BioRxiv*. <https://doi.org/10.1101/2021.02.11.430673>
- Pipaliya, S. V., Schlacht, A., Klinger, C. M., Kahn, R. A., & Dacks, J. (2019). Ancient complement and lineage-specific evolution of the Sec7 ARF GEF proteins in eukaryotes. *Molecular Biology of the Cell*, 30(15), 1846–1863. <https://doi.org/10.1091/mbc.E19-01-0073>
- Pipaliya, S. V., Thompson, L. A., & Dacks, J. B. (2021). The reduced ARF regulatory system in *Giardia intestinalis* pre-dates the transition to parasitism in the lineage Fornicata. *International Journal for Parasitology*. <https://doi.org/10.1016/j.ijpara.2021.02.004>
- Pollo, S. M. J., Reiling, S. J., Wit, J., Workentine, M. L., Guy, R. A., Batoff, G. W., Yee, J., Dixon, B. R., & Wasmuth, J. D. (2020). Benchmarking hybrid assemblies of *Giardia* and prediction of widespread

- intra-isolate structural variation. *Parasites & Vectors*, 13(1), 108. <https://doi.org/10.1186/s13071-020-3968-8>
- Poole, A., & Penny, D. (2007). Engulfed by speculation. *Nature*, 447(7147), 913–913. <https://doi.org/10.1038/447913a>
- Poulin, R., & Randhawa, H. S. (2015). Evolution of parasitism along convergent lines: From ecology to genomics. *Parasitology*, 142(S1), S6–S15. Cambridge Core. <https://doi.org/10.1017/S0031182013001674>
- Poxleitner, M. K., Carpenter, M. L., Mancuso, J. J., Wang, C.-J. R., Dawson, S. C., & Cande, W. Z. (2008). Evidence for Karyogamy and Exchange of Genetic Material in the Binucleate Intestinal Parasite *Giardia intestinalis*. *Science*, 319(5869), 1530. <https://doi.org/10.1126/science.1153752>
- Praefcke, G. J. K., Ford, M. G. J., Schmid, E. M., Olesen, L. E., Gallop, J. L., Peak-Chew, S.-Y., Vallis, Y., Babu, M. M., Mills, I. G., & McMahon, H. T. (2004). Evolving nature of the AP2 alpha-appendage hub during clathrin-coated vesicle endocytosis. *The EMBO Journal*, 23(22), 4371–4383. <https://doi.org/10.1038/sj.emboj.7600445>
- Prakash, A., Jeffryes, M., Bateman, A., & Finn, R. D. (2017). The HMMER Web Server for Protein Sequence Similarity Search. *Current Protocols in Bioinformatics*, 60(1), 3.15.1-3.15.23. <https://doi.org/10.1002/cpbi.40>
- Price, H. P., Stark, M., & Smith, D. F. (2007). *Trypanosoma brucei* ARF1 Plays a Central Role in Endocytosis and Golgi–Lysosome Trafficking. *Molecular Biology of the Cell*, 18, 10.
- Prucca, C. G., & Lujan, H. D. (2009). Antigenic variation in *Giardia lamblia*. *Cellular Microbiology*, 11(12), 1706–1715. <https://doi.org/10.1111/j.1462-5822.2009.01367.x>
- Prucca, C. G., Slavin, I., Quiroga, R., Elías, E. V., Rivero, F. D., Saura, A., Carranza, P. G., & Luján, H. D. (2008). Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature*, 456(7223), 750–754. <https://doi.org/10.1038/nature07585>
- Pruitt, K. D. (2004). NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33(Database issue), D501–D504. <https://doi.org/10.1093/nar/gki025>
- Prystajek, N., Tsui, C. K.-M., Hsiao, W.W. L., Uyaguari-Diaz, M. I., Ho, J., Tang, P., Isaac-Renton, J., & Björkroth, J. (2015). *Giardia* spp. Are Commonly Found in Mixed Assemblages in Surface Water, as Revealed by Molecular and Whole-Genome Characterization. *Applied and Environmental Microbiology*, 81(14), 4827–4834. <https://doi.org/10.1128/AEM.00524-15>
- Pütz, U., & Schwarzmann, G. (1995). Golgi staining by two fluorescent ceramide analogues in cultured fibroblasts requires metabolism. *Eur J Cell Biol*, 68(2), 113–121.
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: Protein domains identifier. *Nucleic Acids Research*, 33(Web Server), W116–W120. <https://doi.org/10.1093/nar/gki442>

- Quilty, D., Gray, F., Summerfeldt, N., Cassel, D., & Melancon, P. (2014). Arf activation at the Golgi is modulated by feed-forward stimulation of the exchange factor GBF1. *Journal of Cell Science*, *127*(2), 354–364. <https://doi.org/10.1242/jcs.130591>
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rai, A. K., & Johnson, P. J. (2019). *Trichomonas vaginalis* extracellular vesicles are internalized by host cells using proteoglycans and caveolin-dependent endocytosis. *Proceedings of the National Academy of Sciences*, *116*(43), 21354. <https://doi.org/10.1073/pnas.1912356116>
- Raiborg, C., & Stenmark, H. (2009). The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature*, *458*(7237), 445–452. <https://doi.org/10.1038/nature07961>
- Rambaut, A. (2010). *FigTree v1.3.1*. <http://tree.bio.ed.ac.uk/software/figtree/>
- Rathod, S. D., Krupp, K., Klausner, J. D., Arun, A., Reingold, A. L., & Madhivanan, P. (2011). Bacterial Vaginosis and Risk for *Trichomonas vaginalis* Infection: A Longitudinal Analysis. *Sexually Transmitted Diseases*, *38*(9). https://journals.lww.com/stdjournal/Fulltext/2011/09000/Bacterial_Vaginosis_and_Risk_for_Trichomonas.19.aspx
- Raychaudhuri, S., & Prinz, W. A. (2008). Nonvesicular phospholipid transfer between peroxisomes and the endoplasmic reticulum. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(41), 15785–15790. <https://doi.org/10.1073/pnas.0808321105>
- Recker, M., Buckee, C. O., Serazin, A., Kyes, S., Pinches, R., Christodoulou, Z., Springer, A. L., Gupta, S., & Newbold, C. I. (2011). Antigenic Variation in *Plasmodium falciparum* Malaria Involves a Highly Structured Switching Pattern. *PLOS Pathogens*, *7*(3), e1001306. <https://doi.org/10.1371/journal.ppat.1001306>
- Redman, C. M., Siekevitz, P., & Palade, G. E. (1966). Synthesis and Transfer of Amylase in Pigeon Pancreatic Microsomes. *Journal of Biological Chemistry*, *241*(5), 1150–1158. [https://doi.org/10.1016/S0021-9258\(18\)96815-8](https://doi.org/10.1016/S0021-9258(18)96815-8)
- Regoes, A., Zourmpanou, D., León-Avila, G., van der Giezen, M., Tovar, J., & Hehl, A. B. (2005). Protein Import, Replication, and Inheritance of a Vestigial Mitochondrion. *Journal of Biological Chemistry*, *280*(34), 30557–30563. <https://doi.org/10.1074/jbc.M500787200>
- Reid, D. W., & Nicchitta, C. V. (2015). Diversity and selectivity in mRNA translation on the endoplasmic reticulum. *Nature Reviews Molecular Cell Biology*, *16*(4), 221–231. <https://doi.org/10.1038/nrm3958>
- Reiner, D. S., Douglas, H., & Gillin, F. D. (1989). Identification and localization of cyst-specific antigens of *Giardia lamblia*. *Infection and Immunity*, *57*(3), 963–968. <https://doi.org/10.1128/iai.57.3.963-968.1989>

