A microscopic, transcriptomic, and phylogenetic investigation of the contractile vacuole of *Reclinomonas americana* (Jakobida, Discoba)

by

Kiran John More

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

Master of Science

in

Systematics and Evolution

Department of Biological Sciences

University of Alberta

Abstract

The origin and timing of characteristic eukaryotic organelles is one of the key questions in the evolution of life on Earth. For the most part, widespread eukaryotic organelles (e.g., Golgi, mitochondria) have been inferred to be present in the last eukaryotic common ancestor (LECA) by examining the distribution of the organelles and their marker genes across the diversity of eukaryotes (the vast majority of which are unicellular protists). Other common organelles, such as the contractile vacuole (CV), are much more poorly understood. The CV is an osmoregulatory organelle that is found across the diversity of eukaryotes, most prominently in protists from freshwater ecosystems. As freshwater is hypoosmotic to the cell, there is continual movement of water across the plasma membrane, which cells must counteract to prevent lysis. The CV gathers and expels this excess water. Due to its absence from many of the model organisms typically used for cell biological study, there is a paucity of molecular data associated with this organelle. The limited data available from four distantly related protist model organisms show a pattern of repeat involvement of specific proteins known from other organelles. The considerable morphological diversity and patchy distribution (as CVs are generally absent from closely related marine and parasitic species) of these organelles makes their homology uncertain. Are CVs a single, homologous organelle with an origin in the LECA, or have there been multiple more recent evolutionary origins of an osmoregulatory organelle?

Through a combination of microscopy, transcriptomics, and phylogenetics I investigated this question within the eukaryotic supergroup Discoba using the CV model organism *Trypanosoma cruzi* as a reference point. I microscopically characterized the CV response of the non-model organism *Reclinomonas americana* to low, control, and high osmolarities. These treatments were then used to induce CV activity, and RNA was harvested for differential expression analysis. Transcripts of the membrane trafficking system that increased as osmolarity decreased were considered putatively CV-

associated. The associated proteins were then used as a proxy to investigate the evolutionary history of the organelle. If CVs have a single origin, I would expect the CV-specific paralogs to form clades in phylogenies to the exclusion of versions found elsewhere in the cell. This was never observed. Though several common CV-associated proteins were detected as differentially expressed in *R. americana*, phylogenetics indicated that these were the result of lineage-specific expansions. These results show no strong support for a single origin of CVs within Discoba, or across eukaryotes. Rather, they support either convergent or parallel evolution of the CVs examined. The reoccurrence of the same proteins in the CVs of diverse eukaryotes raises the question of whether these CVs evolved from the same non-CV organelle with latent osmoregulatory capacity.

Preface

(Mandatory due to collaborative work)

Some of the material in the introduction of this thesis has been published previously. Figure 1 and the text from section 1.2 were reproduced with modifications from an article I co-authored: More K., Klinger C. M., Barlow L. D. & Dacks J. B. 2020. Evolution and natural history of membrane trafficking in eukaryotes. *Current Biology*, 30:R553–R564. doi: 10.1016/j.cub.2020.03.068. Current Biology is an Elsevier journal that allows reuse of published materials by authors, therefore no additional permissions were required for publication within this thesis. Both the writing and figure were my own products, and were checked for errors by all co-authors.

Text from section 1.5 has been modified from and elaborated on an upcoming chapter on contractile vacuoles from a community-sourced virtual protistology textbook, conceptualized and solicited by Alastair G. B. Simpson. I was the primary writer and researcher of this chapter, but structural and conceptual input, as well as manuscript edits, were provided by Alastair G. B. Simpson and Joel B. Dacks.

Acknowledgements

First of all, I would like to thank my supervisor, Joel Dacks—his unending enthusiasm for evolutionary biology, as well as his kindness and encouragement, have been invaluable throughout this degree. This would have been a much different experience without his support. I am grateful to the other members of the Dacks lab (past and present) for enriching my scientific journey with their unique expertise and enthusiasm. In no particular order: Paul, Shweta, Naomi, Beth, Lael, Will, Harpreet, Kacper, Ahbi, Kika, Matthias, and Eve. I would also like to thank our collaborators in the lab of Dr. Franz Lang for providing cultures, data, and guidance.

I would like to thank my committee members Sally Leys and Chris Thompson. Their input greatly improved this project and helped me move past unanticipated roadblocks. I am particularly grateful to Sally, who welcomed me into her group when I arrived at the university: both graciously allowing me in her laboratory space when the Dacks lab space had not yet been set up, and creating an exciting space to talk about all manners of science. I would also like to thank the examining committee (the above and Patrick Hanington), who brought new-to-me perspectives to our conceptual discussion of CV evolution.

Finally, thank you to my family for nurturing my curiosity about the natural world ever since I was a child, and, of course, to my friends who helped me get through an unexpectedly challenging few years.

This work was made possible through funding from the Natural Sciences and Engineering Research Council of Canada, the Government of Alberta, and the University of Alberta.

Table of contents

List of tables

List of figures

1 Introduction

1.1 Overview of the question

The contractile vacuole (CV) is an osmoregulatory organelle found broadly across eukaryotes, most prominently in protists from freshwater ecosystems. As freshwater is hypoosmotic relative to the cytoplasm of the cell, excess water that diffuses across the cell membrane must be removed for the cell to maintain homeostasis and cell size. Despite the widespread distribution of CVs across eukaryote diversity, what little known about this organelle is fragmentary. Most knowledge comes from a few model organisms, and there is little study of how well this can be extrapolated to the rest of eukaryotes, or even between models. In fact, it is currently unclear whether the CV is a single, homologous organelle retained across the breadth of the eukaryotic tree of life, or whether there have been multiple evolutionary origins of an organelle controlling osmoregulation. This thesis contributes to the study of this organelle from an evolutionary perspective by using differential expression analysis in the non-model organism *Reclinomonas americana* (Jakobida, Discoba) to identify putative CV-associated proteins that can then be used as an evolutionary proxy for the organelle. Using the CV model organism *Trypanosoma cruzi* as a reference point, I perform a comparative analysis of CV behaviour and molecular machinery both within the supergroup Discoba, and, where possible, across the other three major CV model eukaryotes.

1.2 The diversity of eukaryotes is largely unicellular[1]

To effectively discuss the evolutionary origins of a eukaryotic organelle, I will first summarize the known diversity of eukaryotes. Though most of the eukaryotes we knowingly interact with in our day-to-day lives are large and multicellular, this belies the fact that the diversity of eukaryotes, and indeed, life in general, is unicellular (del Campo et al. 2014). The diversity of eukaryotes can mostly be divided into a number of larger "supergroups" (Figure 1). Few of these supergroups have defining morphological characteristics — instead, they have been identified through concatenated phylogenies of hundreds of genes. Most supergroups are diverse assemblages of organisms with huge variation in life history, nutritional method, and body plan (for details, see Adl et al. 2019). The following is a summary of the current supergroups; for a thorough review of the concept, see Burki et al. (2019).

[1] The text and figure from this section has been reproduced and modified from Box 1 and Figure 1 of More et al. (2020). Both writing and figure were my own products, and manuscript editing was contributed by all co-authors.

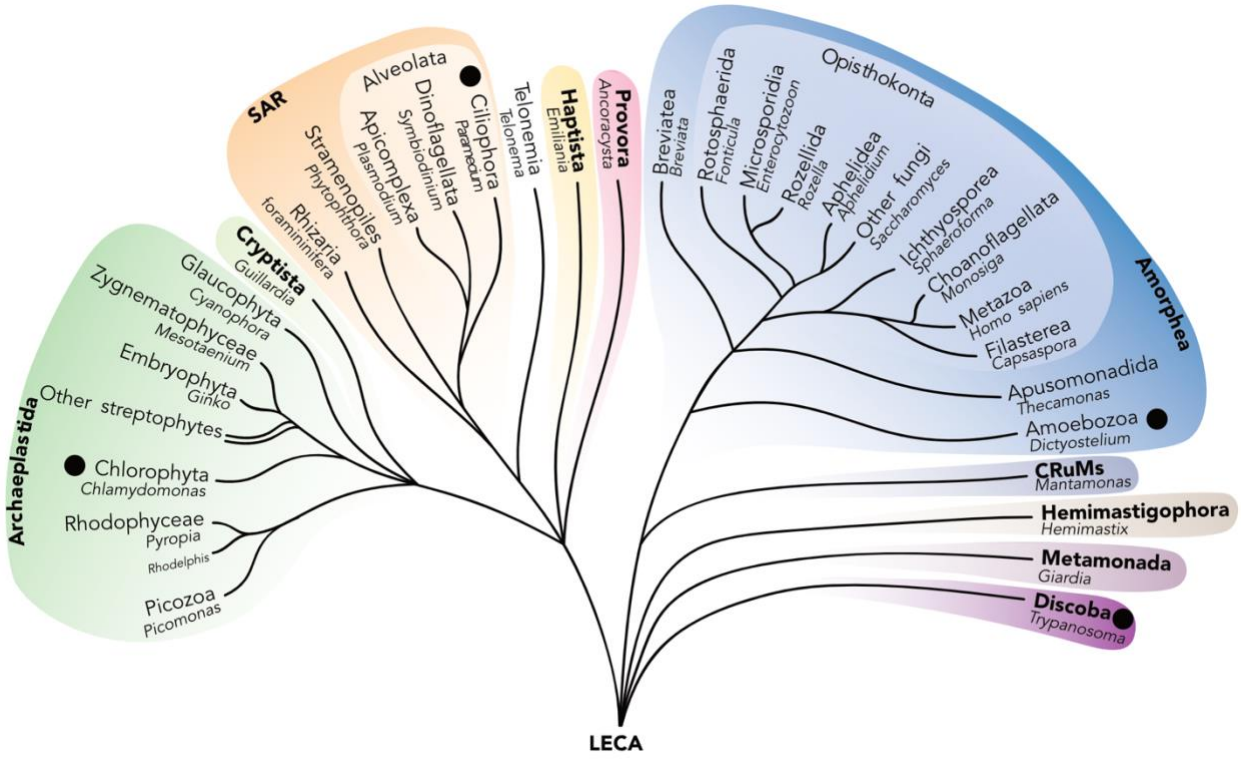


Figure 1 A current view of the eukaryotic tree of life. Supergroups are in bold. Lineages with contractile vacuole model organisms are labelled with black circles. Lineages within Opisthokonta and Chloroplastida are expanded to show the closest unicellular relatives of familiar multicellular organisms. Tree topology is largely based on Adl et al. (2019) and Burki et al. (2020), with Provora from Tikhonenkov et al. (2022), expansions in Opisthokonta from Hehenberger et al. (2017) and Galindo et al. (2022), and expansions in Chloroplastida from de Vries and Archibald (2018). Taxonomy follows the naming conventions set by Adl et al. (2019). This figure is modified from More et al. (2020).

Familiar multicellular organisms are closely related to several unicellular protistan lineages in the supergroups Amorphea and Archaeplastida. Within Amorphea, animals (Metazoa) are related to several lineages of free-living phagotrophic flagellates and amoebae (Choanoflagellata, Filasterea), as well as parasites (Ichthyosporea) (Hehenberger et al. 2017). Fungi, also within Amorphea, encompasses familiar fungi (i.e., yeast and filamentous fungi) as well as the endoparasites Microsporidia and Rozellida, and parasitoids Aphelidea (Galindo et al. 2022). Sister to all fungi are the free-living, phagotrophic amoebae Rotosphaerida (Liu et al. 2009). Both these larger groups (Holozoa and Nucletmycea) form Opisthokonta, which in turn forms Obazoa with the heterotrophic flagellates Apusomonadida and Breviatea (Brown et al. 2013). Amoebozoa is a large group of mostly free-living, heterotrophic amoebae (although it does include the notable human parasite *Entamoeba* and a few flagellated organisms), also within Amorphea (Nikolaev et al. 2006; Lahr et al. 2011). Land plants

(Embryophyta) are found within Archaeplastida. Unicellular and multicellular Rhodophyceae (red algae) and Chloroplastida (land plants and green algae), as well as unicellular Glaucophyta, are all within Archaeplastida. Some analyses show the supergroup Cryptista, which contains phototrophic and phagotrophic flagellates, and the orphan lineage Picozoa, which contains exclusively heterotrophic flagellates, within Archaeplastida as well (e.g., Burki et al. 2016; Schön et al. 2021).

The supergroup SAR comprises the majority of distinct eukaryotic sequences in environmental gene surveys (del Campo et al. 2014) and is an acronym for the lineages it contains: Stramenopiles, Alveolata, and Rhizaria. Stramenopiles are heterotrophic or photosynthetic, including oomycete "fungal" blights that affect many organisms (notably plants), algae that produce a significant amount of atmospheric oxygen, and multicellular seaweeds that form complex ecosystems in the ocean (Derelle et al. 2016). Alveolata includes Dinoflagellata, Ciliophora, and Apicomplexa, which contain important marine algae, predators of other microbial eukaryotes, and symbionts and parasites (e.g., *Plasmodium* and *Toxoplasma*). Rhizaria contains mostly phagotrophic amoebae and flagellates, such as the shell-forming amoeboid foraminifera that are used as pollution indicator species in many environments. The supergroup Metamonada (previously grouped with Discoba as the supergroup "Excavata" based on a number of unusual ultrastructural characteristics and some phylogenomic analyses) contains anaerobic flagellates, including human pathogens (e.g., *Giardia* and *Trichomonas*), free-living heterotrophs, and gut symbionts (Cavalier-Smith 2003; Simpson 2003; Hampl et al. 2009). Discoba contains ecologically significant heterotrophs (e.g., diplonemids), as well as human pathogens (e.g., *Trypanosoma* and *Naegleria*), and phototrophs (e.g., *Euglena*) (Yabuki et al. 2011; Flegontova et al. 2016).

Several new major lineages of eukaryotes have been described in recent years. Some of these are within, or sister to, previously established supergroups (e.g., a few heterotrophic lineages within Cryptista and CRuMs, the phagotrophic *Rhodelphis* sister to Rhodophyceae, and the rappemonads, a novel group of algae within Haptista) (Brown et al. 2018; Gawryluk et al. 2019; Kawachi et al. 2021; Yazaki et al. 2022). Others have been formed from known groups that were previously thought to be unrelated, including the supergroups CRuMs, containing heterotrophic flagellates and amoebae, and Haptista, containing haptophyte algae and centrohelid amoebae (Burki et al. 2016; Brown et al. 2018). Others branch robustly external to previously established supergroups, such as Hemimastigophora and Provora (Janouškovec et al. 2017; Lax et al. 2018; Tikhonenkov et al. 2022). The last group, Provora, may be of particular significance in inferring morphological traits of the last eukaryotic common ancestor (LECA), as species share some unusual characteristics found in the "typical

excavate" taxa within Metamonada and Discoba (i.e., a ventral feeding groove and flagellar vanes; Simpson 2003; Tikhonenkov et al. 2022).

1.3 The endomembrane system is characteristic of eukaryotes

One of the major traits that sets eukaryotes apart from bacteria and archaea is the presence of a complex and interconnected set of membrane-bound organelles (Figure 2). These organelles separate metabolic processes spatially within the cell. The endomembrane system (i.e., membrane-bound organelles excluding mitochondria and plastids) can be conceptualized as two smaller systems: secretory and endocytic. The secretory system is composed of the endoplasmic reticulum (ER), the Golgi, and the *trans*-Golgi network (TGN). The Golgi receives vesicles containing protein cargo from the ER. Through the process of cisternal maturation, this cargo is shuttled from the *cis*-Golgi, proximal to the ER, to the TGN (Glick and Luini 2011). Vesicles are dispatched from the TGN to other organelles within the cell, or to the plasma membrane for secretion. The endolysosomal system is composed of a variety of endosomal organelles, including early endosomes, late endosomes (sometimes known as multivesicular bodies), recycling endosomes (sometimes known as the tubular endosomal network), lysosomes, and autophagosomes. The early endosome receives endocytic vesicles from the plasma membrane and gradually matures into the late endosome, which eventually fuses with the acidic lysosome to digest its contents (Huotari and Helenius 2011). Endocytic contents not destined for degradation may be shuttled to the recycling endosome to be returned to the plasma membrane or the Golgi (Bonifacino and Rojas 2006). Retrograde transport occurs between the endosomal system and TGN, within the Golgi, and between the Golgi and ER. While the autophagosome is not strictly an endocytic organelle, as it is thought to originate from the ER and is involved in the degradation of intracellular contents, it does intersect with endosomes and lysosomes (Tooze et al. 2014). The pre-autophagosome engulfs cytoplasmic contents in a process called macroautophagy to form the double membrane-bound autophagosome, which then fuses with the lysosome to form the autolysosome for final digestion of its contents.

Through a combination of morphological and molecular data, the core set of membrane trafficking organelles described above has been inferred as present in the LECA (Koumandou et al. 2013). However, there is considerable variation even within the well-studied endomembrane systems of plant, yeast, and parasite models. In *Arabidopsis thaliana* (Archaeplastida), the TGN plays the role of the early endosome as well, both receiving endocytic cargo and distributing secretory cargo (Viotti et al. 2010). Other organisms, such as *Giardia lamblia*, have reduced and altered their organellar

complement so much as for familiar compartments to be unrecognizable without the use of molecular markers (Abodeely et al. 2009). Other organisms still have greatly expanded their organellar repertoire, such as the secretory organelles that play a role in host invasion in apicomplexans (SAR) (Gubbels and Duraisingh 2012) or the diverse suite of endolysosome-related organelles in humans (Delevoye et al. 2019). There are still more organelles that are present in a wide array of eukaryotes but have not been studied sufficiently on a molecular level to determine homology (reviewed in More et al. 2020). These include the morphologically diverse extrusomes (involved in a variety of functions including predation or defense), the polyphosphate-dense acidocalcisomes, and the osmoregulatory contractile vacuole (Hausmann 1978; Patterson 1980; Docampo et al. 2010).

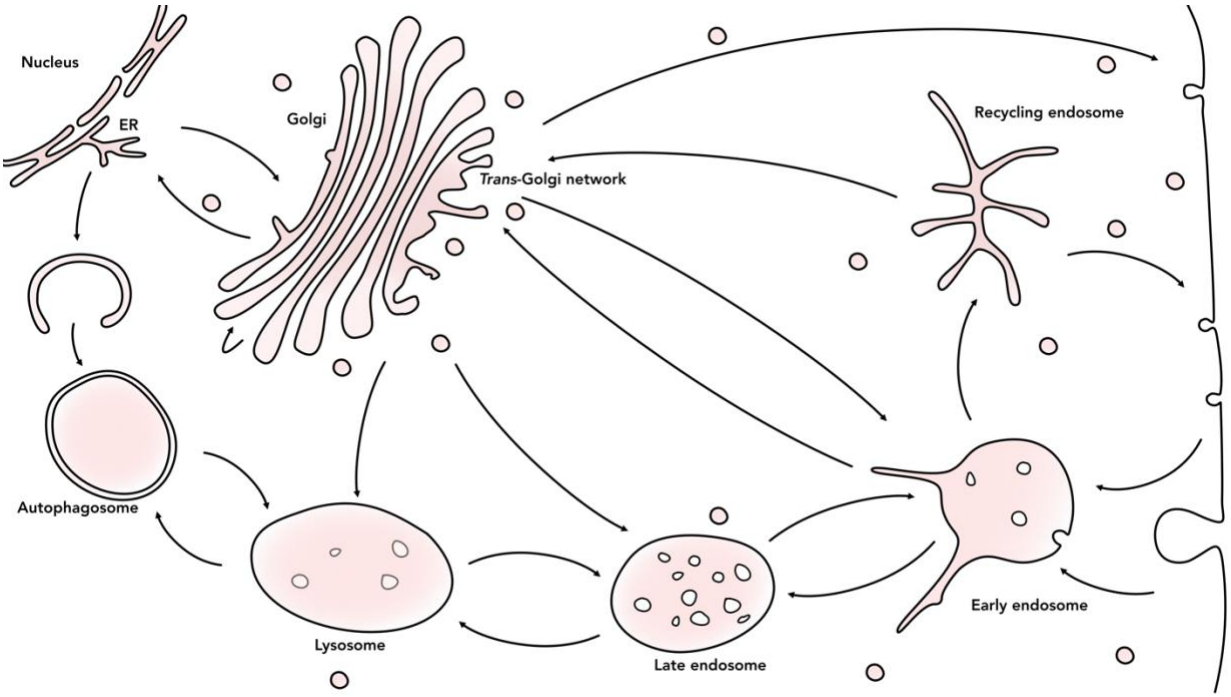


Figure 2 The core organelles of the membrane trafficking system. Redrawn from Schlacht et al. (2014) with modifications from Huotari and Helenius (2011) and Tooze et al. (2014).

1.3.1 Vesicle formation and fusion is a conserved process

The movement of materials between these compartments is largely achieved through the formation of membrane-bound vesicles and/or tubules at the donor membrane, and the fusion of these with an acceptor membrane (Figure 3). A generalized model of this process has been inferred based on extensive cell biological work in animals and yeast (reviewed in Bonifacino and Glick 2004). A summary of this process follows, with emphasis on the protein families relevant to this thesis.

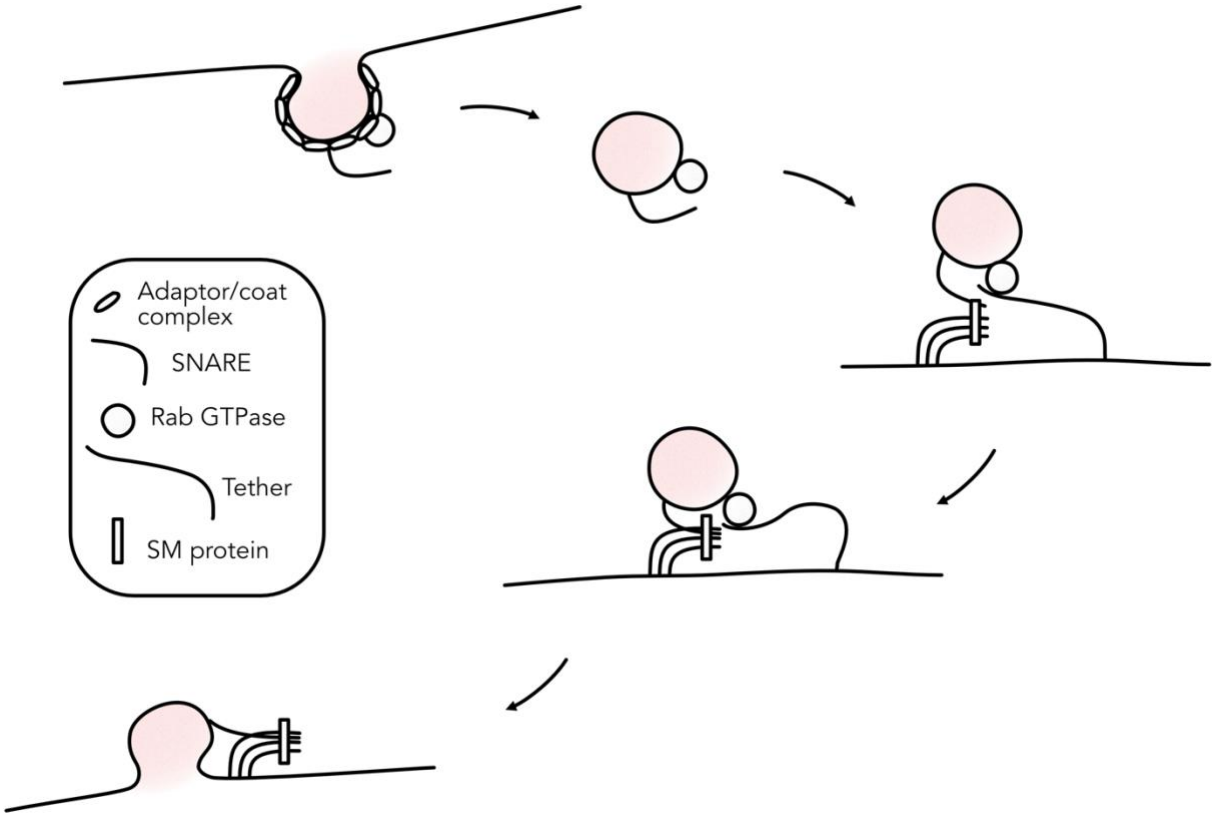


Figure 3 A simplified representation of the process of vesicle budding and fusion.

Vesicle formation is initiated at the donor membrane by a member of the Arf or Sar small GTPase families and their regulators. Cargo-selective adaptor protein complexes are recruited to the membrane (Figure 3), including COPII, COPI, retromer, and adaptor proteins (APs) 1-5. These complexes act as a scaffold for specific membrane-deforming coat proteins (e.g., clathrin in the case of APs 1 and 2, or the COPI α and β′ subunits) (Dacks and Robinson 2017). Scission of the vesicle from the donor membrane can occur in a GTPase-dependent or independent manner (Adolf et al. 2013; Cheng et al. 2021). Rab GTPases act as regulators of diverse activities within the cell, recruiting and activating various effector proteins (Stenmark 2009). While Rab GTPases are generally thought of as a part of the vesicle fusion machinery, increasing evidence places them upstream of this as well, involved in vesicle transport through effector-mediated interactions with the cytoskeleton (Kjos et al. 2018). Rab GDP-GTP exchange factors (Rab GEFs) induce the exchange of a Rab GTPase-bound GDP for GTP, which switches the Rab GTPase "on" and allows the Rab GTPase to associate with a

membrane and its effector proteins (Barr and Lambright 2010). Activated Rab GTPases can have remarkably specific localizations within the cell, and can be used as indicators of membrane identity (e.g., Sönnichsen et al. 2000). Rab GTPase activating proteins (Rab GAPs) catalyze the hydrolysis of GTP to GDP, which disassociates the GTPase from the membrane (Barr and Lambright 2010). During transport to the acceptor membrane, the coat is shed.

Vesicle fusion is initiated once the vesicle is within proximity to the acceptor membrane (Figure 3). Tethering factors, which can either be large coiled coil proteins (up to 3,000 residues) or much smaller multisubunit tethering complexes (MTCs), are thought to bring the vesicle into proximity with the acceptor membrane prior to fusion through interactions with a vesicle-bound Rab GTPase (Bröcker et al. 2010). Mediating vesicle fusion are the soluble N-ethylmaleimide-sensitive factor (NSF) attachment protein (SNAP) receptors (SNAREs). There are four distinct SNARE families, which are named Qa, Qb, Qc, and R SNAREs based on the central residue of the SNARE domain containing a glutamine (Q) or an arginine (R) (Fasshauer et al. 1998). Three Q SNAREs, one from each family, are bundled on one membrane, with the R SNARE on the other. During vesicle fusion, these four SNAREs snap into a coiled coil *trans*-SNARE complex spanning both donor and acceptor membranes, which may provide the energy required to initiate membrane fusion (Risselada and Mayer 2020). Increasing evidence suggests the assembly of the fusogenic trans-SNARE complex is chaperoned by SM proteins (Zhang and Hughson 2021). SM proteins may also interfere with premature disassembly of an intermediate SNARE complex by NSF. After fusion, the *cis*-SNARE complex, now entirely on the acceptor membrane, is disassembled by recruitment of α-SNAP and NSF.

1.3.2 Vesicle formation and fusion is mediated by paralogous protein families

The vast majority of the proteins named above belong to paralogous protein families. Most, but not all, adaptor protein complexes belong to the heterotetrameric adaptor complex (HTAC) family, while their coats share an α-solenoid-β-propellor domain structure that has been interpreted as signs of homology (Field et al. 2011; Dacks and Robinson 2017). Rab GTPases are a single family of proteins, as are almost all known Rab GAPs, which have Tre-2/Bub2/Cdc16 (TBC) domains (Barr and Lambright 2010). SNARE proteins are homologous within the Qa, Qb, Qc, and R SNARE families, and also across the four families (Fasshauer et al. 1998). While evidence for homology within the MTCs is not as strong as for the above, studies of the sequence and structural similarity between protein complexes have demonstrated substantial similarities. The evidence is most robust for the

exocyst, dependent on Sly1-20 (DSL1), Golgi-associated retrograde protein (GARP), endosome-associated recycling protein (EARP), and conserved oligomeric Golgi (COG) complexes, which have been grouped as "complexes associated with tethering containing helical rods" (CATCHR) (Yu and Hughson 2010; Santana-Molina et al. 2021). A notable exception to the tendency of paralogy within the membrane trafficking system are most of the Rab GEFs (Barr and Lambright 2010), which will not be considered further in this thesis.

Individual paralogs of these protein families generally act at specific conserved locations within the cell. For example, within the HTACs there are specific complexes that function in intra-Golgi and retrograde Golgi-to-ER transport (COPI), endocytosis (TSET, AP2), and transport within the endolysosomal system (AP1, 3, 4, 5) (Dacks and Robinson 2017). Similar patterns are also seen for Rab GTPases, MTCs, and SNAREs. The interactions of these compartment specific paralogs seems to confer organellar identity (Cai et al. 2007), and on a broad scale their localization is conserved across the diversity of eukaryotes (Klinger et al. 2016). In a systematic review of cell biologically characterized membrane trafficking proteins in model systems from five of the ten eukaryotic supergroups, Klinger et al. (2016) encountered broadly conserved localizations and inferred functions for each of the seven proteins and complexes examined. Though each model's unique biology had to be considered (the selection included phototrophs, phagotrophs, and parasites, as well as complex unicells and organisms that are constitutively or aggregately multicellular), there was a thread of functional conservation that could be followed across the models. A caveat to this is that in organisms that have greatly expanded complements of specific proteins, not all share the inferred ancestral role and may instead be tissue-, life-stage-, or region-specific (Klinger et al. 2016).

1.4 Organellar evolution and the membrane trafficking system

The observation that the same families of membrane trafficking proteins operate at most of the organelles within the endomembrane system—often with organelle-specific paralogs—has led to the hypothesis that the evolution of differentiated compartments is driven by the duplication and neofunctionalization of genes corresponding to interacting, specificity-encoding proteins (Dacks and Field 2007; Dacks et al. 2008). This idea was formally described as the Organelle Paralogy Hypothesis (OPH) (Figure 4), and is currently the most robust—if not the only—model for the evolution of non-endosymbiotic membrane-bound organelles. Computer simulations support the idea that the duplication of interacting partners is sufficient to drive the evolution of distinct populations of compartments (Ramadas and Thattai 2013), which in turn gives the variation needed for adaptive

evolution. A consequence of the OPH is that by examining the evolutionary history of membrane trafficking proteins, the evolutionary relationships between the compartments they belong to can also be resolved (Figure 4C; Mast et al. 2014). Rab GTPases seem to coalesce into ancient endocytic and exocytic clades, while HTAC phylogenies suggest the diversity of pathways within the endolysosomal system is more recent than endocytosis, intra-Golgi, and retrograde Golgi-to-ER trafficking (notably, anterograde ER-to-Golgi trafficking is performed by a non-HTAC adaptor protein complex, COPII) (Eliáš et al. 2012; Hirst et al. 2014; Schlacht and Dacks 2015).

This figure was removed because of copyright restrictions. It was the schematic of the organelle paralogy hypothesis from Figure 2C of Mast et al. (2014), *Trends in Cell Biology*, 24:435–442 (doi:10.1016/j.tcb.2014.02.003), which was modified from Figure 1B, C of Schlacht et al. (2014), *Cold Spring Harbor Perspectives in Biology*, 6:a016048 (doi:10.1101/cshperspect.a016048).

Figure 4 The organelle paralogy hypothesis. Specificity of the endomembrane system is encoded in interacting sets of membrane trafficking proteins (**A**) which can duplicate and diverge, resulting in paralogous sets of interacting proteins (**B**). The relationships between these proteins corresponds to the relationships between the compartments they belong to (**C**).

Though the OPH was developed to explain the evolution of the pre-LECA compartments, it also has explanatory power for more recent cellular innovations. Glimpses of this were apparent in Dacks et al. (2008), which observed repeated specialization of the endosomal Qa SNARE StxE. In both humans (Metazoa, Amorphea) and *Arabidopsis thaliana* (Archaeplastida), there are distinct paralogs belonging to early and late endosomes, but these resulted from lineage-specific duplications. The best characterized instances of paralogous expansions of membrane trafficking proteins corresponding to the origin of novel organelles are within the Alveolata (SAR). In *Toxoplasma gondii* (Apicomplexa, Alveolata), Rab5C traffics to secretory organelles of the apical complex (Kremer et al. 2013), while Rab11B traffics to the inner membrane complex (IMC) (Agop-Nersesian et al. 2010). Timing the appearance of these paralogs within Alveolata revealed that Rab5C is found across Dinoflagellata and Apicomplexa, while Rab11B is found throughout Alveolata (Klinger et al. 2022). These duplication events correspond to the appearance of apical secretory organelles (Okamoto and Keeling 2014) and the characteristic alveoli of Alveolata, which are homologous to the IMC of *To. gondii*.

1.5 The contractile vacuole[2]

The contractile vacuole (CV) is an osmoregulatory organelle that is broadly distributed across the diversity of freshwater eukaryotes, albeit with considerable morphological variation. It is unclear how the OPH applies to the CV, largely because the homology of the organelle itself is unclear. In this section, I will write about CVs as a single group of organelles (i.e., "*the* contractile vacuole"), but draw attention to the source of the data and where they agree or disagree, because they do not necessarily apply to the CVs of all eukaryotes.

1.5.1 Why is osmoregulation necessary?

Freshwater protists are generally hyperosmotic relative to their environment. While freshwater generally has an osmolarity below 7 mOsm/L (Allen and Naitoh 2002), the intracellular osmolarity of freshwater protists generally measures between 45 and 117 mOsm/L (Prusch 1977). For freshwater organisms, this hyperosmolarity comes with a challenge: plasma membranes are semi-permeable and allow water to pass through, travelling down osmotic gradients. This influx of water means organisms are required either to prevent the entry of that water or to continually remove it, otherwise they will swell and potentially lyse. Many protists overcome this challenge through rigid external coverings (e.g., the frustules of diatoms (Stramenopiles, SAR), or the encasing cell walls of desmids (Zygnematophyceae, Archaeplastida)). These coverings prevent the cell from expanding, which in turn prevents excess water from entering in the first place. This balance of forces—the cytosol pushing the cell membrane against the cell wall—is known as turgor pressure, and is actively maintained by the cell through regulation of osmolyte and ion concentrations (Hellebusi 1976; Zimmermann 1978). However, for a cell to be capable of phagocytosis or movement using pseudopodia or flagella, there needs to be a region of the plasma membrane that is not covered by a rigid cell wall (e.g., the flagellar membrane in the case of flagellates). In many of these organisms, intracellular osmolarity is instead regulated by an organelle known as the contractile vacuole, or contractile vacuole complex (Patterson 1980). Though these terms are often used interchangeably, the contractile vacuole (CV) is the central vacuole that functions in excretion of water, while the contractile vacuole complex (CVC) explicitly

[2] Text from this section will be published as a chapter on contractile vacuoles in an upcoming, community sourced virtual protistology textbook, conceptualized and solicited by Alastair G. B. Simpson. I was the primary writer and researcher of this chapter, but structural and conceptual input, as well as manuscript edits, were provided by Alastair G. B. Simpson and Joel B. Dacks.

includes the extended network of tubules, canals, and/or vacuoles that is involved in water collection, in addition to the central vacuole.

The involvement of the CV in osmoregulation was first convincingly demonstrated by a series of experiments on marine and freshwater ciliates in 1934. Ciliates from both environments were exposed to dilutions of seawater with tap water, which led to an increase or decrease in the water expelled by the CVs of the marine and freshwater ciliates respectively (Kitching 1934). These preliminary experiments have since been confirmed through numerous experiments where organisms with impaired CV function become much more vulnerable to osmotic stressors (e.g., Guillard 1960; Temesvari et al. 1996; Mathavarajah et al. 2018). That cell walls and CVs are different solutions to the same water-influx problem is nicely illustrated by cases where CVs are seen in only some stages of species with complex life histories. Many freshwater organisms have a dominant life stage with an enclosed cell wall that lacks a CV, but have CVs in the unicellular, wall-less reproductive stages. Examples of such CV-bearing stages include the gametes of conjugating desmids and the zygotes that result from conjugation (Ling and Tyler 1972), as well as the flagellated zoospores of the fungi-like oomycetes (Stramenopiles, SAR; Shields and Fuller 1996).

1.5.2 Structure and function of the contractile vacuole

There is a wide array of CVC forms across the diversity of eukaryotes. Nonetheless most, if not all, CVCs share two common components: the bladder (or contractile vacuole, in the narrow sense) and the spongiome (Patterson 1980). The spongiome is composed of membranous tubules and/or vacuoles, and may be differentiated into the decorated spongiome and the smooth spongiome. The decorated spongiome surrounds the smooth spongiome (when present) and is studded with vacuolar ATPases (V-ATPases) (Clarke et al. 2002; Wassmer et al. 2005; Nishihara et al. 2007; Ulrich et al. 2011). These V-ATPases create a proton gradient across the CV membrane that is thought to drive water collection. The spongiome continuously collects water from the cytoplasm and transfers it to the bladder, which swells and eventually expulses its contents into the external environment. In most lineages, the CVC appears to be devoid of visible cytoskeletal components, but the bladder may be anchored and/or reoccurring in the same area of the cell, often near the bases of the flagella (e.g., *Chlamydomonas reinhardtii* (Archaeplastida)*, Leptomonas collosoma* (Discoba); Weiss et al. 1977; Linder and Staehelin 1979) or at the posterior end of the cell (e.g., *Naegleria gruberi* (Discoba); Pittam 1963). Alternatively, it may be transient, occurring in and moving throughout many areas of the cell, with this being typical of large amoebae, for example (e.g., *Amoeba proteus* (Amoebozoa); Wigg et al. 1967).

Perhaps the most complex form of this system is seen in some ciliates (SAR). Here, the CVC is supported by microtubules that anchor it to a specific location in the cell with a permanent pore where the bladder empties, and radial canals are positioned around the CV (McKanna 1973, 1976). The spongiome in these organisms in turn surrounds the radial canals, and transfers water to the canals before it is finally transferred to the CV.

There are two main phases in CV function: the collection of water from the cytoplasm into the contractile vacuole via the spongiome (enlargement of the bladder) and the expulsion of water from the CV to the outside environment (evacuation of the bladder) (Figure 5). This enlargement-evacuation cycle of the bladder is usually visible with light microscopy. The tempo of this cycle varies by organism (e.g., 15 seconds in the green alga *Chlamydomonas reinhardtii*, and 50 seconds in the amoebozoan *Acanthamoeba castellanii*) (Pal 1972; Komsic-Buchmann et al. 2014). In colpodid ciliates, the total amount of water expelled per unit time increases with the volume of the cell (Lynn 1982); presumably this is typical across protists more generally. Shifts in external osmolarity result in changes to the total water expelled per unit time (e.g., Kitching 1934; Cosgrove and Kessel 1958; Pal 1972; Cronkite et al. 1991; Stock et al. 2001; Komsic-Buchmann et al. 2014), reflecting the shift in osmotic pressure and movement of water into the cell. The exact mechanism behind the change in volume of liquid expelled (i.e., CV size, tempo, number) can vary.

1.5.3 There are few common molecular components of contractile vacuoles

Current knowledge of the molecular machinery of CVs is fragmentary, with the majority of information coming from amoebae of the cellular slime mold *Dictyostelium discoideum* (Amoebozoa), the trypanosomatid parasite *Trypanosoma cruzi* (Kinetoplastea, Discoba), *Paramecium* species (Ciliophora, SAR), and the unicellular green alga *Chlamydomonas reinhardtii* (Chlorophyta, Archaeplastida) (Figure 1). There are, as of yet, no distinctive CV-associated proteins identified that are found in all the models. Only a handful of proteins have been identified in more than one model species (compiled by Docampo et al. 2013; Plattner 2013). These include: the V-ATPases and proton pyrophosphatases that create proton gradients across the CV membrane, aquaporins that allow for passive transport of water or other small molecules across the membrane (though whether this is the same aquaporin in all CVs is yet to be determined), the calcium signaling protein calmodulin, and several membrane trafficking proteins. The common membrane trafficking proteins include the GTPase Rab11 (which acts at the recycling endosome; Ullrich et al. 1996), the R SNARE Vamp7 (which is involved in fusion

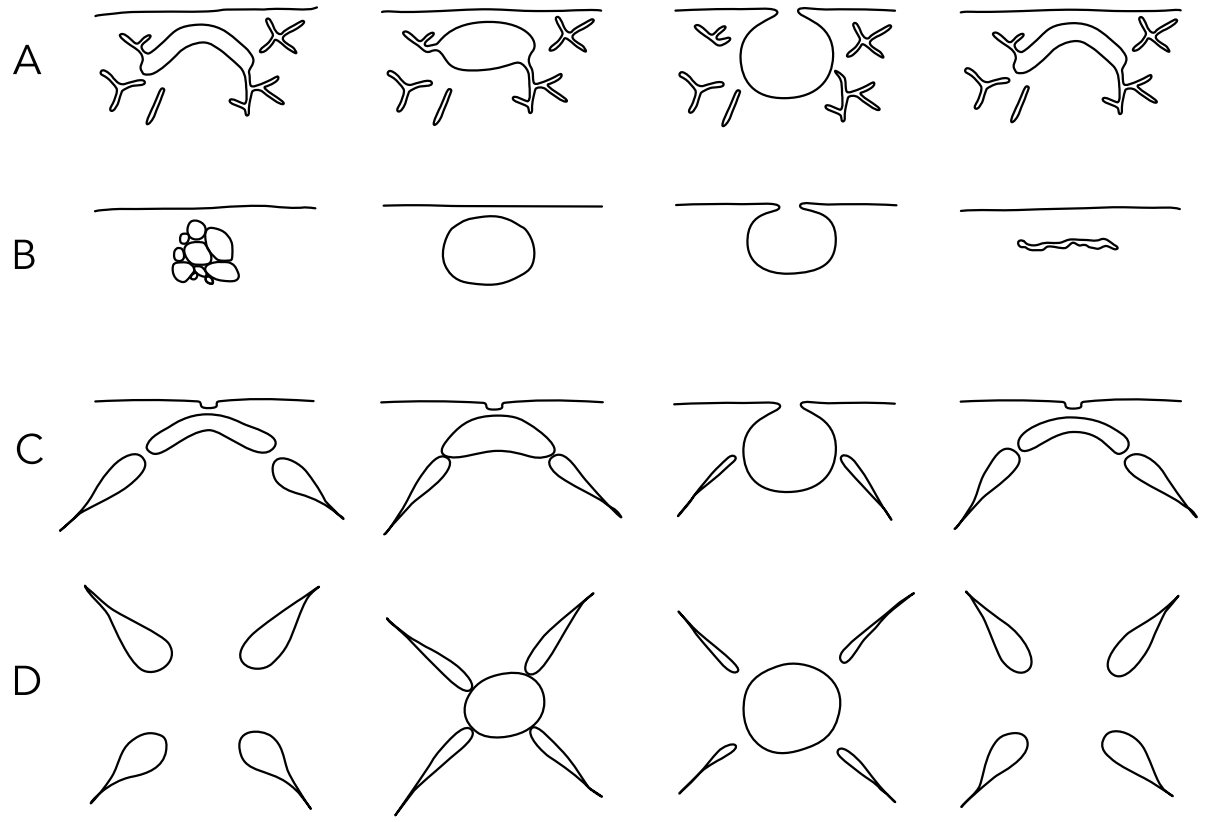


Figure 5 The enlargement-evacuation cycle of three contractile vacuole complex morphologies: a CV with tubular spongiome and central bladder (e.g., *Trypanosoma*) viewed from the side (**A**); a CV with the vacuole-type spongiome, which coalesces into the CV bladder and then collapses (e.g., *Amoeba proteus*) viewed from the side (**B**); and a CV with radial arms and pore (e.g., *Paramecium*) viewed from the side (**C)** and above (**D**). Note that the bladder in (**D**) is only visible from above when filled, which is typical of what is observed with light microscopy. Figure redrawn and modified from Figure 2 of Patterson (1980).

within the endolysosomal system; Chaineau et al. 2009) and the monomeric clathrin adaptor protein AP180 (which is involved in endocytosis; Manna et al. 2015). Additionally, a variety of other membrane trafficking proteins have been identified in one or more systems, including subunits of the multisubunit tethering complex exocyst (involved in exocytosis) and a selection of other Rab GTPases, adaptor proteins, and SNAREs. The involvement of many membrane trafficking proteins is to be expected, as the function of the CV requires, at minimum, two instances of fusion and scission of membrane-bound compartments: 1) the spongiome with the CV bladder, and 2) the bladder with the plasma membrane (see above text and Figure 5).

1.5.4 Contractile vacuoles are broadly—but patchily—distributed

Contractile vacuoles are found in freshwater species of almost all major protist lineages. CVs are also found in a few animals (Metazoa, Amorphea), specifically freshwater sponges and *Hydra* (Benos and Prusch 1972; Brauer and McKanna 1978). These animal lineages diverged prior to the evolution of specialized excretory organs (Andrikou et al. 2021). CVs are, for the most part, absent from parasitic, marine, or hypersaline protists, although some parasitic trypanosomatids (e.g., *Trypanosoma cruzi*) and some marine ciliates and cryptophytes are notable exceptions (e.g., Kitching 1934; Rohloff and Docampo 2008; Hoef-Emden 2014). Most cellular features that are so broadly distributed across eukaryotes are confidently inferred as present in the LECA. However, CVs are both extremely varied in morphology and patchily distributed, which opens the possibility that they are not strictly homologous across eukaryotes and have arisen multiple independent times. As marine, parasitic, and freshwater species of many lineages are closely related (Jamy et al. 2022), a huge number of CV losses would have to be invoked to explain its distribution in freshwater species alone. On the other hand, there are reports that some marine protists have CVs that are induced by introduction of the cell to lower osmolarities (e.g., the heterolobosean amoeba *Vahlkampfia calkensi*; Hogue 1923), and freshwater protists with encasing cell walls have CVs that appear during reproductive phases (Ling and Tyler 1972; Shields and Fuller 1996). The euryhaline alga *Chlamydomonas pulsatilla* (Archaeplastida) has a transient CV that is interpreted as a stable organelle due to the presence of spongiome-like tubules in the typical CV bladder location under marine conditions (Hellebust et al. 1989). This has yet to be studied systematically, but if it is a widespread phenomenon, CVs may be a more universal feature of eukaryotic cells than is generally assumed at present.