- Ren, X., Fariás, G. G., Canagarajah, B. J., Bonifacino, J. S., & Hurley, J. H. (2013). Structural basis for recruitment and activation of the AP-1 clathrin adaptor complex by Arf1. *Cell*, *152*(4), 755–767. <https://doi.org/10.1016/j.cell.2012.12.042>
- Rendtorff, R. C. (1954). The Experimental Transmission of Intestinal Protozoan Parasites: II. *Giardia lamblia* Cysts Given In Capsules. *American Journal of Epidemiology*, *59*(2), 209–222. <https://doi.org/10.1093/oxfordjournals.aje.a119634>
- Reynolds, N. K., Smith, M. E., Tretter, E. D., Gause, J., Heeney, D., Cafaro, M. J., Smith, J. F., Novak, S. J., Bourland, W. A., & White, M. M. (2017). Resolving relationships at the animal-fungal divergence: A molecular phylogenetic study of the protist trichomycetes (Ichthyosporia, Eccrinida). *Molecular Phylogenetics and Evolution*, *109*, 447–464. <https://doi.org/10.1016/j.ympev.2017.02.007>
- Richardson, L. G. L., Clendening, E. A., Sheen, H., Gidda, S. K., White, K. A., Mullen, R. T., & Simon, A. (2014). A Unique N-Terminal Sequence in the Carnation Italian ringspot virus p36 Replicase-Associated Protein Interacts with the Host Cell ESCRT-I Component Vps23. *Journal of Virology*, *88*(11), 6329–6344. <https://doi.org/10.1128/JVI.03840-13>
- Richter, S., Müller, L. M., Stierhof, Y.-D., Mayer, U., Takada, N., Kost, B., Vieten, A., Geldner, N., Koncz, C., & Jürgens, G. (2012). Polarized cell growth in *Arabidopsis* requires endosomal recycling mediated by GBF1-related ARF exchange factors. *Nature Cell Biology*, *14*(1), 80–86. <https://doi.org/10.1038/ncb2389>
- Rivero, M. R., Miras, S. L., Feliziani, C., Zamponi, N., Quiroga, R., Hayes, S. F., Rópolo, A. S., & Touz, M. C. (2012). Vacuolar Protein Sorting Receptor in *Giardia lamblia*. *PLOS ONE*, *7*(8), e43712. <https://doi.org/10.1371/journal.pone.0043712>
- Rivero, M. R., Vranych, C. V., Bisbal, M., Maletto, B. A., Rópolo, A. S., & Touz, M. C. (2010). Adaptor protein 2 regulates receptor-mediated endocytosis and cyst formation in *Giardia lamblia*. *Biochemical Journal*, *428*(1), 33–45. <https://doi.org/10.1042/BJ20100096>
- Robinson, M. S. (2004). Adaptable adaptors for coated vesicles. *Trends in Cell Biology*, *14*(4), 167–174. <https://doi.org/10.1016/j.tcb.2004.02.002>
- Robinson, M. S., & Bonifacino, J. S. (2001). Adaptor-related proteins. *Current Opinion in Cell Biology*, *13*(4), 444–453. [https://doi.org/10.1016/S0955-0674\(00\)00235-0](https://doi.org/10.1016/S0955-0674(00)00235-0)
- Roger, A. J., Svärd, S. G., Tovar, J., Clark, C. G., Smith, M. W., Gillin, F. D., & Sogin, M. L. (1998). A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: Evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proceedings of the National Academy of Sciences*, *95*(1), 229. <https://doi.org/10.1073/pnas.95.1.229>
- Rothman, J. E. (1994). Mechanisms of intracellular protein transport. *Nature*, *372*(6501), 55–63. <https://doi.org/10.1038/372055a0>
- Rothman, J. E., & Fine, R. E. (1980). Coated vesicles transport newly synthesized membrane glycoproteins from endoplasmic reticulum to plasma membrane in two successive stages.

- Proceedings of the National Academy of Sciences*, 77(2), 780.
<https://doi.org/10.1073/pnas.77.2.780>
- Rothschild, L. J. (1989). Protozoa, protista, protoctista: What's in a name? *Journal of the History of Biology*, 22(2), 277–305. <https://doi.org/10.1007/BF00139515>
- Rout, M. P., & Field, M. C. (2017). The Evolution of Organellar Coat Complexes and Organization of the Eukaryotic Cell. *Annual Review of Biochemistry*, 86(1), 637–657. <https://doi.org/10.1146/annurev-biochem-061516-044643>
- Rout, S., Zumthor, J. P., Schraner, E. M., Faso, C., & Hehl, A. B. (2016). An Interactome-Centered Protein Discovery Approach Reveals Novel Components Involved in Mitosome Function and Homeostasis in *Giardia lamblia*. *PLOS Pathogens*, 12(12), e1006036.
<https://doi.org/10.1371/journal.ppat.1006036>
- Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Rueckert, S., Pipaliya, S. V., & Dacks, J. B. (2019). Evolution: Parallel Paths to Parasitism in the Apicomplexa. *Current Biology*, 29(17), R836–R839. <https://doi.org/10.1016/j.cub.2019.07.047>
- Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18(1), 529. <https://doi.org/10.1186/s12859-017-1934-z>
- Ryan, U., & Cacciò, S. M. (2013). Zoonotic potential of *Giardia*. *Zoonoses Special Issue*, 43(12), 943–956. <https://doi.org/10.1016/j.ijpara.2013.06.001>
- Sabe, H., Onodera, Y., Mazaki, Y., & Hashimoto, S. (2006). ArfGAP family proteins in cell adhesion, migration and tumor invasion. *Current Opinion in Cell Biology*, 18(5), 558–564.
<https://doi.org/10.1016/j.ceb.2006.08.002>
- Sagan, L. (1967). On the Origin of Mitosing Clds. *Journal of Theoretical Biology*, 14, 225–274.
- Saha, N., Dutta, S., Datta, S. P., & Sarkar, S. (2018). The minimal ESCRT machinery of *Giardia lamblia* has altered inter-subunit interactions within the ESCRT-II and ESCRT-III complexes. *European Journal of Cell Biology*, 97(1), 44–62. <https://doi.org/10.1016/j.ejcb.2017.11.004>
- Saito-Nakano, Y., Loftus, B. J., Hall, N., & Nozaki, T. (2005). The diversity of Rab GTPases in *Entamoeba histolytica*. *Pathogenesis of Amoebiasis - from Genome to Disease*, 110(3), 244–252.
<https://doi.org/10.1016/j.exppara.2005.02.021>
- Salas-Leiva, D. E., Tromer, E. C., Curtis, B. A., Jerlström-Hultqvist, J., Kolisko, M., Yi, Z., Salas-Leiva, J. S., Gallot-Lavallée, L., Kops, G. J. P. L., Archibald, J. M., Simpson, A. G. B., & Roger, A. J. (2021). A free-living protist that lacks canonical eukaryotic DNA replication and segregation systems. *BioRxiv*, 2021.03.14.435266. <https://doi.org/10.1101/2021.03.14.435266>
- Sato, K., & Nakano, A. (2007). Mechanisms of COPII vesicle formation and protein sorting. *Membrane Trafficking*, 581(11), 2076–2082. <https://doi.org/10.1016/j.febslet.2007.01.091>

- Savioli, L., Smith, H., & Thompson, A. (2006). *Giardia* and *Cryptosporidium* join the 'Neglected Diseases Initiative.' *Trends in Parasitology*, 22(5), 203–208. <https://doi.org/10.1016/j.pt.2006.02.015>
- Scamardella, J. (1999). Not plants or animals: A brief history of the origin of Kingdoms Protozoa, Protista and Protoctista. *Int Microbiol*, 2(4), 207–216.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676–682. <https://doi.org/10.1038/nmeth.2019>
- Schlacht, A., & Dacks, J. B. (2015). Unexpected ancient paralogs and an evolutionary model for the COPII coat complex. *Genome Biology and Evolution*, 7(4), 1098–1109. <https://doi.org/10.1093/gbe/evv045>
- Schlacht, A., Mowbrey, K., Elias, M., Kahn, R. A., & Dacks, J. B. (2013). Ancient Complexity, Opisthokont Plasticity, and Discovery of the 11th Subfamily of Arf GAP Proteins: Arf GAP Evolution. *Traffic*, 14(6), 636–649. <https://doi.org/10.1111/tra.12063>
- Schledzewski, K., Brinkmann, H., & Mendel, R. R. (1999). Phylogenetic Analysis of Components of the Eukaryotic Vesicle Transport System Reveals a Common Origin of Adaptor Protein Complexes 1, 2, and 3 and the F Subcomplex of the Coatomer COPI. *Journal of Molecular Evolution*, 48(6), 770–778. <https://doi.org/10.1007/PL00006521>
- Schmid, S. L., Fuchs, R., Male, P., & Mellman, I. (1988). Two distinct subpopulations of endosomes involved in membrane recycling and transport to lysosomes. *Cell*, 52(1), 73–83. [https://doi.org/10.1016/0092-8674\(88\)90532-6](https://doi.org/10.1016/0092-8674(88)90532-6)
- Schmidt, O., & Teis, D. (2012). The ESCRT machinery. *Current Biology*, 22(4), R116–R120. <https://doi.org/10.1016/j.cub.2012.01.028>
- Seaman, M. N. J. (2004). Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer. *Journal of Cell Biology*, 165(1), 111–122. <https://doi.org/10.1083/jcb.200312034>
- Seaman, M. N. J., Marcusson, E. G., Cereghino, J. L., & Emr, S. D. (1997). Endosome to Golgi Retrieval of the Vacuolar Protein Sorting Receptor, Vps10p, Requires the Function of the VPS29, VPS30, and VPS35 Gene Products. *Journal of Cell Biology*, 137(1), 79–92. <https://doi.org/10.1083/jcb.137.1.79>
- Seitz, K. W., Dombrowski, N., Eme, L., Spang, A., Lombard, J., Sieber, J. R., Teske, A. P., Ettema, T. J. G., & Baker, B. J. (2019). Asgard archaea capable of anaerobic hydrocarbon cycling. *Nature Communications*, 10(1), 1822. <https://doi.org/10.1038/s41467-019-09364-x>
- Serradell, M. C., Saura, A., Rupil, L. L., Gargantini, P. R., Faya, M. I., Furlan, P. J., & Lujan, H. D. (2016). Vaccination of domestic animals with a novel oral vaccine prevents *Giardia* infections, alleviates signs of giardiasis and reduces transmission to humans. *Npj Vaccines*, 1(1), 16018. <https://doi.org/10.1038/npjvaccines.2016.18>

- Shabardina, V., Kischka, T., Kmita, H., Suzuki, Y., & Makalowski, W. (2018). Environmental adaptation of *Acanthamoeba castellanii* and *Entamoeba histolytica* at genome level as seen by comparative genomic analysis. *International Journal of Biological Sciences*, *14*(3), 306–320. <https://doi.org/10.7150/ijbs.23869>
- Shandilya, A., Chacko, S., Jayaram, B., & Ghosh, I. (2013). A plausible mechanism for the antimalarial activity of artemisinin: A computational approach. *Scientific Reports*, *3*(1), 2513. <https://doi.org/10.1038/srep02513>
- Shapiro, T. A., & Englund, P. T. (1995). The Structure And Replication of Kinetoplast DNA. *Annual Review of Microbiology*, *49*(1), 117–143. <https://doi.org/10.1146/annurev.mi.49.100195.001001>
- Sheffield, H. G., & Bjorvatn, B. (1977). Ultrastructure of the Cyst of *Giardia Lamblia*. *The American Journal of Tropical Medicine and Hygiene*, *26*(1), 23–30. <https://doi.org/10.4269/ajtmh.1977.26.23>
- Shin, H.-W., Morinaga, N., Noda, M., & Nakayama, K. (2004). BIG2, A Guanine Nucleotide Exchange Factor for ADP- Ribosylation Factors: Its Localization to Recycling Endosomes and Implication in the Endosome Integrity. *Molecular Biology of the Cell*, *15*, 12.
- Shirley, D.-A. T., Farr, L., Watanabe, K., & Moonah, S. (2018). A Review of the Global Burden, New Diagnostics, and Current Therapeutics for Amebiasis. *Open Forum Infectious Diseases*, *5*(7). <https://doi.org/10.1093/ofid/ofy161>
- Shrivastava, A. K., Kumar, S., Smith, W. A., & Sahu, P. S. (2017). Revisiting the global problem of cryptosporidiosis and recommendations. *Tropical Parasitology*, *7*(1), 8–17. <https://doi.org/10.4103/2229-5070.202290>
- Shumate, A., & Salzberg, S. L. (2021). Liftoff: Accurate mapping of gene annotations. *Bioinformatics*, *btaa1016*. <https://doi.org/10.1093/bioinformatics/btaa1016>
- Sicheritz-Pontén, T., Kurland, C. G., & Andersson, S. G. E. (1998). A phylogenetic analysis of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1365*(3), 545–551. [https://doi.org/10.1016/S0005-2728\(98\)00099-1](https://doi.org/10.1016/S0005-2728(98)00099-1)
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Simmen, T., Höning, S., Icking, A., Tikkanen, R., & Hunziker, W. (2002). AP-4 binds basolateral signals and participates in basolateral sorting in epithelial MDCK cells. *Nature Cell Biology*, *4*(2), 154–159. <https://doi.org/10.1038/ncb745>
- Simpson, A. G. B., & Patterson, D. J. (1999). The ultrastructure of *Carpodionomonas membranifera* (Eukaryota) with reference to the “excavate hypothesis.” *European Journal of Protistology*, *35*(4), 353–370. [https://doi.org/10.1016/S0932-4739\(99\)80044-3](https://doi.org/10.1016/S0932-4739(99)80044-3)
- Simpson, A. G. B., Stevens, J. R., & Lukeš, J. (2006). The evolution and diversity of kinetoplastid flagellates. *Trends in Parasitology*, *22*(4), 168–174. <https://doi.org/10.1016/j.pt.2006.02.006>