1.6 Focus and specific aims of the thesis

This thesis is a part of a larger effort to understand the evolutionary history of the contractile vacuole across eukaryotes. The lack of distinctive marker genes found only in CVs precludes the use of comparative genomics to infer CV origin. All known CV proteins have other functions in the cell (e.g., V-ATPases, which are broadly distributed in the endocytic system, and Rab11, which is canonically involved in recycling endosomes; Nishi and Forgac 2002; Welz et al. 2014). This contrasts with instances where organellar profiling through comparative genomics is possible, exemplified by the eukaryotic flagellum and the Golgi. A number of proteins involved in trafficking to and within the flagellum have been identified based on comparative genomics of organisms with and without flagella (Avidor-Reiss et al. 2004; Merchant et al. 2007; Elias and Archibald 2009; Dewees et al. 2022). On the

other hand, the discovery of a near-complete complement of canonical Golgi trafficking factors in *Naegleria gruberi* prompted the identification of a cryptic Golgi with a previously unknown morphology of small tubules (1 x 4 μm) (Herman et al. 2018). Neither approach is currently possible with CVs as there are no identifiable, CV-exclusive marker genes.

I will approach this problem specifically within the supergroup Discoba, which contains the CV model organism *Trypanosoma cruzi*, as well as emerging molecular model organisms from freshwater ecosystems, *Bodo saltans, N. gruberi* and *Euglena gracilis* (Figure 6). As a starting point, I will examine the jakobid *Reclinomonas americana*. While not a model organism, there is currently a genome project underway, and *R. americana* lacks some of the challenges associated with the discobid models (see section 2.1 below).

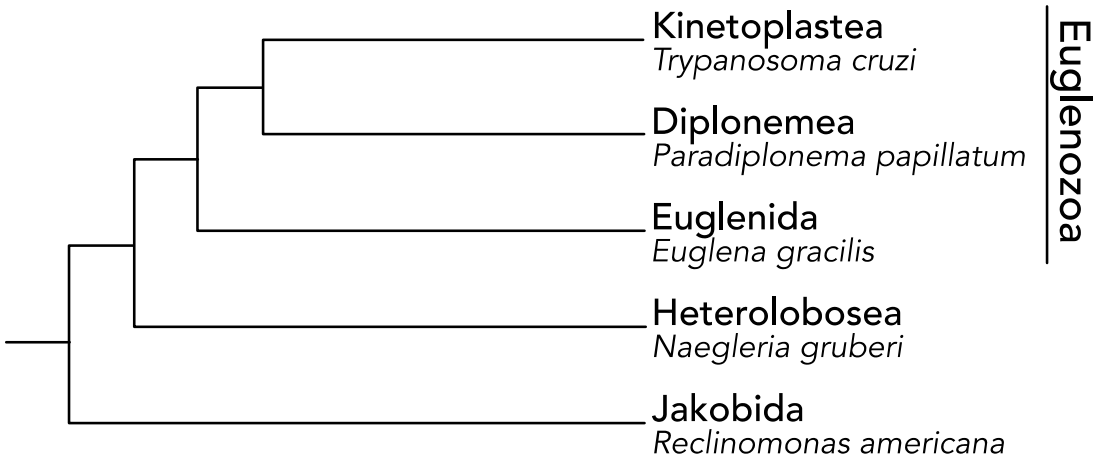


Figure 6 The major lineages within Discoba. Model organisms or representative species are labelled. All shown lineages contain marine and freshwater species; contractile vacuoles have been documented in all shown lineages but Diplonemea. Modified from Simpson et al. (2017).

The goal of this work is to identify putative CV-associated proteins in *R. americana* using differential expression analysis, and then use these as a proxy to infer the evolutionary history of this organelle within Discoba. The following are some of the potential evolutionary scenarios that could arise:

1. **The CVs of *R. americana* and *T. cruzi* are directly homologous.** Many of the same CV-associated proteins are identified (e.g., Rab11, Rab32, Vamp7), and phylogenetics confirms the existence of CV-specific clades of these proteins to the exclusion of versions found

elsewhere in the cell. A contractile vacuole was present in the common ancestor of *R. americana* and *T. cruzi*, and the same organelle is present in the modern organisms.

2. **The CVs of *R. americana* and *T. cruzi* arose through parallel evolution.** Many of the same CV-associated proteins are identified, but phylogenetics does not identify CV-specific clades of these proteins. CV-associated proteins may be directly sister to non-CV-associated versions found in the same organism. The CV in both organisms arose from the same organelle.
3. **The CVs of *R. americana* and *T. cruzi* arose through convergent evolution.** There is little to no overlap in a large selection of CV-associated proteins in both organisms, and these proteins are from distinct compartments, indicating that the CVs in the two organisms do not share a common origin. The CV in each organism arose from a distinct organelle within the endomembrane system.

Subsequent studies will continue this work in the lineages containing other CV models (Archaeplastida, Amoebozoa, and Ciliophora, SAR). If membrane trafficking paralogs that are specific to the CV are identified, they can then be used to infer the evolutionary history of the CV through comparative genomics across the broader diversity of eukaryotes.

2 Methods

2.1 Overview

To investigate the evolution of the CV in Discoba, I examined the evolution of putative CV-associated proteins from my study organism, *Reclinomonas americana*. *R. americana* is an ideal candidate for this work in Discoba, as it: 1) is found in freshwater ecosystems (including rivers, creeks, and eutrophic ponds) and has an observable contractile vacuole using light microscopy (Flavin and Nerad 1993; O'Kelly 1997); 2) is sessile and large enough to image consistently; 3) is thought to have standard transcriptional regulation of protein synthesis, unlike euglenozoans (Vesteg et al. 2019); and 4) does not have different life stages that produce large transcriptional responses and are induced by environmental conditions (i.e., the amoeboflagellate transformation of *Naegleria gruberi* and most other heteroloboseans; Lee and Walsh 1988; Fulton 1993). As information on the CVs of *R. americana* is sparse, I initially investigated the CV morphologically under the low, control, and high osmolarity conditions to determine whether the treatments produce a phenotypic change in CV function. Once CV induction was confirmed microscopically, cultures were harvested in quadruplicate after incubation in the chosen conditions. RNA from these cultures were extracted and sequenced. Reads were assembled into a pooled *de novo* transcriptome, which was then used as a reference to quantify transcript expression. A predicted proteome was generated, which was used to annotate the transcriptome automatically, while genes of interest were annotated manually through similarity searches and phylogenetics. Differential expression analysis was used to identify putative CV-associated genes, which were then used to infer phylogenies and used as a proxy for the evolution of the organelle as a whole.

2.2 Microorganisms and culture conditions

Cultures of the study organism *R. americana* (strain ATCC 50394) and its prey bacterium *Klebsiella aerogenes* (strain ATCC 13048) were kindly provided by B. F. Lang. *K. aerogenes* was grown on LB agar for an initial two days at 22 °C, then transferred to 4 °C for long term storage. *K. aerogenes* was subcultured every month. *R. americana* was grown in vented T25 flasks (Sarstedt) with WCL media (https://megasun.bch.umontreal.ca/People/lang/FMGP/methods/wcl.html) at 22 °C, fed with a suspension of *K. aerogenes* in WCL every 1-2 days as needed, and subcultured every week.

2.3 Characterization of the *R. americana* CV and validation of osmolarity conditions

Previously, it was demonstrated that some CV-associated transcripts are upregulated under low osmolarity conditions when compared to normal culture conditions in *D. discoideum* (Mathavarajah et

al. 2018; Manna et al. 2023), while in studies of *Chlamydomonas reinhardtii* across a broader range of low and high osmolarity conditions CV-associated transcripts either increased or decreased with osmolarity (Komsic-Buchmann et al. 2014). Because *R. americana* has no cell biological protocols available to localize or knock out a protein of interest, I decided to include two treatment conditions, one low osmolarity and one high osmolarity, in addition to the control (WCL) treatment. By looking for transcripts that increased or decreased with osmolarity across more than two conditions, I can be more confident that they are involved in CV function and not a general stress response.

2.3.1 Choice of an inert osmolyte for the high osmolarity condition

To avoid the possibility of ion toxicity on *R. americana*, and to ensure a known osmolarity for the high osmolarity (as initial trials condition using concentrated WCL had salts precipitating out), I decided to use a metabolically inert osmolyte. Small molecules such as sorbitol, mannitol, and sucrose are generally used for experiments like this (e.g., Fulton 1972; Ishida et al. 1996; Slama et al. 2007). Some of these molecules are produced and used as endogenous osmolytes by some eukaryotes (Bonin et al. 2015; Tonon et al. 2017). As genes for mannitol biosynthesis have been identified and characterized in diverse eukaryotes, this seemed like the best initial candidate. Mannitol biosynthesis has largely been considered in the context of eukaryotic algae or other plastid-containing organisms (e.g., apicomplexan parasites), and is generally absent from heterotrophic, aplastidic organisms (Tonon et al. 2017). Mannitol is synthesized from fructose-1-phosphate in two steps: the first to mannitol-1-phosphate, catalyzed by a mannitol-1-phosphate dehydrogenase (M1PDH), and the second to mannitol, catalyzed by a mannitol-1-phosphatase (M1Pase), which can either be an histidine-M1Pase or an haloacid dehalogenase (Liberator et al. 1998; Groisillier et al. 2014; Bonin et al. 2015). Additionally, in some algae (including photosynthetic euglenids), a gene fusion of M1PDH and M1Pase is present (Tonon et al. 2017). Queries representing the full spectrum of known mannitol biosynthesis proteins from *Eutreptiella gymnastica* (Euglenida, Discoba), *Eimeria tenella* (Alveolata, SAR), and *Ectocarpus siliculosus* (Stramenopiles, SAR) were retrieved from the supplementary data of Tonon et al. (2017). The draft *R. americana* predicted proteome (provided by B. F. Lang) was queried for these genes with BLASTp (Altschul et al. 1990) using an unreleased, alpha version of the AMOEBAE workflow by Lael D. Barlow (discussed in section 2.6.2 below). I was unable to identify any mannitol biosynthesis proteins (Supplementary Online Table 1), suggesting that mannitol is not actively produced for use as an osmolyte by *R. americana*.

2.3.2 Microscopy

To ensure conditions chosen induced a change in CV function, and to characterize the *R. americana* CV response, I imaged the cells under each of the conditions. *R americana* was grown for microscopy in plates with #1.5, 25 mm slides (Electron Microscopy Sciences) on the bottom under typical culture conditions. Cells were imaged within four days of culturing using a live cell imaging chamber and differential interference contrast (DIC) optics. The Zeiss AxioObserver.Z1 microscope was equipped with a DIC analyzer, Plan-Neofluar 100X/1.3 oil objective lens, and AxioCam HRm monochrome camera. Two-minute videos were captured at rate of 1 frame per second, ensuring at least one complete bladder enlargement-evacuation cycle in each video. FIJI (Schindelin et al. 2012) was used to measure cell width, cell length, and the diameter of the CV at its largest size, immediately prior to expulsion. To avoid the effect of DIC shadowing on CV diameter measurements, three measurements were taken in different orientations roughly equidistant around the CV, then averaged. Additionally, the duration of complete CV cycle (i.e., from expulsion to expulsion) and the number of CVs visible per individual were recorded.

2.3.2.1 Determining the response time through longitudinal microscopy

Preliminary observations indicated that a 1:3 dilution of WCL with MilliQ water was sufficient to induce a change in CV behaviour; however, the observed response was not consistent, which lead me to investigate how long it took for the CV response to take place. To determine the onset of the osmotic shock response, longitudinal microscopic observations were collected (i.e., individual cells were monitored over a period of time). A slide from an *R. americana* plate culture was transferred to the slide imaging chamber with WCL media, then a cell was identified to image. After a 30-minute rest period, this cell was imaged every five minutes for 30 minutes total under the control conditions (WCL), after which the media was exchanged for a 1:3 dilution of WCL with MilliQ water and the cell was imaged every five minutes for a further 60 minutes. This was repeated 5 times to get a sense of the response time. In one time series a second cell was in focus, leading to a total of 6 individual cells measured. Cell and CV measurements were Z-normalized to make the measurements directly comparable between individuals. Cell lengths and widths were collected to potentially normalize CV size with cell size, as bladder volume varies with cell volume in other organisms (Lynn 1982; Komsic-Buchmann et al. 2014). These data were ultimately uninformative, as the cells were not viewed in the same plane over time (Figure 7A, B). The effect on response variables was visualized with R v. 4.2.0

and package ggplot2 v. 3.3.6 in RStudio v. 1.4.1717 (Wickham 2016; RStudio Team 2020; R Core Team 2022).

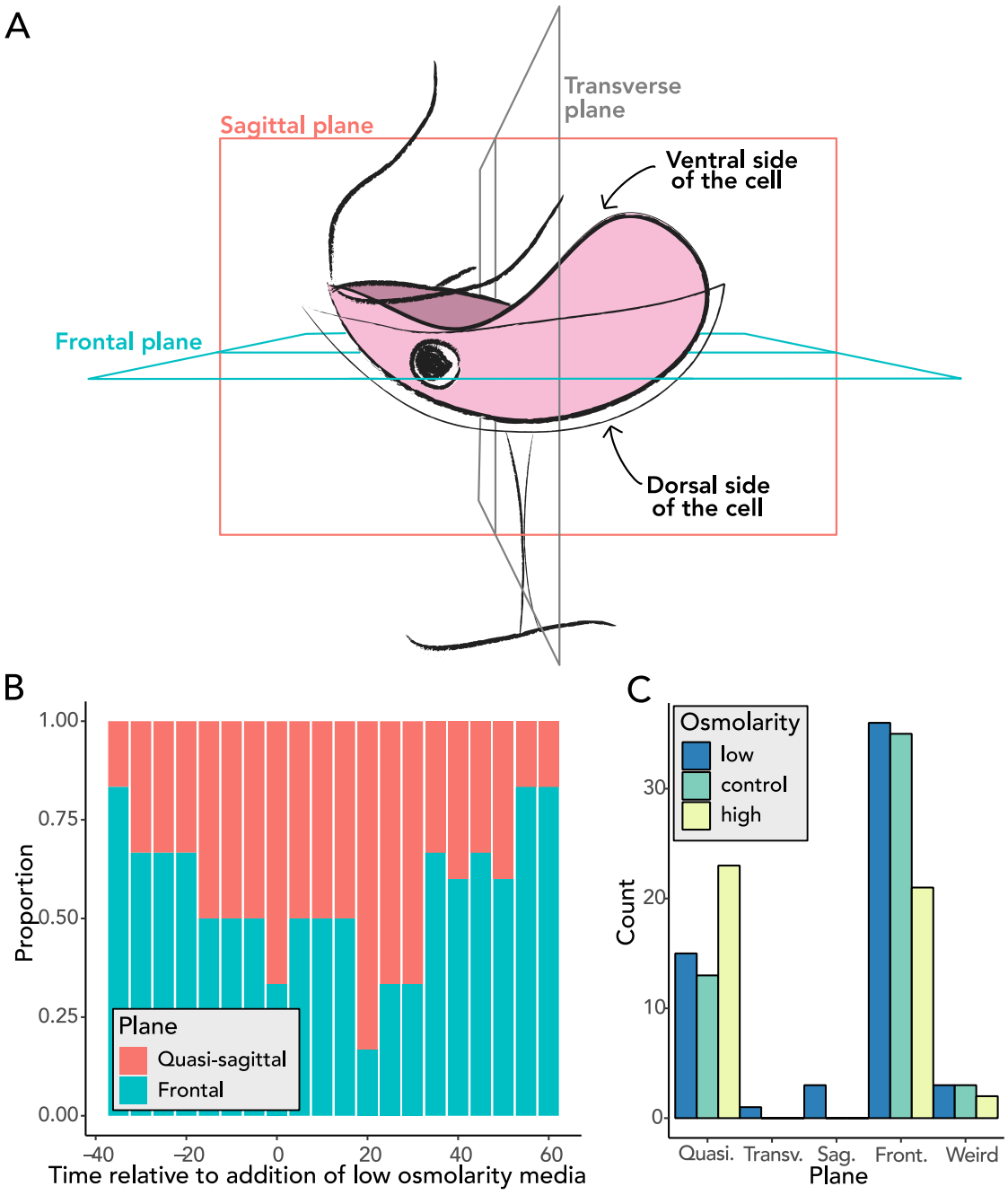


Figure 7 *Reclinomonas americana* could not be imaged in consistent planes in either microscopy experiment. Schematic of *R. americana* showing the various planes of view (**A**). Proportion of the six cells imaged in frontal plane or an off-axis sagittal plane ("quasi-sagittal") in the longitudinal microscopy data (**B**). Counts of cells imaged in each plane (quasi. = quasi-sagittal, transv. = transverse, sag. = sagittal, front. = frontal, weird = none of the above) in the cross-sectional microscopy data (**C**).

2.3.2.2 Determining the response with cross-sectional microscopy

To determine the CV response to the various treatments, cross-sectional data were collected (i.e., many individual cells were imaged once during the response window identified above). A slide from an *R. americana* plate culture was transferred to the slide imaging chamber with WCL and a 30-minute rest period, then the media in the imaging chamber was exchanged for a 1:3 dilution of the original WCL media with either MilliQ water, fresh WCL, or WCL + mannitol (final approximate calculated osmolarities of 3.5, 13.8, and 34.5 mOsm). After a second 30-minute rest, as many cells were imaged as possible in a 30-minute period.

Once CV measurements had been collected, Mahalanobis distances were used to discard multivariate outliers within each treatment, considering the CV diameter and cycle length only. Cell length and width were again uninformative and not included, as *R. americana* was imaged in several different orientations that were not consistent between treatments (Figure 7A,C), and it could not be guaranteed that I was measuring the true length or width of all cells. Treatments were examined visually for normality with ggpubr v. 0.4.0 (Kassambara 2020), and tested for normality and homoscedasticity with the Shapiro-Wilk test and Bartlett's test respectively. Tukey's test was used to test for difference in treatment means of the CV diameter and cycle length, which first requires the fit of a model using an analysis of variance (ANOVA) test, which tests whether there is a difference between any of the means, prior to Tukey's test, which tests each possible pair of means for differences while correcting for multiple tests. All statistics and visualizations were executed with R in RStudio. Initial data visualizations and final graphs were generated with ggplot2.

2.4 Transcriptomics

2.4.1 CV induction experiment

Twelve 35 mL cultures of *R. americana* were grown in vented T175 culturing flasks (Sarstedt). Each culture was seeded initially with 2 mL *R. americana* and fed with 1 mL *K. aerogenes* suspended in WCL. Cells were fed daily until harvest took place on the fourth day. Three-quarters of the media was removed and replaced with either sterilized WCL media (control, final osmolarity approximately 13.8 mOsm), MilliQ water (low osmolarity treatment, approximately 3.5 mOsm), or WCL + mannitol (high osmolarity treatment, approximately 34.5 mOsm) in quadruplicate and gently swirled to mix. After an hour incubation, cells were harvested with all steps performed at 4 °C and cultures kept on ice whenever possible. After incubation, cultures were transferred to the cold room, then the 22 °C media was poured off and replaced with the same volume of chilled media. Flasks were scraped with a cell

scraper and the cell suspension was poured into a chilled 50 mL centrifuge tube. This process was repeated with fresh media to maximize cell yield. Tubes were centrifuged for 10 minutes at 2,500 g. When cell pellets were visible, the supernatant was removed, cells were resuspended in approximately 1 mL of media, which was then transferred to a 2 mL tube. To maximize cell yield, the 50 mL tube was washed three times with the appropriate media and the liquid transferred to the 2 mL tube. Cells were again centrifuged for 5 minutes at 5,000 g. The supernatant was removed using a pipette, and samples were flash frozen in liquid nitrogen and then stored at -80 °C until extraction.

2.4.2 RNA extraction

RNA was extracted using an Animal Tissue RNA Purification Kit (Norgen Biotek) with an on-column DNase I treatment following the manufacturer's instructions. To minimize RNA degradation, prior to extraction samples were kept on dry ice until they were homogenized by passage through a 27-gauge needle 10 times. After extraction, 5 μL of the purified RNA was aliquoted for quantification with a Nanodrop spectrophotometer, and the remainder transferred to -20 °C, where it was stored for quality assessment and sequencing by the High Content Analysis Core (HCAC) at the University of Alberta.

2.4.3 mRNA sequencing (HCAC)

RNA quantification, quality control, and sequencing was performed by the HCAC. RNA libraries were prepared for sequencing according to the Illumina TruSeq RNA sample preparation version 2 guide. Six samples were sequenced per run with Illumina MiSeq technology, generating 150 base pair (bp) paired-end reads.

2.4.4 Read processing

Prior to transcriptome assembly, read quality must be assessed, and, since the *R. americana* culture is monoeukaryotic but not axenic, decontaminated to ensure that the assembled transcriptome belongs solely to the study organism. Trimming low-quality regions from reads greatly decreases computation time, while trimming adaptors is necessary to decrease mis-assemblies (Haas et al. 2013). Illumina data includes quality scores called PHRED scores, which are the -10log(probability of error) (Illumina, Inc 2014). The threshold PHRED 20 is commonly used, which corresponds a 1% probability of a base with this score being miscalled. However, there is some evidence that trimming less stringently increases the completeness of *de novo* assemblies while keeping misassemblies low, so long as

substitution errors are acceptable in the final assembly (MacManes 2014). Reads were trimmed with bbduk (Bushnell 2020) under the following criteria: force trimming of the first 12 bases (due to biased hexamer usage), force trimming by modulo 5 to remove the final, low quality base (base 151 of 150), and quality trimming from either end to remove sequences with a quality of less than 5 or 20 PHRED. Additionally, adaptors were trimmed using the BBMap adaptor set using *k*-mers (i.e., subsequences of length *k*) from 11-22 and a Hamming distance of 1, or overlap of paired sequences. Sequences shorter than 50 bp after trimming were discarded. Since the inclusion of information on read pairing improves assemblies, only pairs where both reads were retained were carried further. Read quality for each sample was assessed initially with FastQC version 0.11.9 (Andrews 2010), then contrasted with the trimmed reads. Trimmed, paired reads for each sample were then decontaminated with Kraken 2 (Wood and Salzberg 2014; Wood et al. 2019) using the full Kraken 2 standard database from 17/05/2021 (archived at https://benlangmead.github.io/aws-indexes/k2), which contains bacteria, archaea, viruses, humans, and common contaminating vector sequences. Briefly, Kraken 2 uses a hashed database containing every *k*-mer present in the selected genomes, and a taxonomically informed approach to identify the last common ancestor (LCA) that would have had each unique *k*-mer (Wood and Salzberg 2014). Reads to be classified by Kraken 2 are used to generate *k*-mers, which are then classified in terms of their likely LCA based on matches to the database, and finally read pairs are classified probabilistically based on the classification of all *k*-mers within the pairs. This information can then be used for filtering. In this case, all classified reads were considered contaminants and filtered out, leaving unclassified reads, the bulk of which should belong to *R. americana*.

Trimming and decontamination was performed on a Digital Resource Alliance of Canada (DRAC) computing cluster using the pre-installed modules bbmap v. 38.86 and kraken2 v. 2.1.2 respectively, and their dependencies. Scripts formatted for the Simple Linux Utility for Research Management (SLURM) job scheduler and containing the commands used are available in Supplementary Online Files 1 and 2.

2.4.5 *De novo* transcriptome assembly

De novo assembly for RNA-seq is typically done on the pooled set of trimmed and decontaminated reads so that the reference transcriptome assembly fully represents the transcriptomic space represented by the reads (Haas et al. 2013). There is a tendency to think that a given assembly of RNA-seq reads is a concrete object that reflects the actual transcriptome present in the organism of interest.

However, when a defined transcriptome is used to simulate reads that are then assembled into a *de novo* transcriptome, assemblers create assemblies that do not accurately correspond to the isoform complexity of the original transcriptome, generating more isoforms than exist (Vijay 2013). Even well-assembled transcriptomes are often fragmented (Freedman et al. 2021). Highly duplicated and fragmented transcriptomes can cause issues with differential expression analysis, as reads that should belong to a single transcript can be split between the multiple contigs, decreasing statistical power (Hsieh et al. 2019). Additionally, poor quality transcriptomes can greatly decrease the number of genes that are suitable for phylogenetic analyses (Spillane et al. 2021).

There is no single best transcriptome assembler for all transcriptome data. The biases of a given program, as well as the assembly parameters selected, can affect the results (Durai and Schulz 2016; Smith-Unna et al. 2016; MacManes 2018). One of the best characterized parameters affecting assembly is the choice of *k*-mer, or length of subsequence that is used to build and refine the hypothetical transcripts. Longer *k*-mer choices require a greater overlap of the reads in order to be assembled, which results in fewer misassemblies and erroneously assembled isoforms, which is preferred, but often fails at assembling low-expression transcripts. Shorter *k*-mers, on the other hand, are better at recovering low-abundance transcripts, but result in an inflated number of false isoforms (Robertson et al. 2010; Schulz et al. 2012). As transcriptome assemblers each come with their own set of biases, assemblies were constructed using Trinity (Grabherr et al. 2011), rnaSPAdes (Bushmanova et al. 2019), and a modified Oyster River Protocol (ORP) multi-assembler (MacManes 2018; Spillane et al. 2021), then assessed with completeness and correctness metrics.

2.4.5.1 Trinity

Trinity was the first *de novo* transcriptome assembler built to deal with the specific challenges of assembling transcriptomes, which largely stem from the unequal expression levels of transcripts and the potential for isoforms of the same gene (Grabherr et al. 2011). At the time of its publishing, Trinity was one of the best in terms of the number of novel features assembled that had previously been missed in other transcriptome assemblies. It uses a single *k*-mer of 25 to assemble reads of similar abundances into contigs, then collapses overlapping contigs into de Bruijin graphs and exports the well-supported paths as transcripts. Trinity was used under default assembly parameters.

The Trinity assemblies were performed on a DRAC computing cluster using the pre-installed module trinity v. 2.12.0 and its dependencies. A script formatted for the SLURM job scheduler and containing the commands used is available in Supplementary Online File 3.

2.4.5.2 RnaSPAdes

RnaSPAdes is a transcriptome-specific modification of the single-cell genome assembler SPAdes, which was built to deal with the uneven read depth coverage in single-cell data due to the use of multiple displacement amplification prior to sequencing, which is a shared challenge of transcriptome assembly (Bushmanova et al. 2019). One of the advantages of rnaSPAdes is the use of multiple *k*-mers to assemble the transcriptome: first with a small *k*-mer, which captures the low-abundance transcripts and creates de Bruijn graphs with many (likely) erroneous isoforms, and then with larger *k*-mers, which refine the graphs, discarding those that are poorly supported and resolving regions with repeats. rnaSPAdes consistently assembles full-length transcripts with fewer misassemblies and duplications than Trinity. Bushmanova et al. (2019) suggest the use of two *k*-mer lengths during assembly: one-third and one-half the read length. RnaSPAdes was used under the default settings, which set the *k*-mers at 45 and 67 for the PHRED 20 trimmed dataset, and 49 and 73 for the PHRED 5 trimmed dataset.

The rnaSPAdes assemblies were performed on a DRAC computing cluster using the pre-installed module spades v. 3.15.3 and its dependencies. A script formatted for the SLURM job scheduler and containing the commands used is available in Supplementary Online File 4.

2.4.5.3 The Oyster River Protocol

The Oyster River Protocol (ORP) is a pipeline which combines and thins the output of multiple assemblers to minimize the trial-and-error involved in generating an acceptable transcriptome assembly (MacManes 2018; Spillane et al. 2021). Briefly, the pipeline error corrects the reads with RCorrector (Song and Florea 2015), trims the reads with Trimmomatic (Bolger et al. 2014), assembles the reads with Trinity with a *k*-mer of 25, Trans-ABySS (Robertson et al. 2010) with a *k*-mer of 32, and twice with rnaSPAdes with *k*-mers of 55 and 75. These assemblies are then combined and grouped into orthogroups with OrthoFinder (Emms and Kelly 2015, 2019). Each contig is assigned a score with TransRate (discussed in section 2.4.6.2 below; Smith-Unna et al. 2016), and the best scoring contig(s) in an orthogroup are retained in the final assembly. Finally, the sequences are clustered using cd-hit-est v. 4.8.1 at a threshold of 98% (Li and Godzik 2006; Fu et al. 2012). The read mappings from TransRate can optionally be used to discard transcripts that are less abundant than a given transcripts per million threshold in the assembly (TPM; typically, the threshold is 1). This method produces assemblies that score consistently better than the Trinity assembly with a variety of metrics (MacManes

2018), and generates more full-length sequences of sufficient quality for phylogenetics (Spillane et al. 2021).

To make the ORP assemblies directly comparable to the Trinity and rnaSPAdes assemblies, I used the previously trimmed and decontaminated reads to assemble the default transcriptomes that the ORP pipeline would assemble automatically (with Trans-ABySS, Trinity, and rnaSPAdes as described above), and fed these assemblies into the pipeline directly prior to the pooling and thinning steps. The TPM threshold was conservatively set at 0.5, and then changed to 1 for a more stringent approach with the 5 PHRED trimmed assembly. ORP discards transcripts smaller than 100 bp.

The individual rnaSPAdes and Trans-ABySS assemblies were performed on a DRAC computing cluster using pre-installed modules spades v. 3.15.3 and trans-abyss v. 2.0.1, respectively, and their dependencies, while the Trinity assemblies from section 2.4.5.1 were repurposed. Scripts formatted for the SLURM job scheduler containing the exact commands used are available in Supplementary Online File 5 and 6. ORP v. 2.3.3 was downloaded from https://hub.docker.com/r/macmaneslab/orp and installed as a Singularity image (Kurtzer et al. 2017) on a DRAC computing cluster as described in Supplementary Online File 7. The ORP image was executed using the pre-installed module singularity v. 3.8. An example ORP script formatted for the SLURM job scheduler is available in Supplementary Online File 8.

2.4.6 *De novo* transcriptome assessment

Assessing the quality of a *de novo* transcriptome without a genome from a close relative with which to compare the assemblies is challenging. Unlike genome assembly, where the most contiguous assembly with even coverage is considered best, there is no easily definable goal with transcriptome assembly. However, we can ask whether the assembly is reasonably complete, and whether it reasonably represents the reads used to assemble it.

2.4.6.1 BUSCO

Benchmarking Universal Single-Copy Orthologs (BUSCO) is a metric for assessing the completeness of a genome or transcriptome using a selection of well-conserved single-copy genes (Simão et al. 2015; Manni et al. 2021). This selection of genes changes depending on the taxonomic identity of the organism of interest. Proteins are predicted from the genome or transcriptome assembly with either MetaEuk under two levels of sensitivity (Levy Karin et al. 2020) or AUGUSTUS (Stanke et al. 2006). The predicted proteome is then queried with HMMER3 (Eddy 2011), and genes are classified as either

complete (if the hit is within two standard deviations of the OrthoDB mean length), fragmented, or missing. BUSCO was used with MetaEuk in transcriptome mode, and, as there is no dataset specifically for Discoba or Jakobida, the most current general eukaryote dataset was used (eukaryota_odb10).

BUSCO analyses were performed on a DRAC computing cluster using the pre-installed module busco v. 5.2.2 and its dependencies. A script formatted for the SLURM job scheduler and containing the commands used is available in Supplementary Online File 9.

2.4.6.2 TransRate

TransRate is a reference-free metric for assessing the quality of a *de novo* transcriptome assembly using read mapping patterns to assess the probability of a selection of assembly errors, including paralog collapse, chimerism, insertions, fragmentation, and redundancy (Smith-Unna et al. 2016). TransRate scores individual contigs out of 1 based on the proportion of read and contig mismatches, and scores the whole assembly using the geometric mean of the contig scores. The assembly score correlates with correctness-based metrics and fewer duplications (Smith-Unna et al. 2016; Bushmanova et al. 2019). The contig scores can be used to filter out low-quality contigs, like in the ORP.

TransRate analyses were performed on a DRAC computing cluster using TransRate v. 1.0.3 packaged in the ORP Docker container (see section 2.4.5.3 above) and executed with Singularity version 3.8. A script formatted for the SLURM job scheduler and containing the commands used is available in Supplementary Online File 10.

2.5 Read mapping

To perform differential expression analysis, the number of reads belonging to each transcript must be determined. The program kallisto was used for this, which uses a *k*-mer based method to rapidly pseudoalign the reads to the transcript contigs. This approximates the results of more resource-intensive alignment-based methods without significantly impacting accuracy (Bray et al. 2016). Kallisto was used to pseudoalign each set of trimmed and decontaminated reads against the ORP PHRED 5 (TPM 1) assembly with correction for sequence-specific biases (--bias flag).

Indexing of the transcriptome and pseudo-alignment of the reads was performed on a DRAC computing cluster using the pre-installed module kallisto v. 0.46.1 and its dependencies. A script formatted for the SLURM job scheduler and containing the commands used is available in Supplementary Online File 11.

2.6 Annotation

2.6.1 Protein prediction and automated annotation

To aid in annotation of the transcriptome, proteins were predicted using GeneMarkS-T (Tang et al. 2015) using the default parameters as a part of the EnTAP annotation pipeline (Hart et al. 2020). EnTAP uses a NCBI taxonomy-informed approach to annotate *de novo* transcriptome assemblies, first by predicting proteins from the transcriptome assembly, then performing similarity searches with DIAMOND (Buchfink et al. 2015) against user-selected protein databases. EnTAP is particularly useful for *de novo* transcriptome annotation of non-model organisms because it filters the top hits for taxonomic relevance (including removing potential contaminants, if desired) and informative sequence names (i.e., removing hits called "hypothetical protein" or similar). Predicted proteins are further annotated using EggNOG 4.1 (Powell et al. 2014). DIAMOND searches were performed against local copies of NCBI RefSeq (release 211), nr (downloaded April 28, 2022), Swissprot (release 2022_01), and *Trypanosoma cruzi* CL Brenner, *Naegleria fowleri* ATCC 30863, and *N. gruberi* NEG-M predicted proteomes downloaded from EuPathDB (release 57) (El-Sayed et al. 2005; Fritz-Laylin et al. 2010; Zysset-Burri et al. 2014; O'Leary et al. 2016; Amos et al. 2022; The UniProt Consortium 2022). Searches were interpreted with "reclinomonas" as the relevant taxon. The cut-offs for a protein to be annotated were an e-value of e-05 and coverage of 50% for both query and target sequences.

EnTAP v. 0.10.8-beta, DIAMOND v. 0.8.31, GeneMarkS-T v. 5.1, and the databases above were installed on a DRAC computing cluster as described in Supplementary Online File 12. Databases were formatted with DIAMOND prior to the execution of the EnTAP workflow. Scripts formatted for the SLURM job scheduler and containing the commands used are available in Supplementary Online File 13 and 14. The EnTAP configuration settings used are available in Supplementary Online File 15.

2.6.2 AMOEBAE

As this study is interested in a very small selection of the potential differentially expressed genes, it is crucial that the genes of interest are annotated reliably and with specificity. Analysis of MOlecular Evolution with BAtch Entry (AMOEBAE), is a semi-automated workflow that is designed to make the process of reciprocal best hit (RBH) similarity searching with BLAST and HMMER both efficient and reproducible (Barlow et al. 2023). RBH searching has two main steps: a "forward search" using a

query from a well-annotated reference genome to search into the database of interest (in this case, the transcriptome assembly), and a "reverse search" of any significant hits back into the reference genome. If the top hits are reciprocal, it is likely that the genes are related, but the exact relationship often requires phylogenetics to determine.

Rather than annotate the entire membrane trafficking system, I selected a subset of protein families that were a) previously identified in CVs in model organisms (SNAREs, Rabs, TBCs); b) well-characterized, paralogous protein families (the HTACs); or protein complexes with characteristic locations (the MTCs). Additionally, some proteins not a part of the core membrane trafficking machinery but previously identified in CVs were annotated (aquaporins, P2X receptors, arrestins). The selection of proteins and naming conventions thereof were based on previously published analyses (Koumandou et al. 2007; Eliáš et al. 2012; Gabernet-Castello et al. 2013; Finn et al. 2014; Hirst et al. 2014; Tesan et al. 2021). Query sequences for forward searches were retrieved from the human genome on RefSeq, with the exception of proteins not found in humans, which were retrieved from NCBI using accessions from the supplementary material of Hirst et al. (2014) (TSET; *N. gruberi*) or Gabernet-Castello et al. (2013) (TBC-RootA and TBC-ExA1 and 2; *T. cruzi*), and the sole Rab GTPase query, an HMM constructed from the pan-eukaryotic Rab GTPase alignment of Eliáš et al. (2012). AMOEBAE searches were performed on a local high performance computing server with a forward search e-value cut-off of 0.05. The human predicted proteome was used as the reference database for most reverse searches (GCF_000001405.40) (International Human Genome Sequencing Consortium 2004); *T. cruzi* CL Brenner (GCF_000209065.1) and *N. gruberi* NEG-M (GCF_000004985.1) were also used (El-Sayed et al. 2005; Fritz-Laylin et al. 2010). Reverse searches were considered RBH if they returned the original hit with an e-value at least 2 orders of magnitude different than the next non-redundant hit, and were not RBH for another query with a higher e-value (--no_overlapping_hits).

AMOEBAE v. 3.0 was installed on a private computing cluster using Miniconda3 v. 4.11.0 (Anaconda Software Distribution 2016) following the instructions on GitHub (https://github.com/laelbarlow/amoebae), and executed using Snakemake v. 7.1.1 (Mölder et al. 2021). The Snakemake workflow containing the AMOEBAE commands used is available in Supplementary Online File 16.

2.6.3 Domain discovery

In some cases, even a meticulous reciprocal best hit approach is not sufficient for classifying a protein. In the case of the Rabs, which are small, diverse, and have closely related protein families, a

combination of domain prediction using InterProScan (Jones et al. 2014; Blum et al. 2021) and reverse hits to any Rab GTPase when NCBI RefSeq was queried were used to identify candidates for phylogenetic analysis (Supplementary Online Table 2). Where the Rab GTPase domain was found in a much larger protein (e.g., 600 residues), the domain was trimmed using extractseq (https://www.bioinformatics.nl/cgi-bin/emboss/extractseq) prior to phylogenetic analysis. In the case of TBC-K, there were 23 positive RBHs in the *R. americana* transcriptome, and domain prediction was used to check for the presence of a TBC domain.

Domain discovery for Rabs and TBC-K was performed on a DRAC computing cluster using the pre-installed module interproscan v. 5.50-84.0 and its dependencies. A script formatted for the SLURM job scheduler and containing the commands used is available in Supplementary Online File 17. All other domain searches were performed with the InterPro online server (https://www.ebi.ac.uk/interpro/).

2.6.4 Phylogenetic inference

Classification of the HTACs, SNAREs, and Rabs were confirmed through phylogenetics using alignments retrieved from previously published analyses in the Dacks lab (Eliáš et al. 2012; Hirst et al. 2014; Arasaki et al. 2015; Venkatesh et al. 2017). Sequences from *R. americana* were aligned to the appropriate alignment with MAFFT v7.471 using the option --addfull (Katoh and Standley 2013); these sequences included results from AMOEBAE searches that met the RBH criteria described above, as well as those with ambiguous identity that had top reverse search hits in the protein families of interest. The new alignments were visually inspected and manually adjusted in SeaView v. 5.0.4 (Gouy et al. 2010), and trimmed with an in-house trimming script (at least 50% of the sequences without gaps, and 50% identity at the site). Models of sequence evolution were chosen with ModelFinder (Kalyaanamoorthy et al. 2017), and maximum likelihood (ML) phylogenies were inferred with IQ-TREE 2 v. 2.0.4+ with 1,000 ultra-fast bootstraps (UFBOOT) (Minh et al. 2019) either locally or on a private computing cluster. As the transcriptome assembly used is highly duplicated, all relevant sequences were included during the first iteration of the phylogenies, but to avoid biasing the model selection with too many closely related sequences in subsequent iterations the initial phylogenies were inspected and full-length representative sequences were chosen for sequences that formed fully supported clades. Phylogenies were visualized with FigTree v. 1.4.4 (Rambaut 2018). An example script formatted for the Portable Batch System (PBS) job scheduler containing the commands used is available in Supplementary Online File 18.

2.7 Transcript to "gene" clustering

To ameliorate the issue of excess isoforms that are likely assembly errors and not biologically relevant, the transcriptome assembly was clustered into gene-equivalents using Corset (Davidson and Oshlack 2014). Corset clusters transcripts together based on a combination of sequence and expression similarity, meaning that if there are isoforms that have the potential to be biologically relevant (e.g., with different expression patterns), they will not be clustered together. Salmon, a pseudo-mapping program similar to kallisto, was used to estimate expression similarity (Patro et al. 2017): first an index of the transcriptome was created with a *k*-mer of 31, then quantification was run with the flags --dumpEq --hardFilter --skipQuant as recommended by the Corset documentation. While Corset *can* use the treatment information to decide where expression levels should be similar between samples, this information was not provided during clustering to avoid circular reasoning (i.e., the program coercing the samples within a treatment to be more similar, creating more consistently clustering samples).

Indexing of the transcriptome and pseudo-mapping of the reads, then clustering of the transcriptome was performed on a DRAC computing cluster using the pre-installed modules salmon v. 1.7.0, corset v. 1.09, and their dependencies. Scripts formatted for the SLURM job scheduler and containing the commands used are available in Supplementary Online Files 19 and 20.

2.8 Differential expression analysis

The R package tximport v. 1.24.0 was used to prepare the read mapping data from kallisto for differential expression (DE) analysis using the gene-to-transcript map generated with Corset (Soneson et al. 2016). All transcripts with fewer than 20 reads mapping across all samples were discarded prior to DE analysis with DESeq2 v. 1.36.0 (Love et al. 2014). The low and high osmolarity comparisons were contrasted directly with the Wald test to obtain log fold change (LFC) and significance values. All transcripts with LFCs that were significantly different than 0 (adjusted p-value <0.05) were then filtered for expression patterns that either increased or decreased with osmolarity across all three treatments. The previously manually annotated genes were used to identify transcripts of interest for phylogenetic analyses; automated annotations were also scanned for protein families of interest. All statistics and visualizations were executed in R within RStudio. The packages ggplot2 and vidger v.

1.16.0 were used to visualize differential expression results (McDermaid et al. 2019; Monier et al. 2022).

2.9 Phylogenetics of putative CV-associated proteins

2.9.1 Identification and classification

As CVs may contain organelle specific versions of membrane trafficking proteins found elsewhere in the cell, the evolutionary history of putative CV-associated proteins was inferred using all orthologs from *R. americana* and the four CV models. For phylogenies to be informative in inferring the evolution of CVs, the proteins must be present in multiple copies in at least two of the organisms of interest. Searches were performed using AMOEBAE for Rab8, Sec1, aquaporins, arrestins, and four LECA R SNAREs (tomosyn, Sec22, Ykt6, and Vamp7) using human proteins as starting queries in the predicted proteomes of *Dictyostelium discoideum* AX4 (GCF_000004695.1), *Trypanosoma cruzi* CL Brenner (GCF_000209065.1), *Paramecium tetraurelia* d4-2 (GCF_000165425.1), and *Chlamydomonas reinhardtii* CC-503 cw92 mt+ (GCF_000002595.2) (Eichinger et al. 2005; El-Sayed et al. 2005; Aury et al. 2006; Merchant et al. 2007). The human predicted proteome was used as the reference database (GCF_000001405.40) (International Human Genome Sequencing Consortium 2004). Additionally, a literature review was performed, and CV-localized versions of these proteins in *Dictyostelium discoideum*, *Trypanosoma cruzi*, *Paramecium* spp., and *Chlamydomonas reinhardtii* were retrieved from the appropriate databases (Supplementary Online Table 3). Once I was confident that I had retrieved all orthologs of the proteins of interest in the CV models and *R. americana*, the protein classifications were confirmed or determined where relevant (Rab8, Vamp7) using the alignments from section 2.6.4 above. Classification of aquaporins was performed using prokaryotic aquaporin sequences retrieved from an analysis by Pommerrenig et al. (2020) and bacterial "aquaporin X" sequences from an analysis by Tesan et al. (2021), which were aligned using MAFFT in local alignment mode with 1,000 iterations. As Vamp7 and some eukaryotic aquaporin families are poorly supported in phylogenies, these groups were resolved by iteratively removing the most divergent fully supported clade, realigning the sequences with MAFFT in local alignment mode with 1,000 iterations, and inferring a new phylogeny. Initial classifications were inferred with maximum likelihood (ML) using 1,000 UFBOOT replicates in IQ-TREE 2.

2.9.2 Model selection for maximum likelihood

With the well resolved Rab GTPase and SNARE classification phylogenies, the default model selected by ModelFinder was used. As there was evidence of long-branch attraction in the initial aquaporin phylogeny, a more complex model was chosen. In this case, phylogenies were inferred using posterior mean site frequency (PMSF) profiles to approximate the more resource-intensive C60 mixture model, which can ameliorate long-branch attraction (Halpern and Bruno 1998; Le et al. 2008; Wang et al. 2018). Briefly, the process of running a PMSF tree includes: 1) inferring an initial ML tree using the model selected by ModelFinder; 2) estimating site-specific amino acid frequency profiles under the selected model + C60 (representing a mixture model with 60 individual within-site amino acid frequencies) using the ML tree as a guide; and 3) resampling bootstraps under the selected model + C60/PMSF, to approximate the mixture model. An example script formatted for the PBS job scheduler containing the commands used is available in Supplementary Online File 21.

2.9.3 CV protein phylogenies

Final, protein-specific ML phylogenies were inferred using 200 non-parametric bootstraps in IQ-TREE 2 under the C60 PMSF model selected by ModelFinder. Bayesian analyses were performed using MrBayes v. 3.2.7a (Ronquist et al. 2012) under an estimated fixed-rate substitution model under a gamma distribution with four categories. Two runs of four chains (one cold, three hot) were run for 1,000,000 generations under the default heat settings and sampled every 1,000 generations, with the first 25% discarded as burn-in. Convergence was assessed through the average standard deviation of split frequencies. If this was more than 0.01 at the end of the 1,000,000 generations, the analysis was run for an additional 500,000 generations. Bayesian inference values were then mapped onto the non-parametric ML tree using Affinity Designer v. 1.9.3 (Serif Europe). An example script formatted for the PBS job scheduler containing the commands used is available in Supplementary Online File 22.

2.9.3.1 Alternate topology testing

In cases where the phylogenies were unresolved, alternate topologies were tested. This involves inferring ML trees under different constraints, estimating the likelihood of each tree with bootstrapping using the resampling estimated log-likelihood (RELL) approximation, then rejecting the trees which are significantly less likely under the specified model of evolution (Kishino et al. 1990; Shimodaira 2002). Four scenarios were considered: 1) the unconstrained tree; 2) lineage-specific expansions in all organisms (i.e., distinct CV origins); 3) an ancestral CV paralog in LECA; and 4) an

ancestral CV paralog in the last discobid common ancestor. Approximately unbiased (AU) tests were performed IQ-TREE 2 under the model selected with BIC in ModelFinder, and using 100,000 RELL bootstraps. An example script formatted for the PBS job scheduler containing the commands used is available in Supplementary Online File 23.

3 Results and Discussion

3.1 *Reclinomonas americana* modulates CV diameter in response to osmotic shock

To characterize the *Reclinomonas americana* contractile vacuole and validate that the chosen treatments produced a change in CV behaviour, the CV was characterized microscopically under the three osmolarity conditions (see Methods section 2.3.2 above). First, an appropriate interval to collect data had to be chosen. To determine the response time to osmotic shock, six cells of *R. americana* were followed for 30 minutes in the control medium, and 60 minutes after the addition of low-osmolarity medium. Two-minute videos were captured every 5 minutes. Cell length and width were measured, as well as CV bladder diameter and the interval of the enlargement-evacuation cycle (Supplementary Online Table 4). The most pronounced effect was in the CV diameter, which increased gradually after the addition of low osmolarity media (Figure 8). The maximum response occurred at roughly 40 minutes post exposure, after which a slight decline in CV diameter was observed. This was used to choose the 30–60-minute post-exposure time to collect cross-sectional data of CV response.

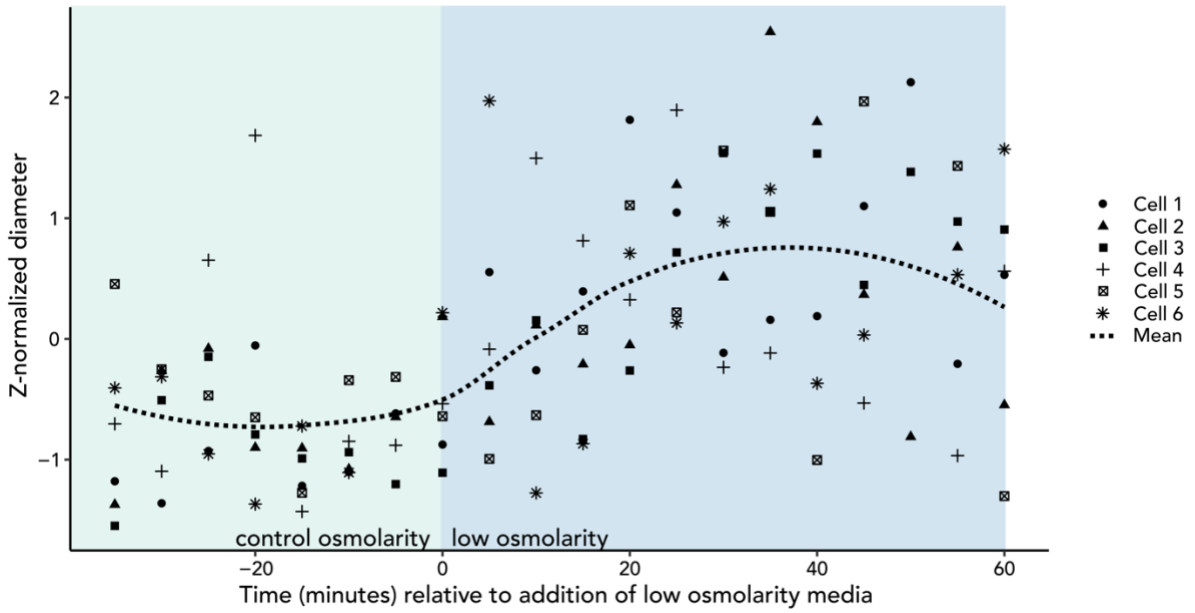


Figure 8 Response of the *Reclinomonas americana* contractile vacuole diameter to hypoosmotic shock. Six individuals were imaged every 5 minutes under control conditions and after the addition of low osmolarity media.

To determine the response to the three conditions, cell dimensions and CV data were collected with a single timepoint from as many cells as possible during the 30–60-minute window. Cell and CV dimensions were measured as described above (Supplementary Online Table 4). As previously observed (O'Kelly 1997), *R. americana* had two visible CVs in most cells (96% cells imaged under

typical culture conditions; Table 1). When observed from the top or the rear, these could be seen on either side of the cell, flanking the central nucleus (Figure 9). In general, the bladder enlargement-evacuation cycle of the two CVs alternated regularly (Table 1, Figure 9, Supplementary Online Video 1), though variations in the pattern, such as vacuoles flipping the order of evacuation, were also observed (Supplementary Online Video 2). Rarely, two "paired" CVs were observed on the same side of the cell expulsing in unison, with a third CV on the other side of the cell alternating with the paired CVs (Table 1, Supplementary Online Video 3).

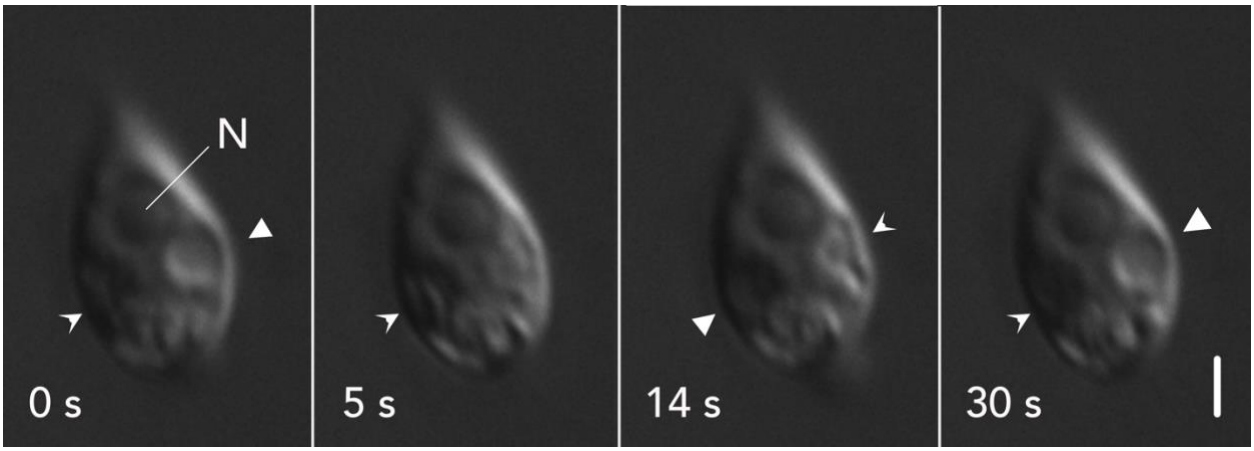


Figure 9 The *Reclinomonas americana* contractile vacuole (CV) enlargement-evacuation cycle; differential interference contrast. Two CVs (full/notched arrowheads) flank the nucleus (N). The enlargement-evacuation cycles alternate: when one CV is about to evacuate (full arrowhead), the other is filling (notched arrowhead). The cell shown is from the low osmolarity treatment for easiest visualization of CVs. Scale bar = 2 μm. The video of this is in Supplementary Online Video 1.

Table 1 Characteristics of the *Reclinomonas americana* contractile vacuole under low, control, and high osmolarity conditions. Bladder diameter, enlargement-evacuation interval, and expulsion rate are shown as mean (standard deviation).

	N	Diameter (μm)	Interval (s)	Expulsion rate (μm^3/min)	Alternating CVs (%)	Paired CVs (#)
Low osmolarity	52	1.256 (0.100)	27.3 (4.2)	4.71 (1.18)	94.2	1
Control	48	1.031 (0.090)	28.5 (4.3)	2.52 (0.75)	95.8	0
High osmolarity	42	0.812 (0.080)	26.8 (5.4)	1.31 (0.37)	83.3	6

Under the typical culture conditions, the CV had a diameter of 1.031 ± 0.090 μm, and an enlargement-evacuation interval of 28.5 ± 4.3 s (N=48; Figure 10, Table 1). In low and high osmolarity

conditions the CV diameter was 1.256 ±0.100 μm and 0.812 ± 0.080 μm, and the enlargement-evacuation interval 27.3 ± 4.2 s and 26.8 ± 5.4 s respectively (N=52, N=42; Figure 10, Table 1). The ANOVA required for model specification indicated a significant difference in CV diameters among conditions ($F(2, 139) = 300.6$, $p < 2e\text{-}16$), while Tukey's test indicated a significant difference between all three diameter means ($p < 0.001$). The ANOVA required for model specification indicated no significant difference in the enlargement-evacuation interval among conditions ($F(2, 139) = 1.677$, $p = 0.191$), and so Tukey's test was not performed.

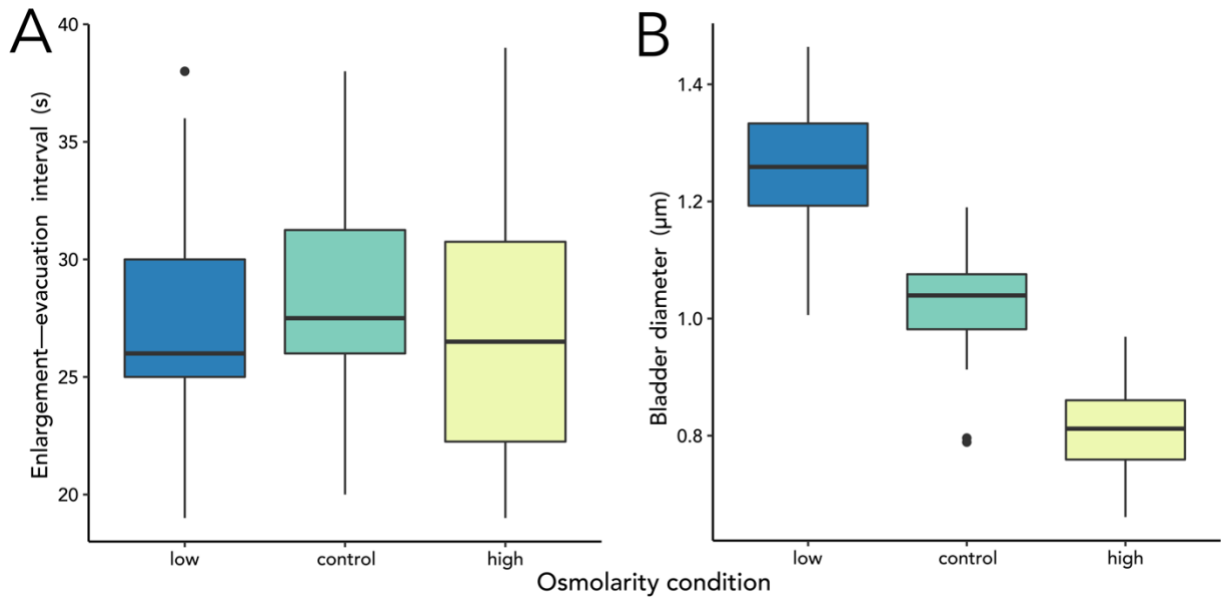


Figure 10 *Reclinomonas americana* contractile vacuole enlargement-evacuation interval (**A**) and bladder diameter (**B**) under low, control, and high osmolarity treatments (N=52, 48, 42 respectively).

There are three potential ways to modulate water efflux: the bladder enlargement-evacuation interval, the diameter of the CV(s), or the number of CVs per cell. In *R. americana,* only a response in CV diameter was observed; however, this could be due to the limited treatments examined. This trait has not been studied in most CV-containing organisms, but from what has been published, it does not seem to be strictly correlated to phylogenetic position. The closest relative to *R. americana* in which this has been studied is *Crithidia fasciculata* (Kinetoplastea, Discoba), which can modulates both CV bladder size and expulsion interval; anecdotally other trypanosomatids modulate both as well (Cosgrove and Kessel 1958; Hajduk 1972). *D. discoideum* also modulates water expulsion through CV bladder size (Mathavarajah et al. 2018), though anecdotally it may modulate the number of CVs per

cell (Gerisch et al. 2002). *Acanthamoeba castellanii,* another well-characterized amoebozoan, seems to modulate both interval and size (Pal 1972). A similar pattern has been observed in ciliates (SAR): *Tetrahymena pyriformis* modulates the CV bladder size, while *Paramecium calkinsi* and *P. multimicronucleatum* can modulate both the CV diameter and interval, depending on the nature of the osmotic shock (Rifkin 1973; Cronkite et al. 1991; Stock et al. 2001). On an even smaller evolutionary timescale, *Chlamydomonas reinhardtii* modulates water efflux through the enlargement-evacuation interval, while *C. pulsatilla* modulates only CV bladder size (Hellebust et al. 1989; Komsic-Buchmann et al. 2014).

3.2 Multi-assemblers are a reasonable approach to *de novo* transcriptome assembly

Since we can be confident that the chosen conditions produced a phenotypic change in CV function, these three conditions were used for the CV induction protocol. Cultures were harvested in quadruplicate one hour after exposure to either low, control, and high osmolarity conditions. This time period was chosen because in a microarray time-series analysis of *D. discoideum* in hyperosmotic conditions the most concerted transcriptomic response occurred at around the 45 minute-1 hour mark (Na et al. 2007), and it was an effective timepoint in the previous studies of *Dictyostelium discoideum* CV transcriptomics (Manna et al. 2023). RNA was extracted and then sequenced at the University of Alberta's High Content Analysis Core with Illumina MiSeq next-generation sequencing technology, generating a mean 11,258,739 paired-end reads (± 1,507,240 SD) per sample.

As the best transcriptome assembler seems to depend on read quality and the properties of the transcriptome (Smith-Unna et al. 2016), and gently trimmed reads may produce more complete assemblies (MacManes 2014), a selection of assemblers was chosen to assemble reads trimmed under two quality thresholds and the results were assessed quantitatively. The best assembly should be determined by the experiment: I wanted to maximize the number of fragments that map well to the transcriptome and the chance that I will recover a full-length protein sequence corresponding to proteins of interest (for use in phylogenetics), and minimize the number of excess isoforms.

Reads were trimmed under two different quality thresholds (PHRED 5 and PHRED 20), decontaminated, then assembled using a selection of transcriptome assemblers. Substantially more reads and bases were retained trimming with a softer threshold: with PHRED 20, an average of 88.4% of reads and 74.8% of bases were retained, while with PHRED 5, an average of 99.9% of reads and 98.7% of bases were retained (Supplementary Online Table 5). The proportion of reads that were removed due to contamination was consistent, with an average of 89.3% and 90.7% discarded from the 5 and 20 PHRED trimmed datasets respectively (Supplementary Online Table 6). Each dataset

was assembled with Trinity (single, fixed *k*-mer), rnaSPAdes (large and small *k*-mers automatically selected based on read length), and the ORP (combining Trinity, Trans-ABySS, and two different rnaSPAdes assemblies, then thinning the results). I attempted to run the rnaSPAdes assembly trimmed at PHRED 5 multiple times with different computational resources, however both attempts ran out of memory. In the future, I would likely use an error corrector (e.g., RCorrector; Song and Florea 2015) prior to assembly. While most transcriptome assemblers have the ability to cope with some level of error in the reads, this should decrease the computation time (Conway and Bromage 2011; Pell et al. 2012) and may decrease the number of duplicates.

The quality of these assemblies was compared with BUSCO as a completeness measure and TransRate as a correctness measure. In general, rnaSPAdes produced the most compact assembly, followed by the ORP PHRED 5 (TPM 1 filtered), while Trinity assemblies were the least compact (Table 2). Even the most compact assembly is inflated in terms of the number of genes we would expect: the well-assembled genomes of free-living heterotrophs within Discoba have between 9,000 and 19,000 predicted proteins (Fritz-Laylin et al. 2010; Jackson et al. 2016; Gray et al. 2020; Richter et al. 2022), with *Andalucia godoyi* (Jakobida, Discoba) having the fewest. All assemblies produced similar BUSCO scores, though the ORP scores were slightly improved in terms of missing and fragmented genes, and the more stringent trimming tended to produce fewer duplicates (Table 2). The BUSCO completeness scores were comparable to the highly contiguous draft genome of *A. godoyi* (88.8% complete, n=303, eukaryote_odb9; Gray et al. 2020). The ORP PHRED 5 assemblies clearly outperformed the other assemblies in terms of the TransRate assembly score (Table 2), which was more than twice that of any other assembly, and was chosen to proceed using the version filtered at 1 TPM. Of the 57,704 contigs in the ORP PHRED 5 assembly, 35.94% (20,741) were predicted to contain complete open reading frames (ORFs), while 37.17% were predicted to contain partial ORFs.

3.3 The *Reclinomonas* membrane trafficking system is relatively complete

To contextualize the results of the differential expression analysis within the *R. americana* membrane trafficking system as a whole and to ensure reliable annotations of key genes of interest, I manually annotated a selection of proteins that are either associated with the CV or potentially informative for inferring the evolution of the organelle (Supplementary Online Table 8). As the transcriptome is duplicated in a way that is unlikely to be related to the true number of paralogs, the results of the searches were considered on a presence/absence basis only (Figure 11). In general, absences in transcriptome data should be treated with caution, as if the genes weren't expressed when the cells

Table 2 Quantitative assessment of the *Reclinomonas americana* transcriptome assemblies. The PHRED 5 rnaSPAdes assembly did not complete. TransRate scores are the probability that reads map well to the transcriptome, the number of potential bridges between real transcripts (~chimeras), the probability that a given contig scores well, and the assembly score (out of 1). BUSCO scores indicate the percentage of complete (C), duplicated (D), fragmented (F) or missing (M) single-copy orthologs. This has been abridged from Supplementary Online Table 7, which contains the full output from TransRate and BUSCO.

				TransRate				BUSCO			
Trimming	**Assembler**	**Filtering**	**Transcripts (#)**	**P(good mapping)**	**Potential bridges (#)**	**P(good contigs)**	**Score**	**C (%)**	**D (%)**	**F (%)**	**M (%)**
PHRED 20	ORP	TPM=0.5	65,965	0.55624	28,774	0.91	0.22	88	45	6	6
PHRED 20	rnaSPAdes	none	52,528	0.55289	24,023	0.99	0.23	85	46	8	7
PHRED 20	Trinity	none	89,048	0.50377	43,200	0.65	0.13	86	57	7	7
PHRED 5	ORP	TPM=0.5	70,762	0.88829	23,306	0.97	0.48	88	56	6	6
	ORP	TPM=1	57,704	0.88707	22,031	0.97	0.48	88	56	6	6
PHRED 5	rnaSPAdes	none	-	-	-	-	-	-	-	-	-
PHRED 5	Trinity	none	104,333	0.00254	409	0	3.00E-05	86	58	8	6

were harvested, they cannot appear in the assembly. Corresponding to the observation that free-living heterotrophs often have retained a near-complete set of membrane trafficking proteins from the LECA (Schlacht et al. 2014), almost all of the proteins I searched for were identified (Figure 11). There were some notable absences in the MTCs (Dsl1p of Dsl1, Sec6 of exocyst), the HTACs (AP5 sigma, half of TSET subunits), TBCs (TBC-L, TBC-RootA, and TBC-ExA1) and in the Rab GTPases (Rab20, 21, and 22). In this study, it would seem that the endocytic Rab GTPases are reduced (Figure 11), which would be unexpected for a bacterivorous organism; however, a similar analysis of the unpublished genome identified both Rab21 and 22 (data not shown). Sec6 is required for typical CV bladder formation and efficient liquid expulsion in *C. reinhardtii*, likely as a part of the full exocyst complex, which is also required for CV exocytosis in *D. discoideum* (Essid et al. 2012; Komsic-Buchmann et al. 2012, 2014). TSET and AP5 are both less well conserved sequence-wise, which may cause issues IDing orthologs with RBH (Hirst et al. 2011, 2014; Weiss et al. 2017).

A near complete set of LECA Qa, Qb, Qc and R SNAREs were identified, as well as their accessory SM proteins (Figure 11). The only absence was synaptobrevin, an R SNARE recently inferred to be in the LECA (Barlow 2019). Notably, the Qc SNARE Use1 has not previously been identified within Discoba (Venkatesh et al. 2017), and the Qbc SNARE has a PH domain upstream of the Qb and Qc domains, which seems to be a novel domain structure. I identified this domain structure in a Qbc SNARE in the *Andalucia godoyi* predicted proteome as well (sequence P000562 in EukProt proteome EP00762; Gray et al. 2020; Richter et al. 2022), indicating that it is probably not an assembly error. At least one instance of a PH domain/SNARE fusion has been reported previously in the vertebrate-specific tomosyn paralog amysin, where it plays a role in the recognition of specific membrane lipids (Kloepper et al. 2008; Kondratiuk et al. 2020).

3.3.1.1 Unanticipated results underscore the importance of human curation

In the case of TBC-K, 23 transcripts were identified through RBH. Since the largest number of TBC-Ks known in a single organism is 4 (Gabernet-Castello et al. 2013), I used domain discovery to check for the presence of a TBC domain in those transcripts. Only a single transcript had an identifiable TBC domain, and instead most of the transcripts had a TBC lysin motif (LysM) domain catalytic (TLDc) domain (Supplementary Online Table 9). A TLDc domain is present in the human TBC-K (Falace et al. 2010), which I used as a starting query. Results like these emphasize the importance of informed human interpretation in protein annotation.

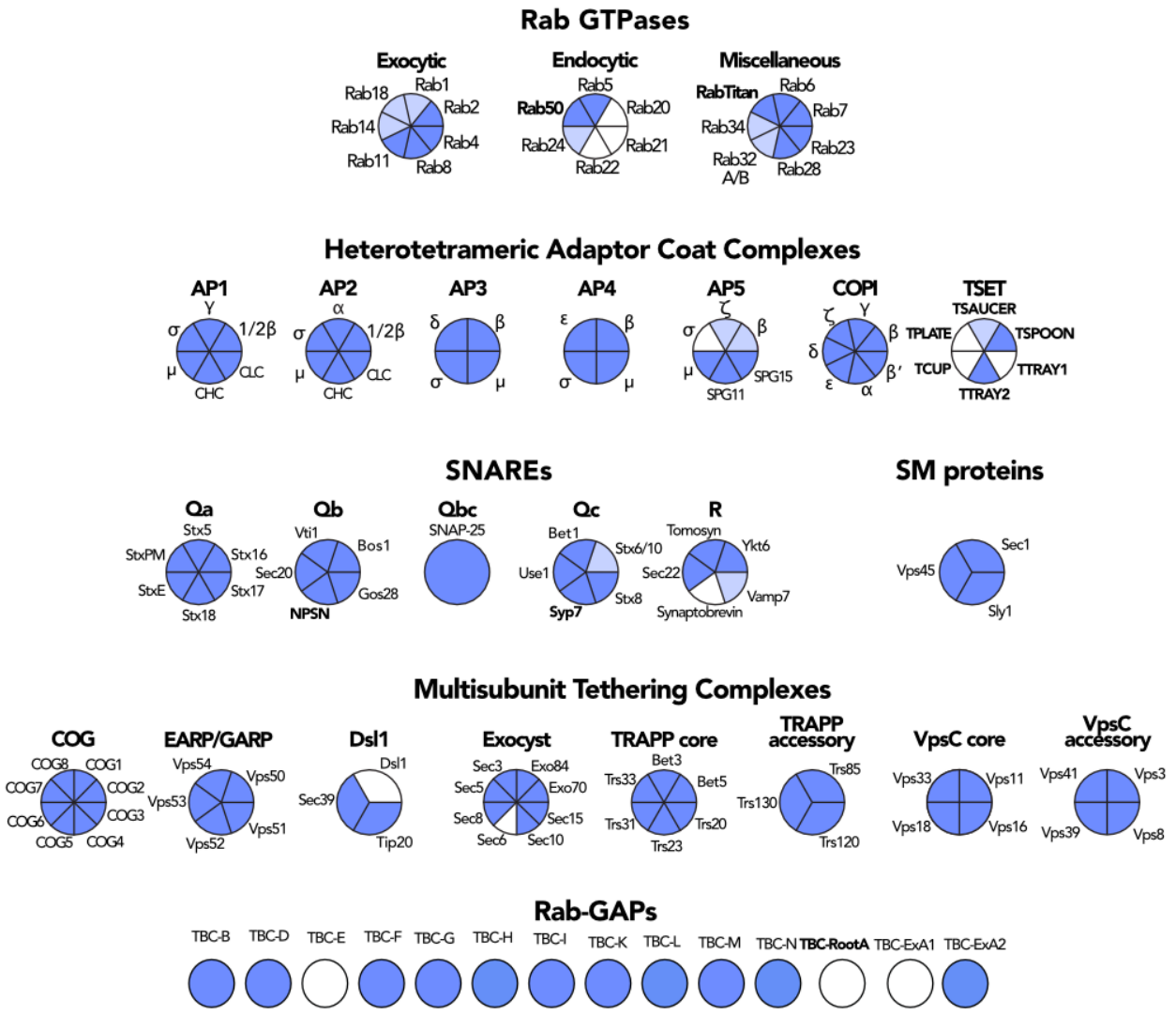


Figure 11 Coulson plot of membrane trafficking proteins identified in the *Reclinomonas americana* transcriptome. Classification for SNAREs, HTACs, and Rabs was confirmed with phylogenetics, which can be viewed in Supplementary Figure 1-3. Lighter colours indicate that the grouping was poorly supported in phylogenies (<90 UFBOOT). Jotnarlogs (in bold) are proteins that are inferred to be in LECA, but are not included in typical models of cell biology because they have been lost from yeast and animal cell biological models (More et al. 2020). Vps33 is also an SM protein but is included in the VpsC core.

3.4 Differential expression analysis identifies proteins previously associated with CVs

Clustering the transcriptome with Corset reduced the number of transcripts to a more reasonable value (39,737 vs 57,669) and resulted in more of the variance being explained by the treatment (Figure 12). After clustering, a total of 5,868 transcripts were detected as differentially expressed between the low and high osmolarity conditions, with 2,712 upregulated and 3179 downregulated under

hypoosmotic conditions (Supplementary Online Table 10). Of these, 211 were manually annotated transcripts of interest, 126 of which were upregulated under hypoosmotic conditions and 85 of which were downregulated under hypoosmotic conditions.

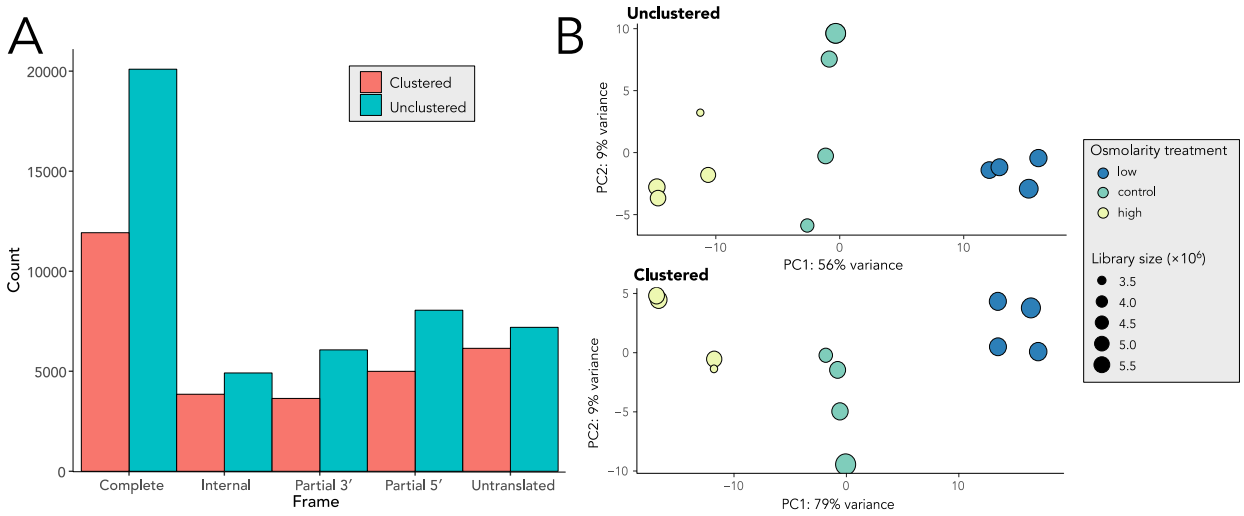


Figure 12 Effect of clustering the *Reclinomonas americana* transcriptome on transcript number and sample behaviour. Numbers of transcripts corresponding to predicted proteins of varying completeness in the ORP PHRED 5 TPM1 transcriptome assembly before and after clustering with Corset (**A**). Principal component analysis of transcript expression levels based on DESeq2 normalized read counts before and after clustering with Corset (**B**). Each dot represents a sample and is scaled to the RNA-seq library size.

To identified putative CV-associated proteins, I looked for a pattern of increasing expression with decreasing osmolarity across the three treatments. This pattern has been previously observed with the exocyst subunit Sec6 in *C. reinhardtii* (Komsic-Buchmann et al. 2014), while in *D. discoideum* 16 of 18 CV-associated genes were significantly upregulated in low osmolarities in comparison to typical culture conditions (Manna et al. 2023). Only 13 transcripts displayed this pattern. Some of these were known previously from studies in CV model organisms, namely Vamp7, aquaporin, Sec1, Rab8, and arrestins (Table 3). It is unclear whether the aquaporins that have been identified are in fact a part of the same aquaporin sub-family. Notably absent from the differentially expressed transcripts was the small GTPase Rab11, which was the only evolutionarily informative membrane trafficking protein identified prior and one of the few CV-associated proteins known from more than one organism (Harris et al. 2001; Niyogi et al. 2014; Manna et al. 2023). AP180 has been localized to both *D. discoideum* and *T. cruzi* CVs, but has only been found in single copies across eukaryotes (Stavrou and O'Halloran 2006;

Wen et al. 2009; Manna et al. 2015), and so was not included in the searches. Other differentially expressed transcripts had no previous CV association (tomosyn, Rab24, and the HTAC subunits). Tomosyn is particularly interesting, as in plants, animals and fungi it is thought to negatively regulate exocytosis, possibly through competitive inhibition of the R SNARE within the fusogenic SNARE tetrad (tomosyn lacks the transmembrane domain that anchors most SNAREs to membranes, and instead has a WD40 domain) (Gracheva et al. 2007; Hattendorf et al. 2007; Kienle et al. 2009; Li et al. 2019). It is possible that the increased CV bladder diameter in low osmolarity conditions is due to a combination of increased homotypic fusion from Vamp7 (as is thought to occur in *D. discoideum*; Wen et al. 2009) and inhibition of exocytosis from tomosyn (as Vamp7 is also involved in lysosomal exocytosis; Rao et al. 2004), though it is unclear where the extra membrane originates.

Table 3 Transcripts of the *Reclinomonas americana* membrane trafficking system that increase as osmolarity decreases. Log fold change (logFC) values are between low and high osmolarity treatments. Canonical locations of proteins include the endolysosomal system (ELS), plasma membrane (PM), and *trans*-Golgi network (TGN). The CV models are *Trypanosoma cruzi* (Tc), *Dictyostelium discoideum* (Dd), *Paramecium* spp. (Pspp.), and *Chlamydomonas reinhardtii* (Cr). Details of CV-association of the various proteins are found in Supplementary Online Table 3.

Protein	Canonical location	Transcript	logFC	p-value (adj.)	CV-associated
Vamp7	ELS	Rame_09154	2.9	6.15E-28	Tc, Dd, Pspp.
		Rame_56134	2.5	1.03E-09	
Aquaporin	various	Rame_22839	1.2	3.29E-07	Dd, Cr, Pspp., Tc
		Rame_08374	0.92	1.66E-14	
Sec1	PM	Rame_20647	0.94	8.96E-04	Dd
Rab8	PM	Rame_13065	0.89	2.52E-03	Dd
AP5 μ	ELS	Rame_22188	0.87	3.40E-03	-
Rab24	ELS	Rame_51353	0.83	3.58E-03	-
Tomosyn	PM	Rame_18337	0.59	1.04E-02	-
RabX7	-	Rame_41998	0.52	4.47E-04	-
Arrestin	Golgi/PM	Rame_00676	0.49	7.01E-03	Dd*
RabX16	-	Rame_19603	0.46	6.81E-03	-
AP4 σ	ELS/TGN	Rame_06827	0.44	4.86E-02	-

* Differentially expressed in a similar experiment in *Dictyostelium discoideum* (Manna et al. 2023).

Though I defined putative CV-involved proteins as increasing expression with decreasing osmolarity, transcripts showing the opposite pattern may give insight into CV function as well.

Interestingly, when I examined the 23 transcripts that showed increased expression with increasing osmolarity, a large number of Golgi and retrograde-transport related factors were identified, including 2/8 COG components (a MTC which acts within the Golgi; Miller and Ungar 2012), 6/7 COPI components (a HTAC involved in transport within the Golgi and to the ER; Beck et al. 2009), Vps52

Table 4 Transcripts of the *Reclinomonas americana* membrane trafficking system that decrease as osmolarity decreases. Log fold change (logFC) values are between low and high osmolarity treatments. Canonical locations of proteins include the endolysosomal system (ELS), early endosome (EE), endoplasmic reticulum (ER), and *trans*-Golgi network (TGN). The CV models are *Trypanosoma cruzi* (Tc)*, Dictyostelium discoideum* (Dd), *Paramecium* spp. (Pspp.), and *Chlamydomonas reinhardtii* (Cr). Details of CV-association of the various proteins are found in Supplementary Online Table 3.

Protein	**Canonical location**	**Transcript**	**logFC**	***p*-value (adj.)**	**CV-associated**
Trs23 (TRAPP core)	Golgi	Rame_21743	-1.3	7.50E-17	-
COG4	Golgi	Rame_45641	-0.8	1.09E-13	-
COG3	Golgi	Rame_10014	-0.7	3.67E-02	-
COPI β	Golgi	Rame_19821	-0.59	2.36E-04	-
TBC-F	EE	Rame_36811	-0.58	5.88E-03	-
Vps52 (GARP/EARP)	Golgi/EE	Rame_43267	-0.58	1.22E-02	-
COP γ	Golgi	Rame_05314	-0.57	1.61E-05	-
Sec22	ER/Golgi	Rame_52261	-0.54	2.15E-05	Tc
COPI δ	Golgi	Rame_50137	-0.54	1.73E-06	-
COPI β	Golgi	Rame_36176	-0.4	1.23E-06	-
Vamp7	ELS	Rame_03403	-0.4	8.76E-03	-
		Rame_13020	-0.38	1.65E-02	-
		Rame_00583	-0.19	1.13E-02	-
COPI α	Golgi	Rame_18418	-0.37	1.27E-06	-
COPI β′	Golgi	Rame_52214	-0.35	2.37E-03	-
AP1 μ	EE/TGN	Rame_01817	-0.33	2.49E-02	-
StxE	ELS	Rame_57702	-0.27	4.80E-03	-
RabX11	-	Rame_52151	-0.25	1.24E-02	-
COPI ζ	Golgi	Rame_07935	-0.25	1.29E-03	-
Rab23	Flagella	Rame_01031	-0.23	1.12E-02	-
Vps45	ELS	Rame_36268	-0.23	1.01E-02	-
AP12 β	AP1 or AP2 loc.	Rame_19578	-0.2	3.23E-03	*
AP2 α	PM	Rame_04400	-0.17	2.60E-02	Dd

(which could act in retrograde transport within the endosome or to the Golgi as a part of the Golgi-associated retrograde protein (GARP) or endosome-associated recycling protein (EARP) complexes; Bonifacino and Hierro 2011; Schindler et al. 2015), and TBC-F (a Rab GAP associated with retrograde transport from the endosome to the Golgi; Jia et al. 2016). The R SNARE Sec22 is also differentially expressed. In *T. cruzi,* a kinetoplastid inparalog of Sec22 colocalizes with markers of the CV spongiome and the Golgi body, though Sec22 localizes to the ER and Golgi in model systems (Newman et al. 1990; Hay et al. 1998; Ulrich et al. 2011). Additionally, the most highly expressed Vamp7 paralog was upregulated under hyperosmotic conditions. I speculate that this is the lysosomal Vamp7, as it has been observed in *D. discoideum* that Vamp7 is located at both the CV and at the lysosome, but in higher concentrations at the lysosome (Wen et al. 2009), and in humans the lysosome and autophagy are upregulated under hyperosmotic stress to deal with high osmolarity-induced protein agglutination (López-Hernández et al. 2020). Though significantly differentially expressed, the change in expression for these transcripts was generally small, with most LFCs between 0.17 and 0.6, corresponding to between 1.125 and 1.516 times the abundance (Table 4).

3.4.1 Corset and phylogenetically relevant clusters are not the same

While clustering with Corset was effective in reducing the number of total transcripts (Figure 12), the reduction did not correspond exactly to the reductions I made manually while doing phylogenetic classifications. For the membrane trafficking proteins examined, Corset clusters were consistent within classifications (i.e., a cluster would contain transcripts manually identified as belonging to the same protein), but there were often multiple clusters that belonged to a single phylogenetically relevant group. This is likely due to assembly errors, as many of these sequences were truncated when viewed in alignments, and Corset clusters are informed both by sequence similarity and expression similarity. It is also possible that these multiple clusters could be due to transcript isoforms. This inflated number of clusters could affect the sensitivity of the DE results, as splitting reads between similar transcripts decreases the statistical power of the analyses (Hsieh et al. 2019). As well, it complicated the interpretation of the DE results, as multiple clusters of the same phylogenetically relevant protein were sometimes identified as differentially expressed (e.g., multiple Vamp7 transcripts that corresponded to a protein in my phylogeny were upregulated in hypoosmotic conditions; Table 3; Supplementary Figure 1).

3.5 Four proteins are phylogenetically informative for CV evolution

If CVs within Discoba share a common origin, I would expect a shared set of proteins to be found in all the examined CVs, with CV-specific version of proteins paralogous to those from other membrane trafficking organelles (Figure 13A). If there are multiple independent origins of CV-like organelles in discobids, there are two possible patterns that could emerge: 1) Parallel evolution: If the evolutionary origin of the organelles is the same membrane-bound compartment, there may be some shared set of proteins in contractile vacuoles, but these would not form clades in phylogenies (Figure 13B); or 2) Convergent evolution: If the organelles have different evolutionary origins, they may have different sets of associated molecular machinery altogether.

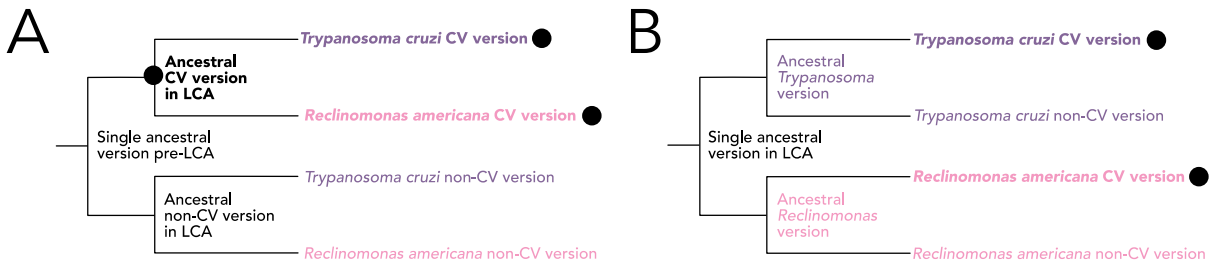


Figure 13 Example phylogenies that would support a single origin (**A**) or parallel origins (**B**) of contractile vacuoles (CVs) in Discoba. CV-associated proteins are labelled with black circles. LCA = last common ancestor.

To be phylogenetically informative in CV evolution, the putative CV-associated proteins identified through DE analysis must be: 1) found in CV model organisms; and 2) present in multiple copies (i.e., CV and non-CV versions). A combination of similarity searches and literature review was used to determine the distribution of the putative CV-associated proteins in the CV models (Figure 14; Supplementary Online Table 11). Searches were also performed for arrestins (data not shown); however, they are too poorly conserved on a sequence level to be reliably detected with homology searching (Baile et al. 2019) and a phylogeny would not be meaningful. Some differentially expressed transcripts were impossible to classify as orthologs of a known protein (RabX7, RabX16), and were not considered further. Other transcripts (AP5 mu, Rab24, tomosyn) were only present in the transcriptome assembly in a single phylogenetically relevant group. Others, such as Rab8, Vamp7, aquaporin, arrestin, and Sec1 were present in multiple copies in more than one organism (Figure 14).

The differentially expressed transcripts that were identified as present in more than one copy in more than one organism were then investigated phylogenetically.

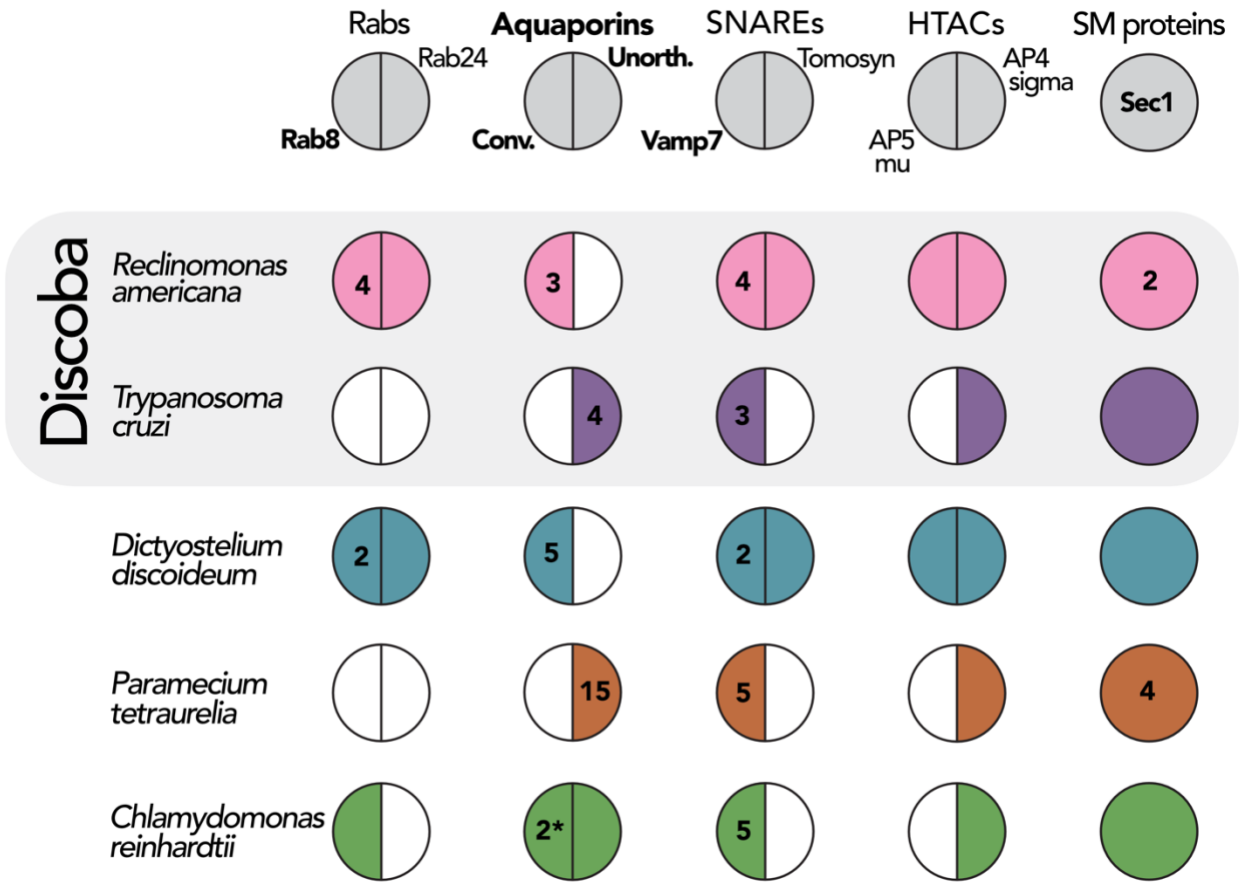


Figure 14 Coulson plot of putative contractile vacuole (CV) associated proteins in *Reclinomonas americana* and the CV model organisms. Data for the HTACs are taken from Hirst et al. (2011), Manna et al. (2013), Richardson and Dacks (2022), Dictybase.org, or a manual BLAST search. Classification for SNAREs, Rabs, and aquaporins was confirmed through phylogenetics, which can be viewed in Supplementary Figure 4-6. Numbers indicate copy numbers where more than one paralog is present. *R. americana* copy numbers are based on the number of phylogenetically distinct proteins after thinning duplicates manually. Proteins that are considered phylogenetically informative for CV evolution in the context of this study are in bold. Aquaporins are conventional (conv.) or unorthodox (unorth.). Additionally, *R. americana* has a single aquaglyceroporin, but is not included in the plot as aquaglyceroporins are not present in any CV model organisms and have not been implicated in CV function.