- Singer, S. M., Elmendorf, H. G., Conrad, J. T., & Nash, T. E. (2001). Biological Selection of Variant-Specific Surface Proteins in *Giardia lamblia*. *The Journal of Infectious Diseases*, 183(1), 119–124. <https://doi.org/10.1086/317659>
- Singer, S. M., Yee, J., & Nash, T. E. (1998). Episomal and integrated maintenance of foreign DNA in *Giardia lamblia*. *Molecular and Biochemical Parasitology*, 92(1), 59–69. [https://doi.org/10.1016/S0166-6851\(97\)00225-9](https://doi.org/10.1016/S0166-6851(97)00225-9)
- Singh, B. N. (1993). Liposphoglycan-like glycoconjugate of *Tritrichomonas foetus* and *Trichomonas vaginalis*. *Molecular and Biochemical Parasitology*, 57(2), 281–294. [https://doi.org/10.1016/0166-6851\(93\)90204-B](https://doi.org/10.1016/0166-6851(93)90204-B)
- Siniosoglou, S., Peak-Chew, S. Y., & Pelham, H. R. B. (2000). Ric1p and Rgp1p form a complex that catalyses nucleotide exchange on Ypt6p. *The EMBO Journal*, 19(18), 4885–4894. <https://doi.org/10.1093/emboj/19.18.4885>
- Smirnov, A. V., Chao, E., Nassonova, E. S., & Cavalier-Smith, T. (2011). A Revised Classification of Naked Lobose Amoeboae (Amoebozoa: Lobosa). *Protist*, 162(4), 545–570. <https://doi.org/10.1016/j.protis.2011.04.004>
- Smith, B. J. (1994). SDS Polyacrylamide Gel Electrophoresis of Proteins. In J. M. Walker (Ed.), *Basic Protein and Peptide Protocols* (pp. 23–34). Humana Press. <https://doi.org/10.1385/0-89603-268-X:23>
- Söllner, T., Whiteheart, S. W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., & Rothman, J. E. (1993). SNAP receptors implicated in vesicle targeting and fusion. *Nature*, 362(6418), 318–324. <https://doi.org/10.1038/362318a0>
- Soltys, B. J., Falah, M., & Gupta, R. S. (1996). Identification of endoplasmic reticulum in the primitive eukaryote *Giardia lamblia* using cryoelectron microscopy and antibody to BiP. *Journal of Cell Science*, 109, 1909–1917.
- Sosa, R. T., Weber, M. M., Wen, Y., & O'Halloran, T. J. (2012). A single β adaptin contributes to AP1 and AP2 complexes and clathrin function in *Dictyostelium*. *Traffic (Copenhagen, Denmark)*, 13(2), 305–316. PubMed. <https://doi.org/10.1111/j.1600-0854.2011.01310.x>
- Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L., & Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521(7551), 173–179. <https://doi.org/10.1038/nature14447>
- Spang, A., Shiba, Y., & Randazzo, P. A. (2010). Arf GAPs: Gatekeepers of vesicle generation. *FEBS Letters*, 584(12), 2646–2651. <https://doi.org/10.1016/j.febslet.2010.04.005>
- Spang, A., Stairs, C. W., Dombrowski, N., Eme, L., Lombard, J., Caceres, E. F., Greening, C., Baker, B. J., & Ettema, T. J. G. (2019). Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nature Microbiology*, 4(7), 1138–1148. <https://doi.org/10.1038/s41564-019-0406-9>

- Sparvoli, D., & Lebrun, M. (2021). Unraveling the Elusive Rhopty Exocytic Mechanism of Apicomplexa. *Trends in Parasitology*, 37(7), 622–637. <https://doi.org/10.1016/j.pt.2021.04.011>
- Sprangers, J., & Rabouille, C. (2015). SEC16 in COPII coat dynamics at ER exit sites. *Biochemical Society Transactions*, 43(1), 97–103. <https://doi.org/10.1042/BST20140283>
- Stadelmann, B., Merino, M. C., Persson, L., & Svärd, S. G. (2012). Arginine Consumption by the Intestinal Parasite *Giardia intestinalis* Reduces Proliferation of Intestinal Epithelial Cells. *PLOS ONE*, 7(9), e45325. <https://doi.org/10.1371/journal.pone.0045325>
- Stairs, C. W., Kokla, A., Ástvaldsson, Á., Jerlström-Hultqvist, J., Svärd, S., & Ettema, T. J. G. (2019). Oxygen induces the expression of invasion and stress response genes in the anaerobic salmon parasite *Spiroplasma salmonicida*. *BMC Biology*, 17(1), 19. <https://doi.org/10.1186/s12915-019-0634-8>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stamnes, M. A., & Rothman, J. E. (1993). The binding of AP-1 clathrin adaptor particles to Golgi membranes requires ADP-ribosylation factor, a small GTP-binding protein. *Cell*, 73(5), 999–1005. [https://doi.org/10.1016/0092-8674\(93\)90277-W](https://doi.org/10.1016/0092-8674(93)90277-W)
- Stefanic, S., Morf, L., Kulangara, C., Regos, A., Sonda, S., Schraner, E., Spycher, C., Wild, P., & Hehl, A. B. (2009). Neogenesis and maturation of transient Golgi-like cisternae in a simple eukaryote. *Journal of Cell Science*, 122(16), 2846–2856. <https://doi.org/10.1242/jcs.049411>
- Stenmark, H. (2009). Rab GTPases as coordinators of vesicle traffic. *Nature Reviews Molecular Cell Biology*, 10(8), 513–525. <https://doi.org/10.1038/nrm2728>
- Stenmark, H., & Olkkonen, V. M. (2001). The Rab GTPase family. *Genome Biology*, 2(5), reviews3007.1. <https://doi.org/10.1186/gb-2001-2-5-reviews3007>
- Sterud, E., Mo, T. A., & Poppe, T. T. (1997). Ultrastructure of *Spiroplasma barkhanus* N. Sp. (Diplomonadida: Hexamitidae) from Grayling *Thymallus thymallus* (L.) (Salmonidae) and Atlantic Salmon *Salmo salar* L. (Salmonidae). *Journal of Eukaryotic Microbiology*, 44(5), 399–407. <https://doi.org/10.1111/j.1550-7408.1997.tb05715.x>
- Stevens, A. N. P. (2010). Predation, Herbivory, and Parasitism. *Nature Education Knowledge*, 3(10), 36.
- Stiller, J. W. (2007). Plastid endosymbiosis, genome evolution and the origin of green plants. *Trends in Plant Science*, 12(9), 391–396. <https://doi.org/10.1016/j.tplants.2007.08.002>
- Struck, N. S., de Souza Dias, S., Langer, C., Marti, M., Pearce, J. A., Cowman, A. F., & Gilberger, T. W. (2005). Re-defining the Golgi complex in *Plasmodium falciparum* using the novel Golgi marker PfGRASP. *Journal of Cell Science*, 118(23), 5603–5613. <https://doi.org/10.1242/jcs.02673>
- Su, M.-Y., Fromm, S. A., Remis, J., Toso, D. B., & Hurley, J. H. (2021). Structural basis for the ARF GAP activity and specificity of the C9orf72 complex. *Nature Communications*, 12(1), 3786. <https://doi.org/10.1038/s41467-021-24081-0>