3.5.1 Rab8

Rab8 is a Rab GTPase primarily involved in trafficking from the TGN to the plasma membrane in animal and yeast model systems through interactions with the exocyst (Huber et al. 1993; Guo et al.

1999). In *D. discoideum,* a Rab8 paralog is part of a well characterized model of CV exocytosis, and is required for recruitment of the exocyst complex to the CV bladder prior to water expulsion (Essid et al. 2012). Rab8 is broadly distributed across eukaryotes (Eliáš et al. 2012). Out of the four CV models, Rab8 was only found in multiple copies in *D. discoideum* and *R. americana*, and was not identified in *P. tetraurelia* or *T. cruzi* (Figure 14). Although RBH identified 11 putative Rab8s in *P. tetraurelia*, during phylogenetic classification all sequences were found outside the Rab8 clade with full support (Supplementary Figure 4), and so these were not included in the Rab8 phylogeny. The Rab8 phylogeny had very little support apart from the two Rab8 paralogs from *D. discoideum,* which group together with strong support (83/1.00), indicating that this is likely a lineage-specific expansion (Figure 15). There is not enough resolution in the rest of the tree to comment on the relationships between the four *R. americana* Rab8s.

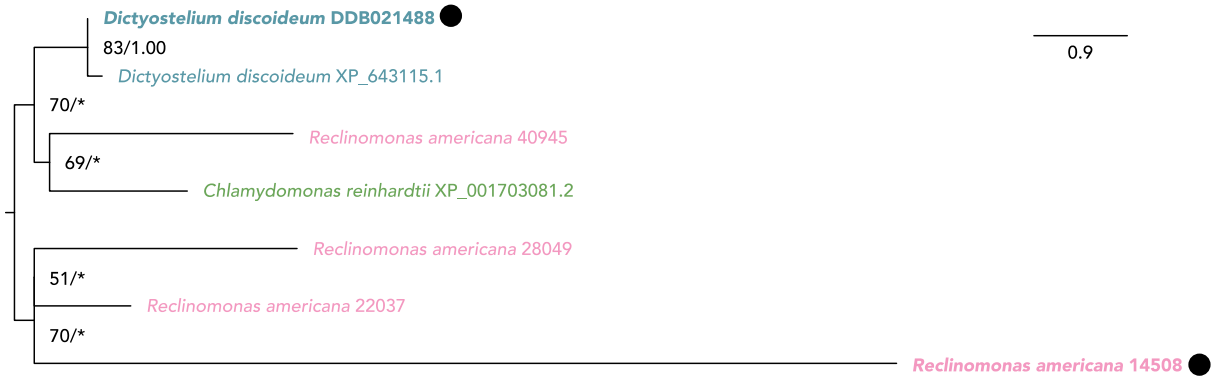


Figure 15 Maximum likelihood (ML) phylogeny of Rab8 in contractile vacuole (CV) model organisms and *Reclinomonas americana*. CV-associated proteins are labelled with black circles. CV involvement is this study or references in Supplementary Online Table 3. Support values from the ML analysis and Bayesian inference (BI) are shown as (ML/BI); support values < 50/0.90 ML/BI are considered uninformative and are displayed as asterisks or not shown. Scalebar represents the number of expected substitutions per site. Fully supported branches are in bold.

3.5.2 Sec1

Sec1 is an SM protein primarily thought to be involved in the templating of the *trans*-SNARE complex formed prior to exocytosis (Hashizume et al. 2009; Morgera et al. 2012; Zhang and Hughson 2021). Sec1 is required for CV function in *D. discoideum*; however, blocking the function of Sec1 blocks exocytosis as a whole, and not specifically that of the CV (Zanchi et al. 2010). Correspondingly, only a single copy of the gene is found in the *D. discoideum* genome (Figure 14). SM proteins are broadly

conserved in eukaryotes and are thought to be duplicated rarely in comparison to other MTS genes (Dacks and Field 2007; Koumandou et al. 2007). Multiple copies of Sec1 were only identified in *R. americana* and *P. tetraurelia* (Figure 14). The expansions in both organisms are lineage-specific with full support (100/1.00) (Figure 16). The expansion of Sec1 in *P. tetraurelia* has previously been characterized as occurring at the base of Oligohymenophorea, and has been suggested to be involved in the *P. tetraurelia* CV due to the concurrent duplication of a CV-specific Syntaxin PM, the Qa SNARE involved in exocytosis (Dacks and Doolittle 2004; Kissmehl et al. 2007; Kaur et al. 2022). Whether any of the *P. tetraurelia* Sec1 inparalogs are indeed CV-specific has yet to be determined.

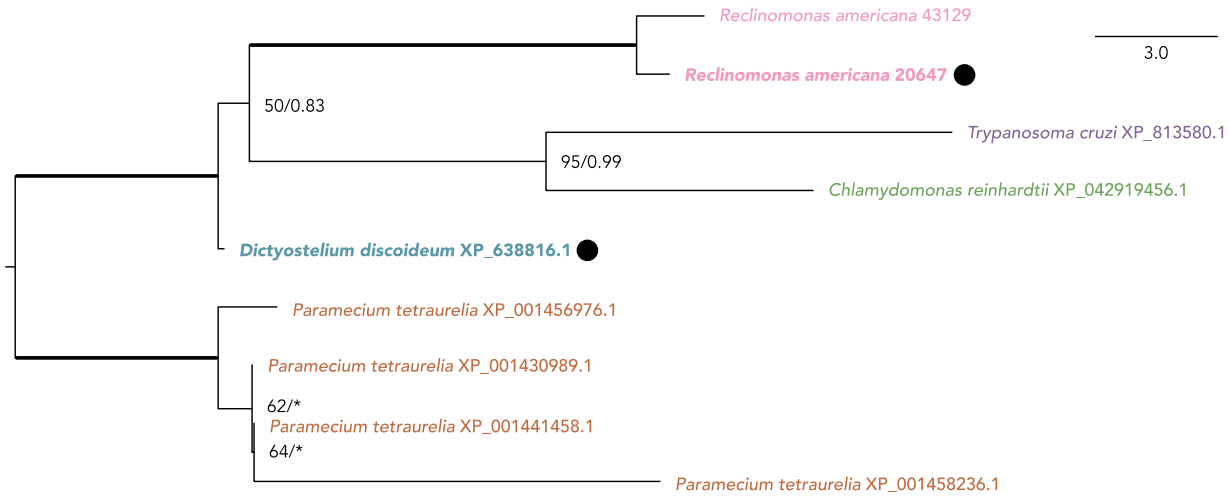


Figure 16 Maximum likelihood (ML) phylogeny of Sec1 in contractile vacuole (CV) model organisms and *Reclinomonas americana*. CV-associated proteins are labelled with black circles. CV involvement is this study or references in Supplementary Online Table 3. Support values from the ML analysis and Bayesian inference (BI) are shown as (ML/BI); support values < 50/0.90 ML/BI are considered uninformative and are displayed as asterisks or not shown. Scalebar represents the number of expected substitutions per site. Fully supported branches (100/1.00) are in bold.

3.5.3 Aquaporins

Aquaporins are small transmembrane proteins that act as channels for water and other small molecules (Gomes et al. 2009). There are five well-supported groups of prokaryotic aquaporins: AqpM, AqpN, GlpF, AqpZ, and AqpX (Finn et al. 2014; Pommerrenig et al. 2020; Tesan et al. 2021). Eukaryotic aquaporins were thought to primarily be related to prokaryotic AqpZ and GlpF, with a HGT of AqpN within Streptophyta (Finn and Cerdà 2015; Pommerrenig et al. 2020). A recent analysis identified not only a novel group of bacterial aquaporins, AqpX, but also a potential HGT of AqpX occurring within Discoba (Tesan et al. 2021). Aquaporins are one of the most broadly identified CV-associated proteins

and have been localized to the CV of *P. multimicronucleatum*, *T. cruzi*, *Leishmania major* (Kinetoplastea, Discoba), and *Amoeba proteus* (Amoebozoa, Amorphea); however, the classification of most of these has been unclear (Montalvetti et al. 2004; Figarella et al. 2007; Nishihara et al. 2008; Komsic-Buchmann et al. 2012; Ishida et al. 2021). In *P. multimicronucleatum*, silencing of the CV-associated aquaporin gene with RNAi significantly slowed the CV bladder enlargement-evacuation cycle, indicating its involvement in osmoregulation (Ishida et al. 2021).

All aquaporins identified through RBH were classified phylogenetically in a large phylogeny including all five prokaryotic aquaporin families, as well as all human aquaporins. The eukaryotic sequences fell into three larger groups: those associated with prokaryotic GlpF, those associated with the poorly characterized and largely archaeal AqpZx and AqpN, and those associated with bacterial AqpX (Supplementary Figure 5A, Figure 17). Following the naming conventions of Finn et al. (2014), I refer to these as aquaglyceroporins (human aquaporins 3, 7, and 9; GlpF-associated), unorthodox aquaporins (human aquaporin 11 and 12; AqpX-associated), and conventional aquaporins (human aquaporin 0, 1, 2, 4, 5, and 6; archaeal AqpZx- and AqpN-associated). Though human aquaporin 8 is considered a distinct family in animals (Finn et al. 2014), it is not phylogenetically distinct from conventional aquaporins in a paneukaryotic context (Figure 17). The presence of unorthodox aquaporins in Discoba was previously thought to be due to a Discoba-specific HGT (Tesan et al. 2021). However, this analysis has demonstrated that not only is this aquaporin found in Discoba (though it is absent in *R. americana)*, but that there are orthologs in *P. tetraurelia* (SAR), *H. sapiens* (Amorphea), and *C. reinhardtii* (Archaeplastida), making it a potential pan-eukaryotic aquaporin and not a Discoba-specific HGT (Figure 17). *C. reinhardtii* was the only CV model to contain both conventional and unorthodox aquaporins; all other organisms had either one or the other (Figure 14).

Notably, all *T. cruzi* aquaporins, including the CV version, belong to the unorthodox aquaporin group, while *R. americana* has an aquaglyceroporin and three conventional aquaporins (Supplementary Figure 5A, Figure 17). Two conventional aquaporins are differentially expressed in *R. americana*; though it should be noted that these are upregulated in low osmolarity conditions, and the single datapoint where both CV localization and expression data of an aquaporin exists shows decreasing expression of the aquaporin with decreasing osmolarity (*C. reinhardtii*; Komsic-Buchmann et al. 2014). Additionally, aquaporins are thought to be evolutionarily labile in response to environmental challenges (Finn and Cerdà 2015; Von Bülow and Beitz 2015), so it is possible that they may be uninformative in terms of inferring CV evolution despite their ubiquity in the organelle. Nevertheless,

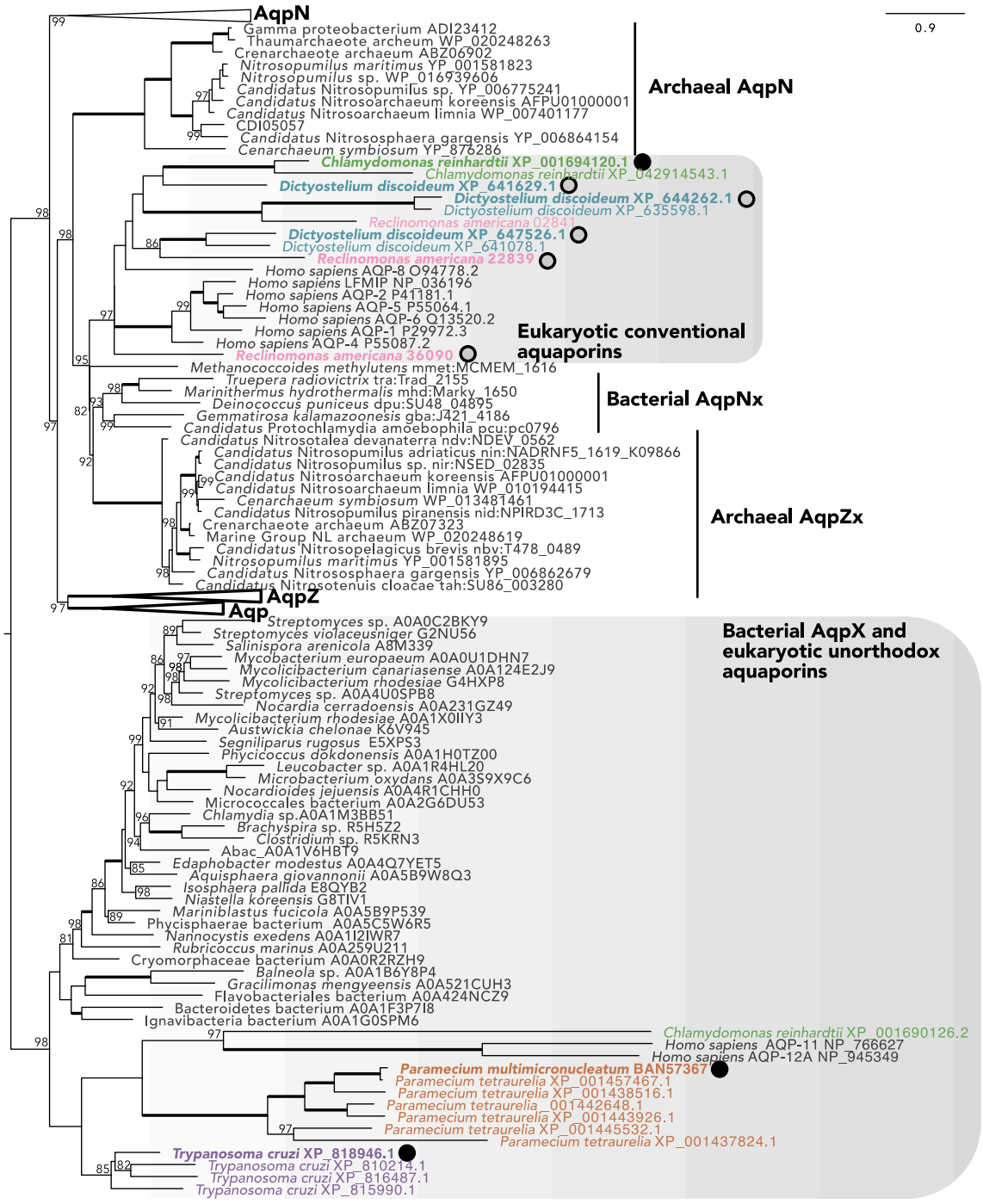


Figure 17 Maximum likelihood (ML) phylogeny of eukaryotic aquaporins in a prokaryotic context. CV-associated proteins with localization data are labelled with black circles, those with differential expression data only are labelled with grey circles. CV involvement is this study or references in Supplementary Online Table 3. Support values from UFBOOT ML analysis; values < 80 are considered uninformative and not shown. Scalebar represents the number of expected substitutions per site. Fully supported branches (100) are in bold.

I inferred a phylogeny for the conventional aquaporins with the CV model organisms and representative human sequences in case there is paralogy that could correspond to a CV-specific conventional aquaporin. There are no well-supported groups of aquaporins in the phylogeny apart from the human conventional aquaporins (95/1.00), two *D. discoideum* lineage-specific expansions, and the grouping of two *D. discoideum* aquaporins with a *R. americana* aquaporin (100/1.00) (Figure 18). The differentially expressed *D. discoideum, R. americana,* and *C. reinhardtii* aquaporins are dispersed throughout the phylogeny. Two of the differentially expressed *D. discoideum* aquaporins were sister to another *D. discoideum* aquaporin whose expression was not affected by osmolarity (89/1.00 and 99/1.00) (Figure 18). The CV-associated unorthodox aquaporins clearly belong to lineage-specific expansions (Figure 17), and so no further phylogenies were inferred.

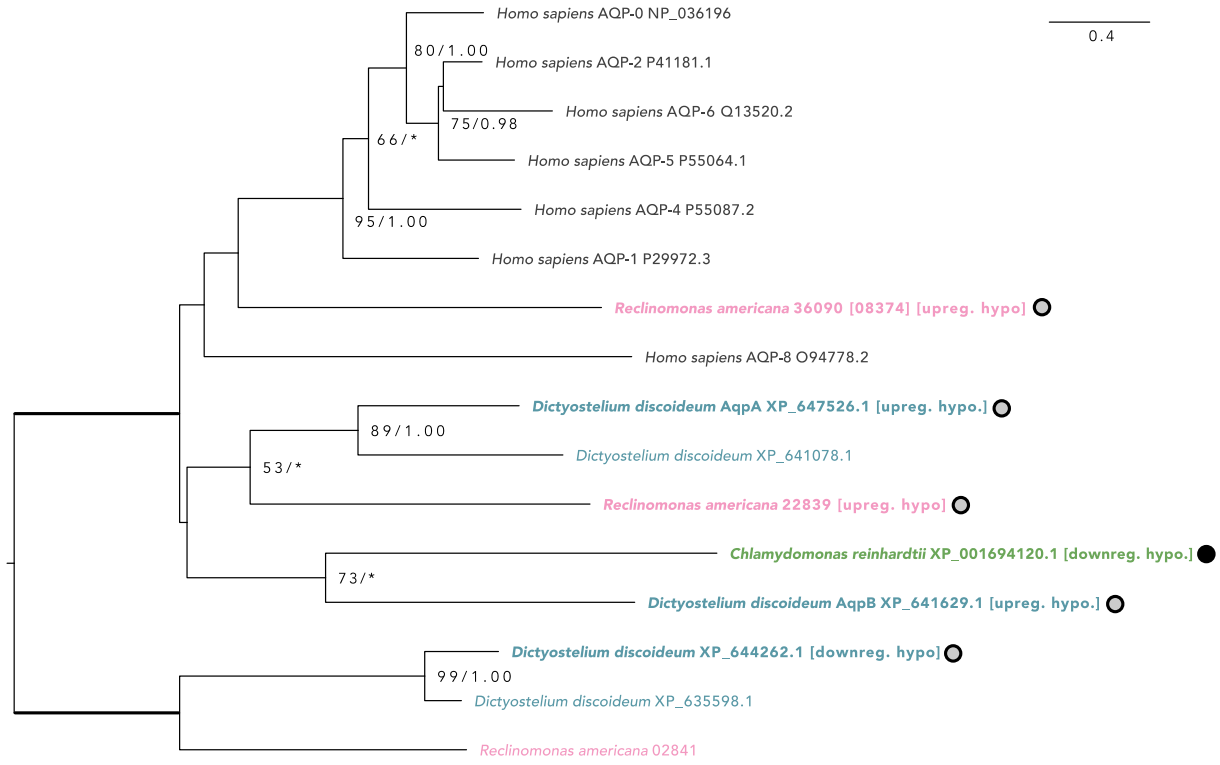


Figure 18 Maximum likelihood (ML) phylogeny of classical aquaporins in contractile vacuole (CV) model organisms, *Reclinomonas americana*, and humans. CV-associated proteins with localization data are labelled with black circles, those with differential expression data only are labelled with grey circles. CV involvement is this study or references in Supplementary Online Table 3. Support values from the ML analysis and Bayesian inference (BI) are shown as (ML/BI); support values < 50/0.90 ML/BI are considered uninformative and are displayed as asterisks or not shown. Scalebar represents the number of expected substitutions per site. Fully supported branches (100/1.00) are in bold.

3.5.3.1 Long-branch attraction obscures relationships between eukaryotic aquaporins

Classification of eukaryotic aquaporins is challenging across large evolutionary distances. The proteins are small, with few informative sites (~200 amino acids) and the evolutionary timescales are large. Additionally, the prokaryotic sequences are relatively well conserved in comparison to the eukaryotic sequences. This was particularly evident in the first iteration of the aquaporin phylogeny, where I used a relatively simple model of sequence evolution: most eukaryotic conventional and unorthodox aquaporins formed a poorly supported and extremely long-branching group with the archaeal AqpN and AqpZx, except for *T. cruzi* unorthodox aquaporins, which fell within the bacterial AqpX clade (data not shown). After I selected a mixture model to specifically deal with the issue of long-branch attraction (the ModelFinder selected model + C60/PMSF; Methods section 2.9.2) (Wang et al. 2018), this grouping disappeared (Supplementary Figure 5A). The phylogeny resolved further when the most divergent aquaporin family, GlpF, was removed (Figure 17).

3.5.4 Vamp7

Vamp7 is an R SNARE that is involved in several membrane fusion processes within the endolysosomal system, including lysosomal exocytosis, fusion of late endosomes and lysosomes, and homotypic fusion of autophagosome precursors (e.g. Advani et al. 1999; Rao et al. 2004; Moreau et al. 2011; reviewed in Chaineau et al. 2009). It is the most commonly identified membrane trafficking protein in CVs. This has been previously obscured by the inconsistent naming conventions within the SNARE family, as the CV-specific versions have been known as VAMP1 and SNARE2.2 (*T. cruzi*; Ulrich et al. 2011), Syb2 (*P. tetraurelia*; Schilde et al. 2006), and Vamp7B (*D. discoideum*; Wen et al. 2009). Functional characterization for the involvement of Vamp7 in CVs has only been done indirectly in *D. discoideum*, where Vamp7B is speculatively thought to be involved in homotypic fusion (Wen et al. 2009).

Because RBH is an unreliable method of classification within SNAREs, all R SNAREs from the CV models were classified using phylogenetics (Supplementary Figure 6). The best supported groups were iteratively removed until just Vamp7 and synaptobrevin remained, which was then used to root the phylogeny (Barlow 2019). The backbone of this phylogeny was very poorly supported (Figure 19A). As the relationships between most CV-associated Vamp7 genes were ambiguous, AU tests were used to test several constraints against the original version of the phylogeny: lineage-specific expansions of Vamp7 in all organisms, a single origin of CV-associated Vamp7 in Discoba, and a

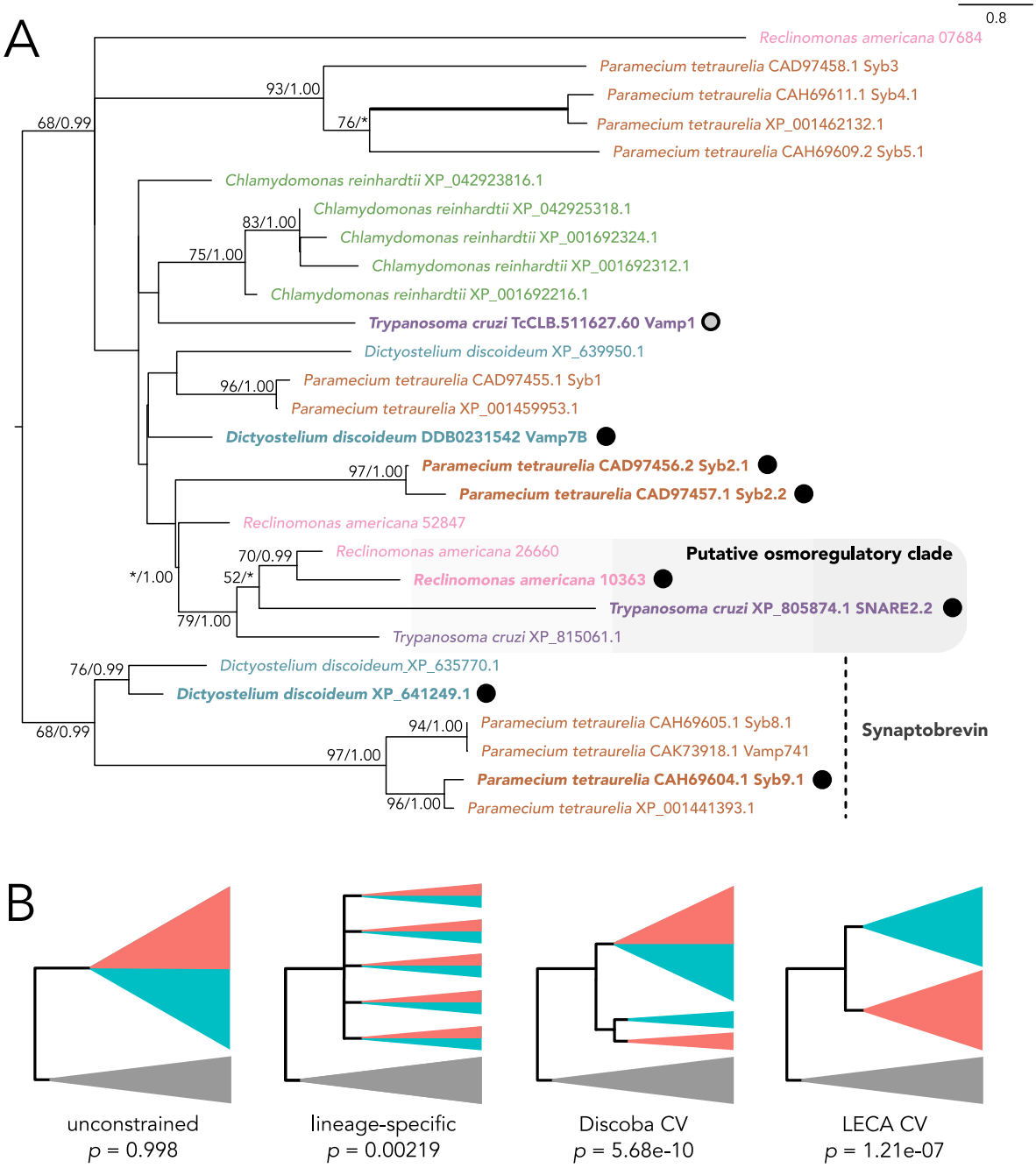


Figure 19 Maximum likelihood (ML) phylogeny of Vamp7 in contractile vacuole (CV) model organisms and *Reclinomonas americana* (**A**) and the alternate topologies tested with approximately unbiased (AU) tests (**B**). CV-associated proteins are labelled with black circles; a *T. cruzi* Vamp7 paralog that is transiently associated with the CV under hypoosmotic conditions is labelled with a grey circle. CV involvement is this study or references in Supplementary Online Table 3. Support values from the ML analysis and Bayesian inference (BI) are shown as (ML/BI); support values < 50/0.90 ML/BI are considered uninformative and are displayed as asterisks or not shown. Scalebar represents the number of expected substitutions per site. The *p*-values indicate significant difference from the most likely tree (in this case, the unconstrained tree) assessed with AU tests with 100,000 RELL bootstraps.

single origin of CV-associated Vamp7 across eukaryotes (Figure 19B). Synaptobrevin orthologs were unconstrained. All constrained phylogenies were rejected as being significantly worse than the most likely (unconstrained) phylogeny ($p < 0.05$) (Figure 19B). Interestingly, a moderately supported clade (79/1.00) in the unconstrained phylogeny contained both CV- and non-CV-associated Vamp7 paralogs from both *R. americana* and *T. cruzi*; however, the two Vamp7 paralogs from *R. americana* were sister to each other in this phylogeny with moderate support (70/0.99) (Figure 19A). A phylogeny was inferred with just those four proteins, which grouped the proteins by species, not by CV-association with strong support (88/1.00) (Figure 20). Notably, the other *T. cruzi* Vamp7 in this clade has been found in the reservosome proteome of epimastigotes (the sole endocytic life stage of *T. cruzi*) (Sant'Anna et al. 2009). Reservosomes are organelles which are considered the endpoint of the *T. cruzi* endocytic pathway prior to fusion with lysosomes (Soares et al. 1992), and have been proposed to be lysosome-related organelles (Sant'Anna et al. 2008). The other *R. americana* Vamp7 in this clade, as previously noted, is the most highly expressed Vamp7 in the dataset.

Figure 20 Maximum likelihood (ML) phylogeny of Vamp7 proteins from the putative osmoregulatory clade in Discoba. CV-associated proteins are labelled with black circles. CV involvement is this study or references in Supplementary Online Table 3. Support values from the ML analysis and Bayesian inference (BI) are shown as (ML/BI). Scalebar represents the number of expected substitutions per site.

4 Perspectives

In this final section, I will begin with an assessment of the data and my approach, speculate on the origin of contractile vacuoles both within Discoba and across eukaryotes, and end with some intriguing avenues for future research.

4.1 Advantages and limitations of the approach

This work was a thorough examination of the *Reclinomonas americana* contractile vacuole using light microscopy, transcriptomics, and phylogenetics. The use of three osmotic treatments (low, control, and high osmolarities) meant that I could search specifically for transcripts that consistently increased or decreased with osmolarity across all three treatments; this is the pattern that has been seen in cell-biologically-validated CV proteins when multiple treatments have been examined (Komsic-Buchmann et al. 2014). When assembling the transcriptome, I selected the assembly based on a number of quantitative *de novo* assessment metrics, which, without a genome to use as reference, is the most robust approach currently possible. I considered the common artefacts of transcriptome assembly (inflated isoforms, different genes being assembled by different assemblers, and the biases of low and high *k*-mer selections) and attempted to ameliorate them bioinformatically using a multi-assembler and clustering the genes based on expression and sequence similarity. The validation of the osmotic treatments with microscopy gave an additional dimension with which to interpret the transcriptomic response. Tomosyn, a negative regulator of exocytosis, was upregulated in the low osmolarity condition along with an R SNARE that can be involved with fusion at the plasma membrane—a result that may not have made sense without the additional context that the bladder size increased, but the enlargement–evacuation interval stayed constant. Additionally, the careful classification of proteins of interest in a paneukaryotic (and sometimes prokaryotic) context using phylogenetics revealed the previously overlooked presence of the R SNARE Vamp7 in three out of four CV models as well as the presence of distinct origins of the conventional and unorthodox aquaporin families in eukaryotes.

The largest advantage (and disadvantage) to the *de novo* transcriptomic approach specifically is being able to do this research in a non-model system. *De novo* transcriptomics opens a much broader availability of taxa, and can result in a larger volume of data than is generally attainable with conventional cell biological methods. While the availability of protist genomes and models is improving rapidly, there are still few models among the free-living, heterotrophic protists. The model systems within Discoba come with not insignificant challenges: 1) trypanosomatids (Kinetoplastea) are thought to primarily regulate protein synthesis through post-transcriptional modification (Jensen et al. 2014), making it extremely complicated to draw protein expression-level inferences from

transcriptomic data; 2) *Euglena* (Euglenida) seems to also regulate protein synthesis post-transcriptionally (Schwartzbach 2017), which would parsimoniously suggest that the entirety of Euglenozoa use this approach (though some level of DE analysis within Diplonemea seems to be possible; Vesteg et al. 2019; Prokopchuk et al. 2022); and 3) heteroloboseans (including the model *Naegleria gruberi*) generally have a biphasic lifestyle and transform between amoeba and flagellate life stages, which can be triggered by osmotic shifts (Fulton 1993). However, the use of a non-model organism does come with drawbacks. The lack of a publicly available genome made this challenging. The initial transcriptome assembly was unsatisfactory, and transcriptome assembly became a significant and unanticipated part of this project. This project required a transcriptome assembly optimized for both differential expression (i.e., few excess isoforms) and phylogenetics (i.e., full length sequences of all proteins of interest). Common artefacts of transcriptome assembly include incomplete transcript recovery and inflated isoform numbers (Vijay et al. 2013; Freedman et al. 2021), which can substantially affect the sensitivity of DE analyses and the availability of genes for phylogenetics (Hsieh et al. 2019; Spillane et al. 2021). Though attempts were made to decrease these artefacts, the final clustered assembly was still quite duplicated (phylogenetically relevant clusters sometimes included multiple Corset clusters; data not shown). Despite this, three phylogenetically informative proteins for CV evolution were identified. However, transcriptomics can only generate hypotheses for a protein's involvement. While the involvement of similar proteins in CVs in models which have been validated with cell biological work can increase the confidence in some of the CV associations in the data, the logic is circular, and confirmation of the CV involvement of these proteins can only be obtained through further molecular characterization.

As this study was only the second to identify putative CV-associated proteins through differential expression analysis, it was unclear what conditions would be appropriate to obtain a robust CV transcriptomic response. When I chose the conditions, transcriptomic data from the *Dictyostelium discoideum* CV induction experiment had not yet been analyzed, and I was concerned that the direct transition from the culture medium to MilliQ purified water (as performed in Manna et al. 2023) would be too harsh and come with additional stress responses that could obscure the CV involvement. The medium used, HL5, has an approximate osmolarity of 100 mOsm (Gamper et al. 1999). A similar experiment of osmotic stressors in *D. discoideum* used a 1:3 dilution of HL5 with distilled water (Mathavarajah et al. 2018), which was the same dilution I used in this experiment. A factor that I did not consider is that *D. discoideum* is grown axenically in a nutrient-replete medium. The difference between HL5 medium and MilliQ water (or even HL5 and a 1:3 dilution of HL5) is much greater than

the difference between my treatments (3.5, 13.8, and 34.5 mOsm). While assessing the phenotypic response of the CV was a logical starting point to determine appropriate osmolarities, and the response in CV bladder diameter was statistically significant, the transcriptomic response between conditions was not strong, at least within the membrane trafficking system. Combined with the issue of excess isoforms in the transcriptome assembly obscuring signal, I expect that the gentle osmotic contrast of the conditions has resulted in missing a selection of evolutionarily informative, truly differentially expressed genes. In the future, experiments should use more extreme contrasts between treatments, with as low osmolarities as the cells are able to tolerate over a few hours without being obviously distressed. Additionally, while the three treatments gave increased confidence in the CV-association of the proteins, hypoosmotic-to-control contrasts were sufficient to generate hypotheses for CV involvement in *D. discoideum* (Manna et al. 2023), so I may be missing potentially informative proteins with the current approach.

4.2 Thoughts on the evolution of contractile vacuoles

4.2.1 Evolution of the contractile vacuole within Discoba

There is no strong evidence for a single origin of CVs within Discoba. A recent study inferred a non-marine (i.e., soil or freshwater) origin for Discoba (Jamy et al. 2022). One might expect this to correspond to an ancestral CV within the supergroup, but this is not borne out by the data. The data are currently compatible with both parallel and convergent evolution of the *R. americana* and *Trypanosoma cruzi* CVs. *T. cruzi* (and Kinetoplastea as a whole) has lost many of the genes identified as differentially expressed in *R. americana* (Rab8, Rab24, AP5) (Venkatesh et al. 2018), while the genes that can be used to infer organellar origin in *T. cruzi* (Rab11, Rab32) were not differentially expressed in *R. americana*. This evokes the idea of convergent evolution of the two CVs from distinct endomembrane compartments. On the other hand, while Vamp7, the sole protein implicated in CVs in both organisms, is broadly the endolysosomal R SNARE, the Vamp7 paralogs in *T. cruzi* show signs of specialization. *T. cruzi* has one Vamp7 paralog that has been localized to the CV spongiome (Ulrich et al. 2011), one that has been detected in the proteome of the endpoint of its endocytic system, the reservosome (Sant'Anna et al. 2009), and one that has been localized to the acidocalcisome and the reservosome under basal conditions, and the CV under hypoosmotic stress (Sant'Anna et al. 2009; Ulrich et al. 2011; Niyogi et al. 2015). It is striking that the differentially expressed paralog in *R. americana* and the CV spongiome-localized paralog in *T. cruzi* are found within the only supported multi-organism groupings in the Vamp7 phylogeny, together with the most highly expressed *R.*

americana Vamp7 and the reservosome-localized *T. cruzi* paralog. This points towards parallel evolution of a putative ancestrally osmoregulatory late endosomal organelle. Conservation of the *T. cruzi* paralog localizations within cells of other discobids would strengthen this argument. While orthologs of the acidocalcisome and reservosome Vamp7s have been localized in *T. brucei*, the evidence for localization of these proteins at specific organelles within the endolysosomal system is not strong (Venkatesh et al. 2017; Billington et al. 2022). An ortholog of the spongiome Vamp7 is not found in the *T. brucei* genome, nor is a CV present in its endomembrane system (Supplementary Figure 6; Overath and Engstler 2004).

The exact organellar origins of both *R. americana* and *T. cruzi* contractile vacuoles are unclear. There are only a few proteins identified in each organism that can give hints of the potential origins. Previous hypotheses on the origin of CV-like organelles include the recycling endosome, due to the involvement of Rab11 in *D. discoideum* (Harris et al. 2001), or the Golgi, due to the presence of clathrin-coated regions of the CV in green algae (Komsic-Buchmann and Becker 2012). The presence of clathrin doesn't require that the CV derives from the Golgi, as clathrin coats can be found on a number of endolysosomal organelles (reviewed in Klumperman and Raposo 2014). The recycling endosome hypothesis is more plausible for *T. cruzi*. However, it should be noted that the original hypothesis of Harris et al. (2001) proposed that the *D. discoideum* CV was a *precursor* to the mammalian recycling endosome. This specific relationship is unlikely, as Rab11 has a role in recycling in *T. brucei*, which does not have a CV (i.e., might have canonical recycling endosomes) (Overath and Engstler 2004). Rab11 and Rab32 have been localized to the *T. cruzi* CV (Ulrich et al. 2011; Niyogi et al. 2014, 2015). In model systems, Rab11 is involved in recycling endocytosed materials from the recycling endosome back to the plasma membrane (Ullrich et al. 1996; Takahashi et al. 2012; Choi et al. 2013; Umaer et al. 2018). Rab32 has not been studied as broadly, but is involved in the trafficking of proteins from the recycling endosome to a model animal lysosome-related organelle, the melanosome, as well as formation of autophagosomes and recruitment of an effector protein involved in mitochondrial division at the ER (Hirota and Tanaka 2009; Bultema et al. 2012; Ortiz-Sandoval et al. 2014). While the distribution of Rab32 is paneukaryotic (Eliáš et al. 2012; Ortiz-Sandoval et al. 2014), it is unclear whether there is a protistan equivalent to animal Rab32 lysosome-related organelles (Delevoye et al. 2019), or what the role of Rab32 is outside Metazoa. Neither of these proteins were identified as differentially expressed in *R. americana*.

The *R. americana* CV instead shows a slight signal for secretion and the late endosomes/autophagosomes. A secretory signal (in this case Rab8, Sec1, and tomosyn) is not

unexpected for an organelle whose most apparent function is exocytosis. Aspects of the typical secretory pathway have been identified in the CV of *D. discoideum* and *Chlamydomonas reinhardtii* as well (Essid et al. 2012; Komsic-Buchmann et al. 2012). Sec1 and Rab8 are both part of an exocytic pathway with the tethering exocyst complex in yeast (Guo et al. 1999; Hashizume et al. 2009; Morgera et al. 2012), and in the CV of *D. discoideum* (Du et al. 2008; Zanchi et al. 2010; Essid et al. 2012). In *D. discoideum,* this is thought to be initiated by a CV-specific Rab11 paralog (Essid et al. 2012). It should be noted that the exocyst complex is most prominently involved in endocytosis in *T. brucei*, and contains a recently discovered ninth subunit that is specific to Discoba (not queried here) (Boehm et al. 2017; Boehm and Field 2019). It is unclear whether the exocyst maintains its role in exocytosis in trypanosomatids. The autophagosomal signal (Rab24, AP5 mu, AP4 sigma) is more unexpected, and shares no similarity to the protein complement identified in the CVs of other organisms outside the involvement of Rab32 in *T. cruzi*. Though the apparent roles of these three proteins are only starting to be understood in the best studied human cell lines, all seem to localize to the latter part of the endolysosomal system. Rab24 localizes to late endosomes, lysosomes, and autophagosomes, and is involved in the clearance of autophagosomes under basal conditions, though whether this is through fusion with the plasma membrane or reformation of lysosomes is unknown (Ylä-Anttila et al. 2015; Amaya et al. 2016). Rab24 has been coprecipitated with key plasma membrane fusion proteins, and has been proposed to be involved with fusion at the plasma membrane (Schardt et al. 2009; Ylä-Anttila and Eskelinen 2018). AP5 is involved in the transport of protein cargo between the late endosome/lysosome and the Golgi (Hirst et al. 2018; Khundadze et al. 2019), and is involved in the reformation of lysosomes from autolysosomes (Chang et al. 2014; Khundadze et al. 2019), while AP4 is involved in the transport of protein cargo between the TGN and endosomes (Burgos et al. 2010) or autophagosomes (Mattera et al. 2017). Vamp7 is a non-specific signal, as there are only four fusogenic R SNAREs in LECA and Vamp7 has been implicated in many fusion processes within the endolysosomal system (Advani et al. 1999; Rao et al. 2004; Moreau et al. 2011; Barlow 2019). Without knowledge of whether there are specific paralogs of Vamp7 with conserved roles across Discoba, this protein remains uninformative in terms of organellar origin.

4.2.2 Implications for the origin of the contractile vacuole

There was more overlap of the *R. americana* and *T. cruzi* CV associated proteins with the non-discobid CV models than between the two discobids. In each case, phylogenetics demonstrates strong evidence for lineage-specific expansions in at least some of the organisms, and no strong evidence for a single

origin of a CV-specific version. This corresponds with the previous observations of CV-specific Rab11 paralogs in *D. discoideum* and *T. cruzi*: both are the result of lineage specific expansions (Manna et al. 2023). This overlapping protein complement has, without the use of phylogenetics, been used to infer a single origin of CVs in the LECA (e.g., Hoef-Emden 2014) or invoke involvement of proteins from other CV model organisms in models of function (such as the exocyst in the CV of *T. cruzi;* Jimenez et al. 2022). An alternate explanation for this pattern is that the CVs that have been studied are more recent innovations derived from the same ancestral organelle present in the LECA that was, crucially, not a CV. An implication of this is that *if* there is a consistent organellar source from which an osmoregulatory organelle evolved, this could be due to a latent capacity for osmoregulation in the source organelle.

Intriguingly, lysosomes have been documented as responding to osmotic shock in some model organisms that do not have CVs. There is a sparsely documented phenomenon of human lysosome swelling and exocytosis in response to hypoosmotic shock (Okada et al. 1992; Li et al. 2020). When this response was impaired through knocking out a subunit of a lysosome-associated anion channel, the survival rate of the cells introduced to hypotonic challenges decreased significantly (Li et al. 2020). Analogy of this response with the *D. discoideum* CV has been proposed (Hu et al. 2022), primarily due to the role of calcium stimulations in the exocytosis of lysosomes in non-secretory human cells and the *D. discoideum* CV (Reddy et al. 2001; Jaiswal et al. 2002; Fountain et al. 2007; Parkinson et al. 2014). The yeast vacuole (the fungal lysosome equivalent) is also involved in osmotic stress responses. Though vacuolar exocytosis is not involved, fusion and fission of vacuoles takes place under hypo- and hyper-osmotic stress respectively (Latterich and Watson 1993; Bone et al. 1998), and mutants deficient in vacuolar fusion have increased susceptibility to osmotic stress (Banta et al. 1988; Latterich and Watson 1993). Notably, the V-ATPase is key in the response to hyperosmotic stress, as it creates a proton gradient that drives sequestration of ions in the vacuole (Li et al. 2012). Additionally, there are a variety of transient CVs found in marine organisms that are exposed to low osmolarities (Hogue 1923; Meyer and Pienaar 1984; Hellebust et al. 1989; Hauer et al. 2001) and freshwater organisms that have cell walls throughout most of their lifecycle (Ling and Tyler 1972; Shields and Fuller 1996; Mitchell and Hardham 1999). Are these stable organelles, with distinct protein complements, or a general response of the endolysosomal system?

4.3 Future directions

The most obvious next step is to continue this work within Discoba. A few phylogenetically informative genes from two datapoints is not sufficient to make confident conclusions about the evolution of CVs across such broad evolutionary timescales. A datapoint within Heterolobosea would be the next priority, particularly because the recent inference of a non-marine origin of Discoba by Jamy et al. (2022) only sequenced a single jakobid, and so their results might better be interpreted as a non-marine origin of Heterolobosea + Euglenozoa. If CV induction is possible in the model *N. gruberi* without also inducing the flagellate differentiation, this would be the obvious choice, as there is a well characterized genome and recently developed molecular transformation protocols (Fritz-Laylin et al. 2010; Herman et al. 2018; Faktorová et al. 2020b). If the amoeboflagellate differentiation is impossible to avoid upon osmotic shock, there is the potential to use *de novo* transcriptomics on one of the few heterolobosean where differentiation has not been documented (e.g., the genus *Vahlkampfia*, for which flagellate stages have not been observed; Brown and De Jonckheere 1999). Additionally, the monospecific, deepest branching lineage within Discoba, *Tsukubamonas*, is from freshwater (Yabuki et al. 2011). Though a CV was not identified in the original description, this bears reinvestigation and could provide an interesting datapoint for the study of CVs within Discoba. Transcriptomics within these lineages can be used as hypothesis-generating for proteins with potential CV involvement in model systems such as *N. gruberi, B. saltans,* or *E. gracilis*. This was recently demonstrated in *D. discoideum* with the SNARE novel plant syntaxin (NPSN) (Manna et al. 2023). NPSN was first discovered in *Arabidopsis thaliana*, where it localizes to the plasma membrane and is involved in membrane fusion at the cell plate during cytokinesis (Zheng et al. 2002; Uemura et al. 2004; El Kasmi et al. 2013). In *D. discoideum,* NPSN was upregulated in hypoosmotic conditions, and further examination showed that it localizes to the plasma membrane and seems to be involved in exocytosis of the CV (Manna et al. 2023).

From a paneukaryotic perspective, the exploration of transient or inducible contractile vacuoles is the most interesting area of future study. Investigation into whether lysosomal exocytosis in response to hypoosmotic shock is a derived trait in primates or a characteristic of the eukaryotic endolysosomal system as a whole could provide important context for inferring the evolutionary origin of CVs. There are several potential approaches. A quick approach to identify inducible CVs in marine organisms could be exposure of the organism to lower osmolarities and careful examination in light and transmission electron microscopy. While most marine organisms cannot be grown in freshwater, many can tolerate a broad range of salinities (e.g., Hellebust et al. 1989; Arndt et al. 2000; Hauer et al.