- Su, M.-Y., Fromm, S. A., Zoncu, R., & Hurley, J. H. (2020). Structure of the C9orf72 ARF GAP complex that is haploinsufficient in ALS and FTD. *Nature*, *585*(7824), 251–255. <https://doi.org/10.1038/s41586-020-2633-x>
- Szempruch, A. J., Sykes, S. E., Kieft, R., Dennison, L., Becker, A. C., Gartrell, A., Martin, W. J., Nakayasu, E. S., Almeida, I. C., Hajduk, S. L., & Harrington, J. M. (2016). Extracellular Vesicles from *Trypanosoma brucei* Mediate Virulence Factor Transfer and Cause Host Anemia. *Cell*, *164*(1–2), 246–257. <https://doi.org/10.1016/j.cell.2015.11.051>
- Sztul, E., Chen, P.-W., Casanova, J. E., Cherfils, J., Dacks, J. B., Lambright, D. G., Lee, F.-J. S., Randazzo, P. A., Santy, L. C., Schürmann, A., Wilhelmi, I., Yohe, M. E., & Kahn, R. A. (2019). ARF GTPases and their GEFs and GAPs: Concepts and challenges. *Molecular Biology of the Cell*, *30*(11), 1249–1271. <https://doi.org/10.1091/mbc.E18-12-0820>
- Tachezy, J., Sánchez, L. B., & Müller, M. (2001). Mitochondrial Type Iron-Sulfur Cluster Assembly in the Amitochondriate Eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as Indicated by the Phylogeny of IscS. *Molecular Biology and Evolution*, *18*(10), 1919–1928. <https://doi.org/10.1093/oxfordjournals.molbev.a003732>
- Takishita, K., Kolisko, M., Komatsuzaki, H., Yabuki, A., Inagaki, Y., Cepicka, I., Smejkalová, P., Silberman, J. D., Hashimoto, T., Roger, A. J., & Simpson, A. G. B. (2012a). Multigene Phylogenies of Diverse *Carpediemonas*-like Organisms Identify the Closest Relatives of 'Amitochondriate' Diplomonads and Retortamonads. *Protist*, *163*(3), 344–355. <https://doi.org/10.1016/j.protis.2011.12.007>
- Takishita, K., Kolisko, M., Komatsuzaki, H., Yabuki, A., Inagaki, Y., Cepicka, I., Smejkalová, P., Silberman, J. D., Hashimoto, T., Roger, A. J., & Simpson, A. G. B. (2012b). Multigene Phylogenies of Diverse *Carpediemonas*-like Organisms Identify the Closest Relatives of 'Amitochondriate' Diplomonads and Retortamonads. *Protist*, *163*(3), 344–355. <https://doi.org/10.1016/j.protis.2011.12.007>
- Taku, I., Hirai, T., Makiuchi, T., Shinzawa, N., Iwanaga, S., Annoura, T., Nagamune, K., Nozaki, T., & Saito-Nakano, Y. (2021). Rab5b-Associated Arf1 GTPase Regulates Export of N-Myristoylated Adenylate Kinase 2 From the Endoplasmic Reticulum in *Plasmodium falciparum*. *Frontiers in Cellular and Infection Microbiology*, *10*, 610200. <https://doi.org/10.3389/fcimb.2020.610200>
- Tamaki, H., & Yamashina, S. (2002). The Stack of the Golgi Apparatus. *Archives of Histology and Cytology*, *65*(3), 209–218. <https://doi.org/10.1679/aohc.65.209>
- Tanabe, K., Kon, S., Natsume, W., Torii, T., Watanabe, T., & Satake, M. (2006). Involvement of a novel ADP-ribosylation factor GTPase-activating protein, SMAP, in membrane trafficking: Implications in cancer cell biology. *Cancer Science*, *97*(9), 801–806. <https://doi.org/10.1111/j.1349-7006.2006.00251.x>
- Tanabe, K., Torii, T., Natsume, W., Braesch-Andersen, S., Watanabe, T., & Satake, M. (2005). A Novel GTPase-activating Protein for ARF6 Directly Interacts with Clathrin and Regulates Clathrin-

- dependent Endocytosis. *Molecular Biology of the Cell*, 16(4), 1617–1628.
<https://doi.org/10.1091/mbc.e04-08-0683>
- Tang, S., Lomsadze, A., & Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. *Nucleic Acids Research*, 43(12), e78–e78. <https://doi.org/10.1093/nar/gkv227>
- Tanifuji, G., Takabayashi, S., Kume, K., Takagi, M., Nakayama, T., Kamikawa, R., Inagaki, Y., & Hashimoto, T. (2018). The draft genome of *Kipferlia bialata* reveals reductive genome evolution in fornicate parasites. *PLOS One*, 13(3), e0194487. <https://doi.org/10.1371/journal.pone.0194487>
- Tanna, C., Goss, L., Ludwig, C., & Chen, P.-W. (2019). ArfGAPs as Regulators of the Actin Cytoskeleton—An Update. *International Journal of Molecular Sciences*, 20(2), 442. <https://doi.org/10.3390/ijms20020442>
- Tao, L., Han, L., Li, X., Gao, Q., Pan, L., Wu, L., Luo, Y., Wang, W., Zheng, Z., & Guo, X. (2014). Prevalence and risk factors for cervical neoplasia: A cervical cancer screening program in Beijing. *BMC Public Health*, 14(1), 1185. <https://doi.org/10.1186/1471-2458-14-1185>
- Taylor, J. W. (2014). Evolutionary Perspectives on Human Fungal Pathogens. *Cold Spring Harbor Perspectives in Medicine*, 5(9), a019588. <https://doi.org/10.1101/cshperspect.a019588>
- Temkin, P., Lauffer, B., Jäger, S., Cimermanic, P., Krogan, N. J., & von Zastrow, M. (2011). SNX27 mediates retromer tubule entry and endosome-to-plasma membrane trafficking of signalling receptors. *Nature Cell Biology*, 13(6), 715–721. <https://doi.org/10.1038/ncb2252>
- The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), 796–815. <https://doi.org/10.1038/35048692>
- Thomarat, F., Vivarès, C. P., & Gouy, M. (2004). Phylogenetic Analysis of the Complete Genome Sequence of *Encephalitozoon cuniculi* Supports the Fungal Origin of Microsporidia and Reveals a High Frequency of Fast-Evolving Genes. *Journal of Molecular Evolution*, 59(6), 780–791. <https://doi.org/10.1007/s00239-004-2673-0>
- Tikhonenkov, D. V., Jhin, S. H., Eglit, Y., Miller, K., Plotnikov, A., Simpson, A. G. B., & Park, J. S. (2019). Ecological and evolutionary patterns in the enigmatic protist genus *Percolomonas* (Heterolobosea; Discoba) from diverse habitats. *PLOS ONE*, 14(8), e0216188. <https://doi.org/10.1371/journal.pone.0216188>
- Timmis, J. N., Ayliffe, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*, 5(2), 123–135. <https://doi.org/10.1038/nrg1271>
- Togneri, J., Cheng, Y.-S., Munson, M., Hughson, F. M., & Carr, C. M. (2006). Specific SNARE complex binding mode of the Sec1/Munc-18 protein, Sec1p. *Proceedings of the National Academy of Sciences of the United States of America*, 103(47), 17730–17735. <https://doi.org/10.1073/pnas.0605448103>
- Touz, M. C., Kulakova, L., & Nash, T. E. (2004). Adaptor Protein Complex 1 Mediates the Transport of Lysosomal Proteins from a Golgi-like Organelle to Peripheral Vacuoles in the Primitive Eukaryote