2001; Hauer and Rogerson 2005; Cowie and Hannah 2006; More et al. 2019). The mechanisms behind this phenomenon are poorly understood in heterotrophic protists (Suescún-Bolívar and Thomé 2015), though some inducible CVs have been reported (Hogue 1923; Hauer et al. 2001). Unfortunately, as far as I can tell, there are no marine model protists with well characterized endolysosomal systems, which would be the best choice for this type of experiment. However, the recent coordinated efforts to establish a number of marine model protists could provide a good starting point for selecting potential datapoints for attempts at CV induction that could be followed with transcriptomics or molecular cell biological methods (Waller et al. 2018; Faktorová et al. 2020b). In particular, the emerging model *Paradiplonema papillatum* (Diplonemea) would be an excellent datapoint within Discoba, as transformation tools have recently been established (Faktorová et al. 2020a). In species that possess a transient CV, this could be examined through transcriptomics, and, ideally, molecular cell biological methods to confirm localization of the proteins. Transcriptomic analyses would likely be impossible or uninformative for CVs that appear in distinct life stages of freshwater organisms; luckily, these are found in lineages that have a wealth of transformation protocols available (Zhou and von Schwartzenberg 2020; Ghimire et al. 2022). Differentiating whether these transient CVs are distinct organelles, or whether they are simply an uncharacterized response of the lysosome, as seen in human cells (Li et al. 2020) could help distinguish between whether the diversity of CVs in modern eukaryotes is due to parallel evolution, convergent evolution, or common descent. With only a handful of CVs investigated on a molecular level, it is impossible to make claims about a CV *not* being present in the LECA—it could be possible that the organisms examined are those with recent reinventions of the concept. However, if CV-like behaviour is a general property of the lysosome, this supports the hypothesis of convergent evolution of modern CVs, and potentially explains why a collection of organelles with similar functions and protein complements have no (currently) recoverable signal for a single origin in the LECA.

4.4 Conclusions

In this thesis I characterized the *Reclinomonas americana* contractile vacuole using light microscopy, transcriptomics, and phylogenetics. *R. americana* modulates the CV bladder size in response to environmental osmolarity, but not the enlargement–evacuation interval. Neither of these traits seem to be related to phylogeny, even across relatively small evolutionary distances. Using a *de novo* transcriptome assembly, I characterized the *R. americana* membrane trafficking system as relatively complete in terms of the protein complement inferred to be found in the LECA. Through differential

expression analyses, I identified an additional four genes that are evolutionarily informative for the history of the CV, where previously only a single gene was known. Comparative genomics and phylogenetic analyses demonstrated no strong evidence for a single origin of CVs in the ancestor of Discoba, or across CV model eukaryotes. The lack of phylogenetic signal for a single origin of CVs but repeat involvement of some genes raise the question of whether CVs, when present as a distinct organelle, have evolved from an organelle within the endomembrane system with latent osmoregulatory capacity.

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*. *Eukaryot. Cell*, 8:1665–1676. doi:10.1128/EC.00123-09

Adl S. M., Bass D., Lane C. E., Lukeš J., Schoch C. L., Smirnov A., Agatha S., Berney C., Brown M. W., Burki F., Cárdenas P., Čepička I., Chistyakova L., del Campo J., Dunthorn M., Edvardsen B., Eglit Y., Guillou L., Hampl V., Heiss A. A., Hoppenrath M., James T. Y., Karnkowska A., Karpov S., Kim E., Kolisko M., Kudryavtsev A., Lahr D. J. G., Lara E., Lynn D. H., Le Gall L., Mann D. G., Massana R., Mitchell E. A. D., Morrow C., Park J. S., Pawlowski J. W., Powell M. J., Richter D. J., Rueckert S., Shadwick L., Shimano S., Spiegel F. W., Torruella G., Youssef N., Zlatogursky V. & Zhang Q. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.*, 66:4–119. doi:10.1111/jeu.12691

Adolf F., Herrmann A., Hellwig A., Beck R., Brügger B. & Wieland F. T. 2013. Scission of COPI and COPII vesicles is independent of GTP hydrolysis. *Traffic*, 14:922–932. doi:10.1111/tra.12084

Advani R. J., Yang B., Prekeris R., Lee K. C., Klumperman J. & Scheller R. H. 1999. Vamp-7 mediates vesicular transport from endosomes to lysosomes. *J. Cell Biol.*, 146:765–776. doi:10.1083/jcb.146.4.765

Agop-Nersesian C., Egarter S., Langsley G., Foth B. J., Ferguson D. J. P. & Meissner M. 2010. Biogenesis of the inner membrane complex is dependent on vesicular transport by the alveolate specific GTPase Rab11B. *PLOS Pathog.*, 6:e1001029. doi:10.1371/journal.ppat.1001029

Allen R. D. & Naitoh Y. 2002. Osmoregulation and contractile vacuoles of protozoa. *In*: International Review of Cytology. Vol. 215. Elsevier. p. 351–394. doi:10.1016/S0074-7696(02)15015-7

Altschul S. F., Gish W., Miller W., Myers E. W. & Lipman D. J. 1990. Basic Local Alignment Search Tool. *J. Mol. Biol.*, 215:403–410.

Amaya C., Militello R. D., Calligaris S. D. & Colombo M. I. 2016. Rab24 interacts with the Rab7/Rab interacting lysosomal protein complex to regulate endosomal degradation. *Traffic*, 17:1181–1196. doi:10.1111/tra.12431

Amos B., Aurrecoechea C., Barba M., Barreto A., Basenko E. Y., Bażant W., Belnap R., Blevins A. S., Böhme U., Brestelli J., Brunk B. P., Caddick M., Callan D., Campbell L., Christensen M. B., Christophides G. K., Crouch K., Davis K., DeBarry J., Doherty R., Duan Y., Dunn M., Falke D., Fisher S., Flicek P., Fox B., Gajria B., Giraldo-Calderón G. I., Harb O. S., Harper E., Hertz-Fowler C., Hickman M. J., Howington C., Hu S., Humphrey J., Iodice J., Jones A., Judkins J., Kelly S. A., Kissinger J. C., Kwon D. K., Lamoureux K., Lawson D., Li W., Lies K., Lodha D., Long J., MacCallum R. M., Maslen G., McDowell M. A., Nabrzyski J., Roos D. S., Rund S. S. C., Schulman S. W., Shanmugasundram A., Sitnik V., Spruill D., Starns D., Stoeckert C. J. Jr, Tomko S. S., Wang H., Warrenfeltz S., Wieck R., Wilkinson P. A., Xu L. & Zheng J. 2022. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res.*, 50:D898–D911. doi:10.1093/nar/gkab929

Anaconda Software Distribution. 2016. Miniconda3. anaconda.com

Andrews S. 2010. FastQC. bioinformatics.babraham.ac.uk/projects/fastqc

Andrikou C., Gasiorowski L. & Hejnol A. 2021. Cell types, morphology, and evolution of animal excretory organs. *In*: Leys S. & Hejnol A. (eds.), Origin and Evolution of Metazoan Cell Types. 1st ed. Boca Raton, FL, CRC Press. p. 129–164. doi:10.1201/b21831

Arasaki K., Shimizu H., Mogari H., Nishida N., Hirota N., Furuno A., Kudo Y., Baba M., Baba N., Cheng J., Fujimoto T., Ishihara N., Ortiz-Sandoval C., Barlow L. D., Raturi A., Dohmae N., Wakana Y., Inoue H., Tani K., Dacks J. B., Simmen T. & Tagaya M. 2015. A role for

the ancient SNARE syntaxin 17 in regulating mitochondrial division. *Dev. Cell*, 32:304–317. doi:10.1016/j.devcel.2014.12.011

Arndt H., Cleven E.-J. & Mylnikov A. P. 2000. Functional diversity of heterotrophic flagellates in aquatic ecosystems. *In*: Leadbeater B. S. C. & Green J. C. (eds.), The Flagellates. London, Taylor & Francis. p. 240–268. doi:10.1201/9781482268225-18

Aury J.-M., Jaillon O., Duret L., Noel B., Jubin C., Porcel B. M., Ségurens B., Daubin V., Anthouard V., Aiach N., Arnaiz O., Billaut A., Beisson J., Blanc I., Bouhouche K., Câmara F., Duharcourt S., Guigo R., Gogendeau D., Katinka M., Keller A.-M., Kissmehl R., Klotz C., Koll F., Le Mouël A., Lepère G., Malinsky S., Nowacki M., Nowak J. K., Plattner H., Poulain J., Ruiz F., Serrano V., Zagulski M., Dessen P., Bétermier M., Weissenbach J., Scarpelli C., Schächter V., Sperling L., Meyer E., Cohen J. & Wincker P. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, 444:171–178. doi:10.1038/nature05230

Avidor-Reiss T., Maer A. M., Koundakjian E., Polyanovsky A., Keil T., Subramaniam S. & Zuker C. S. 2004. Decoding cilia function: Defining specialized genes required for compartmentalized cilia biogenesis. *Cell*, 117:527–539. doi:10.1016/S0092-8674(04)00412-X

Baile M. G., Guiney E. L., Sanford E. J., MacGurn J. A., Smolka M. B. & Emr S. D. 2019. Activity of a ubiquitin ligase adaptor is regulated by disordered insertions in its arrestin domain. *Mol. Biol. Cell*, 30:3057–3072. doi:10.1091/mbc.E19-08-0451

Banta L. M., Robinson J. S., Klionsky D. J. & Emr S. D. 1988. Organelle assembly in yeast: characterization of yeast mutants defective in vacuolar biogenesis and protein sorting. *J. Cell Biol.*, 107:1369–1383. doi:10.1083/jcb.107.4.1369

Barlow L. D. 2019. Evolution of the eukaryotic membrane trafficking system as revealed by comparative genomic and phylogenetic analysis of adaptin, golgin, and SNARE proteins. Edmonton, Alberta, Canada, University of Alberta.

Barlow L. D., Maciejowski W., More K., Terry K., Vargová R., Záhonová K. & Dacks J. B. 2023. Comparative genomics for evolutionary cell biology using AMOEBAE: Understanding the Golgi and beyond. *In*: Wang Y., Lupashin V. V. & Graham T. R. (eds.), Golgi: Methods and Protocols. Vol. 2557. Methods in Molecular Biology. New York, NY, Springer US. p. 431–452. doi:10.1007/978-1-0716-2639-9_26

Barr F. & Lambright D. G. 2010. Rab GEFs and GAPs. *Curr. Opin. Cell Biol.*, 22:461–470. doi:10.1016/j.ceb.2010.04.007

Beck R., Ravet M., Wieland F. T. & Cassel D. 2009. The COPI system: Molecular mechanisms and function. *FEBS Lett.*, 583:2701–2709. doi:10.1016/j.febslet.2009.07.032

Benos D. J. & Prusch R. D. 1972. Osmoregulation in fresh-water *Hydra*. *Comp. Biochem. Physiol. A Physiol.*, 43:165–171. doi:10.1016/0300-9629(72)90478-1

Billington K., Halliday C., Madden R., Dyer P., Carrington M., Vaughan S., Hertz-Fowler C., Dean S., Sunter J. D., Wheeler R. J. & Gull K. 2022. Genome-wide subcellular protein localisation in the flagellate parasite *Trypanosoma brucei*. :2022.06.09.495287. doi:10.1101/2022.06.09.495287

Blum M., Chang H.-Y., Chuguransky S., Grego T., Kandasaamy S., Mitchell A., Nuka G., Paysan-Lafosse T., Qureshi M., Raj S., Richardson L., Salazar G. A., Williams L., Bork P., Bridge A., Gough J., Haft D. H., Letunic I., Marchler-Bauer A., Mi H., Natale D. A., Necci M., Orengo C. A., Pandurangan A. P., Rivoire C., Sigrist C. J. A., Sillitoe I., Thanki N., Thomas P. D., Tosatto S. C. E., Wu C. H., Bateman A. & Finn R. D. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.*, 49:D344–D354. doi:10.1093/nar/gkaa977

Boehm C. & Field M. C. 2019. Evolution of late steps in exocytosis: conservation and specialization of the exocyst complex [version 2; peer review: 3 approved]. *Wellcome Open Res.*, 4. doi:10.12688/wellcomeopenres.15142.2

Boehm C. M., Obado S., Gadelha C., Kaupisch A., Manna P. T., Gould G. W., Munson M., Chait B. T., Rout M. P. & Field M. C. 2017. The trypanosome exocyst: A conserved structure revealing a new role in endocytosis. *PLOS*

Pathog., 13:e1006063. doi:10.1371/journal.ppat.1006063

Bolger A. M., Lohse M. & Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30:2114–2120. doi:10.1093/bioinformatics/btu170

Bone N., Millar J. B. A., Toda T. & Armstrong J. 1998. Regulated vacuole fusion and fission in *Schizosaccharomyces pombe*: an osmotic response dependent on MAP kinases. *Curr. Biol.*, 8:135–144. doi:10.1016/S0960-9822(98)00060-8

Bonifacino J. S. & Glick B. S. 2004. The mechanisms of vesicle budding and fusion. *Cell*, 116:153–166. doi:10.1016/S0092-8674(03)01079-1

Bonifacino J. S. & Hierro A. 2011. Transport according to GARP: Receiving retrograde cargo at the *trans*-Golgi network. *Trends Cell Biol.*, 21:159–167. doi:10.1016/j.tcb.2010.11.003

Bonifacino J. S. & Rojas R. 2006. Retrograde transport from endosomes to the *trans*-Golgi network. *Nat. Rev. Mol. Cell Biol.*, 7:568–579. doi:10.1038/nrm1985

Bonin P., Groisillier A., Raimbault A., Guibert A., Boyen C. & Tonon T. 2015. Molecular and biochemical characterization of mannitol-1-phosphate dehydrogenase from the model brown alga *Ectocarpus* sp. *Phytochemistry*, 117:509–520. doi:10.1016/j.phytochem.2015.07.015

Brauer ElizabethB. & McKanna JamesA. 1978. Contractile vacuoles in cells of a fresh water sponge, *Spongilla lacustris*. *Cell Tissue Res.*, 192. doi:10.1007/BF00220748

Bray N. L., Pimentel H., Melsted P. & Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, 34:525–527. doi:10.1038/nbt.3519

Bröcker C., Engelbrecht-Vandré S. & Ungermann C. 2010. Multisubunit tethering complexes and their role in membrane fusion. *Curr. Biol.*, 20:R943–R952. doi:10.1016/j.cub.2010.09.015

Brown M. W., Heiss A. A., Kamikawa R., Inagaki Y., Yabuki A., Tice A. K., Shiratori T., Ishida K.-I., Hashimoto T., Simpson A. G. B. & Roger A. J. 2018. Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. *Genome Biol. Evol.*, 10:427–433. doi:10.1093/gbe/evy014

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

Brown S. & De Jonckheere J. F. 1999. A reevaluation of the amoeba genus *Vahlkampfia* based on SSUrDNA sequences. *Eur. J. Protistol.*, 35:49–54. doi:10.1016/S0932-4739(99)80021-2

Buchfink B., Xie C. & Huson D. H. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, 12:59–60. doi:10.1038/nmeth.3176

Bultema J. J., Ambrosio A. L., Burek C. L. & Di Pietro S. M. 2012. BLOC-2, AP-3, and AP-1 proteins function in concert with Rab38 and Rab32 proteins to mediate protein trafficking to lysosome-related organelles. *J. Biol. Chem.*, 287:19550–19563. doi:10.1074/jbc.M112.351908

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. *Dev. Cell*, 18:425–436. doi:10.1016/j.devcel.2010.01.015

Burki F., Kaplan M., Tikhonenkov D. V., Zlatogursky V., Minh B. Q., Radaykina L. V., Smirnov A., Mylnikov A. P. & Keeling P. J. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B Biol. Sci.*, 283:20152802. doi:10.1098/rspb.2015.2802

Burki F., Roger A. J., Brown M. W. & Simpson A. G. B. 2020. The new tree of eukaryotes. *Trends Ecol. Evol.*, 35:43–55. doi:10.1016/j.tree.2019.08.008

Bushmanova E., Antipov D., Lapidus A. & Prjibelski A. D. 2019. rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. *GigaScience*, 8:giz100. doi:10.1093/gigascience/giz100

Bushnell B. 2020. BBMap. sourceforge.net/projects/bbmap/

Cai H., Reinisch K. & Ferro-Novick S. 2007. Coats, tethers, Rabs, and SNAREs work together to mediate the intracellular destination of a transport vesicle. *Dev. Cell*, 12:671–682. doi:10.1016/j.devcel.2007.04.005

del Campo J., Sieracki M. E., Molestina R., Keeling P., Massana R. & Ruiz-Trillo I. 2014. The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.*, 29:252–259. doi:10.1016/j.tree.2014.03.006

Cavalier-Smith T. Y. 2003. 2003. The excavate protozoan phyla Metamonada Grassé emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): Their evolutionary affinities and new higher taxa. *Int. J. Syst. Evol. Microbiol.*, 53:1741–1758. doi:10.1099/ijs.0.02548-0

Chaineau M., Danglot L. & Galli T. 2009. Multiple roles of the vesicular-SNARE TI-VAMP in post-Golgi and endosomal trafficking. *FEBS Lett.*, 583:3817–3826. doi:10.1016/j.febslet.2009.10.026

Chang J., Lee S. & Blackstone C. 2014. Spastic paraplegia proteins spastizin and spatacsin mediate autophagic lysosome reformation. *J. Clin. Invest.*, 124:5249–5262. doi:10.1172/JCI77598

Cheng X., Chen K., Dong B., Yang M., Filbrun S. L., Myoung Y., Huang T.-X., Gu Y., Wang G. & Fang N. 2021. Dynamin-dependent vesicle twist at the final stage of clathrin-mediated endocytosis. *Nat. Cell Biol.*, 23:859–869. doi:10.1038/s41556-021-00713-x

Choi S., Tamaki T., Ebine K., Uemura T., Ueda T. & Nakano A. 2013. RABA members act in distinct steps of subcellular trafficking of the FLAGELLIN SENSING2 receptor. *Plant Cell*, 25:1174–1187. doi:10.1105/tpc.112.108803

Clarke M., Köhler J., Arana Q., Liu T., Heuser J. & Gerisch G. 2002. Dynamics of the vacuolar H^+-ATPase in the contractile vacuole complex and the endosomal pathway of *Dictyostelium* cells. *J. Cell Sci.*, 115:2893–2905.

Conway T. C. & Bromage A. J. 2011. Succinct data structures for assembling large genomes. *Bioinformatics*, 27:479–486. doi:10.1093/bioinformatics/btq697

Cosgrove W. B. & Kessel R. G. 1958. The activity of the contractile vacuole of *Crithidia fasciculata*. *J. Protozool.*, 5:296–298. doi:10.1111/j.1550-7408.1958.tb02569.x

Cowie P. R. & Hannah F. 2006. Responses of four isolates of marine naked amoebae to reductions in salinity. *J. Exp. Mar. Biol. Ecol.*, 337:196–204. doi:10.1016/j.jembe.2006.06.031

Cronkite D. L., Neuman J., Walker D. & Pierce S. K. 1991. The response of contractile and non-contractile vacuoles of *Paramecium calkinsi* to widely varying salinities. *J. Protozool.*, 38:565–573. doi:10.1111/j.1550-7408.1991.tb06081.x

Dacks J. B. & Doolittle W. F. 2004. Molecular and phylogenetic characterization of syntaxin genes from parasitic protozoa. *Mol. Biochem. Parasitol.*, 136:123–136. doi:10.1016/j.molbiopara.2004.02.014

Dacks J. B. & Field M. C. 2007. Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J. Cell Sci.*, 120:2977–2985. doi:10.1242/jcs.013250

Dacks J. B., Poon P. P. & Field M. C. 2008. Phylogeny of endocytic components yields insight into the process of nonendosymbiotic organelle evolution. *Proc. Natl. Acad. Sci.*, 105:588–593. doi:10.1073/pnas.0707318105

Dacks J. B. & Robinson M. S. 2017. Outerwear through the ages: Evolutionary cell biology of vesicle coats. *Curr. Opin. Cell Biol.*, 47:108–116. doi:10.1016/j.ceb.2017.04.001

Davidson N. M. & Oshlack A. 2014. Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes. *Genome Biol.*, 15:410. doi:10.1186/s13059-014-0410-6

Delevoye C., Marks M. S. & Raposo G. 2019. Lysosome-related organelles as functional adaptations of the endolysosomal system. *Curr. Opin. Cell Biol.*, 59:147–158. doi:10.1016/j.ceb.2019.05.003

Derelle R., López-García P., Timpano H. & Moreira D. 2016. A phylogenomic framework to study the diversity and evolution of Stramenopiles (=Heterokonts). *Mol. Biol. Evol.*, 33:2890–2898. doi:10.1093/molbev/msw168

Dewees S. I., Vargová R., Hardin K. R., Turn R. E., Devi S., Linnert J., Wolfrum U., Caspary T., Eliáš M. & Kahn R. A. 2022. Phylogenetic profiling and cellular analyses of ARL16 reveal roles in traffic of IFT140 and INPP5E. *Mol. Biol. Cell*, 33:ar33. doi:10.1091/mbc.E21-10-0509-T

Docampo R., Jimenez V., Lander N., Li Z.-H. & Niyogi S. 2013. New insights into roles of acidocalcisomes and contractile vacuole complex in osmoregulation in protists. *In*: Jeon K. W. (ed.), International Review of Cell and Molecular Biology. Vol. 305. Academic Press. p. 69–113. doi:10.1016/B978-0-12-407695-2.00002-0

Docampo R., Ulrich P. & Moreno S. N. J. 2010. Evolution of acidocalcisomes and their role in polyphosphate storage and osmoregulation in eukaryotic microbes. *Philos. Trans. R. Soc. B Biol. Sci.*, 365:775–784. doi:10.1098/rstb.2009.0179

Du F., Edwards K., Shen Z., Sun B., De Lozanne A., Briggs S. & Firtel R. A. 2008. Regulation of contractile vacuole formation and activity in *Dictyostelium*. *EMBO J.*, 27:2064–2076. doi:10.1038/emboj.2008.131

Durai D. A. & Schulz M. H. 2016. Informed *k*mer selection for *de novo* transcriptome assembly. *Bioinformatics*, 32:1670–1677. doi:10.1093/bioinformatics/btw217

Eddy S. R. 2011. Accelerated profile HMM searches. *PLoS Comput. Biol.*, 7:e1002195. doi:10.1371/journal.pcbi.1002195

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., Gaudet P., Fey P., Pilcher K., Chen G., Saunders D., Sodergren E., Davis P., Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N., Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C., Cooper J., Haydock S., Driessche N. van, Cronin A., Goodhead I., Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R., Hauser H., James K., Quiles M., Babu M. M., Saito T., Buchrieser C., Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A., Loulseged H., Mungall K., Oliver K., Price C., Quail M. A., Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M., Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A., Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y., Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E. C., Chisholm R. L., Gibbs R., Loomis W. F., Platzer M., Kay R. R., Williams J., Dear P. H., Noegel A. A., Barrell B. & Kuspa A. 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature*, 435:43–57. doi:10.1038/nature03481

El Kasmi F., Krause C., Hiller U., Stierhof Y.-D., Mayer U., Conner L., Kong L., Reichardt I., Sanderfoot A. A. & Jürgens G. 2013. SNARE complexes of different composition jointly mediate membrane fusion in *Arabidopsis* cytokinesis. *Mol. Biol. Cell*, 24:1593–1601. doi:10.1091/mbc.E13-02-0074

Elias M. & Archibald J. M. 2009. The RJL family of small GTPases is an ancient eukaryotic invention probably functionally associated with the flagellar apparatus. *Gene*, 442:63–72. doi:10.1016/j.gene.2009.04.011

Eliáš 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. *J. Cell Sci.*, 125:2500–2508. doi:10.1242/jcs.101378

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., Bontempi E., Bringaud F., Burton P., Cadag E., Campbell D. A., Carrington M., Crabtree J., Darban H., Silveira J. F. da, Jong P. de, Edwards K., Englund P. T., Fazelina G., Feldblyum T., Ferella M., Frasch A. C., Gull K., Horn D., Hou L., Huang Y., Kindlund E., Klingbeil M., Kluge S., Koo H., Lacerda D., Levin M. J., Lorenzi H., Louie T., Machado C. R., McCulloch R., McKenna A., Mizuno Y., Mottram J. C., Nelson S., Ochaya S., Osoegawa K., Pai G., Parsons M., Pentony M., Pettersson U., Pop M., Ramirez J. L., Rinta J., Robertson L.,

Salzberg S. L., Sanchez D. O., Seyler A., Sharma R., Shetty J., Simpson A. J., Sisk E., Tammi M. T., Tarleton R., Teixeira S., Aken S. V., Vogt C., Ward P. N., Wickstead B., Wortman J., White O., Fraser C. M., Stuart K. D. & Andersson B. 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas Disease. *Science*, 309:409–415. doi:10.1126/science.1112631

Emms D. M. & Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, 16:157. doi:10.1186/s13059-015-0721-2

Emms D. M. & Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.*, 20:238. doi:10.1186/s13059-019-1832-y

Essid M., Gopaldass N., Yoshida K., Merrifield C. & Soldati T. 2012. Rab8a regulates the exocyst-mediated kiss-and-run discharge of the *Dictyostelium* contractile vacuole. *Mol. Biol. Cell*, 23:1267–1282. doi:10.1091/mbc.e11-06-0576

Faktorová D., Kaur B., Valach M., Graf L., Benz C., Burger G. & Lukeš J. 2020a. Targeted integration by homologous recombination enables in situ tagging and replacement of genes in the marine microeukaryote *Diplonema papillatum*. *Environ. Microbiol.*, 22:3660–3670. doi:10.1111/1462-2920.15130

Faktorová D., Nisbet R. E. R., Robledo J. A. F., Casacuberta E., Sudek L., Allen A. E., Ares M., Aresté C., Balestreri C., Barbrook A. C., Beardslee P., Bender S., Booth D. S., Bouget F.-Y., Bowler C., Breglia S. A., Brownlee C., Burger G., Cerutti H., Cesaroni R., Chiurillo M. A., Clemente T., Coles D. B., Collier J. L., Cooney E. C., Coyne K., Docampo R., Dupont C. L., Edgcomb V., Einarsson E., Elustondo P. A., Federici F., Freire-Beneitez V., Freyria N. J., Fukuda K., García P. A., Girguis P. R., Gomaa F., Gornik S. G., Guo J., Hampl V., Hanawa Y., Haro-Contreras E. R., Hehenberger E., Highfield A., Hirakawa Y., Hopes A., Howe C. J., Hu I., Ibañez J., Irwin N. A. T., Ishii Y., Janowicz N. E., Jones A. C., Kachale A., Fujimura-Kamada K., Kaur B., Kaye J. Z., Kazana E., Keeling P. J., King N., Klobutcher L. A., Lander N., Lassadi I., Li Z., Lin S., Lozano J.-C., Luan F., Maruyama S., Matute T., Miceli C., Minagawa J., Moosburner M., Najle S. R., Nanjappa D., Nimmo I. C., Noble L., Vanclová A. M. G. N., Nowacki M., Nuñez I., Pain A., Piersanti A., Pucciarelli S., Pyrih J., Rest J. S., Rius M., Robertson D., Ruaud A., Ruiz-Trillo I., Sigg M. A., Silver P. A., Slamovits C. H., Smith G. J., Sprecher B. N., Stern R., Swart E. C., Tsaousis A. D., Tsypin L., et al. 2020b. Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nat. Methods*, 17:481–494. doi:10.1038/s41592-020-0796-x

Falace A., Filipello F., La Padula V., Vanni N., Madia F., De Pietri Tonelli D., de Falco F. A., Striano P., Dagna Bricarelli F., Minetti C., Benfenati F., Fassio A. & Zara F. 2010. TBC1D24, an ARF6-interacting protein, is mutated in familial infantile myoclonic epilepsy. *Am. J. Hum. Genet.*, 87:365–370. doi:10.1016/j.ajhg.2010.07.020

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. *Proc. Natl. Acad. Sci.*, 95:15781–15786. doi:10.1073/pnas.95.26.15781

Field M. C., Sali A. & Rout M. P. 2011. On a bender—BARs, ESCRTs, COPs, and finally getting your coat. *J. Cell Biol.*, 193:963–972. doi:10.1083/jcb.201102042

Figarella K., Uzcategui N. L., Zhou Y., LeFurgey A., Ouellette M., Bhattacharjee H. & Mukhopadhyay R. 2007. Biochemical characterization of *Leishmania major* aquaglyceroporin LmAQP1: possible role in volume regulation and osmotaxis. *Mol. Microbiol.*, 65:1006–1017. doi:10.1111/j.1365-2958.2007.05845.x

Finn R. N. & Cerdà J. 2015. Evolution and functional diversity of aquaporins. *Biol. Bull.*, 229:6–23. doi:10.1086/BBLv229n1p6

Finn R. N., Chauvigné F., Hlidberg J. B., Cutler C. P. & Cerdà J. 2014. The lineage-specific evolution of aquaporin gene clusters facilitated tetrapod terrestrial adaptation. *PLOS ONE*, 9:e113686. doi:10.1371/journal.pone.0113686

Flavin M. & Nerad T. A. 1993. *Reclinomonas americana* n.g., n. sp., a new freshwater heterotrophic flagellate. *J. Eukaryot. Microbiol.*, 40:172–179. doi:10.1111/j.1550-7408.1993.tb04900.x

Flegontova O., Flegontov P., Malviya S., Audic S., Wincker P., de Vargas C., Bowler C., Lukeš J. & Horák A. 2016. Extreme diversity of diplonemid eukaryotes in the ocean. *Curr. Biol.*, 26:3060–3065. doi:10.1016/j.cub.2016.09.031

Fountain S. J., Parkinson K., Young M. T., Cao L., Thompson C. R. L. & North R. A. 2007. An intracellular P2X receptor required for osmoregulation in *Dictyostelium discoideum*. *Nature*, 448:200–203. doi:10.1038/nature05926

Freedman A. H., Clamp M. & Sackton T. B. 2021. Error, noise and bias in *de novo* transcriptome assemblies. *Mol. Ecol. Resour.*, 21:18–29. doi:10.1111/1755-0998.13156

Fritz-Laylin L. K., Prochnik S. E., Ginger M. L., Dacks J. B., Carpenter M. L., Field M. C., Kuo A., Paredez A., Chapman J., Pham J., Shu S., Neupane R., Cipriano M., Mancuso J., Tu H., Salamov A., Lindquist E., Shapiro H., Lucas S., Grigoriev I. V., Cande W. Z., Fulton C., Rokhsar D. S. & Dawson S. C. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell*, 140:631–642. doi:10.1016/j.cell.2010.01.032

Fu L., Niu B., Zhu Z., Wu S. & Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28:3150–3152. doi:10.1093/bioinformatics/bts565

Fulton C. 1972. Early events of cell differentiation in *Naegleria gruberi*. Synergistic control by electrolytes and a factor from yeast extract. *Dev. Biol.*, 28:603–619. doi:10.1016/0012-1606(72)90006-1

Fulton C. 1993. *Naegleria*: A research partner for cell and developmental biology. *J. Eukaryot. Microbiol.*, 40:520–532. doi:10.1111/j.1550-7408.1993.tb04945.x

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. *Mol. Biol. Cell*, 24:1574–1583. doi:10.1091/mbc.e12-07-0557

Galindo L. J., Torruella G., López-García P., Ciobanu M., Gutiérrez-Preciado A., Karpov S. A. & Moreira D. 2022. Phylogenomics supports the monophyly of aphelids and fungi and identifies new molecular synapomorphies. *Syst. Biol.*, :syac054. doi:10.1093/sysbio/syac054

Gamper M., Kim E., Howard P. K., Ma H., Hunter T. & Firtel R. A. 1999. Regulation of *Dictyostelium* protein-tyrosine phosphatase-3 (PTP3) through osmotic shock and stress stimulation and identification of pp130 as a PTP3 substrate. *J. Biol. Chem.*, 274:12129–12138. doi:10.1074/jbc.274.17.12129

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

Gerisch G., Heuser J. & Clarke M. 2002. Tubular-vesicular transformation in the contractile vacuole system of *Dictyostelium*. *Cell Biol. Int.*, 26:845–852. doi:10.1006/cbir.2002.0938

Ghimire B., Saraiva M., Andersen C. B., Gogoi A., Saleh M., Zic N., van West P. & Brurberg M. B. 2022. Transformation systems, gene silencing and gene editing technologies in oomycetes. *Fungal Biol. Rev.*, 40:37–52. doi:10.1016/j.fbr.2021.11.001

Glick B. S. & Luini A. 2011. Models for Golgi traffic: A critical assessment. *Cold Spring Harb. Perspect. Biol.*, 3:a005215. doi:10.1101/cshperspect.a005215

Gomes D., Agasse A., Thiébaud P., Delrot S., Gerós H. & Chaumont F. 2009. Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. *Biochim. Biophys. Acta BBA - Biomembr.*, 1788:1213–1228. doi:10.1016/j.bbamem.2009.03.009

Gouy M., Guindon S. & Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.*, 27:221–224. doi:10.1093/molbev/msp259

Grabherr M. G., Haas B. J., Yassour M., Levin J. Z., Thompson D. A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B. W., Nusbaum C., Lindblad-Toh K., Friedman N. & Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, 29:644–652. doi:10.1038/nbt.1883

Gracheva E. O., Burdina A. O., Touroutine D., Berthelot-Grosjean M., Parekh H. & Richmond J. E. 2007. Tomosyn negatively regulates both synaptic transmitter and neuropeptide release at the *C. elegans* neuromuscular junction. *J. Physiol.*, 585:705–709. doi:10.1113/jphysiol.2007.138321

Gray M. W., Burger G., Derelle R., Klimeš V., Leger M. M., Sarrasin M., Vlček Č., Roger A. J., Eliáš M. & Lang B. F. 2020. The draft nuclear genome sequence and predicted mitochondrial proteome of *Andalucia godoyi*, a protist with the most gene-rich and bacteria-like mitochondrial genome. *BMC Biol.*, 18:22. doi:10.1186/s12915-020-0741-6

Groisillier A., Shao Z., Michel G., Goulitquer S., Bonin P., Krahulec S., Nidetzky B., Duan D., Boyen C. & Tonon T. 2014. Mannitol metabolism in brown algae involves a new phosphatase family. *J. Exp. Bot.*, 65:559–570. doi:10.1093/jxb/ert405

Gubbels M.-J. & Duraisingh M. T. 2012. Evolution of apicomplexan secretory organelles. *Int. J. Parasitol.*, 42:1071–1081. doi:10.1016/j.ijpara.2012.09.009

Guillard R. R. L. 1960. A mutant of *Chlamydomonas moewusii* lacking contractile vacuoles. *J. Protozool.*, 7:262–268. doi:10.1111/j.1550-7408.1960.tb00740.x

Guo W., Roth D., Walch-Solimena C. & Novick P. 1999. The exocyst is an effector for Sec4p, targeting secretory vesicles to sites of exocytosis. *EMBO J.*, 18:1071–1080. doi:10.1093/emboj/18.4.1071

Haas B. J., Papanicolaou A., Yassour M., Grabherr M., Blood P. D., Bowden J., Couger M. B., Eccles D., Li B., Lieber M., MacManes M. D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C. N., Henschel R., LeDuc R. D., Friedman N. & Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.*, 8:1494–1512. doi:10.1038/nprot.2013.084

Hajduk S. L. 1972. Occurance of contractile vacuoles in trypanosomatids. *J. Protozool.*, 19(Supl.):37.

Halpern A. L. & Bruno W. J. 1998. Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. *Mol. Biol. Evol.*, 15:910–917. doi:10.1093/oxfordjournals.molbev.a025995

Hampl V., Hug L., Leigh J. W., Dacks J. B., Lang B. F., Simpson A. G. B. & Roger A. J. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." *Proc. Natl. Acad. Sci.*, 106:3859–3864. doi:10.1073/pnas.0807880106

Harris E., Yoshida K., Cardelli J. & Bush J. 2001. Rab11-like GTPase associates with and regulates the structure and function of the contractile vacuole system in *Dictyostelium*. *J. Cell Sci.*, 114:3035–3045. doi:10.1242/jcs.114.16.3035

Hart A. J., Ginzburg S., Xu M. (Sam), Fisher C. R., Rahmatpour N., Mitton J. B., Paul R. & Wegrzyn J. L. 2020. EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Mol. Ecol. Resour.*, 20:591–604. doi:10.1111/1755-0998.13106

Hashizume K., Cheng Y.-S., Hutton J. L., Chiu C. & Carr C. M. 2009. Yeast Sec1p functions before and after vesicle docking. *Mol. Biol. Cell*, 20:4673–4685. doi:10.1091/mbc.e09-02-0172

Hattendorf D. A., Andreeva A., Gangar A., Brennwald P. J. & Weis W. I. 2007. Structure of the yeast polarity protein Sro7 reveals a SNARE regulatory mechanism. *Nature*, 446:567–571. doi:10.1038/nature05635

Hauer G. & Rogerson A. 2005. Remarkable salinity tolerance of seven species of naked amoebae (gymnamoebae). *Hydrobiologia*, 549:33–42. doi:10.1007/s10750-005-2210-1

Hauer G., Rogerson A. & Anderson O. R. 2001. *Platyamoeba pseudovannellida* n. sp., a naked amoeba with wide salt tolerance isolated from the Salton Sea, California. *J. Eukaryot. Microbiol.*, 48:663–669. doi:10.1111/j.1550-7408.2001.tb00206.x

Hausmann K. 1978. Extrusive organelles in protists. *In*: Bourne G. H., Danielli J. F. & Jeon K. W. (eds.), International Review of Cytology. Vol. 52. Academic Press. p. 197–276. doi:10.1016/S0074-7696(08)60757-3

Hay J. C., Klumperman J., Oorschot V., Steegmaier M., Kuo C. S. & Scheller R. H. 1998. Localization, dynamics, and protein interactions reveal distinct roles for ER and Golgi SNAREs. *J. Cell Biol.*, 141:1489–1502. doi:10.1083/jcb.141.7.1489

Hehenberger E., Tikhonenkov D. V., Kolisko M., del Campo J., Esaulov A. S., Mylnikov A. P. & Keeling P. J. 2017. Novel predators reshape holozoan phylogeny and reveal the presence of a two-component signaling system in the ancestor of animals. *Curr. Biol.*, 27:2043-2050.e6. doi:10.1016/j.cub.2017.06.006

Hellebusi J. A. 1976. Osmoregulation. *Annu. Rev. Plant Physiol.*, 27:485–505. doi:10.1146/annurev.pp.27.060176.002413

Hellebust J. A., Mérida T. & Ahmad I. 1989. Operation of contractile vacuoles in the euryhaline green flagellate *Chlamydomonas pulsatilla* (Chlorophyceae) as a function of salinity. *Mar. Biol.*, 100:373–379. doi:10.1007/BF00391153

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*. *J. Cell Sci.*, 131:jcs213306. doi:10.1242/jcs.213306

Hirota Y. & Tanaka Y. 2009. A small GTPase, human Rab32, is required for the formation of autophagic vacuoles under basal conditions. *Cell. Mol. Life Sci.*, 66:2913–2932. doi:10.1007/s00018-009-0080-9

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 Biol.*, 9:e1001170. doi:10.1371/journal.pbio.1001170

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 Biol.*, 16:e2004411. doi:10.1371/journal.pbio.2004411

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

Hoef-Emden K. 2014. Osmotolerance in the Cryptophyceae: Jacks-of-all-trades in the *Chroomonas* clade. *Protist*, 165:123–143. doi:10.1016/j.protis.2014.01.001

Hogue M. J. 1923. Contractile vacuoles in amoebae—factors influencing their formation and rate of contraction. *Elisha Mitchell Sci. Soc.*, 39:49–55.

Hsieh P.-H., Oyang Y.-J. & Chen C.-Y. 2019. Effect of *de novo* transcriptome assembly on transcript quantification. *Sci. Rep.*, 9:8304. doi:10.1038/s41598-019-44499-3

Hu M., Zhou N., Cai W. & Xu H. 2022. Lysosomal solute and water transport. *J. Cell Biol.*, 221:e202109133. doi:10.1083/jcb.202109133

Huber L. A., Pimplikar S., Parton R. G., Virta H., Zerial M. & Simons K. 1993. Rab8, a small GTPase involved in vesicular traffic between the TGN and the basolateral plasma membrane. *J. Cell Biol.*, 123:35–45. doi:10.1083/jcb.123.1.35

Huotari J. & Helenius A. 2011. Endosome maturation. *EMBO J.*, 30:3481–3500. doi:10.1038/emboj.2011.286

Illumina, Inc. 2014. Understanding Illumina quality scores. illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431:931–945. doi:10.1038/nature03001

Ishida M., Fok A. K., Aihara M. S. & Allen R. D. 1996. Hyperosmotic stress leads to reversible dissociation of the proton pump-bearing tubules from the contractile vacuole complex in *Paramecium*. *J. Cell Sci.*, 109:229–237.

Ishida M., Hori M., Ooba Y., Kinoshita M., Matsutani T., Naito M., Hagimoto T., Miyazaki K., Ueda S., Miura K. & Tominaga T. 2021. A functional Aqp1 gene product localizes on the contractile vacuole complex in *Paramecium multimicronucleatum*. *J. Eukaryot. Microbiol.*, 68:e12843. doi:10.1111/jeu.12843

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. *Curr. Biol.*, 26:161–172. doi:10.1016/j.cub.2015.11.055

Jaiswal J. K., Andrews N. W. & Simon S. M. 2002. Membrane proximal lysosomes are the major vesicles responsible for calcium-dependent exocytosis in nonsecretory cells. *J. Cell Biol.*, 159:625–635. doi:10.1083/jcb.200208154

Jamy M., Biwer C., Vaulot D., Obiol A., Jing H., Peura S., Massana R. & Burki F. 2022. Global patterns and rates of habitat transitions across the eukaryotic tree of life. *Nat. Ecol. Evol.*, 6:1458–1470. doi:10.1038/s41559-022-01838-4

Janouškovec J., Tikhonenkov D. V., Burki F., Howe A. T., Rohwer F. L., Mylnikov A. P. & Keeling P. J. 2017. A new lineage of eukaryotes illuminates early mitochondrial genome reduction. *Curr. Biol.*, 27:3717-3724.e5. doi:10.1016/j.cub.2017.10.051

Jensen B. C., Ramasamy G., Vasconcelos E. J. R., Ingolia N. T., Myler P. J. & Parsons M. 2014. Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. *BMC Genomics*, 15:911. doi:10.1186/1471-2164-15-911

Jia D., Zhang J.-S., Li F., Wang J., Deng Z., White M. A., Osborne D. G., Phillips-Krawczak C., Gomez T. S., Li H., Singla A., Burstein E., Billadeau D. D. & Rosen M. K. 2016. Structural and mechanistic insights into regulation of the retromer coat by TBC1d5. *Nat. Commun.*, 7:13305. doi:10.1038/ncomms13305

Jones P., Binns D., Chang H.-Y., Fraser M., Li W., McAnulla C., McWilliam H., Maslen J., Mitchell A., Nuka G., Pesseat S., Quinn A. F., Sangrador-Vegas A., Scheremetjew M., Yong S.-Y., Lopez R. & Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30:1236–1240. doi:10.1093/bioinformatics/btu031

Kalyaanamoorthy S., Minh B. Q., Wong T. K. F., Haeseler A. von & Jermiin L. S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*, 14:587–589. doi:10.1038/nmeth.4285

Kassambara A. 2020. ggpubr: "ggplot2" Based Publication Ready Plots.

Katoh K. & Standley D. M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.*, 30:772–780. doi:10.1093/molbev/mst010

Kaur H., Richardson E., Kamra K. & Dacks J. B. 2022. Molecular evolutionary analysis of the SM and SNARE vesicle fusion machinery in ciliates shows concurrent expansions in late secretory machinery. *J. Eukaryot. Microbiol.*, 69:e12919. doi:10.1111/jeu.12919

Kawachi M., Nakayama T., Kayama M., Nomura M., Miyashita H., Bojo O., Rhodes L., Sym S., Pienaar R. N., Probert I., Inouye I. & Kamikawa R. 2021. Rappemonads are haptophyte phytoplankton. *Curr. Biol.*, 31:2395-2403.e4. doi:10.1016/j.cub.2021.03.012

Khundadze M., Ribaudo F., Hussain A., Rosentreter J., Nietzsche S., Thelen M., Winter D., Hoffmann B., Afzal M. A., Hermann T., de Heus C., Piskor E.-M., Kosan C., Franzka P., von Kleist L., Stauber T., Klumperman J., Damme M., Proikas-Cezanne T. & Hübner C. A. 2019. A mouse model for SPG48 reveals a block of autophagic flux upon disruption of adaptor protein complex five. *Neurobiol. Dis.*, 127:419–431. doi:10.1016/j.nbd.2019.03.026

Kienle N., Kloepper T. H. & Fasshauer D. 2009. Phylogeny of the SNARE vesicle fusion machinery yields insights into the conservation of the secretory pathway in fungi. *BMC Evol. Biol.*, 9:19. doi:10.1186/1471-2148-9-19

Kishino H., Miyata T. & Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.*, 31:151–160. doi:10.1007/BF02109483

Kissmehl R., Schilde C., Wassmer T., Danzer C., Nuehse K., Lutter K. & Plattner H. 2007. Molecular identification of 26 syntaxin genes and their assignment to the different trafficking pathways in *Paramecium*. *Traffic*, 8:523–542. doi:10.1111/j.1600-0854.2007.00544.x

Kitching J. A. 1934. The physiology of contractile vacuoles: I. Osmotic relations. *J. Exp. Biol.*, 11:264–381.