- Giardia lamblia*. *Molecular Biology of the Cell*, 15(7), 3053–3060.
<https://doi.org/10.1091/mbc.e03-10-0744>
- Touz, M. C., Rivero, M. R., Miras, S. L., & Bonifacino, J. S. (2012). Lysosomal protein trafficking in *Giardia lamblia*: Common and distinct features. *Frontiers in Bioscience (Elite Edition)*, 4, 1898–1909. <https://doi.org/10.2741/511>
- Touz, M. C., & Zamponi, N. (2017). Sorting without a Golgi complex. *Traffic*, 18(10), 637–645.
<https://doi.org/10.1111/tra.12500>
- Tovar, J., León-Avila, G., Sánchez, L. B., Sutak, R., Tachezy, J., van der Giezen, M., Hernández, M., Müller, M., & Lucocq, J. M. (2003). Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature*, 426(6963), 172–176. <https://doi.org/10.1038/nature01945>
- Tsui, C. K.-M., Miller, R., Uyaguari-Diaz, M., Tang, P., Chauve, C., Hsiao, W., Isaac-Renton, J., & Prystajek, N. (2018). Beaver Fever: Whole-Genome Characterization of Waterborne Outbreak and Sporadic Isolates To Study the Zoonotic Transmission of Giardiasis. *MSphere*, 3(2), e00090-18. <https://doi.org/10.1128/mSphere.00090-18>
- Tůmová, P., Hofštetřová, K., Nohýnková, E., Hovorka, O., & Král, J. (2007). Cytogenetic evidence for diversity of two nuclei within a single diplomonad cell of *Giardia*. *Chromosoma*, 116(1), 65–78. <https://doi.org/10.1007/s00412-006-0082-4>
- Twu, O., de Miguel, N., Lustig, G., Stevens, G. C., Vashisht, A. A., Wohlschlegel, J. A., & Johnson, P. J. (2013). *Trichomonas vaginalis* exosomes deliver cargo to host cells and mediate host:parasite interactions. *PLOS Pathogens*, 9(7), e1003482–e1003482.
<https://doi.org/10.1371/journal.ppat.1003482>
- Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M. B., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., ... Boore, J. L. (2006). *Phytophthora* Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. *Science*, 313(5791), 1261. <https://doi.org/10.1126/science.1128796>
- Ulrich, P. N., Jimenez, V., Park, M., Martins, V. P., Atwood, J., III, Moles, K., Collins, D., Rohloff, P., Tarleton, R., Moreno, S. N. J., Orlando, R., & Docampo, R. (2011). Identification of Contractile Vacuole Proteins in *Trypanosoma cruzi*. *PLOS ONE*, 6(3), e18013.
<https://doi.org/10.1371/journal.pone.0018013>
- Upcroft, J. A., Chen, N., & Upcroft, P. (1996). Mapping variation in chromosome homologues of different *Giardia* strains. *Molecular and Biochemical Parasitology*, 76(1), 135–143.
[https://doi.org/10.1016/0166-6851\(95\)02554-5](https://doi.org/10.1016/0166-6851(95)02554-5)
- Upcroft, J. A., Krauer, K. G., & Upcroft, P. (2010). Chromosome sequence maps of the *Giardia lamblia* assemblage A isolate WB. *Trends in Parasitology*, 26(10), 484–491.
<https://doi.org/10.1016/j.pt.2010.07.002>

- van der Laan, M., Bohnert, M., Wiedemann, N., & Pfanner, N. (2012). Role of MINOS in mitochondrial membrane architecture and biogenesis. *Trends in Cell Biology*, 22(4), 185–192. <https://doi.org/10.1016/j.tcb.2012.01.004>
- van Valen, L. (1973). *A New Evolutionary Law*. PASCAL.
- van Weering, J. R. T., Verkade, P., & Cullen, P. J. (2012). SNX–BAR-Mediated Endosome Tubulation is Co-ordinated with Endosome Maturation. *Traffic*, 13(1), 94–107. <https://doi.org/10.1111/j.1600-0854.2011.01297.x>
- Vance, J. E. (2014). MAM (mitochondria-associated membranes) in mammalian cells: Lipids and beyond. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1841(4), 595–609. <https://doi.org/10.1016/j.bbailip.2013.11.014>
- Vargová, R., Wideman, J. G., Derelle, R., Klimeš, V., Kahn, R. A., Dacks, J. B., & Eliáš, M. (2021). A eukaryote-wide perspective on the diversity and evolution of the ARF GTPase protein family. *Genome Biology and Evolution*, *evab157*. <https://doi.org/10.1093/gbe/evab157>
- Venkatesh, D., Boehm, C., Barlow, L. D., Nankisoor, N. N., O'Reilly, A., Kelly, S., Dacks, J. B., & Field, M. C. (2017). Evolution of the endomembrane systems of trypanosomatids – conservation and specialisation. *Journal of Cell Science*, 130(8), 1421–1434. <https://doi.org/10.1242/jcs.197640>
- Vezi, F., Narzisi, G., & Mishra, B. (2012). Reevaluating assembly evaluations with feature response curves: GAGE and assemblathons. *PLOS One*, 7(12), e52210–e52210. <https://doi.org/10.1371/journal.pone.0052210>
- Vicidomini, G., Bianchini, P., & Diaspro, A. (2018). STED super-resolved microscopy. *Nature Methods*, 15(3), 173–182. <https://doi.org/10.1038/nmeth.4593>
- Vidugiriene, J., & Menon, A. K. (1993). Early lipid intermediates in glycosyl-phosphatidylinositol anchor assembly are synthesized in the ER and located in the cytoplasmic leaflet of the ER membrane bilayer. *Journal of Cell Biology*, 121(5), 987–996. <https://doi.org/10.1083/jcb.121.5.987>
- Vietri, M., Radulovic, M., & Stenmark, H. (2020). The many functions of ESCRTs. *Nature Reviews Molecular Cell Biology*, 21(1), 25–42. <https://doi.org/10.1038/s41580-019-0177-4>
- Visvesvara, G. S., Moura, H., & Schuster, F. L. (2007). Pathogenic and opportunistic free-living amoebae: *Acanthamoeba* spp., *Balamuthia mandrillaris*, *Naegleria fowleri*, and *Sappinia diploidea*. *FEMS Immunology & Medical Microbiology*, 50(1), 1–26. <https://doi.org/10.1111/j.1574-695X.2007.00232.x>
- Voleman, L., Najdrová, V., Ástvaldsson, Á., Tůmová, P., Einarsson, E., Švindrych, Z., Hagen, G. M., Tachezy, J., Svárd, S. G., & Doležal, P. (2017). *Giardia intestinalis* mitosomes undergo synchronized fission but not fusion and are constitutively associated with the endoplasmic reticulum. *BMC Biology*, 15(1), 27. <https://doi.org/10.1186/s12915-017-0361-y>
- Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A., & Woese, C. R. (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature*, 326(6111), 411–414. <https://doi.org/10.1038/326411a0>

- Wampfler, P. B., Faso, C., & Hehl, A. B. (2014). The Cre/loxP system in *Giardia lamblia*: Genetic manipulations in a binucleate tetraploid protozoan. *International Journal for Parasitology*, *44*(8), 497–506. <https://doi.org/10.1016/j.ijpara.2014.03.008>
- Wang, J., Yperman, K., Grones, P., Jiang, Q., Dragwidge, J., Mylle, E., Mor, E., Nolf, J., Eeckhout, D., De Jaeger, G., De Rybel, B., Pleskot, R., & Van Damme, D. (2021). Conditional destabilization of the TPLATE complex impairs endocytic internalization. *Proceedings of the National Academy of Sciences*, *118*(15), e2023456118. <https://doi.org/10.1073/pnas.2023456118>
- Wang, P., Pleskot, R., Zang, J., Winkler, J., Wang, J., Yperman, K., Zhang, T., Wang, K., Gong, J., Guan, Y., Richardson, C., Duckney, P., Vandorpe, M., Mylle, E., Fiserova, J., Van Damme, D., & Hussey, P. J. (2019). Plant AtEH/Pan1 proteins drive autophagosome formation at ER-PM contact sites with actin and endocytic machinery. *Nature Communications*, *10*(1), 5132. <https://doi.org/10.1038/s41467-019-12782-6>
- Watkins, R. R., & Eckmann, L. (2014). Treatment of Giardiasis: Current Status and Future Directions. *Current Infectious Disease Reports*, *16*(2), 396. <https://doi.org/10.1007/s11908-014-0396-y>
- Weigel, A. V., Chang, C.-L., Shtengel, G., Xu, C. S., Hoffman, D. P., Freeman, M., Iyer, N., Aaron, J., Khuon, S., Bogovic, J., Qiu, W., Hess, H. F., & Lippincott-Schwartz, J. (2021). ER-to-Golgi protein delivery through an interwoven, tubular network extending from ER. *Cell*, *184*(9), 2412-2429.e16. <https://doi.org/10.1016/j.cell.2021.03.035>
- Weimer, C., Beck, R., Eckert, P., Reckmann, I., Moelleken, J., Brügger, B., & Wieland, F. (2008). Differential roles of ArfGAP1, ArfGAP2, and ArfGAP3 in COPI trafficking. *Journal of Cell Biology*, *183*(4), 725–735. <https://doi.org/10.1083/jcb.200806140>
- Weisz, F., Lalle, M., Nohynkova, E., Sannella, A. R., Dluhošová, J., & Cacciò, S. M. (2019). Testing the impact of Whole Genome Amplification on genome comparison using the polyploid flagellated *Giardia duodenalis* as a model. *Experimental Parasitology*, *207*, 107776. <https://doi.org/10.1016/j.exppara.2019.107776>
- Welte, M. A., & Gould, A. P. (2017). Lipid droplet functions beyond energy storage. *Biochimica et Biophysica Acta. Molecular and Cell Biology of Lipids*, *1862*(10 Pt B), 1260–1272. <https://doi.org/10.1016/j.bbalip.2017.07.006>
- Wendeler, M. W., Paccaud, J.-P., & Hauri, H.-P. (2007). Role of Sec24 isoforms in selective export of membrane proteins from the endoplasmic reticulum. *EMBO Reports*, *8*(3), 258–264. <https://doi.org/10.1038/sj.embor.7400893>
- Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.-J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X., Liu, Y., ... Rothberg, J. M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature*, *452*(7189), 872–876. <https://doi.org/10.1038/nature06884>