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. *Biochim. Biophys. Acta BBA - Mol. Cell Res.*, 1865:1397–1409. doi:10.1016/j.bbamcr.2018.07.009

Klinger C. M., Jimenez-Ruiz E., Mourier T., Klingl A., Lemgruber L., Pain A., Dacks J. B. & Meissner M. 2022. Evolution of lineage-specific trafficking proteins and a novel post-Golgi trafficking pathway in Apicomplexa. :2022.12.12.520010. doi:10.1101/2022.12.12.520010

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

Kloepper T. H., Kienle C. N. & Fasshauer D. 2008. SNAREing the basis of multicellularity: Consequences of protein family expansion during evolution. *Mol. Biol. Evol.*, 25:2055–2068. doi:10.1093/molbev/msn151

Klumperman J. & Raposo G. 2014. The complex ultrastructure of the endolysosomal system. *Cold Spring Harb. Perspect. Biol.*, 6:a016857. doi:10.1101/cshperspect.a016857

Komsic-Buchmann K. & Becker B. 2012. Contractile vacuoles in green algae – structure and function. *In*: Heimann K. & Katsaros C. (eds.), Advances in Algal Cell Biology. Berlin, Boston, DE GRUYTER. doi:10.1515/9783110229615.123

Komsic-Buchmann K., Stephan L. M. & Becker B. 2012. The SEC6 protein is required for contractile vacuole function in *Chlamydomonas reinhardtii*. *J. Cell Sci.*, 125:2885–2895. doi:10.1242/jcs.099184

Komsic-Buchmann K., Wöstehoff L. & Becker B. 2014. The contractile vacuole as a key regulator of cellular water flow in *Chlamydomonas reinhardtii*. *Eukaryot. Cell*, 13:1421–1430. doi:10.1128/EC.00163-14

Kondratiuk I., Jakhanwal S., Jin J., Sathyanarayanan U., Kroppen B., Pobbati A. V., Krisko A., Ashery U., Meinecke M., Jahn R., Fasshauer D. & Milosevic I. 2020. PI(4,5)P2-dependent regulation of exocytosis by amisyn, the vertebrate-specific competitor of synaptobrevin 2. *Proc. Natl. Acad. Sci.*, 117:13468–13479. doi:10.1073/pnas.1908232117

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

Koumandou V. L., Wickstead B., Ginger M. L., van der Giezen M., Dacks J. B. & Field M. C. 2013. Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.*, 48:373–396. doi:10.3109/10409238.2013.821444

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 Pathog.*, 9:e1003213. doi:10.1371/journal.ppat.1003213

Kurtzer G. M., Sochat V. & Bauer M. W. 2017. Singularity: Scientific containers for mobility of compute. *PLOS ONE*, 12:e0177459. doi:10.1371/journal.pone.0177459

Lahr D. J. G., Grant J., Nguyen T., Lin J. H. & Katz L. A. 2011. Comprehensive phylogenetic reconstruction of Amoebozoa based on concatenated analyses of SSU-rDNA and actin genes. *PLOS ONE*, 6:e22780. doi:10.1371/journal.pone.0022780

Latterich M. & Watson M. D. 1993. Evidence for a dual osmoregulatory mechanism in the yeast *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.*, 191:1111–1117. doi:10.1006/bbrc.1993.1331

Lax G., Eglit Y., Eme L., Bertrand E. M., Roger A. J. & Simpson A. G. B. 2018. Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature*, 564:410–414. doi:10.1038/s41586-018-0708-8

Le S. Q., Gascuel O. & Lartillot N. 2008. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*, 24:2317–2323. doi:10.1093/bioinformatics/btn445

Lee J. H. & Walsh C. J. 1988. Transcriptional regulation of coordinate changes in flagellar mRNAs during differentiation of *Naegleria gruberi* amebae into flagellates. *Mol. Cell. Biol.*, 8:2280–2287. doi:10.1128/mcb.8.6.2280-2287.1988

Levy Karin E., Mirdita M. & Söding J. 2020. MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome*, 8:48. doi:10.1186/s40168-020-00808-x

Li B., Li Yanbin, Liu F., Tan X., Rui Q., Tong Y., Qiao L., Gao R., Li G., Shi R., Li Yan & Bao Y. 2019. Overexpressed tomosyn binds syntaxins and blocks secretion during pollen development. *Plant Physiol.*, 181:1114–1126. doi:10.1104/pp.19.00965

Li P., Hu M., Wang C., Feng X., Zhao Z., Yang Ying, Sahoo N., Gu M., Yang Yexin, Xiao S., Sah R., Cover T. L., Chou J., Geha R., Benavides F., Hume R. I. & Xu H. 2020. LRRC8 family proteins within lysosomes regulate cellular osmoregulation and enhance cell survival to multiple physiological stresses. *Proc. Natl. Acad. Sci.*, 117:29155–29165. doi:10.1073/pnas.2016539117

Li S. C., Diakov T. T., Rizzo J. M. & Kane P. M. 2012. Vacuolar H^+-ATPase works in parallel with the HOG pathway to adapt *Saccharomyces cerevisiae* cells to osmotic stress. *Eukaryot. Cell*, 11:282–291. doi:10.1128/EC.05198-11

Li W. & Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22:1658–1659. doi:10.1093/bioinformatics/btl158

Liberator P., Anderson J., Feiglin M., Sardana M., Griffin P., Schmatz D. & Myers R. W. 1998. Molecular cloning and functional expression of mannitol-1-phosphatase from the apicomplexan parasite *Eimeria tenella*. *J. Biol. Chem.*, 273:4237–4244. doi:10.1074/jbc.273.7.4237

Linder J. C. & Staehelin L. A. 1979. A novel model for fluid secretion by the trypanosomatid contractile vacuole apparatus. *J. Cell Biol.*, 83:371–382. doi:10.1083/jcb.83.2.371

Ling H. U. & Tyler P. A. 1972. The process and morphology of conjugation in desmids, especially the genus *Pleurotaenium*. *Br. Phycol. J.*, 7:65–79. doi:10.1080/00071617200650091

Liu Y., Steenkamp E. T., Brinkmann H., Forget L., Philippe H. & Lang B. F. 2009. Phylogenomic analyses predict sistergroup relationship of nucleariids and Fungi and paraphyly of zygomycetes with significant support. *BMC Evol. Biol.*, 9:272. doi:10.1186/1471-2148-9-272

López-Hernández T., Puchkov D., Krause E., Maritzen T. & Haucke V. 2020. Endocytic regulation of cellular ion homeostasis controls lysosome biogenesis. *Nat. Cell Biol.*, 22:815–827. doi:10.1038/s41556-020-0535-7

Love M. I., Huber W. & Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15:550. doi:10.1186/s13059-014-0550-8

Lynn D. H. 1982. Dimensionality and contractile vacuole function in ciliated protozoa. *J. Exp. Zool.*, 223:219–229. doi:10.1002/jez.1402230304

MacManes M. 2014. On the optimal trimming of high-throughput mRNA sequence data. *Front. Genet.*, 5:13. doi:10.3389/fgene.2014.00013

MacManes M. D. 2018. The Oyster River Protocol: A multi-assembler and *k*mer approach for *de novo* transcriptome assembly. *PeerJ*, 6:e5428. doi:10.7717/peerj.5428

Manna P. T., Barlow L. D., Ramirez-Macias I., Herman E. K. & Dacks J. B. 2023. Endosomal vesicle fusion machinery is involved with the contractile vacuole in *Dictyostelium discoideum*. *J. Cell Sci.*, 136:jcs260477. doi:10.1242/jcs.260477

Manna P. T., Gadelha C., Puttick A. E. & Field M. C. 2015. ENTH and ANTH domain proteins participate in AP2-independent clathrin-mediated endocytosis. *J. Cell Sci.*, 128:2130–2142. doi:10.1242/jcs.167726

Manna P. T., Kelly S. & Field M. C. 2013. Adaptin evolution in kinetoplastids and emergence of

the variant surface glycoprotein coat in African trypanosomatids. *Mol. Phylogenet. Evol.*, 67:123–128. doi:10.1016/j.ympev.2013.01.002

Manni M., Berkeley M. R., Seppey M., Simão F. A. & Zdobnov E. M. 2021. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.*, 38:4647–4654. doi:10.1093/molbev/msab199

Mast F. D., Barlow L. D., Rachubinski R. A. & Dacks J. B. 2014. Evolutionary mechanisms for establishing eukaryotic cellular complexity. *Trends Cell Biol.*, 24:435–442. doi:10.1016/j.tcb.2014.02.003

Mathavarajah S., Meagan D. McLaren, & Robert J. Huber. 2018. Cln3 function is linked to osmoregulation in a *Dictyostelium* model of Batten disease. *BBA - Mol. Basis Dis.*, 1864:3559–3573. doi:10.1016/j.bbadis.2018.08.013

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. *Proc. Natl. Acad. Sci.*, 114:E10697–E10706. doi:10.1073/pnas.1717327114

McDermaid A., Monier B., Zhao J., Liu B. & Ma Q. 2019. Interpretation of differential gene expression results of RNA-seq data: review and integration. *Brief. Bioinform.*, 20:2044–2054. doi:10.1093/bib/bby067

McKanna J. A. 1973. Fine structure of the contractile vacuole pore in *Paramecium*. *J. Protozool.*, 20:631–638. doi:10.1111/j.1550-7408.1973.tb03587.x

McKanna J. A. 1976. Fine structure of fluid segregation organelles of *Paramecium* contractile vacuoles. *J. Ultrastruct. Res.*, 54:1–10. doi:10.1016/S0022-5320(76)80002-0

Merchant S. S., Prochnik S. E., Vallon O., Harris E. H., Karpowicz S. J., Witman G. B., Terry A., Salamov A., Fritz-Laylin L. K., Maréchal-Drouard L., Marshall W. F., Qu L.-H., Nelson D. R., Sanderfoot A. A., Spalding M. H., Kapitonov V. V., Ren Q., Ferris P., Lindquist E., Shapiro H., Lucas S. M., Grimwood J., Schmutz J., Cardol P., Cerutti H., Chanfreau G., Chen C.-L., Cognat V., Croft M. T., Dent R., Dutcher S., Fernández E., Fukuzawa H., González-Ballester D., González-Halphen D., Hallmann A., Hanikenne M., Hippler M., Inwood W., Jabbari K., Kalanon M., Kuras R., Lefebvre P. A., Lemaire S. D., Lobanov A. V., Lohr M., Manuell A., Meier I., Mets L., Mittag M., Mittelmeier T., Moroney J. V., Moseley J., Napoli C., Nedelcu A. M., Niyogi K., Novoselov S. V., Paulsen I. T., Pazour G., Purton S., Ral J.-P., Riaño-Pachón D. M., Riekhof W., Rymarquis L., Schroda M., Stern D., Umen J., Willows R., Wilson N., Zimmer S. L., Allmer J., Balk J., Bisova K., Chen C.-J., Elias M., Gendler K., Hauser C., Lamb M. R., Ledford H., Long J. C., Minagawa J., Page M. D., Pan J., Pootakham W., Roje S., Rose A., Stahlberg E., Terauchi A. M., Yang P., Ball S., Bowler C., Dieckmann C. L., Gladyshev V. N., Green P., Jorgensen R., Mayfield S., Mueller-Roeber B., Rajamani S., et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*, 318:245–250. doi:10.1126/science.1143609

Meyer S. R. & Pienaar R. N. 1984. The microanatomy of *Chroomonas africana* sp. nov. (Cryptophyceae). *South Afr. J. Bot.*, 3:306–319. doi:10.1016/S0022-4618(16)30019-5

Miller V. J. & Ungar D. 2012. Re'COG'nition at the Golgi. *Traffic*, 13:891–897. doi:10.1111/j.1600-0854.2012.01338.x

Minh B. Q., Schmidt H. A., Chernomor O., Schrempf D., Woodhams M. D., von Haeseler A. & Lanfear R. 2019. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* doi:10.1093/molbev/msaa015

Mitchell H. J. & Hardham A. R. 1999. Characterisation of the water expulsion vacuole in *Phytophthora nicotianae* zoospores. *Protoplasma*, 206:118–130. doi:10.1007/BF01279258

Mölder F., Jablonski K. P., Letcher B., Hall M. B., Tomkins-Tinch C. H., Sochat V., Forster J., Lee S., Twardziok S. O., Kanitz A., Wilm A., Holtgrewe M., Rahmann S., Nahnsen S. & Köster J. 2021. Sustainable data analysis with Snakemake. doi:10.12688/f1000research.29032.2

Monier B., McDermaid A., Zhao J. & Ma Q. 2022. vidger: Create rapid visualizations of RNAseq data in R.

Montalvetti A., Rohloff P. & Docampo R. 2004. A functional aquaporin co-localizes with the vacuolar proton pyrophosphatase to acidocalcisomes and the contractile vacuole complex of *Trypanosoma cruzi*. *J. Biol. Chem.*, 279:38673–38682. doi:10.1074/jbc.M406304200

More K., Klinger C. M., Barlow L. D. & Dacks J. B. 2020. Evolution and natural history of membrane trafficking in eukaryotes. *Curr. Biol.*, 30:R553–R564. doi:10.1016/j.cub.2020.03.068

More K., Simpson A. G. B. & Hess S. 2019. Two new marine species of *Placopus* (Vampyrellida, Rhizaria) that perforate the theca of *Tetraselmis* (Chlorodendrales, Viridiplantae). *J. Eukaryot. Microbiol.*, 66:560–573. doi:10.1111/jeu.12698

Moreau K., Ravikumar B., Renna M., Puri C. & Rubinsztein D. C. 2011. Autophagosome precursor maturation requires homotypic fusion. *Cell*, 146:303–317. doi:10.1016/j.cell.2011.06.023

Morgera F., Sallah M. R., Dubuke M. L., Gandhi P., Brewer D. N., Carr C. M. & Munson M. 2012. Regulation of exocytosis by the exocyst subunit Sec6 and the SM protein Sec1. *Mol. Biol. Cell*, 23:337–346. doi:10.1091/mbc.e11-08-0670

Na J., Tunggal B. & Eichinger L. 2007. STATc is a key regulator of the transcriptional response to hyperosmotic shock. *BMC Genomics*, 8:123. doi:10.1186/1471-2164-8-123

Newman A. P., Shim J. & Ferro-Novick S. 1990. BET1, BOS1, and SEC22 are members of a group of interacting yeast genes required for transport from the endoplasmic reticulum to the Golgi complex. *Mol. Cell. Biol.*, 10:3405–3414. doi:10.1128/mcb.10.7.3405-3414.1990

Nikolaev S. I., Berney C., Petrov N. B., Mylnikov A. P., Fahrni J. F. & Pawlowski J. 2006. 2006. Phylogenetic position of *Multicilia marina* and the evolution of Amoebozoa. *Int. J. Syst. Evol. Microbiol.*, 56:1449–1458. doi:10.1099/ijs.0.63763-0

Nishi T. & Forgac M. 2002. The vacuolar (H^+)-ATPases — Nature's most versatile proton pumps. *Nat. Rev. Mol. Cell Biol.*, 3:94–103. doi:10.1038/nrm729

Nishihara E., Shimmen T. & Sonobe S. 2007. New aspects of membrane dynamics of *Amoeba proteus* contractile vacuole revealed by vital staining with FM 4-64. *Protoplasma*, 231:25–30. doi:10.1007/s00709-007-0247-x

Nishihara E., Yokota E., Tazaki A., Orii H., Katsuhara M., Kataoka K., Igarashi H., Moriyama Y., Shimmen T. & Sonobe S. 2008. Presence of aquaporin and V-ATPase on the contractile vacuole of *Amoeba proteus*. *Biol. Cell*, 100:179–188. doi:10.1042/BC20070091

Niyogi S., Jimenez V., Girard-Dias W., Souza W. de, Miranda K. & Docampo R. 2015. Rab32 is essential for maintaining functional acidocalcisomes, and for growth and infectivity of *Trypanosoma cruzi*. *J. Cell Sci.*, 128:2363–2373. doi:10.1242/jcs.169466

Niyogi S., Mucci J., Campetella O. & Docampo R. 2014. Rab11 regulates trafficking of trans-sialidase to the plasma membrane through the contractile vacuole complex of *Trypanosoma cruzi*. *PLOS Pathog.*, 10:e1004224. doi:10.1371/journal.ppat.1004224

Okada Y., Hazama A., Hashimoto A., Maruyama Y. & Kubo M. 1992. Exocytosis upon osmotic swelling in human epithelial cells. *Biochim. Biophys. Acta BBA - Biomembr.*, 1107:201–205. doi:10.1016/0005-2736(92)90348-P

Okamoto N. & Keeling P. J. 2014. The 3D structure of the apical complex and association with the flagellar apparatus revealed by serial TEM tomography in *Psammosa pacifica*, a distant relative of the apicomplexa. *PLOS ONE*, 9:e84653. doi:10.1371/journal.pone.0084653

O'Kelly C. J. 1997. Ultrastructure of trophozoites, zoospores and cysts of *Reclinomonas americana* Flavin & Nerad, 1993 (Protista *incertae sedis*: Histionidae). *Eur. J. Protistol.*, 33:337–348. doi:10.1016/S0932-4739(97)80045-4

O'Leary N. A., Wright M. W., Brister J. R., Ciufo S., Haddad D., McVeigh R., Rajput B., Robbertse B., Smith-White B., Ako-Adjei D., Astashyn A., Badretdin A., Bao Y., Blinkova O., Brover V., Chetvernin V., Choi J., Cox E., Ermolaeva O., Farrell C. M., Goldfarb T., Gupta T., Haft D., Hatcher E., Hlavina W., Joardar V. S., Kodali V. K., Li W., Maglott D., Masterson P., McGarvey K. M., Murphy M. R., O'Neill K., Pujar S., Rangwala S. H., Rausch D., Riddick L. D., Schoch C., Shkeda A., Storz S.

S., Sun H., Thibaud-Nissen F., Tolstoy I., Tully R. E., Vatsan A. R., Wallin C., Webb D., Wu W., Landrum M. J., Kimchi A., Tatusova T., DiCuccio M., Kitts P., Murphy T. D. & Pruitt K. D. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, 44:D733–D745. doi:10.1093/nar/gkv1189

Ortiz-Sandoval C. G., Hughes S. C., Dacks J. B. & Simmen T. 2014. Interaction with the effector dynamin-related protein 1 (Drp1) is an ancient function of Rab32 subfamily proteins. *Cell. Logist.*, 4:e986399. doi:10.4161/21592799.2014.986399

Overath P. & Engstler M. 2004. Endocytosis, membrane recycling and sorting of GPI-anchored proteins: *Trypanosoma brucei* as a model system. *Mol. Microbiol.*, 53:735–744. doi:10.1111/j.1365-2958.2004.04224.x

Pal R. A. 1972. The osmoregulatory system of the amoeba, *Acanthamoeba castellanii*. *J. Exp. Biol.*, 57:55–76. doi:10.1242/jeb.57.1.55

Parkinson K., Baines A. E., Keller T., Gruenheit N., Bragg L., North R. A. & Thompson C. R. L. 2014. Calcium-dependent regulation of Rab activation and vesicle fusion by an intracellular P2X ion channel. *Nat. Cell Biol.*, 16:87–98. doi:10.1038/ncb2887

Patro R., Duggal G., Love M. I., Irizarry R. A. & Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, 14:417–419. doi:10.1038/nmeth.4197

Patterson D. J. 1980. Contractile vacuoles and associated structures: Their organization and function. *Biol. Rev.*, 55:1–46. doi:10.1111/j.1469-185X.1980.tb00686.x

Pell J., Hintze A., Canino-Koning R., Howe A., Tiedje J. M. & Brown C. T. 2012. Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. *Proc. Natl. Acad. Sci.*, 109:13272–13277. doi:10.1073/pnas.1121464109

Pittam M. D. 1963. Studies of an amoebo-flagellate, *Naegleria gruberi*. *Q. J. Microsc. Sci.*, 104:17.

Plattner H. 2013. Contractile vacuole complex—Its expanding protein inventory. *In*: International Review of Cell and Molecular Biology. Vol. 306. Elsevier. p. 371–416. doi:10.1016/B978-0-12-407694-5.00009-2

Pommerrenig B., Diehn T. A., Bernhardt N., Bienert M. D., Mitani-Ueno N., Fuge J., Bieber A., Spitzer C., Bräutigam A., Ma J. F., Chaumont F. & Bienert G. P. 2020. Functional evolution of nodulin 26-like intrinsic proteins: from bacterial arsenic detoxification to plant nutrient transport. *New Phytol.*, 225:1383–1396. doi:10.1111/nph.16217

Powell S., Forslund K., Szklarczyk D., Trachana K., Roth A., Huerta-Cepas J., Gabaldón T., Rattei T., Creevey C., Kuhn M., Jensen L. J., von Mering C. & Bork P. 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.*, 42:D231–D239. doi:10.1093/nar/gkt1253

Prokopchuk G., Korytář T., Juricová V., Majstorović J., Horák A., Šimek K. & Lukeš J. 2022. Trophic flexibility of marine diplonemids - switching from osmotrophy to bacterivory. *ISME J.*, 16:1409–1419. doi:10.1038/s41396-022-01192-0

Prusch R. D. 1977. Protozoan osmotic and ionic regulation. *In*: Gupta B. L., Moreton R. B., Oschman J. L. & Wall B. J. (eds.), Transport of Ions and Water in Animals. U.S. New York, NY, Academic Press Incorporated. p. 363–377.

R Core Team. 2022. R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Ramadas R. & Thattai M. 2013. New organelles by gene duplication in a biophysical model of eukaryote endomembrane evolution. *Biophys. J.*, 104:2553–2563. doi:10.1016/j.bpj.2013.03.066

Rambaut A. 2018. FigTree. tree.bio.ed.ac.uk/software/figtree

Rao S. K., Huynh C., Proux-Gillardeaux V., Galli T. & Andrews N. W. 2004. Identification of SNAREs involved in synaptotagmin VII-regulated lysosomal exocytosis. *J. Biol. Chem.*, 279:20471–20479. doi:10.1074/jbc.M400798200

Reddy A., Caler E. V. & Andrews N. W. 2001. Plasma membrane repair is mediated by Ca^{2+}-

regulated exocytosis of lysosomes. *Cell*, 106:157–169. doi:10.1016/S0092-8674(01)00421-4

Richardson E. & Dacks J. B. 2022. Distribution of membrane trafficking system components across ciliate diversity highlights heterogenous organelle-associated machinery. *Traffic*, 23:208–220. doi:10.1111/tra.12834

Richter D. J., Berney C., Strassert J. F. H., Poh Y.-P., Herman E. K., Muñoz-Gómez S. A., Wideman J. G., Burki F. & de Vargas C. 2022. EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Community J.*, 2:e56. doi:10.24072/pcjournal.173

Rifkin J. L. 1973. The role of the contractile vacuole in the osmoregulation of *Tetrahymena pyriformis*. *J. Protozool.*, 20:108–114. doi:10.1111/j.1550-7408.1973.tb06012.x

Risselada H. J. & Mayer A. 2020. SNAREs, tethers and SM proteins: how to overcome the final barriers to membrane fusion? *Biochem. J.*, 477:243–258. doi:10.1042/BCJ20190050

Robertson G., Schein J., Chiu R., Corbett R., Field M., Jackman S. D., Mungall K., Lee S., Okada H. M., Qian J. Q., Griffith M., Raymond A., Thiessen N., Cezard T., Butterfield Y. S., Newsome R., Chan S. K., She R., Varhol R., Kamoh B., Prabhu A.-L., Tam A., Zhao Y., Moore R. A., Hirst M., Marra M. A., Jones S. J. M., Hoodless P. A. & Birol I. 2010. *De novo* assembly and analysis of RNA-seq data. *Nat. Methods*, 7:909–912. doi:10.1038/nmeth.1517

Rohloff P. & Docampo R. 2008. A contractile vacuole complex is involved in osmoregulation in *Trypanosoma cruzi*. *Exp. Parasitol.*, 118:17–24. doi:10.1016/j.exppara.2007.04.013

Ronquist F., Teslenko M., van der Mark P., Ayres D. L., Darling A., Höhna S., Larget B., Liu L., Suchard M. A. & Huelsenbeck J. P. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.*, 61:539–542. doi:10.1093/sysbio/sys029

RStudio Team. 2020. RStudio: Integrated Development Environment for R. Boston, MA, RStudio, PBC.

Santana-Molina C., Gutierrez F. & Devos D. P. 2021. Homology and modular evolution of CATCHR at the origin of the eukaryotic endomembrane system. *Genome Biol. Evol.*, 13:evab125. doi:10.1093/gbe/evab125

Sant'Anna C., Nakayasu E. S., Pereira M. G., Lourenço D., de Souza W., Almeida I. C. & Cunha-e-Silva N. L. 2009. Subcellular proteomics of *Trypanosoma cruzi* reservosomes. *Proteomics*, 9:1782–1794. doi:10.1002/pmic.200800730

Sant'Anna C., Parussini F., Lourenço D., de Souza W., Cazzulo J. J. & Cunha-e-Silva N. L. 2008. All *Trypanosoma cruzi* developmental forms present lysosome-related organelles. *Histochem. Cell Biol.*, 130:1187–1198. doi:10.1007/s00418-008-0486-8

Schardt A., Brinkmann B. G., Mitkovski M., Sereda M. W., Werner H. B. & Nave K.-A. 2009. The SNARE protein SNAP-29 interacts with the GTPase Rab3A: Implications for membrane trafficking in myelinating glia. *J. Neurosci. Res.*, 87:3465–3479. doi:10.1002/jnr.22005

Schilde C., Wassmer T., Mansfeld J., Plattner H. & KissmehI R. 2006. A multigene family encoding R-SNAREs in the ciliate *Paramecium tetraurelia*. *Traffic*, 7:440–455. doi:10.1111/j.1600-0854.2006.00397.x

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. *Nat. Methods*, 9:676–682.

Schindler C., Chen Y., Pu J., Guo X. & Bonifacino J. S. 2015. EARP is a multisubunit tethering complex involved in endocytic recycling. *Nat. Cell Biol.*, 17:639–650. doi:10.1038/ncb3129

Schlacht A. & Dacks J. B. 2015. Unexpected ancient paralogs and an evolutionary model for the COPII coat complex. *Genome Biol. Evol.*, 7:1098–1109. doi:10.1093/gbe/evv045

Schlacht A., Herman E. K., Klute M. J., Field M. C. & Dacks J. B. 2014. Missing pieces of an ancient puzzle: Evolution of the eukaryotic membrane-trafficking system. *Cold Spring Harb. Perspect. Biol.*, 6:a016048. doi:10.1101/cshperspect.a016048

Schön M. E., Zlatogursky V. V., Singh R. P., Poirier C., Wilken S., Mathur V., Strassert J. F. H., Pinhassi J., Worden A. Z., Keeling P. J., Ettema T. J. G., Wideman J. G. & Burki F. 2021. Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. *Nat. Commun.*, 12:6651. doi:10.1038/s41467-021-26918-0

Schulz M. H., Zerbino D. R., Vingron M. & Birney E. 2012. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 28:1086–1092. doi:10.1093/bioinformatics/bts094

Schwartzbach S. D. 2017. Photo and nutritional regulation of *Euglena* organelle development. *In*: Schwartzbach S. D. & Shigeoka S. (eds.), Euglena: Biochemistry, Cell and Molecular Biology. Advances in Experimental Medicine and Biology. Cham, Springer International Publishing. p. 159–182. doi:10.1007/978-3-319-54910-1_9

Shields J. P. & Fuller M. S. 1996. Ultrastructure of chytridiomycete and oomycete zoospores using spray-freeze fixation. *Protoplasma*, 191:84–95. doi:10.1007/BF01280828

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.*, 51:492–508.

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*, 31:3210–3212. doi:10.1093/bioinformatics/btv351

Simpson A. G. B. 2003. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). *Int. J. Syst. Evol. Microbiol.*, 53:1759–1777. doi:10.1099/ijs.0.02578-0

Simpson A. G. B., Slamovits C. H. & Archibald J. M. 2017. Protist Diversity and Eukaryote Phylogeny. *In*: Archibald J. M., Simpson A. G. B. & Slamovits C. H. (eds.), Handbook of the Protists. Cham, Springer International Publishing. p. 1–21. doi:10.1007/978-3-319-28149-0_45

Slama I., Ghnaya T., Hessini K., Messedi D., Savouré A. & Abdelly C. 2007. Comparative study of the effects of mannitol and PEG osmotic stress on growth and solute accumulation in *Sesuvium portulacastrum*. *Environ. Exp. Bot.*, 61:10–17. doi:10.1016/j.envexpbot.2007.02.004

Smith-Unna R., Boursnell C., Patro R., Hibberd J. M. & Kelly S. 2016. TransRate: Reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Res.*, 26:1134–1144. doi:10.1101/gr.196469.115

Soares M. J., Souto-Padron T. & De Souza W. 1992. Identification of a large pre-lysosomal compartment in the pathogenic protozoon *Trypanosoma cruzi*. *J. Cell Sci.*, 102:157–167. doi:10.1242/jcs.102.1.157

Soneson C., Love M. I. & Robinson M. D. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. doi:10.12688/f1000research.7563.2

Song L. & Florea L. 2015. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*, 4:48. doi:10.1186/s13742-015-0089-y

Sönnichsen B., De Renzis S., Nielsen E., Rietdorf J. & Zerial M. 2000. Distinct membrane domains on endosomes in the recycling pathway visualized by multicolor imaging of Rab4, Rab5, and Rab11. *J. Cell Biol.*, 149:901–914. doi:10.1083/jcb.149.4.901

Spillane J. L., LaPolice T. M., MacManes M. D. & Plachetzki D. C. 2021. Signal, bias, and the role of transcriptome assembly quality in phylogenomic inference. *BMC Ecol. Evol.*, 21:43. doi:10.1186/s12862-021-01772-2

Stanke M., Keller O., Gunduz I., Hayes A., Waack S. & Morgenstern B. 2006. AUGUSTUS: *Ab initio* prediction of alternative transcripts. *Nucleic Acids Res.*, 34:W435–W439. doi:10.1093/nar/gkl200

Stavrou I. & O'Halloran T. J. 2006. The monomeric clathrin assembly protein, AP180, regulates contractile vacuole size in *Dictyostelium discoideum*. *Mol. Biol. Cell*, 17:5381–5389.

Stenmark H. 2009. Rab GTPases as coordinators of vesicle traffic. *Nat. Rev. Mol. Cell Biol.*, 10:513–525. doi:10.1038/nrm2728

Stock C., Allen R. D. & Naitoh Y. 2001. How external osmolarity affects the activity of the contractile vacuole complex, the cytosolic osmolarity and the water permeability of the plasma membrane in *Paramecium multimicronucleatum*. *J. Exp. Biol.*, 204:291–304.

Suescún-Bolívar L. P. & Thomé P. E. 2015. Osmosensing and osmoregulation in unicellular eukaryotes. *World J. Microbiol. Biotechnol.*, 31:435–443. doi:10.1007/s11274-015-1811-8

Takahashi S., Kubo K., Waguri S., Yabashi A., Shin H.-W., Katoh Y. & Nakayama K. 2012. Rab11 regulates exocytosis of recycling vesicles at the plasma membrane. *J. Cell Sci.*, 125:4049–4057. doi:10.1242/jcs.102913

Tang S., Lomsadze A. & Borodovsky M. 2015. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.*, 43:e78. doi:10.1093/nar/gkv227

Temesvari L. A., Rodriguez-Paris J. M., Bush J. M., Zhang L. & Cardelli J. A. 1996. Involvement of the vacuolar proton-translocating ATPase in multiple steps of the endo-lysosomal system and in the contractile vacuole system of *Dictyostelium discoideum*. *J. Cell Sci.*, 109:1479–1495. doi:10.1242/jcs.109.6.1479

Tesan F. C., Lorenzo R., Alleva K. & Fox A. R. 2021. AQPX-cluster aquaporins and aquaglyceroporins are asymmetrically distributed in trypanosomes. *Commun. Biol.*, 4:1–14. doi:10.1038/s42003-021-02472-9

The UniProt Consortium. 2022. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.*, :gkac1052. doi:10.1093/nar/gkac1052

Tikhonenkov D. V., Mikhailov K. V., Gawryluk R. M. R., Belyaev A. O., Mathur V., Karpov S. A., Zagumyonnyi D. G., Borodina A. S., Prokina K. I., Mylnikov A. P., Aleoshin V. V. & Keeling P. J. 2022. Microbial predators form a new supergroup of eukaryotes. *Nature*. doi:10.1038/s41586-022-05511-5

Tonon T., Li Y. & McQueen-Mason S. 2017. Mannitol biosynthesis in algae: More widespread and diverse than previously thought. *New Phytol.*, 213:1573–1579.

Tooze S. A., Abada A. & Elazar Z. 2014. Endocytosis and autophagy: Exploitation or cooperation? *Cold Spring Harb. Perspect. Biol.*, 6:a018358. doi:10.1101/cshperspect.a018358

Uemura T., Ueda T., Ohniwa R. L., Nakano A., Takeyasu K. & Sato M. H. 2004. Systematic analysis of SNARE molecules in *Arabidopsis*: Dissection of the post-Golgi network in plant cells. *Cell Struct. Funct.*, 29:49–65. doi:10.1247/csf.29.49

Ullrich O., Reinsch S., Urbé S., Zerial M. & Parton R. G. 1996. Rab11 regulates recycling through the pericentriolar recycling endosome. *J. Cell Biol.*, 135:913–924. doi:10.1083/jcb.135.4.913

Ulrich P. N., Jimenez V., Park M., Martins V. P., Atwood J., 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:e18013. doi:10.1371/journal.pone.0018013

Umaer K., Bush P. J. & Bangs J. D. 2018. Rab11 mediates selective recycling and endocytic trafficking in *Trypanosoma brucei*. *Traffic*, 19:406–420. doi:10.1111/tra.12565

Venkatesh D., Boehm C., Barlow L. D., Nankissoor N. N., O'Reilly A., Kelly S., Dacks J. B. & Field M. C. 2017. Evolution of the endomembrane systems of trypanosomatids – conservation and specialisation. *J. Cell Sci.*, 130:1421–1434. doi:10.1242/jcs.197640

Venkatesh D., Zhang N., Zoltner M., del Pino R. C. & Field M. C. 2018. Evolution of protein trafficking in kinetoplastid parasites: Complexity and pathogenesis. *Traffic*, 19:803–812. doi:10.1111/tra.12601

Vesteg M., Hadariová L., Horváth A., Estraño C. E., Schwartzbach S. D. & Krajčovič J. 2019. Comparative molecular cell biology of phototrophic euglenids and parasitic trypanosomatids sheds light on the ancestor of Euglenozoa. *Biol. Rev.*, 94:1701–1721. doi:10.1111/brv.12523

Vijay N., Poelstra J. W., Künstner A. & Wolf J. B. W. 2013. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Mol. Ecol.*, 22:620–634. doi:10.1111/mec.12014

Viotti C., Bubeck J., Stierhof Y.-D., Krebs M., Langhans M., van den Berg W., van Dongen W., Richter S., Geldner N., Takano J., Jürgens G., de Vries S. C., Robinson D. G. & Schumacher K. 2010. Endocytic and secretory traffic in *Arabidopsis* merge in the *trans*-Golgi network/early endosome, an independent and highly dynamic organelle. *Plant Cell*, 22:1344–1357. doi:10.1105/tpc.109.072637

Von Bülow J. & Beitz E. 2015. Number and regulation of protozoan aquaporins reflect environmental complexity. *Biol. Bull.*, 229:38–46. doi:10.1086/BBLv229n1p38

de Vries J. & Archibald J. M. 2018. Plant evolution: landmarks on the path to terrestrial life. *New Phytol.*, 217:1428–1434. doi:10.1111/nph.14975

Waller R. F., Cleves P. A., Rubio-Brotons M., Woods A., Bender S. J., Edgcomb V., Gann E. R., Jones A. C., Teytelman L., Dassow P. von, Wilhelm S. W. & Collier J. L. 2018. Strength in numbers: Collaborative science for new experimental model systems. *PLOS Biol.*, 16:e2006333. doi:10.1371/journal.pbio.2006333

Wang H.-C., Minh B. Q., Susko E. & Roger A. J. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.*, 67:216–235. doi:10.1093/sysbio/syx068

Wassmer T., Froissard M., Plattner H., Kissmehl R. & Cohen J. 2005. The vacuolar proton-ATPase plays a major role in several membrane-bounded organelles in *Paramecium*. *J. Cell Sci.*, 118:2813–2825. doi:10.1242/jcs.02405

Weiss R. L., Goodenough D. A. & Goodenough U. W. 1977. Membrane particle arrays associated with the basal body and with contractile vacuole secretion in *Chlamydomonas*. *J. Cell Biol.*, 72:133–143. doi:10.1083/jcb.72.1.133

Weiss S., Xu Z. Z., Peddada S., Amir A., Bittinger K., Gonzalez A., Lozupone C., Zaneveld J. R., Vázquez-Baeza Y., Birmingham A., Hyde E. R. & Knight R. 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5:27. doi:10.1186/s40168-017-0237-y

Welz T., Wellbourne-Wood J. & Kerkhoff E. 2014. Orchestration of cell surface proteins by Rab11. *Trends Cell Biol.*, 24:407–415. doi:10.1016/j.tcb.2014.02.004

Wen Y., Stavrou I., Bersuker K., Brady R. J., De Lozanne A. & O'Halloran T. J. 2009. AP180-mediated trafficking of Vamp7B limits homotypic fusion of *Dictyostelium* contractile vacuoles. *Mol. Biol. Cell*, 20:4278–4288. doi:10.1091/mbc.e09-03-0243

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wigg D., Bovee E. C. & Jahn T. L. 1967. The evacuation mechanism of the water expulsion vesicle ("contractile vacuole") of *Amoeba proteus*. *J. Protozool.*, 14:104–108. doi:10.1111/j.1550-7408.1967.tb01453.x

Wood D. E., Lu J. & Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol.*, 20:257. doi:10.1186/s13059-019-1891-0

Wood D. E. & Salzberg S. L. 2014. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.*, 15:R46. doi:10.1186/gb-2014-15-3-r46

Yabuki A., Nakayama T., Yubuki N., Hashimoto T., Ishida K.-I. & Inagaki Y. 2011. *Tsukubamonas globosa* n. gen., n. sp., a novel excavate flagellate possibly holding a key for the early evolution in "Discoba." *J. Eukaryot. Microbiol.*, 58:319–331. doi:10.1111/j.1550-7408.2011.00552.x

Yazaki E., Yabuki A., Imaizumi A., Kume K., Hashimoto T. & Inagaki Y. 2022. The closest lineage of Archaeplastida is revealed by phylogenomics analyses that include *Microheliella maris*. *Open Biol.*, 12:210376. doi:10.1098/rsob.210376

Ylä-Anttila P. & Eskelinen E.-L. 2018. Roles for RAB24 in autophagy and disease. *Small GTPases*, 9:57–65. doi:10.1080/21541248.2017.1317699

Ylä-Anttila P., Mikkonen E., Happonen K. E., Holland P., Ueno T., Simonsen A. & Eskelinen E.-L. 2015. RAB24 facilitates clearance of autophagic compartments during basal conditions. *Autophagy*, 11:1833–1848. doi:10.1080/15548627.2015.1086522

Yu I.-M. & Hughson F. M. 2010. Tethering factors as organizers of intracellular vesicular traffic. *Annu. Rev. Cell Dev. Biol.*, 26:137–156. doi:10.1146/annurev.cellbio.042308.113327

Zanchi R., Howard G., Bretscher M. S. & Kay R. R. 2010. The exocytic gene secA is required for *Dictyostelium* cell motility and osmoregulation. *J. Cell Sci.*, 123:3226–3234. doi:10.1242/jcs.072876

Zhang Y. & Hughson F. M. 2021. Chaperoning SNARE folding and assembly. *Annu. Rev. Biochem.*, 90:581–603. doi:10.1146/annurev-biochem-081820-103615

Zheng H., Bednarek S. Y., Sanderfoot A. A., Alonso J., Ecker J. R. & Raikhel N. V. 2002. NPSN11 Is a cell plate-associated SNARE protein that interacts with the syntaxin KNOLLE. *Plant Physiol.*, 129:530–539. doi:10.1104/pp.003970

Zhou H. & von Schwartzenberg K. 2020. Zygnematophyceae: From living algae collections to the establishment of future models. *J. Exp. Bot.*, 71:3296–3304. doi:10.1093/jxb/eraa091

Zimmermann U. 1978. Physics of turgor- and osmoregulation. *Annu. Rev. Plant Physiol.*, 29:121–148. doi:10.1146/annurev.pp.29.060178.001005

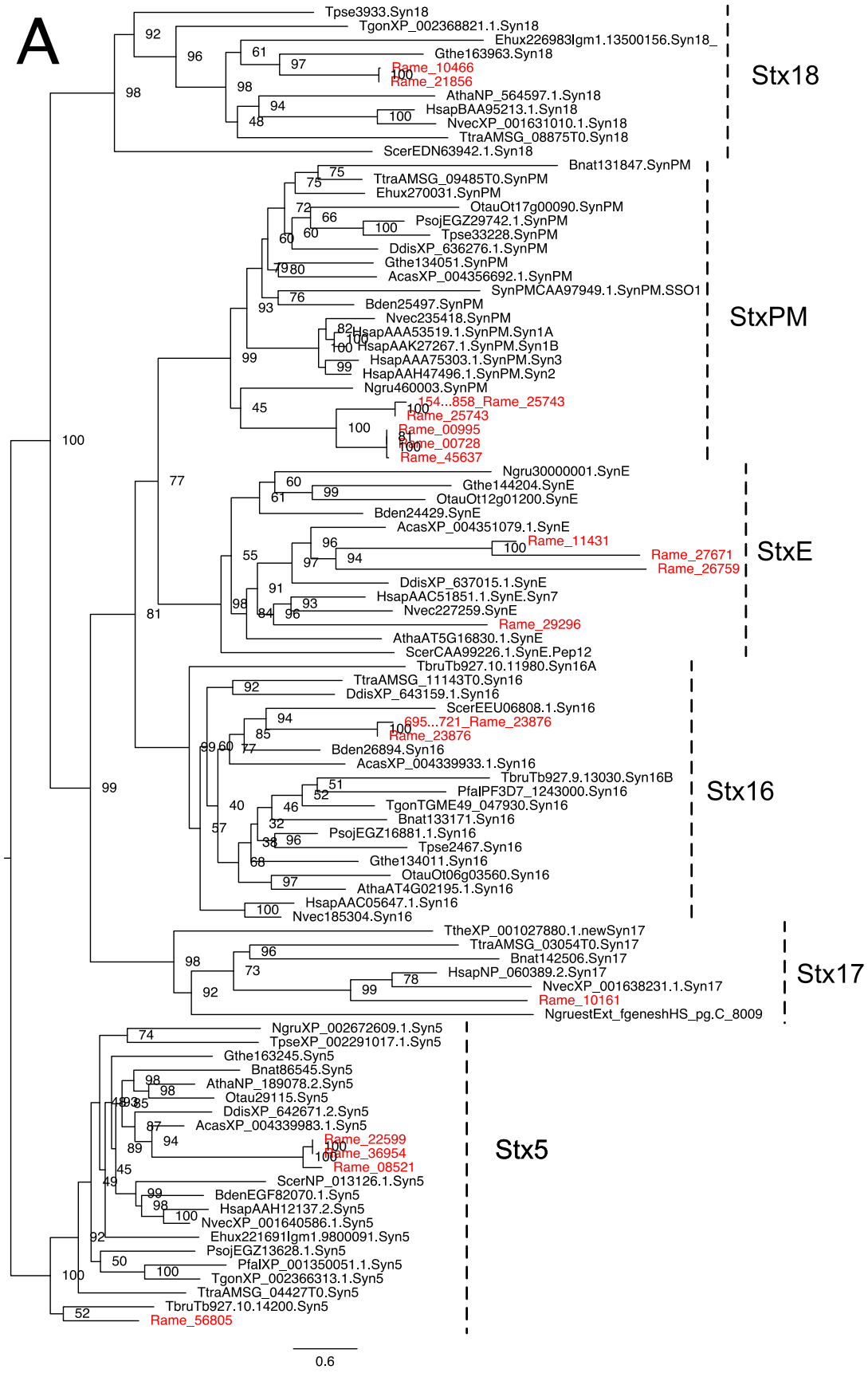
Zysset-Burri D. C., Müller N., Beuret C., Heller M., Schürch N., Gottstein B. & Wittwer M. 2014. Genome-wide identification of pathogenicity factors of the free-living amoeba *Naegleria fowleri*. *BMC Genomics*, 15:496. doi:10.1186/1471-2164-15-496

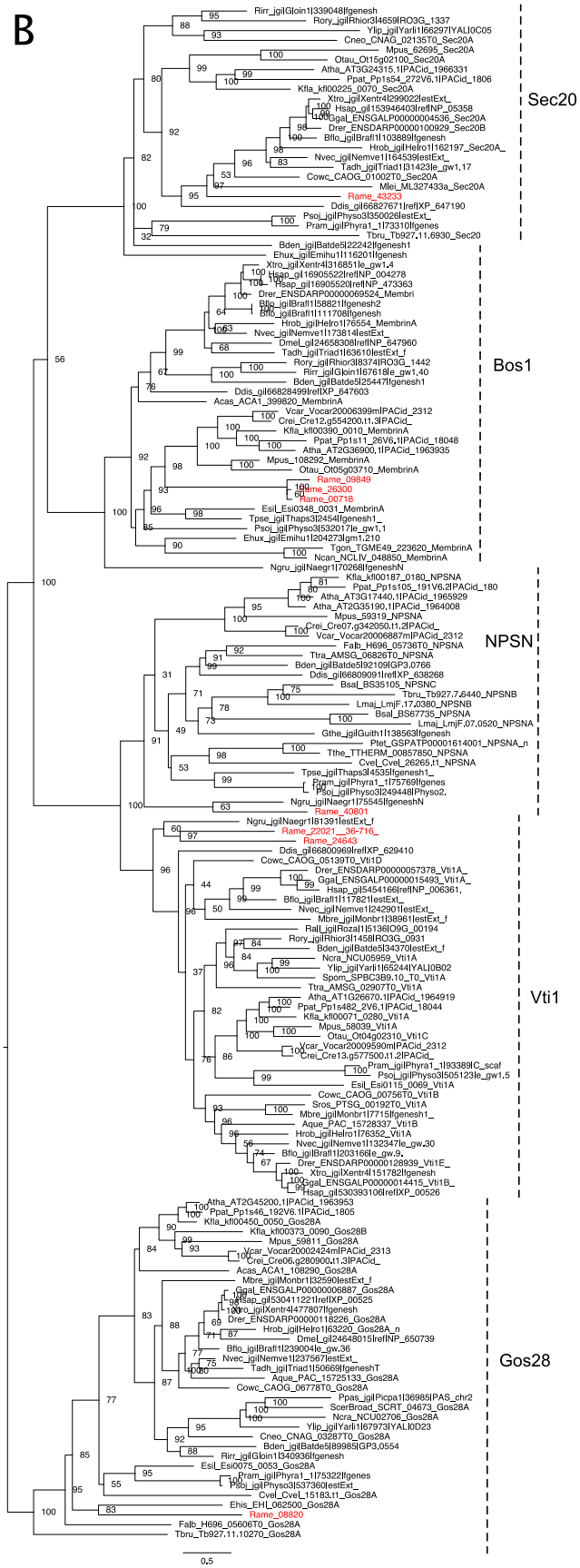
Appendix A

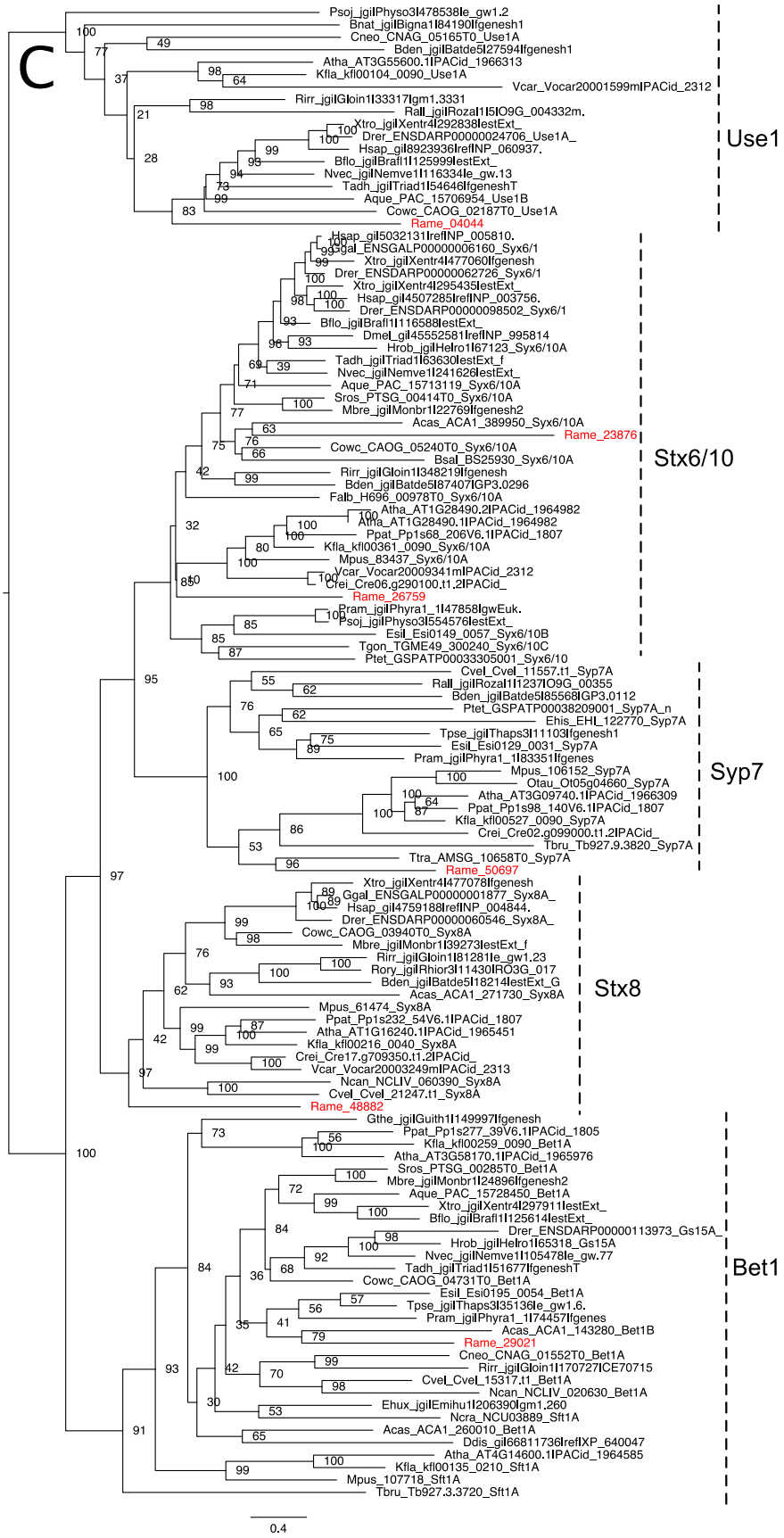
Supplementary Online Data

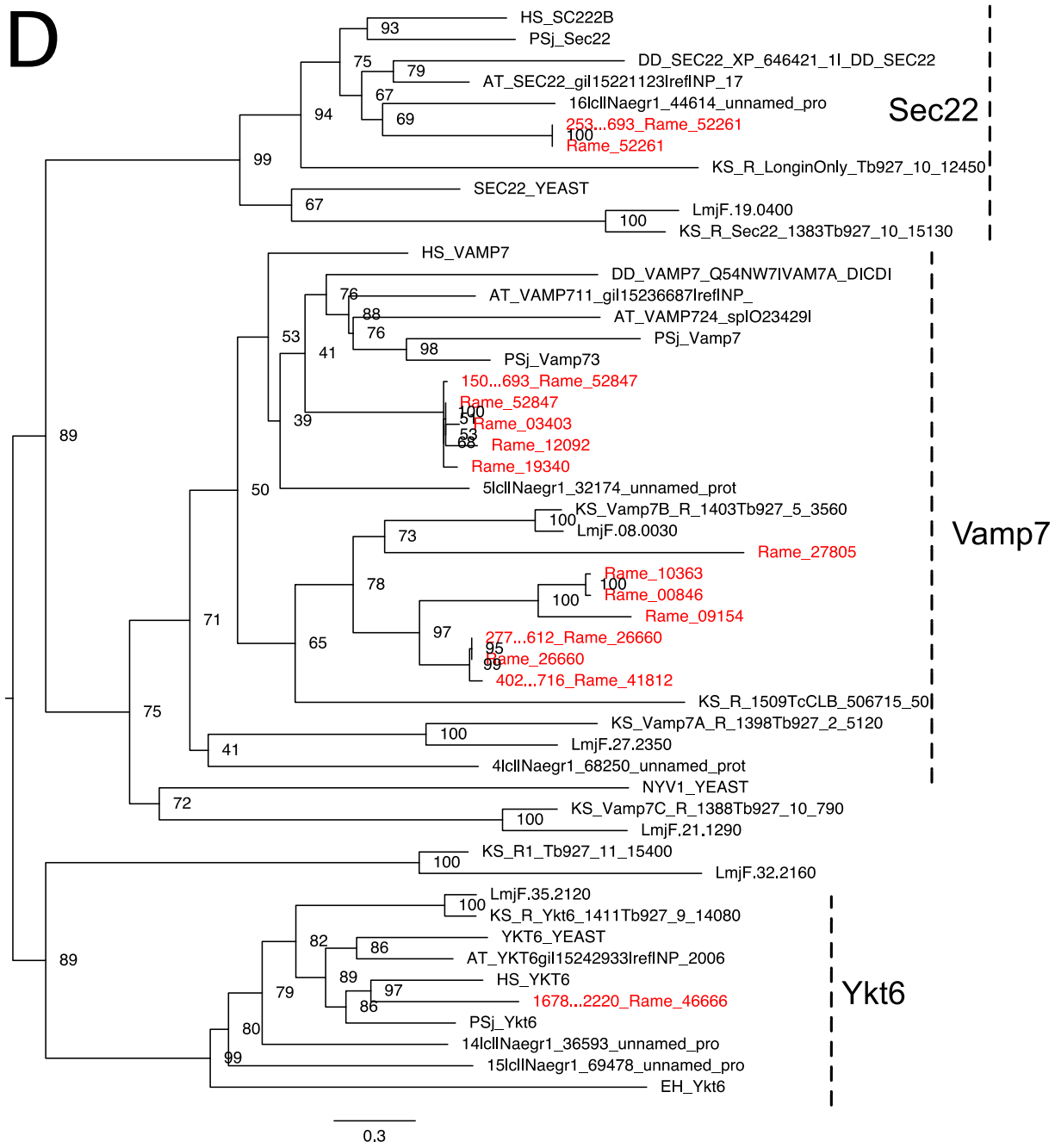
The supplementary online videos, files, and tables, and full-size PDF copies of the supplementary figures are available at the following URL:
https://drive.google.com/drive/folders/17W_f0MhSYwSBp4_wu42jF87sL7ioCwm5?usp=share_link

Supplementary figures follow.

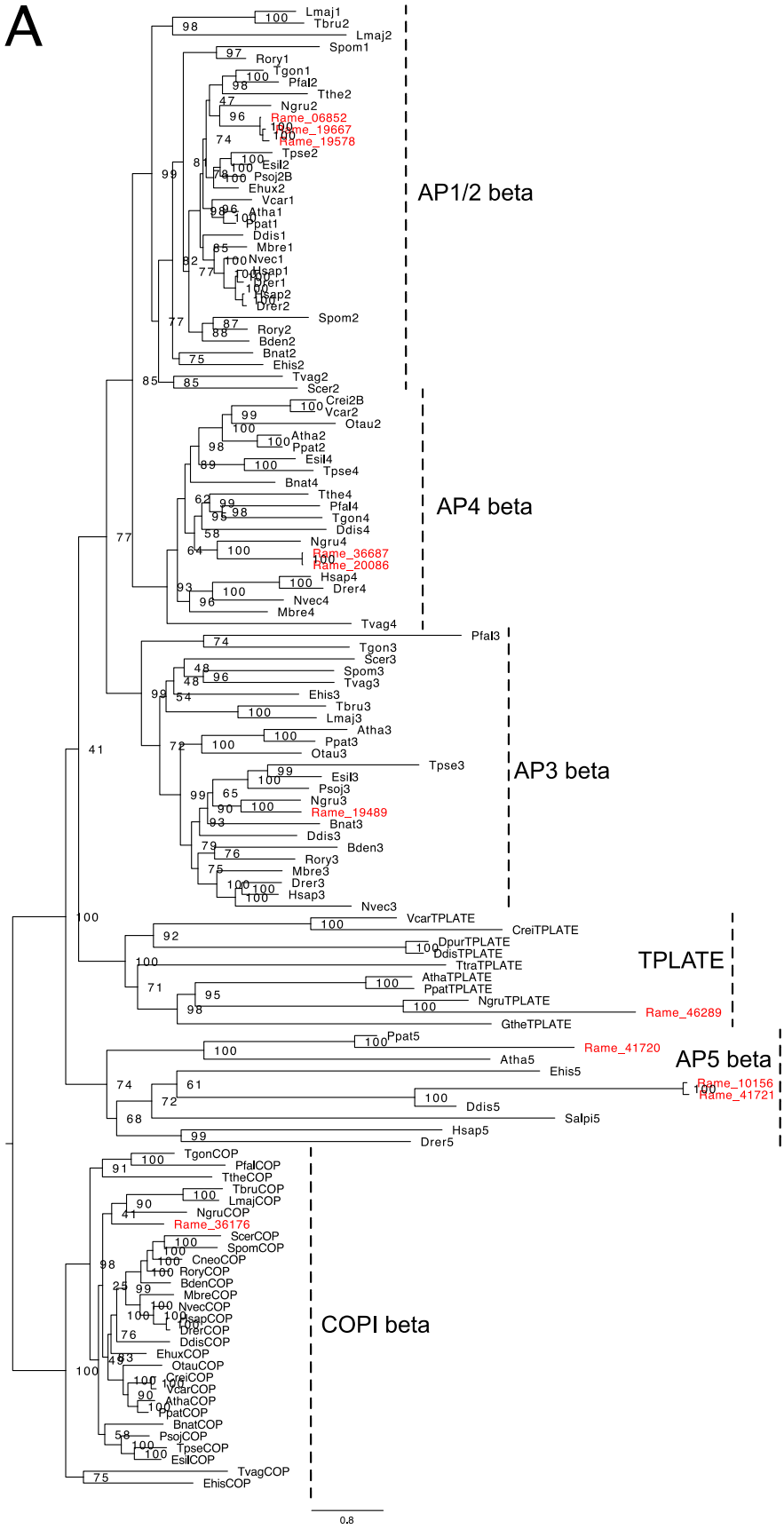

A
Tpse3933.Syn18
TgonXP_002368821.1.Syn18
Ehux226983Igm1.13500156.Syn18_
Gthe163963.Syn18
Rame_10466
Rame_21856
AthaNP_564597.1.Syn18
HsapBAA95213.1.Syn18
NvecXP_001631010.1.Syn18
TtraAMSG_08875T0.Syn18
ScerEDN63942.1.Syn18
Stx18
Bnat131847.SynPM
TtraAMSG_09485T0.SynPM
Ehux270031.SynPM
OtauOt17g00090.SynPM
PsojEGZ29742.1.SynPM
Tpse33228.SynPM
DdisXP_636276.1.SynPM
Gthe134051.SynPM
AcasXP_004356692.1.SynPM
SynPMCAA97949.1.SynPM.SSO1
Bden25497.SynPM
Nvec235418.SynPM
HsapAAA53519.1.SynPM.Syn1A
HsapAAK27267.1.SynPM.Syn1B
HsapAAA75303.1.SynPM.Syn3
HsapAAH47496.1.SynPM.Syn2
Ngru460003.SynPM
154...858_Rame_25743
Rame_25743
Rame_00995
Rame_00728
Rame_45637
StxPM
Ngru30000001.SynE
Gthe144204.SynE
OtauOt12g01200.SynE
Bden24429.SynE
AcasXP_004351079.1.SynE
Rame_11431
Rame_27671
Rame_26759
DdisXP_637015.1.SynE
HsapAAC51851.1.SynE.Syn7
Nvec227259.SynE
Rame_29296
AthaAT5G16830.1.SynE
ScerCAA99226.1.SynE.Pep12
StxE
TbruTb927.10.11980.Syn16A
TtraAMSG_11143T0.Syn16
DdisXP_643159.1.Syn16
ScerEEU06808.1.Syn16
695...721_Rame_23876
Rame_23876
Bden26894.Syn16
AcasXP_004339933.1.Syn16
TbruTb927.9.13030.Syn16B
PfalPF3D7_1243000.Syn16
TgonTGME49_047930.Syn16
Bnat133171.Syn16
PsojEGZ16881.1.Syn16
Tpse2467.Syn16
Gthe134011.Syn16
OtauOt06g03560.Syn16
AthaAT4G02195.1.Syn16
HsapAAC05647.1.Syn16
Nvec185304.Syn16
Stx16
TtheXP_001027880.1.newSyn17
TtraAMSG_03054T0.Syn17
Bnat142506.Syn17
HsapNP_060389.2.Syn17
NvecXP_001638231.1.Syn17
Rame_10161
NgruestExt_fgeneshHS_pg.C_8009
Stx17
NgruXP_002672609.1.Syn5
TpseXP_002291017.1.Syn5
Gthe163245.Syn5
Bnat86545.Syn5
AthaNP_189078.2.Syn5
Otau29115.Syn5
DdisXP_642671.2.Syn5
AcasXP_004339983.1.Syn5
Rame_22599
Rame_36954
Rame_08521
ScerNP_013126.1.Syn5
BdenEGF82070.1.Syn5
HsapAAH12137.2.Syn5
NvecXP_001640586.1.Syn5
Ehux221691Igm1.9800091.Syn5
PsojEGZ13628.1.Syn5
PfalXP_001350051.1.Syn5
TgonXP_002366313.1.Syn5
TtraAMSG_04427T0.Syn5
TbruTb927.10.14200.Syn5
Rame_56805
Stx5
0.6

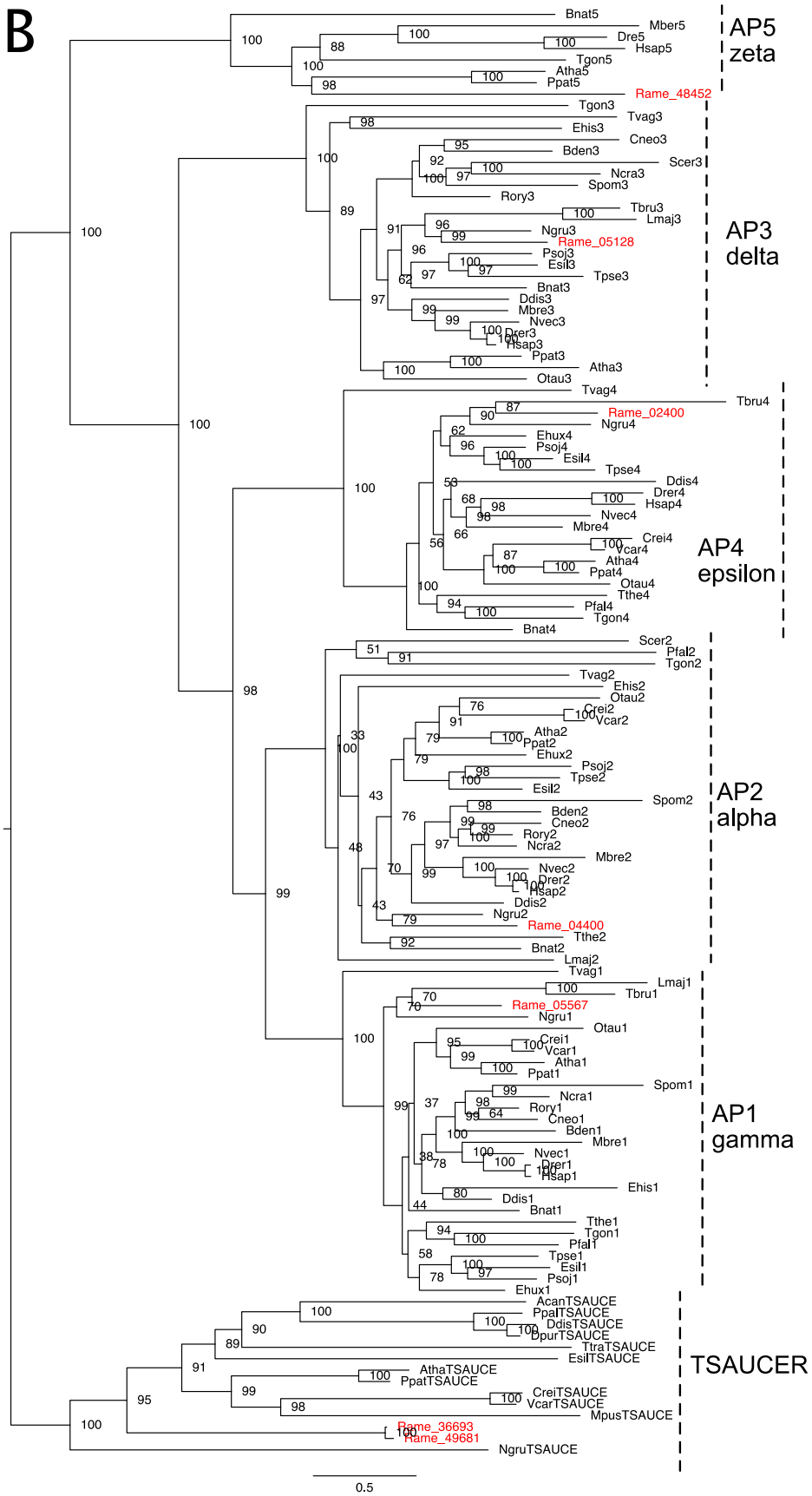

B
Sec20
Bos1
NPSN
Vti1
Gos28
Rame_43233
Rame_09849
Rame_26300
Rame_00718
Rame_40801
Rame_22021_36-716_
Rame_24643
Rame_08820
0.5

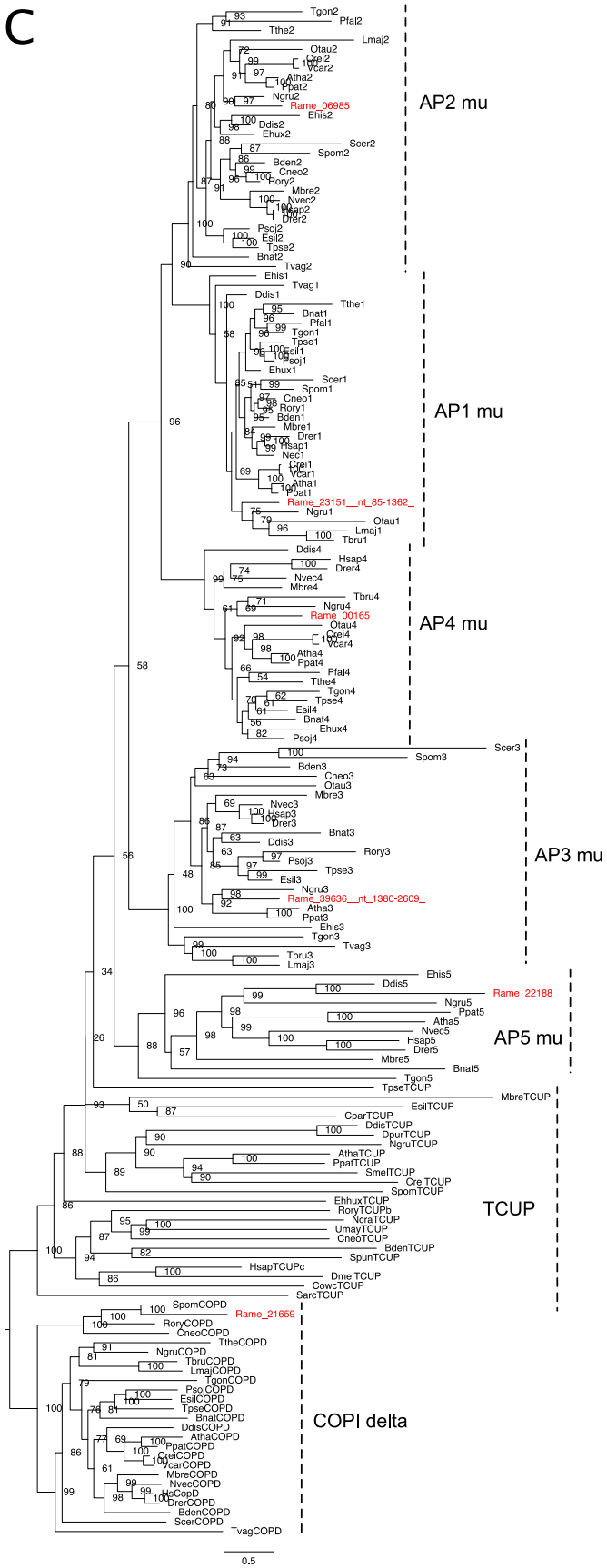

C
Use1
Stx6/10
Syp7
Stx8
Bet1
Psoj_jgi|Physo3|478538|e_gw1.2
Bnat_jgi|Bigna1|84190|fgenesh1
Cneo_CNAG_05165T0_Use1A
Bden_jgi|Batde5|27594|fgenesh1
Atha_AT3G55600.1|PACid_1966313
Kfla_kfl00104_0090_Use1A
Vcar_Vocar20001599m|PACid_2312
Rirr_jgi|Gloin1|33317|gm1.3331
Rall_jgi|Rozal1|5|O9G_004332m.
Xtro_jgi|Xentr4|292838|estExt_
Drer_ENSDARP00000024706_Use1A_
Hsap_gi|8923936|ref|NP_060937.
Bflo_jgi|Brafl1|125999|estExt_
Nvec_jgi|Nemve1|116334|e_gw.13
Tadh_jgi|Triad1|54646|fgeneshT
Aque_PAC_15706954_Use1B
Cowc_CAOG_02187T0_Use1A
Rame_04044
Hsap_gi|5032131|ref|NP_005810.
Ggal_ENSGALP00000006160_Syx6/1
Xtro_jgi|Xentr4|477060|fgenesh
Drer_ENSDARP00000062726_Syx6/1
Xtro_jgi|Xentr4|295435|estExt_
Hsap_gi|4507285|ref|NP_003756.
Drer_ENSDARP00000098502_Syx6/1
Bflo_jgi|Brafl1|116588|estExt_
Dmel_gi|45552581|ref|NP_995814
Hrob_jgi|Helro1|67123_Syx6/10A
Tadh_jgi|Triad1|63630|estExt_f
Nvec_jgi|Nemve1|241626|estExt_
Aque_PAC_15713119_Syx6/10A
Sros_PTSG_00414T0_Syx6/10A
Mbre_jgi|Monbr1|22769|fgenesh2
Acas_ACA1_389950_Syx6/10A
Rame_23876
Cowc_CAOG_05240T0_Syx6/10A
Bsal_BS25930_Syx6/10A
Rirr_jgi|Gloin1|348219|fgenesh
Bden_jgi|Batde5|87407|GP3.0296
Falb_H696_00978T0_Syx6/10A
Atha_AT1G28490.2|PACid_1964982
Atha_AT1G28490.1|PACid_1964982
Ppat_Pp1s68_206V6.1|PACid_1807
Kfla_kfl00361_0090_Syx6/10A
Mpus_83437_Syx6/10A
Vcar_Vocar20009341m|PACid_2312
Crei_Cre06.g290100.t1.2|PACid_
Rame_26759
Pram_jgi|Phyra1_1|47858|gwEuk.
Psoj_jgi|Physo3|554576|estExt_
Esil_Esi0149_0057_Syx6/10B
Tgon_TGME49_300240_Syx6/10C
Ptet_GSPATP00033305001_Syx6/10
Cvel_Cvel_11557.t1_Syp7A
Rall_jgi|Rozal1|1|237|O9G_00355
Bden_jgi|Batde5|85568|GP3.0112
Ptet_GSPATP00038209001_Syp7A_n
Ehis_EHI_122770_Syp7A
Tpse_jgi|Thaps3|11103|fgenesh1
Esil_Esi0129_0031_Syp7A
Pram_jgi|Phyra1_1|83351|fgenes
Mpus_106152_Syp7A
Otau_Ot05g04660_Syp7A
Atha_AT3G09740.1|PACid_1966309
Ppat_Pp1s98_140V6.1|PACid_1807
Kfla_kfl00527_0090_Syp7A
Crei_Cre02.g099000.t1.2|PACid_
Tbru_Tb927.9.3820_Syp7A
Ttra_AMSG_10658T0_Syp7A
Rame_50697
Xtro_jgi|Xentr4|477078|fgenesh
Ggal_ENSGALP00000001877_Syx8A_
Hsap_gi|4759188|ref|NP_004844.
Drer_ENSDARP00000060546_Syx8A_
Cowc_CAOG_03940T0_Syx8A
Mbre_jgi|Monbr1|39273|estExt_f
Rirr_jgi|Gloin1|81281|e_gw1.23
Rory_jgi|Rhior3|11430|RO3G_017
Bden_jgi|Batde5|18214|estExt_G
Acas_ACA1_271730_Syx8A
Mpus_61474_Syx8A
Ppat_Pp1s232_54V6.1|PACid_1807
Atha_AT1G16240.1|PACid_1965451
Kfla_kfl00216_0040_Syx8A
Crei_Cre17.g709350.t1.2|PACid_
Vcar_Vocar20003249m|PACid_2313
Ncan_NCLIV_060390_Syx8A
Cvel_Cvel_21247.t1_Syx8A
Rame_48882
Gthe_jgi|Guith1|149997|fgenesh
Ppat_Pp1s277_39V6.1|PACid_1805
Kfla_kfl00259_0090_Bet1A
Atha_AT3G58170.1|PACid_1965976
Sros_PTSG_00285T0_Bet1A
Mbre_jgi|Monbr1|24896|fgenesh2
Aque_PAC_15728450_Bet1A
Xtro_jgi|Xentr4|297911|estExt_
Bflo_jgi|Brafl1|125614|estExt_
Drer_ENSDARP00000113973_Gs15A_
Hrob_jgi|Helro1|65318_Gs15A
Nvec_jgi|Nemve1|105478|e_gw.77
Tadh_jgi|Triad1|51677|fgeneshT
Cowc_CAOG_04731T0_Bet1A
Esil_Esi0195_0054_Bet1A
Tpse_jgi|Thaps3|35136|e_gw1.6.
Pram_jgi|Phyra1_1|74457|fgenes
Acas_ACA1_143280_Bet1B
Rame_29021
Cneo_CNAG_01552T0_Bet1A
Rirr_jgi|Gloin1|170727|CE70715
Cvel_Cvel_15317.t1_Bet1A
Ncan_NCLIV_020630_Bet1A
Ehux_jgi|Emihu1|206390|gm1.260
Ncra_NCU03889_Sft1A
Acas_ACA1_260010_Bet1A
Ddis_gi|66811736|ref|XP_640047
Atha_AT4G14600.1|PACid_1964585
Kfla_kfl00135_0210_Sft1A
Mpus_107718_Sft1A
Tbru_Tb927.3.3720_Sft1A
0.4



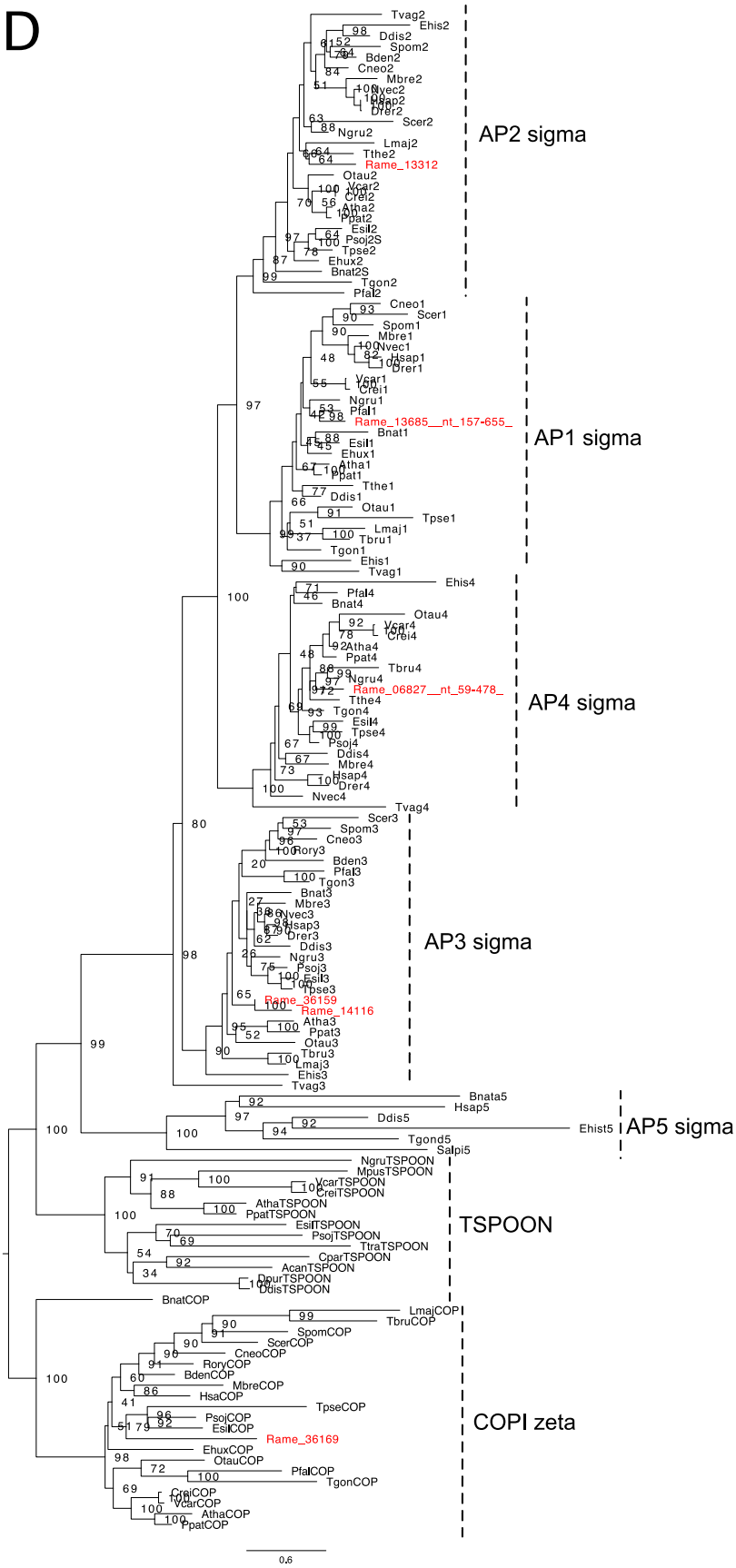
Supplementary Figure 1 Ultrafast bootstrap maximum likelihood phylogenies for classification of Qa (**A**), Qb (**B**), Qc (**C**), and R (**D**) SNAREs in *Reclinomonas americana*. *R. americana* sequences in red. The base alignment for Qa SNAREs is from Arasaki et al. (2017); the rest are from Venkatesh et al. (2017).

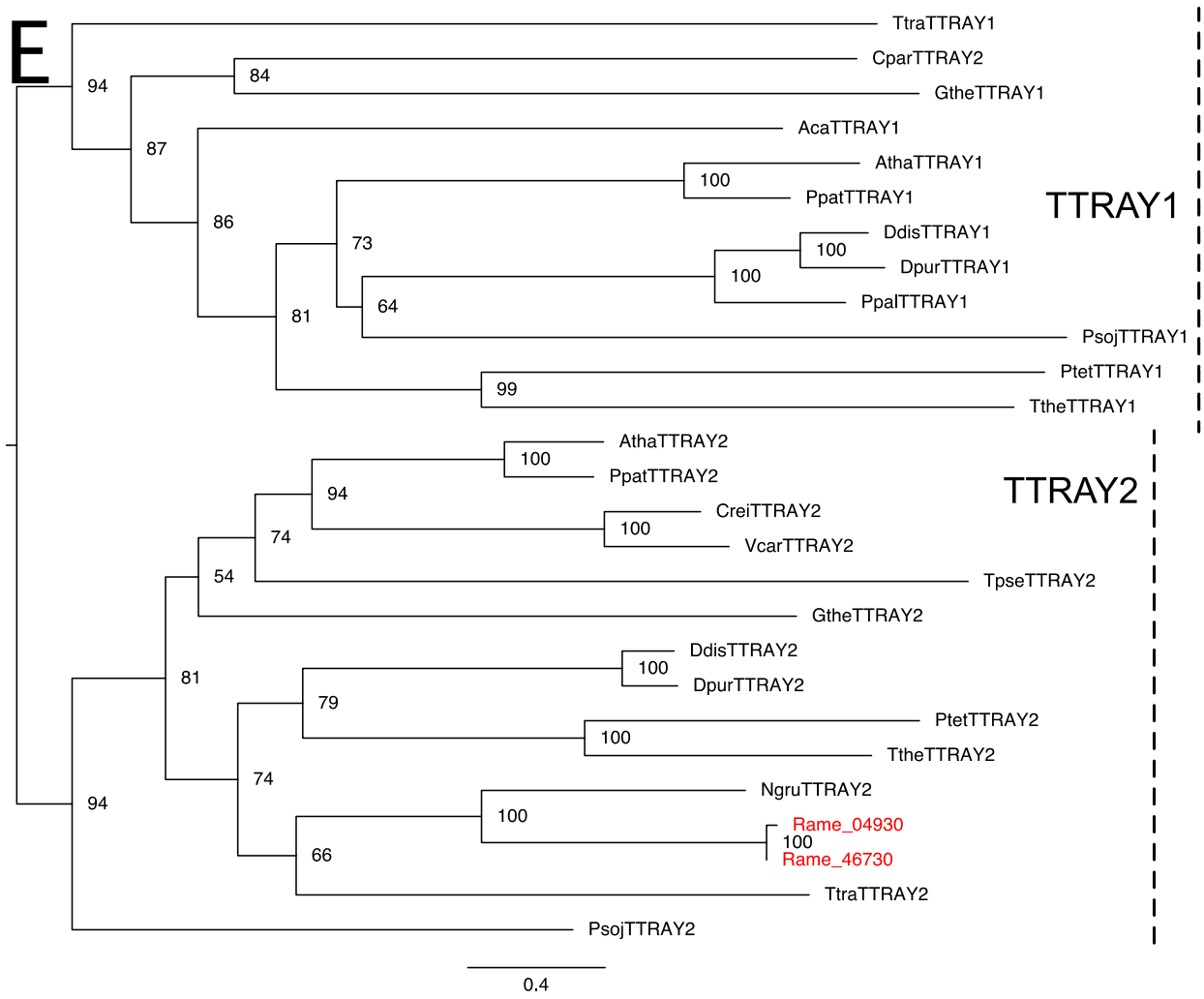

A
AP1/2 beta
AP4 beta
AP3 beta
TPLATE
AP5 beta
COPI beta
Lmaj1
Tbru2
Lmaj2
Spom1
Rory1
Tgon1
Pfal2
Tthe2
Ngru2
Rame_06852
Rame_19667
Rame_19578
Tpse2
Esil2
Psoj2B
Ehux2
Vcar1
Atha1
Ppat1
Ddis1
Mbre1
Nvec1
Hsap1
Drer1
Hsap2
Drer2
Spom2
Rory2
Bden2
Bnat2
Ehis2
Tvag2
Scer2
Crei2B
Vcar2
Otau2
Atha2
Ppat2
Esil4
Tpse4
Bnat4
Tthe4
Pfal4
Tgon4
Ddis4
Ngru4
Rame_36687
Rame_20086
Hsap4
Drer4
Nvec4
Mbre4
Tvag4
Pfal3
Tgon3
Scer3
Spom3
Tvag3
Ehis3
Tbru3
Lmaj3
Atha3
Ppat3
Otau3
Tpse3
Esil3
Psoj3
Ngru3
Rame_19489
Bnat3
Ddis3
Bden3
Rory3
Mbre3
Drer3
Hsap3
Nvec3
VcarTPLATE
CreiTPLATE
DpurTPLATE
DdisTPLATE
TtraTPLATE
AthaTPLATE
PpatTPLATE
NgruTPLATE
Rame_46289
GtheTPLATE
Ppat5
Rame_41720
Atha5
Ehis5
Rame_10156
Rame_41721
Ddis5
Salpi5
Hsap5
Drer5
TgonCOP
PfalCOP
TtheCOP
TbruCOP
LmajCOP
NgruCOP
Rame_36176
ScerCOP
SpomCOP
CneoCOP
RoryCOP
BdenCOP
MbreCOP
NvecCOP
HsapCOP
DrerCOP
DdisCOP
EhuxCOP
OtauCOP
CreiCOP
VcarCOP
AthaCOP
PpatCOP
BnatCOP
PsojCOP
TpseCOP
EsilCOP
TvagCOP
EhisCOP
0.8


B
AP5 zeta
AP3 delta
AP4 epsilon
AP2 alpha
AP1 gamma
TSAUCER
Bnat5
Mber5
Dre5
Hsap5
Tgon5
Atha5
Ppat5
Rame_48452
Tgon3
Tvag3
Ehis3
Cneo3
Bden3
Scer3
Ncra3
Spom3
Rory3
Tbru3
Lmaj3
Ngru3
Rame_05128
Psoj3
Esil3
Tpse3
Bnat3
Ddis3
Mbre3
Nvec3
Drer3
Hsap3
Ppat3
Atha3
Otau3
Tvag4
Tbru4
Rame_02400
Ngru4
Ehux4
Psoj4
Esil4
Tpse4
Ddis4
Drer4
Hsap4
Nvec4
Mbre4
Crei4
Vcar4
Atha4
Ppat4
Otau4
Tthe4
Pfal4
Tgon4
Bnat4
Scer2
Pfal2
Tgon2
Tvag2
Ehis2
Otau2
Crei2
Vcar2
Atha2
Ppat2
Ehux2
Psoj2
Tpse2
Esil2
Spom2
Bden2
Cneo2
Rory2
Ncra2
Mbre2
Nvec2
Drer2
Hsap2
Ddis2
Ngru2
Rame_04400
Tthe2
Bnat2
Lmaj2
Tvag1
Lmaj1
Tbru1
Rame_05567
Ngru1
Otau1
Crei1
Vcar1
Atha1
Ppat1
Spom1
Ncra1
Rory1
Cneo1
Bden1
Mbre1
Nvec1
Drer1
Hsap1
Ehis1
Ddis1
Bnat1
Tthe1
Tgon1
Pfal1
Tpse1
Esil1
Psoj1
Ehux1
AcanTSAUCE
PpalTSAUCE
DdisTSAUCE
BpurTSAUCE
TtraTSAUCE
EsilTSAUCE
AthaTSAUCE
PpatTSAUCE
CreiTSAUCE
VcarTSAUCE
MpusTSAUCE
Rame_36693
Rame_49681
NgruTSAUCE
0.5

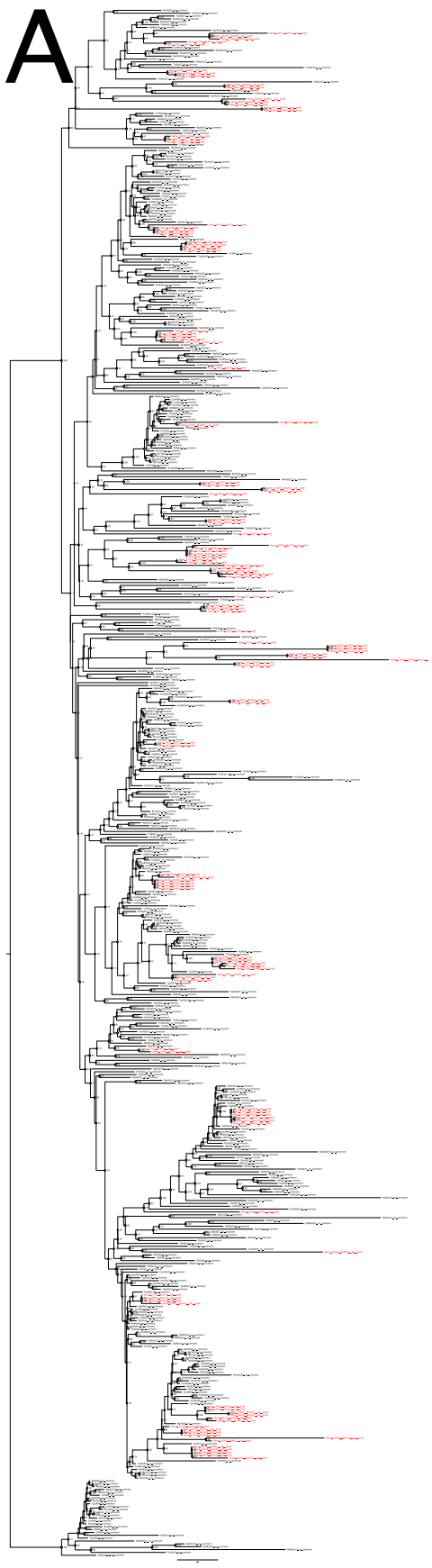

C
AP2 mu
AP1 mu
AP4 mu
AP3 mu
AP5 mu
TCUP
COPI delta
Rame_06985
Rame_23151__nt_85-1362_
Rame_00165
Rame_39636__nt_1380-2609_
Rame_22188
Rame_21659
0.5

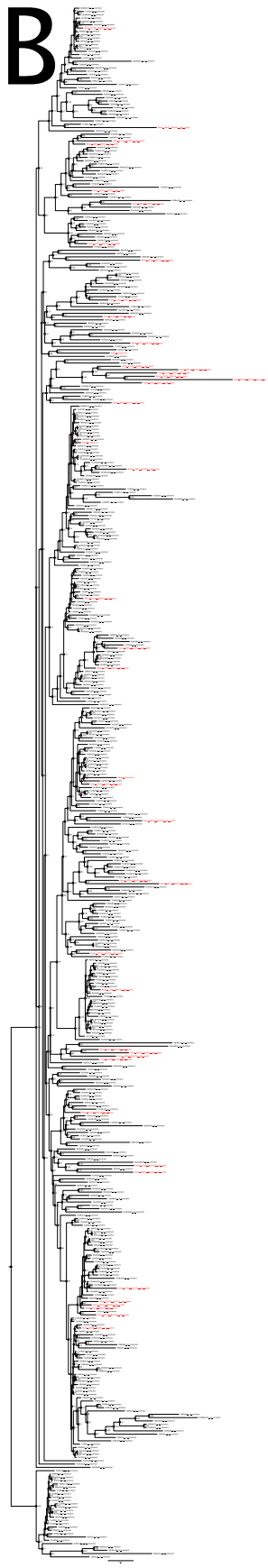
D

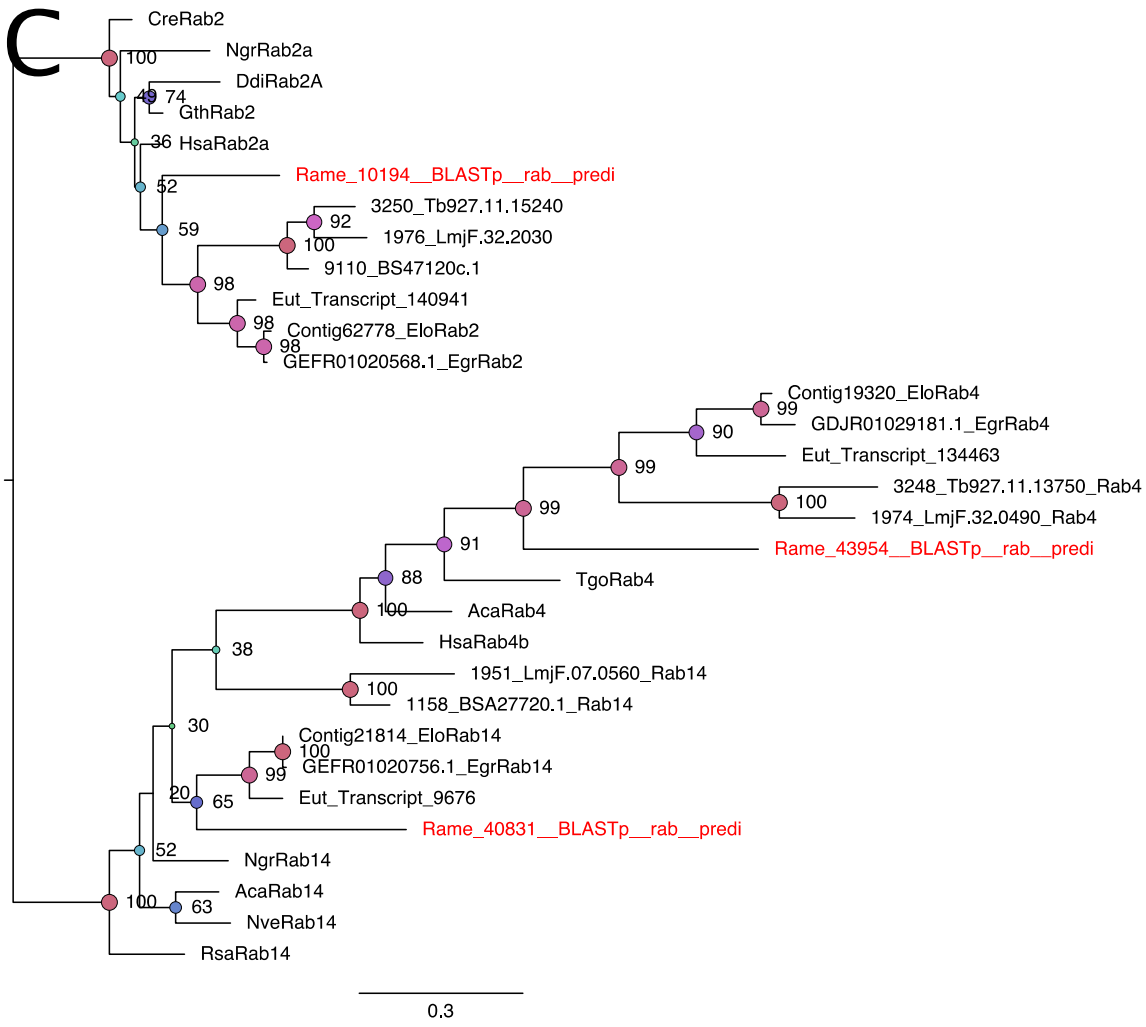

AP2 sigma
AP1 sigma
AP4 sigma
AP3 sigma
AP5 sigma
TSPOON
COPI zeta
Rame_13312
Rame_13685__nt_157-655_
Rame_06827__nt_59-478_
Rame_36159
Rame_14116
Rame_36169
0.6



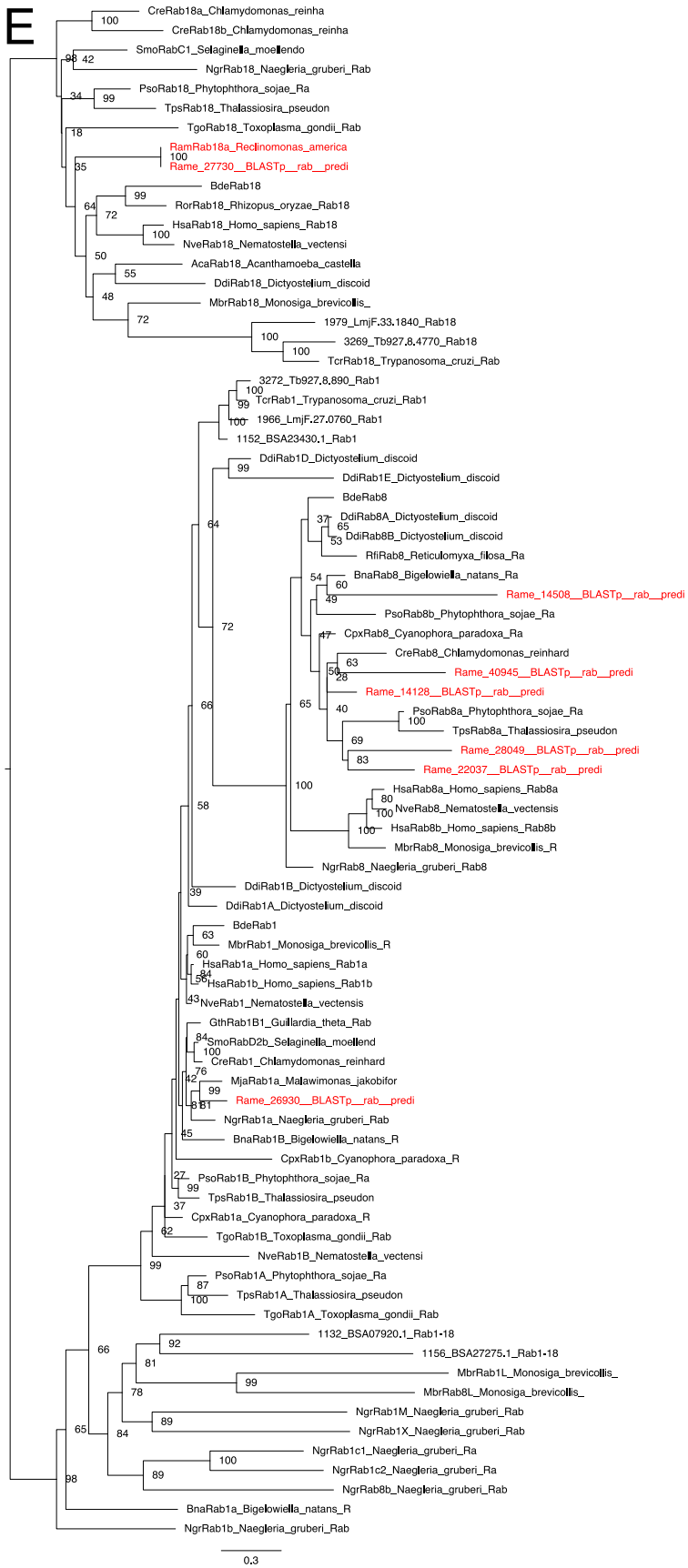
Supplementary Figure 2 Ultrafast bootstrap maximum likelihood phylogenies for classification of heterotetrameric adaptor protein complex subunits β (**A**), EGADZ (**B**), μ (**C**), and σ (**D**), and the two β-propellor/α-solenoid subunits of TSET (**E**) in *Reclinomonas americana*. *R. americana* sequences are in red. Alignments from Hirst et al. (2014).