- Whitley, P., Reaves, B. J., Hashimoto, M., Riley, A. M., Potter, B. V. L., & Holman, G. D. (2003). Identification of Mammalian Vps24p as an Effector of Phosphatidylinositol 3,5-Bisphosphate-dependent Endosome Compartmentalization. *Journal of Biological Chemistry*, 278(40), 38786–38795. <https://doi.org/10.1074/jbc.M306864200>
- Whittle, J. R. R., & Schwartz, T. U. (2010). Structure of the Sec13–Sec16 edge element, a template for assembly of the COPII vesicle coat. *Journal of Cell Biology*, 190(3), 347–361. <https://doi.org/10.1083/jcb.201003092>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickner, W., & Schekman, R. (2008). Membrane fusion. *Nature Structural & Molecular Biology*, 15(7), 658–664. <https://doi.org/10.1038/nsmb.1451>
- Wielinga, C., Thompson, R. C. A., Monis, P., & Ryan, U. (2015). Identification of polymorphic genes for use in assemblage B genotyping assays through comparative genomics of multiple assemblage B *Giardia duodenalis* isolates. *Molecular and Biochemical Parasitology*, 201(1), 1–4. <https://doi.org/10.1016/j.molbiopara.2015.05.002>
- Wilking, H., Thamm, M., Stark, K., Aebischer, T., & Seeber, F. (2016). Prevalence, incidence estimations and risk factors of *Toxoplasma gondii* infection in Germany: A representative, cross-sectional, serological study. *Scientific Reports*, 6(1), 22551. <https://doi.org/10.1038/srep22551>
- Williams, B. A. P., Hirt, R. P., Lucocq, J. M., & Embley, T. M. (2002). A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature*, 418(6900), 865–869. <https://doi.org/10.1038/nature00949>
- Wilson, M. E. (1995). Travel and the emergence of infectious diseases. *Emerging Infectious Diseases*, 1(2), 39–46. <https://doi.org/10.3201/eid0102.950201>
- Wiringa, A. E., Ness, R. B., Darville, T., Beigi, R. H., & Haggerty, C. L. (2019). *Trichomonas vaginalis*, endometritis and sequelae among women with clinically suspected pelvic inflammatory disease. *Sexually Transmitted Infections*, sextrans-2019-054079. <https://doi.org/10.1136/sextrans-2019-054079>
- Woessner, D. J., & Dawson, S. C. (2012). The *Giardia* Median Body Protein Is a Ventral Disc Protein That Is Critical for Maintaining a Domed Disc Conformation during Attachment. *Eukaryotic Cell*, 11(3), 292–301. <https://doi.org/10.1128/EC.05262-11>
- Woo, Y. H., Ansari, H., Otto, T. D., Klinger, C. M., Kolisko, M., Michálek, J., Saxena, A., Shanmugam, D., Tayyrov, A., Veluchamy, A., Ali, S., Bernal, A., del Campo, J., Cihlář, J., Flegontov, P., Gornik, S. G., Hajdušková, E., Horák, A., Janouškovec, J., ... Pain, A. (2015). Chromerid genomes reveal

- the evolutionary path from photosynthetic algae to obligate intracellular parasites. *eLife*, 4, e06974. <https://doi.org/10.7554/eLife.06974>
- Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20(1), 257. <https://doi.org/10.1186/s13059-019-1891-0>
- World Health Organization. (2013). *First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases*. https://apps.who.int/iris/bitstream/handle/10665/44440/9789241564090_eng.pdf
- World Health Organization. (2020). *World malaria report 2020—20 years of global progress & challenges*. 300.
- Xu, F., Jerlström-Hultqvist, J., & Andersson, J. O. (2012). Genome-Wide Analyses of Recombination Suggest That *Giardia intestinalis* Assemblages Represent Different Species. *Molecular Biology and Evolution*, 29(10), 2895–2898. <https://doi.org/10.1093/molbev/mss107>
- Xu, F., Jerlström-Hultqvist, J., Einarsson, E., Ástvaldsson, Á., Svärd, S. G., & Andersson, J. O. (2014). The Genome of *Spironucleus salmonicida* Highlights a Fish Pathogen Adapted to Fluctuating Environments. *PLOS Genetics*, 10(2), e1004053. <https://doi.org/10.1371/journal.pgen.1004053>
- Xu, F., Jerlström-Hultqvist, J., Kolisko, M., Simpson, A. G. B., Roger, A. J., Svärd, S. G., & Andersson, J. O. (2016). On the reversibility of parasitism: Adaptation to a free-living lifestyle via gene acquisitions in the diplomonad *Trepomonas* sp. PC1. *BMC Biology*, 14(1), 62. <https://doi.org/10.1186/s12915-016-0284-z>
- Xu, F., Jex, A., & Svärd, S. G. (2020). A chromosome-scale reference genome for *Giardia intestinalis* WB. *Scientific Data*, 7(1), 38. <https://doi.org/10.1038/s41597-020-0377-y>
- Xu, F., Jiménez-González, A., Einarsson, E., Ástvaldsson, Á., Peirasmaki, D., Eckmann, L., Andersson, J. O., Svärd, S. G., & Jerlström-Hultqvist, J. (2020). The compact genome of *Giardia muris* reveals important steps in the evolution of intestinal protozoan parasites. *Microbial Genomics*, 6(8). <https://doi.org/10.1099/mgen.0.000402>
- Yahara, N., Ueda, T., Sato, K., & Nakano, A. (2001). Multiple Roles of Arf1 GTPase in the Yeast Exocytic and Endocytic Pathways. *Molecular Biology of the Cell*, 12(1), 221–238. <https://doi.org/10.1091/mbc.12.1.221>
- Yan, D., Wang, Y., Murakami, T., Shen, Y., Gong, J., Jiang, H., Smith, D. R., Pombert, J.-F., Dai, J., & Wu, Q. (2015). *Auxenochlorella protothecoides* and *Prototheca wickerhamii* plastid genome sequences give insight into the origins of non-photosynthetic algae. *Scientific Reports*, 5(1), 14465. <https://doi.org/10.1038/srep14465>
- Yang, J.-S., Lee, S. Y., Gao, M., Bourgoïn, S., Randazzo, P. A., Premont, R. T., & Hsu, V. W. (2002). ARFGAP1 promotes the formation of COPI vesicles, suggesting function as a component of the coat. *The Journal of Cell Biology*, 159(1), 69–78. PubMed. <https://doi.org/10.1083/jcb.200206015>
- Yang, Z., Vild, C., Ju, J., Zhang, X., Liu, J., Shen, J., Zhao, B., Lan, W., Gong, F., Liu, M., Cao, C., & Xu, Z. (2012). Structural Basis of Molecular Recognition between ESCRT-III-like Protein Vps60 and

- AAA-ATPase Regulator Vta1 in the Multivesicular Body Pathway. *Journal of Biological Chemistry*, 287(52), 43899–43908. <https://doi.org/10.1074/jbc.M112.390724>
- Yarwood, R., Hellicar, J., Woodman, P. G., & Lowe, M. (2020). Membrane trafficking in health and disease. *Disease Models & Mechanisms*, 13(4). <https://doi.org/10.1242/dmm.043448>
- Yazaki, E., Kume, K., Shiratori, T., Eglit, Y., Tanifuji, G., Harada, R., Simpson, A. G. B., Ishida, K., Hashimoto, T., & Inagaki, Y. (2020). Barthelonids represent a deep-branching metamonad clade with mitochondrion-related organelles predicted to generate no ATP. *Proceedings of the Royal Society B: Biological Sciences*, 287(1934), 20201538. <https://doi.org/10.1098/rspb.2020.1538>
- Yoshimura, S., Gerondopoulos, A., Linford, A., Rigden, D. J., & Barr, F. A. (2010). Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors. *Journal of Cell Biology*, 191(2), 367–381. <https://doi.org/10.1083/jcb.201008051>
- Yperman, K., Wang, J., Eeckhout, D., Winkler, J., Vu, L. D., Vandorpe, M., Grones, P., Mylle, E., Kraus, M., Merceron, R., Nolf, J., Mor, E., De Bruyn, P., Loris, R., Potocký, M., Savvides, S. N., De Rybel, B., De Jaeger, G., Van Damme, D., & Pleskot, R. (2021). Molecular architecture of the endocytic TPLATE complex. *Science Advances*, 7(9), eabe7999. <https://doi.org/10.1126/sciadv.abe7999>
- Yubuki, N., Huang, S. S. C., & Leander, B. S. (2016). Comparative Ultrastructure of Fornicate Excavates, Including a Novel Free-living Relative of Diplomonads: *Aduncisulcus paluster* gen. et sp. nov. *Protist*, 167(6), 584–596. <https://doi.org/10.1016/j.protis.2016.10.001>
- Yubuki, N., Inagaki, Y., Nakayama, T., & Inouye, I. (2007). Ultrastructure and Ribosomal RNA Phylogeny of the Free-Living Heterotrophic Flagellate *Dysnectes brevis* n. Gen., n. Sp., a New Member of the Fornicata. *The Journal of Eukaryotic Microbiology*, 54(2), 191–200. <https://doi.org/10.1111/j.1550-7408.2007.00252.x>
- Yubuki, N., Simpson, A. G. B., & Leander, B. S. (2013). Comprehensive Ultrastructure of *Kipferlia bialata* Provides Evidence for Character Evolution within the Fornicata (Excavata). *Protist*, 164(3), 423–439. <https://doi.org/10.1016/j.protis.2013.02.002>
- Yüzbaşıoğlu, E., Dalyan, E., Memon, A., Öz, G., & Yüksel, B. (2017). Functional specialization of Arf paralogs in nodules of model legume, *Medicago truncatula*. *Plant Growth Regulation*, 81(3), 501–510. <https://doi.org/10.1007/s10725-016-0227-2>
- Záhonová, K., Lax, G., Sinha, S. D., Leonard, G., Richards, T. A., Lukeš, J., & Wideman, J. G. (2021). Single-cell genomics unveils a canonical origin of the diverse mitochondrial genomes of euglenozoans. *BMC Biology*, 19(1), 103. <https://doi.org/10.1186/s12915-021-01035-y>
- Zamponi, N., Zamponi, E., Mayol, G. F., Lanfredi-Rangel, A., Svärd, S. G., & Touz, M. C. (2017). Endoplasmic reticulum is the sorting core facility in the Golgi-lacking protozoan *Giardia lamblia*. *Traffic*, 18(9), 604–621. <https://doi.org/10.1111/tra.12501>
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T.,

- Banfield, J. F., Schramm, A., Baker, B. J., Spang, A., & Ettema, T. J. G. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, *541*(7637), 353–358. <https://doi.org/10.1038/nature21031>
- Žárský, V., Klimeš, V., Pačes, J., Vlček, Č., Hradilová, M., Beneš, V., Nývltová, E., Hrdý, I., Pyrih, J., Mach, J., Barlow, L., Stairs, C. W., Eme, L., Hall, N., Eliáš, M., Dacks, J. B., Roger, A., & Tachezy, J. (2021). The *Mastigamoeba balamuthi* Genome and the Nature of the Free-Living Ancestor of *Entamoeba*. *Molecular Biology and Evolution*, *38*(6), 2240–2259. <https://doi.org/10.1093/molbev/msab020>
- Zhang, Z.-F., & Begg, C. B. (1994). Is *Trichomonas vaginalis* a Cause of Cervical Neoplasia? Results from a Combined Analysis of 24 Studies. *International Journal of Epidemiology*, *23*(4), 682–690. <https://doi.org/10.1093/ije/23.4.682>
- Zhen, Y., Spangenberg, H., Munson, M. J., Brech, A., Schink, K. O., Tan, K.-W., Sørensen, V., Wenzel, E. M., Radulovic, M., Engedal, N., Simonsen, A., Raiborg, C., & Stenmark, H. (2020). ESCRT-mediated phagophore sealing during mitophagy. *Autophagy*, *16*(5), 826–841. <https://doi.org/10.1080/15548627.2019.1639301>
- Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*(21), 2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>
- Zimorski, V., Ku, C., Martin, W. F., & Gould, S. B. (2014). Endosymbiotic theory for organelle origins. *Growth and Development: Eukaryotes/ Prokaryotes*, *22*, 38–48. <https://doi.org/10.1016/j.mib.2014.09.008>
- Zumthor, J. P., Cernikova, L., Rout, S., Kaech, A., Faso, C., & Hehl, A. B. (2016). Static Clathrin Assemblies at the Peripheral Vacuole—Plasma Membrane Interface of the Parasitic Protozoan *Giardia lamblia*. *PLoS Pathogens*, *12*(7), e1005756. <https://doi.org/10.1371/journal.ppat.1005756>
- Zwiewka, M., Feraru, E., Möller, B., Hwang, I., Feraru, M. I., Kleine-Vehn, J., Weijers, D., & Friml, J. (2011). The AP-3 adaptor complex is required for vacuolar function in *Arabidopsis*. *Cell Research*, *21*(12), 1711–1722. <https://doi.org/10.1038/cr.2011.99>

APPENDIX A

Online supplemental material

All appendix material referenced in Chapters 2 through 6 has been made available online and is accessible using the following link:

https://drive.google.com/drive/folders/1S_AQXPe884An0XMsi86llZIZQPbm3tBN?usp=sharing