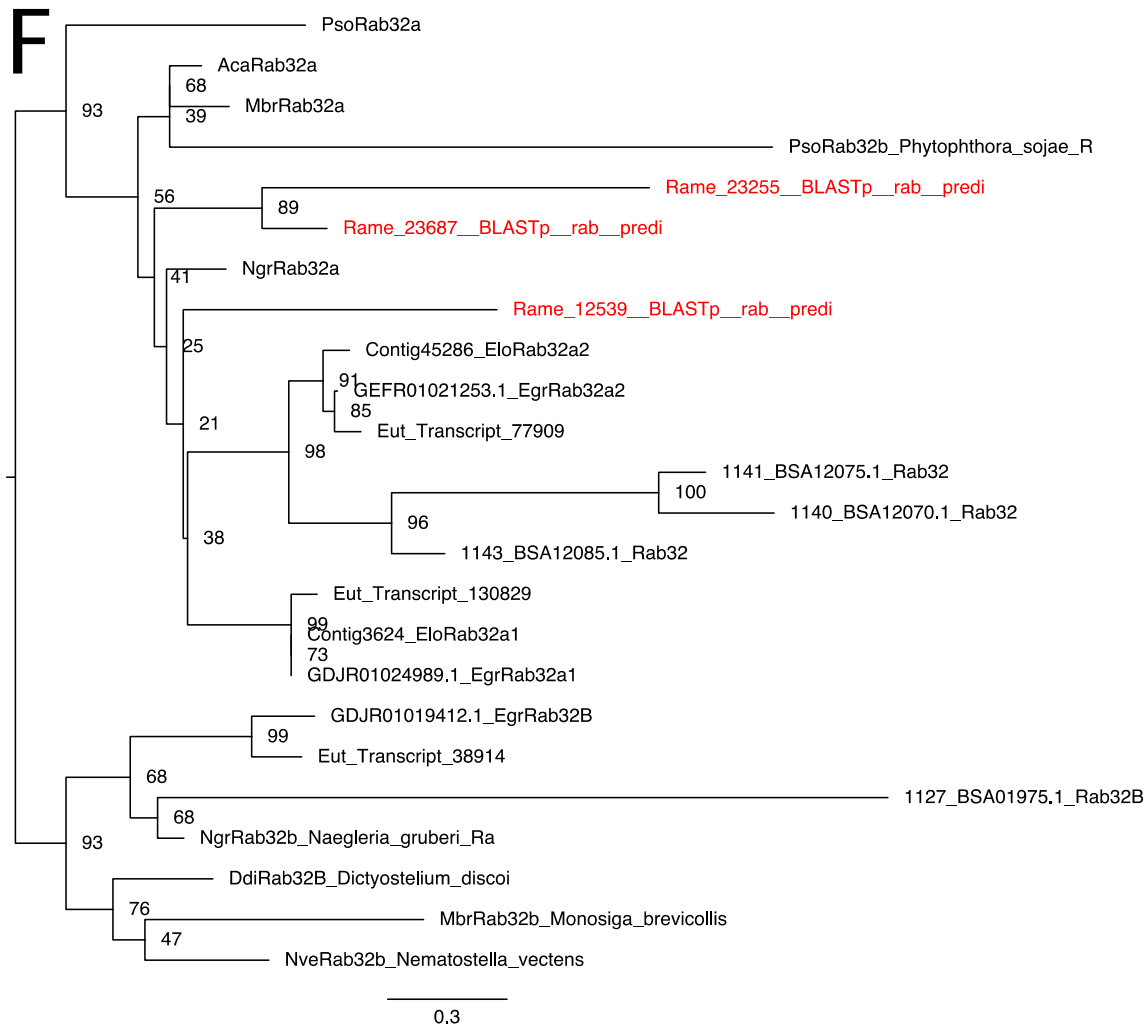

A


B

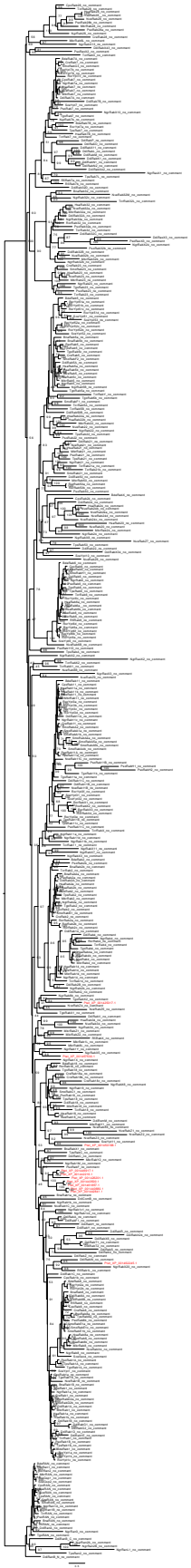

C
CreRab2
NgrRab2a
DdiRab2A
GthRab2
HsaRab2a
Rame_10194__BLASTp__rab__predi
3250_Tb927.11.15240
1976_LmjF.32.2030
9110_BS47120c.1
Eut_Transcript_140941
Contig62778_EloRab2
GEFR01020568.1_EgrRab2
Contig19320_EloRab4
GDJR01029181.1_EgrRab4
Eut_Transcript_134463
3248_Tb927.11.13750_Rab4
1974_LmjF.32.0490_Rab4
Rame_43954__BLASTp__rab__predi
TgoRab4
AcaRab4
HsaRab4b
1951_LmjF.07.0560_Rab14
1158_BSA27720.1_Rab14
Contig21814_EloRab14
GEFR01020756.1_EgrRab14
Eut_Transcript_9676
Rame_40831__BLASTp__rab__predi
NgrRab14
AcaRab14
NveRab14
RsaRab14
0.3


D
Contig2217_EloRab1C
GDJR01024919.1_EgrRab1C
EuL_Transcript_110921
Contig68016_EloRab1D
GDJR01028264.1_EgrRab1D
Contig29950_EloRab1B
GDJR01006599.1_EgrRab1B
Eut_Transcript_92486
NgrRab1c1_Naegleria_gruberi_Ra
NgrRab1c2_Naegleria_gruberi_Ra
NgrRab8b_Naegleria_gruberi_Rab
Contig62994_EloRab1L
GDJR01007661.1_EgrRab1L
1131_BSA07915.1_Rab1-18
1132_BSA07920.1_Rab1-18
NgrRab1M_Naegleria_gruberi_Rab
NgrRab1X_Naegleria_gruberi_Rab
MbrRab1L_Monosiga_brevicollis_
MbrRab8L_Monosiga_brevicollis_
Contig13773_EloRab18
GDJR01011494.1_EgrRab18
Eut_Transcript_31271
TgoRab18_Toxoplasma_gondii_Rab
RamRab18a_Rediimonas_america
Rame_27730__BLASTp__rab__predi
NgrRab18_Naegleria_gruberi_Rab
SmoRabC1_Selaginella_moellendo
CreRab18a_Chlamydomonas_reinha
CreRab18b_Chlamydomonas_reinha
CreRab18c_Chlamydomonas_reinha
PsoRab18_Phytophthora_sojae_Ra
TpsRab18_Thalassiosira_pseudon
AcaRab18_Acanthamoeba_castella
MbrRab18_Monosiga_brevicollis_
BdeRab18
RorRab18_Rhizopus_oryzae_Rab18
HsaRab18_Homo_sapiens_Rab18
NveRab18_Nematostella_vectensi
DdiRab18_Dictyostelium_discoid
3269_Tb927.8.4770_Rab18
TcrRab18_Trypanosoma_cruzi_Rab
1979_LmjF.33.1840_Rab18
1156_BSA27275.1_Rab1-18
NgrRab1Y_Naegleria_gruberi_Rab
NgrRab1b_Naegleria_gruberi_Rab
NveRab1C_Nematostella_vectensi
BnaRab1a_Bigelowiella_natans_R
3277_Tb927.9.15930_Rab1
PsoRab1A_Phytophthora_sojae_Ra
TpsRab1A_Thalassiosira_pseudon
TgoRab1A_Toxoplasma_gondii_Rab
DdiRab1D_Dictyostelium_discoid
DdiRab1E_Dictyostelium_discoid
PsoRab1B_Phytophthora_sojae_Ra
TpsRab1B_Thalassiosira_pseudon
BdeRab1
DdiRab1A_Dictyostelium_discoid
CpxRab1b_Cyanophora_paradoxa_R
NveRab1B_Nematostella_vectensi
DdiRab1B_Dictyostelium_discoid
NveRab1_Nematostella_vectensis
HsaRab1a_Homo_sapiens_Rab1a
HsaRab1b_Homo_sapiens_Rab1b
MbrRab1_Monosiga_brevicollis_R
CpxRab1a_Cyanophora_paradoxa_R
TgoRab1B_Toxoplasma_gondii_Rab
GthRab1B1_Guillardia_theta_Rab
SmoRabD2b_Selaginella_moellend
RtiRab1L_Reticulomyxa_filosa_R
CreRab1_Chlamydomonas_reinhard
MjaRab1a_Malawimonas_jakobifor
Rame_26930__BLASTp__rab__predi
NgrRab1a_Naegleria_gruberi_Rab
BnaRab1B_Bigelowiella_natans_R
3272_Tb927.8.890_Rab1
TcrRab1_Trypanosoma_cruzi_Rab1
1966_LmjF.27.0760_Rab1
1152_BSA23430.1_Rab1
DdiRab1C_Dictyostelium_discoid
Eut_Transcript_46125
Eut_Transcript_127021
Contig18959_EloRab1A
GDJR01012556.1_EgrRab1A
BdeRab8
DdiRab8A_Dictyostelium_discoid
DdiRab8B_Dictyostelium_discoid
RtiRab8_Reticulomyxa_filosa_Ra
HsaRab8a_Homo_sapiens_Rab8a
NveRab8_Nematostella_vectensis
HsaRab8b_Homo_sapiens_Rab8b
MbrRab8_Monosiga_brevicollis_R
BnaRab8_Bigelowiella_natans_Ra
NveRab8B_Nematostella_vectensi
Rame_14508__BLASTp__rab__predi
PsoRab8b_Phytophthora_sojae_Ra
CpxRab8_Cyanophora_paradoxa_Ra
NgrRab8_Naegleria_gruberi_Rab8
Rame_22037__BLASTp__rab__predi
Rame_14128__BLASTp__rab__predi
Rame_28049__BLASTp__rab__predi
Rame_40945__BLASTp__rab__predi
CreRab6_Chlamydomonas_reinhard
PsoRab8a_Phytophthora_sojae_Ra
TpsRab8a_Thalassiosira_pseudon
0.4


E
CreRab18a_Chlamydomonas_reinha
CreRab18b_Chlamydomonas_reinha
SmoRabC1_Selaginella_moellendo
NgrRab18_Naegleria_gruberi_Rab
PsoRab18_Phytophthora_sojae_Ra
TpsRab18_Thalassiosira_pseudon
TgoRab18_Toxoplasma_gondii_Rab
RamRab18a_Reclinomonas_america
Rame_27730__BLASTp__rab__predi
BdeRab18
RorRab18_Rhizopus_oryzae_Rab18
HsaRab18_Homo_sapiens_Rab18
NveRab18_Nematostella_vectensi
AcaRab18_Acanthamoeba_castella
DdiRab18_Dictyostelium_discoid
MbrRab18_Monosiga_brevicollis_
1979_LmjF.33.1840_Rab18
3269_Tb927.8.4770_Rab18
TcrRab18_Trypanosoma_cruzi_Rab
3272_Tb927.8.890_Rab1
TcrRab1_Trypanosoma_cruzi_Rab1
1966_LmjF.27.0760_Rab1
1152_BSA23430.1_Rab1
DdiRab1D_Dictyostelium_discoid
DdiRab1E_Dictyostelium_discoid
BdeRab8
DdiRab8A_Dictyostelium_discoid
DdiRab8B_Dictyostelium_discoid
RfiRab8_Reticulomyxa_filosa_Ra
BnaRab8_Bigelowiella_natans_Ra
Rame_14508__BLASTp__rab__predi
PsoRab8b_Phytophthora_sojae_Ra
CpxRab8_Cyanophora_paradoxa_Ra
CreRab8_Chlamydomonas_reinhard
Rame_40945__BLASTp__rab__predi
Rame_14128__BLASTp__rab__predi
PsoRab8a_Phytophthora_sojae_Ra
TpsRab8a_Thalassiosira_pseudon
Rame_28049__BLASTp__rab__predi
Rame_22037__BLASTp__rab__predi
HsaRab8a_Homo_sapiens_Rab8a
NveRab8_Nematostella_vectensis
HsaRab8b_Homo_sapiens_Rab8b
MbrRab8_Monosiga_brevicollis_R
NgrRab8_Naegleria_gruberi_Rab8
DdiRab1B_Dictyostelium_discoid
DdiRab1A_Dictyostelium_discoid
BdeRab1
MbrRab1_Monosiga_brevicollis_R
HsaRab1a_Homo_sapiens_Rab1a
HsaRab1b_Homo_sapiens_Rab1b
NveRab1_Nematostella_vectensis
GthRab1B1_Guillardia_theta_Rab
SmoRabD2b_Selaginella_moellend
CreRab1_Chlamydomonas_reinhard
MjaRab1a_Malawimonas_jakobifor
Rame_26930__BLASTp__rab__predi
NgrRab1a_Naegleria_gruberi_Rab
BnaRab1B_Bigelowiella_natans_R
CpxRab1b_Cyanophora_paradoxa_R
PsoRab1B_Phytophthora_sojae_Ra
TpsRab1B_Thalassiosira_pseudon
CpxRab1a_Cyanophora_paradoxa_R
TgoRab1B_Toxoplasma_gondii_Rab
NveRab1B_Nematostella_vectensi
PsoRab1A_Phytophthora_sojae_Ra
TpsRab1A_Thalassiosira_pseudon
TgoRab1A_Toxoplasma_gondii_Rab
1132_BSA07920.1_Rab1-18
1156_BSA27275.1_Rab1-18
MbrRab1L_Monosiga_brevicollis_
MbrRab8L_Monosiga_brevicollis_
NgrRab1M_Naegleria_gruberi_Rab
NgrRab1X_Naegleria_gruberi_Rab
NgrRab1c1_Naegleria_gruberi_Ra
NgrRab1c2_Naegleria_gruberi_Ra
NgrRab8b_Naegleria_gruberi_Rab
BnaRab1a_Bigelowiella_natans_R
NgrRab1b_Naegleria_gruberi_Rab
0.3

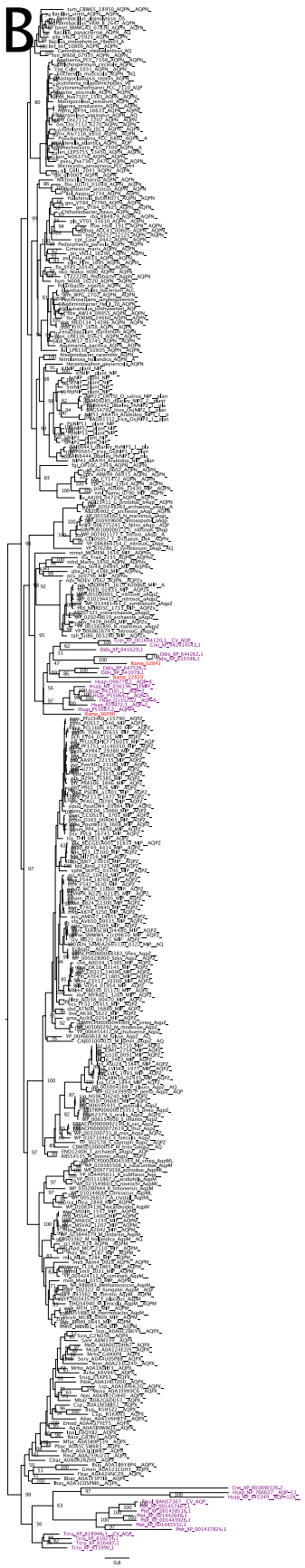


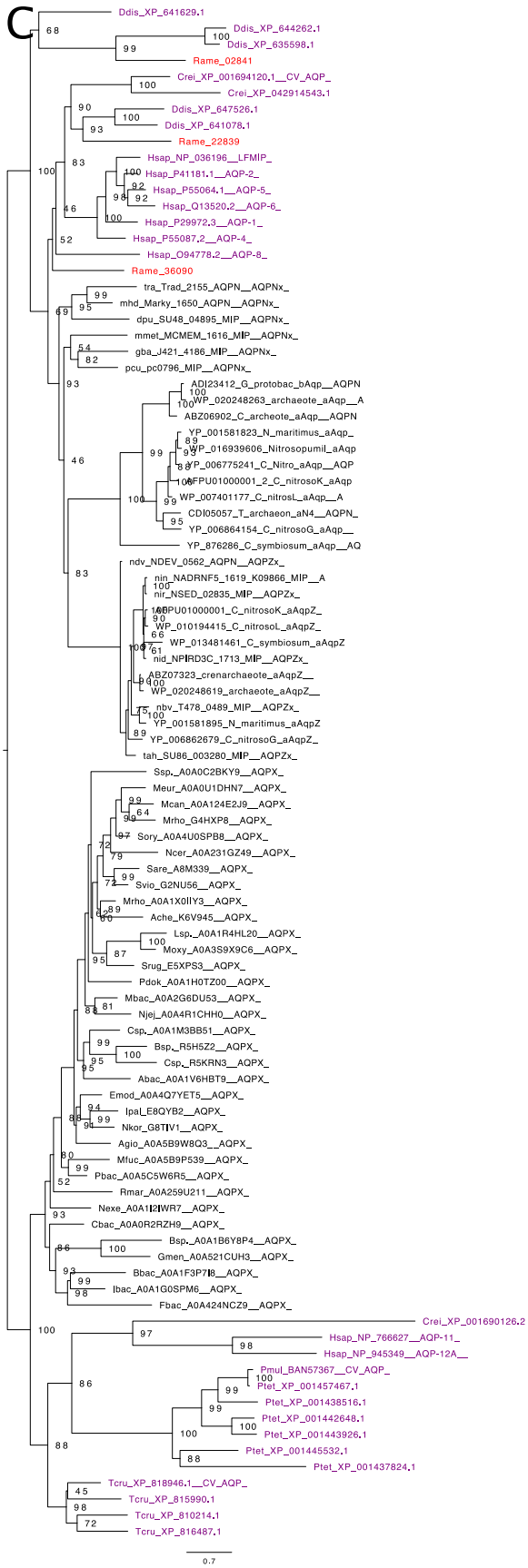
Supplementary Figure 3 Ultrafast bootstrap maximum likelihood (ML) phylogenies for classification of Rab GTPases in *Reclinomonas americana*. Sequences were initially classified to paneukaryotic Rab GTPases (**A**), and subsequently confirmed in restricted phylogenies containing the endocytic Rab GTPases (**B**), Rab2/Rab4/Rab14 (**C**), Rab1/Rab8/Rab18 (**D**, **E**), and Rab32 (**F**). *R. americana* sequences are in red. Alignments modified from Eliáš et al. (2012).

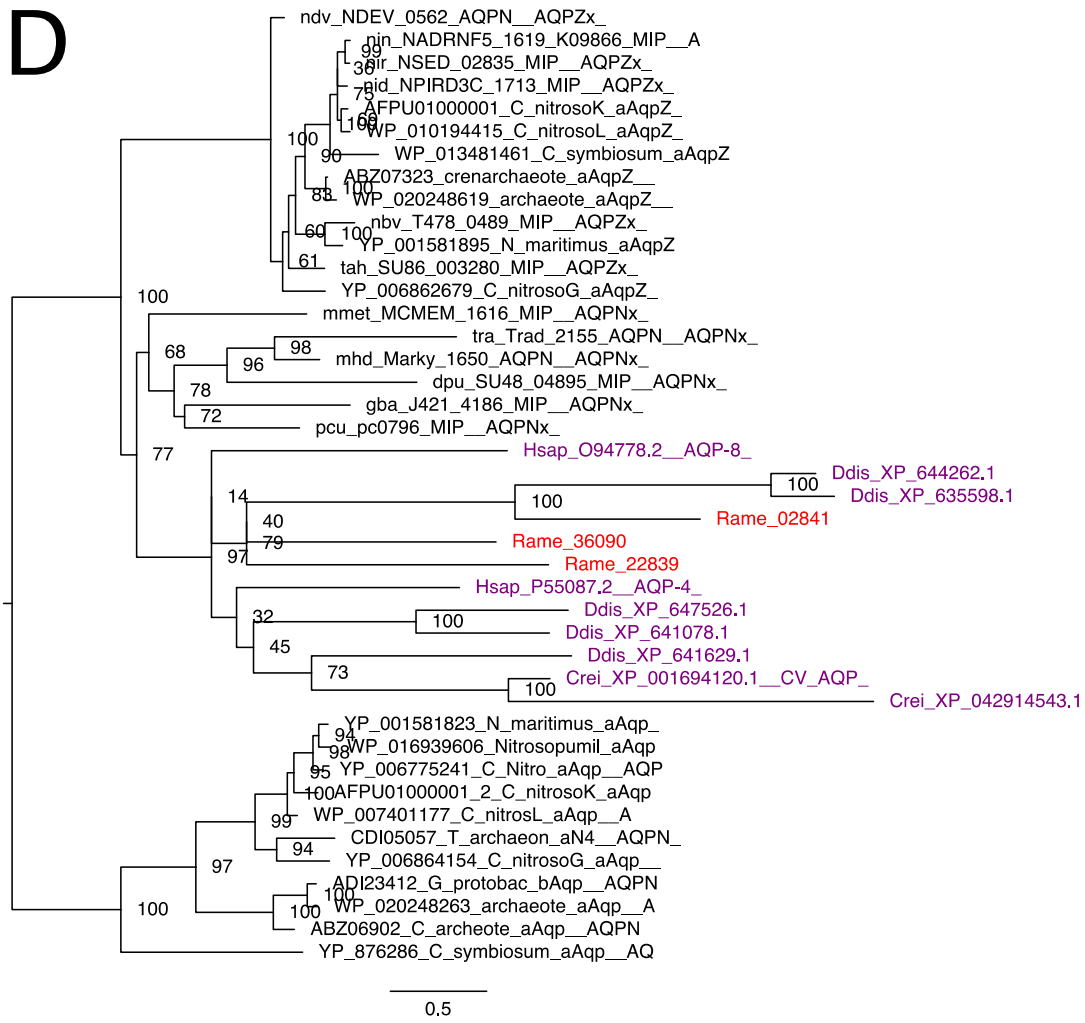


Supplementary Figure 4 Ultrafast bootstrap maximum likelihood (ML) phylogenies for classification of the Rab GTPase Rab8 in *Paramecium tetraurelia*. *P. tetraurelia* sequences are in red. Alignment modified from Eliáš et al. (2012).

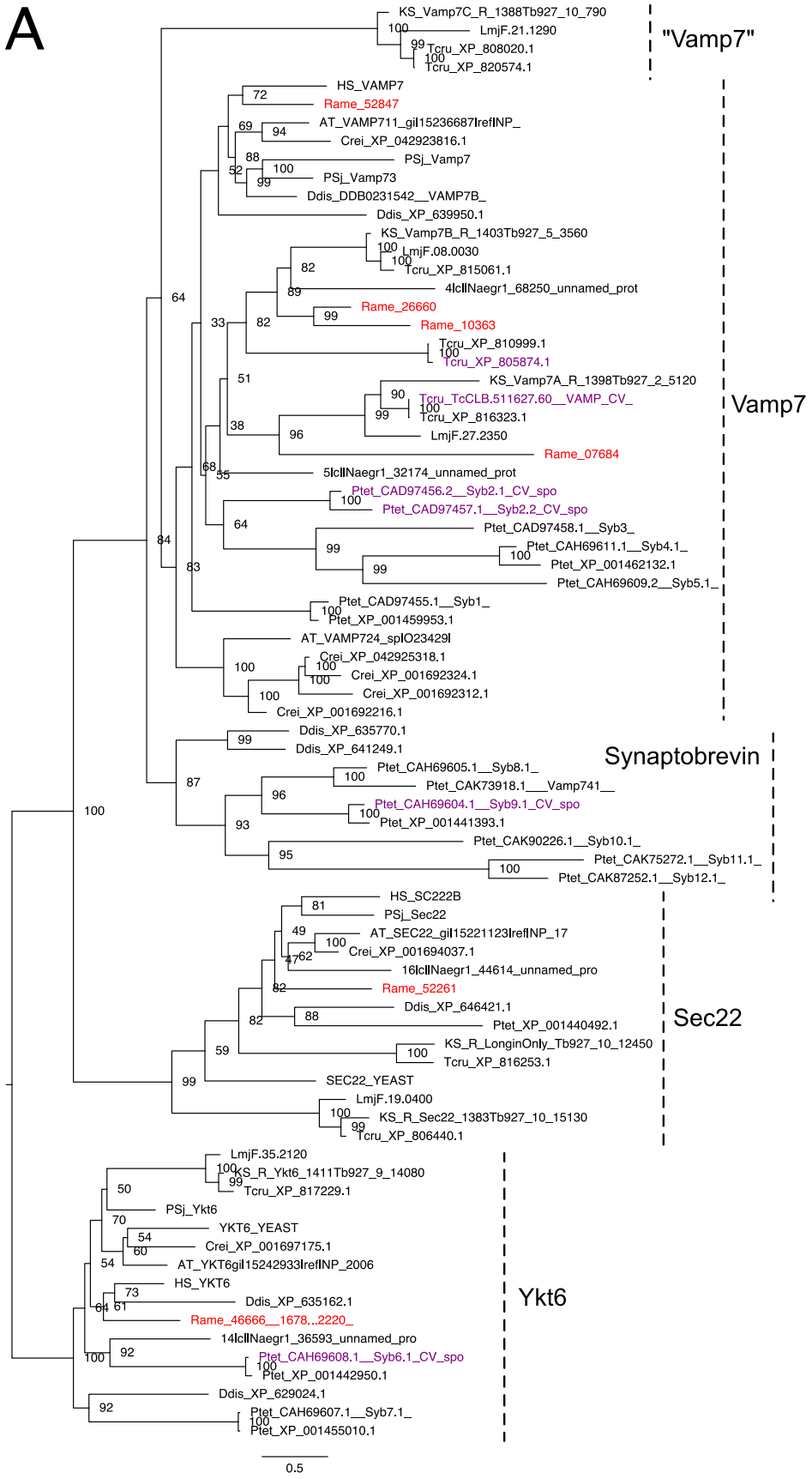

A

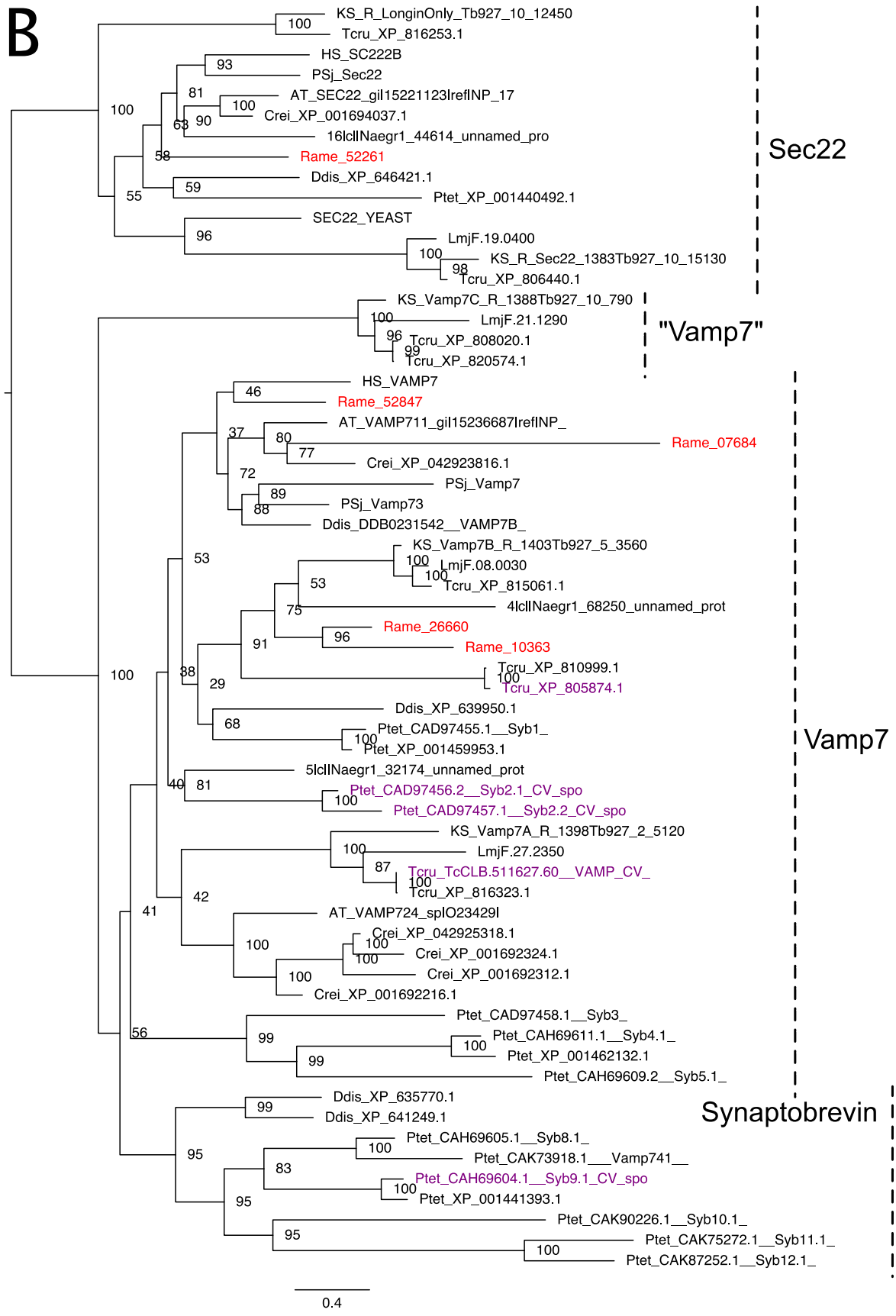

B

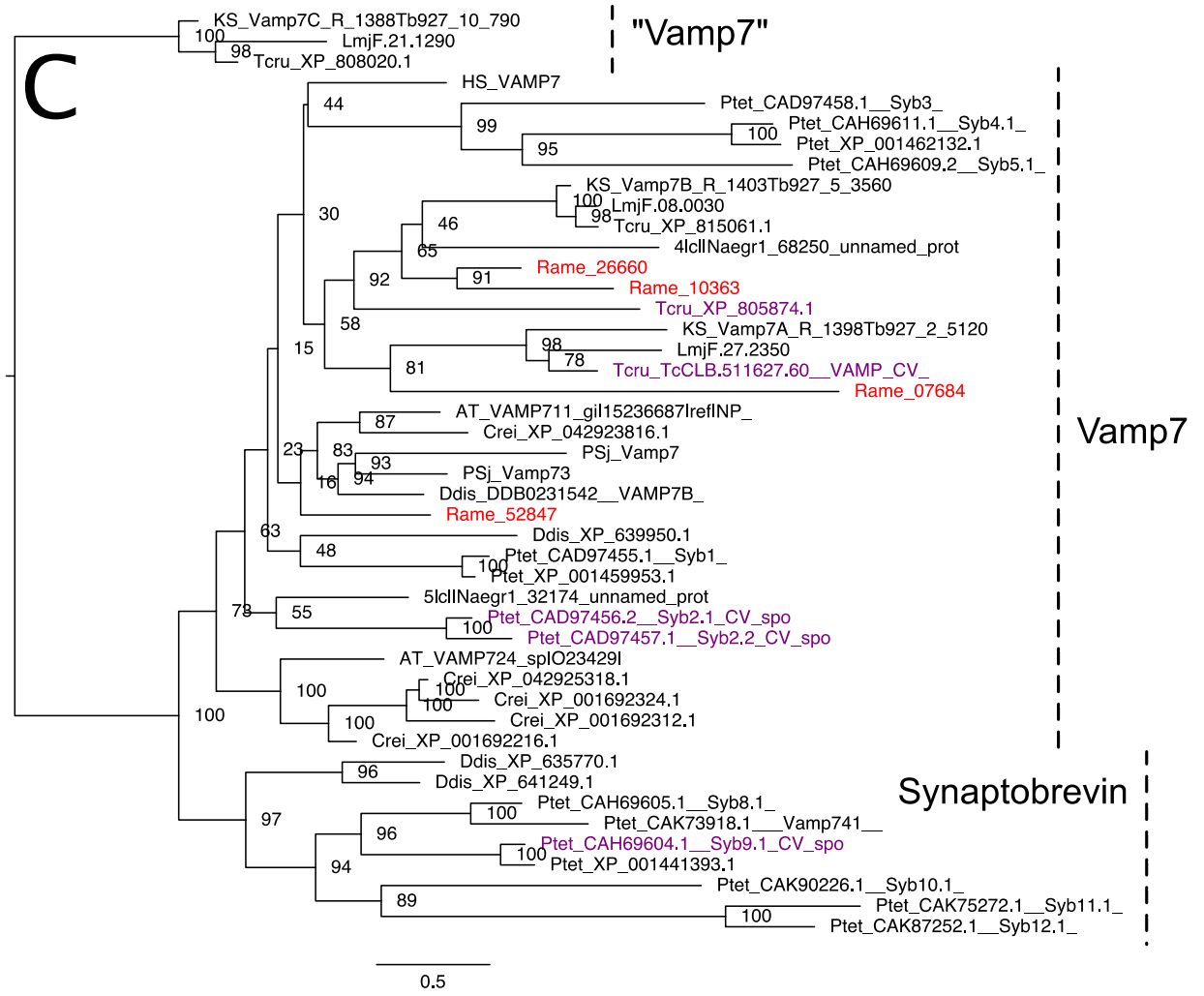

C
Ddis_XP_641629.1
Ddis_XP_644262.1
Ddis_XP_635598.1
Rame_02841
Crei_XP_001694120.1__CV_AQP_
Crei_XP_042914543.1
Ddis_XP_647526.1
Ddis_XP_641078.1
Rame_22839
Hsap_NP_036196__LFMIP_
Hsap_P41181.1__AQP-2_
Hsap_P55064.1__AQP-5_
Hsap_Q13520.2__AQP-6_
Hsap_P29972.3__AQP-1_
Hsap_P55087.2__AQP-4_
Hsap_O94778.2__AQP-8_
Rame_36090
tra_Trad_2155_AQPN__AQPNx_
mhd_Marky_1650_AQPN__AQPNx_
dpu_SU48_04895_MIP__AQPNx_
mmet_MCMEM_1616_MIP__AQPNx_
gba_J421_4186_MIP__AQPNx_
pcu_pc0796_MIP__AQPNx_
ADJ23412_G_protobac_bAqp__AQPN
WP_020248263_archaeote_aAqp__A
ABZ06902_C_archeote_aAqp__AQPN
YP_001581823_N_maritimus_aAqp_
WP_016939606_Nitrosopumil_aAqp
YP_006775241_C_Nitro_aAqp__AQP
WP_007401177_C_nitrosL_aAqp__A
CDI05057_T_archaeon_aN4__AQPN_
YP_006864154_C_nitrosoG_aAqp__
YP_876286_C_symbiosum_aAqp__AQ
ndv_NDEV_0562_AQPN__AQPZx_
nin_NADRNF5_1619_K09866_MIP__A
nir_NSED_02835_MIP__AQPZx_
WP_010194415_C_nitrosoL_aAqpZ_
WP_013481461_C_symbiosum_aAqpZ
nid_NPIRD3C_1713_MIP__AQPZx_
ABZ07323_crenarchaeote_aAqpZ__
WP_020248619_archaeote_aAqpZ__
nbv_T478_0489_MIP__AQPZx_
YP_001581895_N_maritimus_aAqpZ
YP_006862679_C_nitrosoG_aAqpZ_
tah_SU86_003280_MIP__AQPZx_
Ssp._A0A0C2BKY9__AQPX_
Meur_A0A0U1DHN7__AQPX_
Mcan_A0A124E2J9__AQPX_
Mrho_G4HXP8__AQPX_
Sory_A0A4U0SPB8__AQPX_
Ncer_A0A231GZ49__AQPX_
Sare_A8M339__AQPX_
Svio_G2NU56__AQPX_
Mrho_A0A1X0IIY3__AQPX_
Ache_K6V945__AQPX_
Lsp._A0A1R4HL20__AQPX_
Moxy_A0A3S9X9C6__AQPX_
Srug_E5XPS3__AQPX_
Pdok_A0A1H0TZ00__AQPX_
Mbac_A0A2G6DU53__AQPX_
Njej_A0A4R1CHH0__AQPX_
Csp._A0A1M3BB51__AQPX_
Bsp._R5H5Z2__AQPX_
Csp._R5KRN3__AQPX_
Abac_A0A1V6HBT9__AQPX_
Emod_A0A4Q7YET5__AQPX_
Ipal_E8QYB2__AQPX_
Nkor_G8TIV1__AQPX_
Agio_A0A5B9W8Q3__AQPX_
Mfuc_A0A5B9P539__AQPX_
Pbac_A0A5C5W6R5__AQPX_
Rmar_A0A259U211__AQPX_
Nexe_A0A1I2IWR7__AQPX_
Cbac_A0A0R2RZH9__AQPX_
Bsp._A0A1B6Y8P4__AQPX_
Gmen_A0A521CUH3__AQPX_
Bbac_A0A1F3P7I8__AQPX_
Ibac_A0A1G0SPM6__AQPX_
Fbac_A0A424NCZ9__AQPX_
Crei_XP_001690126.2
Hsap_NP_766627__AQP-11_
Hsap_NP_945349__AQP-12A__
Pmul_BAN57367__CV_AQP_
Ptet_XP_001457467.1
Ptet_XP_001438516.1
Ptet_XP_001442648.1
Ptet_XP_001443926.1
Ptet_XP_001445532.1
Ptet_XP_001437824.1
Tcru_XP_818946.1__CV_AQP_
Tcru_XP_815990.1
Tcru_XP_810214.1
Tcru_XP_816487.1
0.7



Supplementary Figure 5 Ultrafast bootstrap maximum likelihood (ML) phylogenies for classification of eukaryotic aquaporins in a prokaryotic context. From the initial phylogeny (**A**), maximally supported clades were iteratively removed to resolve the more divergent eukaryotic clades (**B-D**). *R. americana* sequences are in red, other eukaryotes are in purple. Prokaryotic sequences from Pommerrenig et al. (2020) and Tesan et al. (2021).


A
KS_Vamp7C_R_1388Tb927_10_790
LmjF.21.1290
Tcru_XP_808020.1
Tcru_XP_820574.1
"Vamp7"
HS_VAMP7
Rame_52847
AT_VAMP711_gi|15236687|ref|NP_
Crei_XP_042923816.1
PSj_Vamp7
PSj_Vamp73
Ddis_DDB0231542__VAMP7B_
Ddis_XP_639950.1
KS_Vamp7B_R_1403Tb927_5_3560
LmjF.08.0030
Tcru_XP_815061.1
4|cl|Naegr1_68250_unnamed_prot
Rame_26660
Rame_10363
Tcru_XP_810999.1
Tcru_XP_805874.1
KS_Vamp7A_R_1398Tb927_2_5120
Tcru_TcCLB.511627.60__VAMP_CV_
Tcru_XP_816323.1
LmjF.27.2350
Rame_07684
5|cl|Naegr1_32174_unnamed_prot
Ptet_CAD97456.2__Syb2.1_CV_spo
Ptet_CAD97457.1__Syb2.2_CV_spo
Ptet_CAD97458.1__Syb3_
Ptet_CAH69611.1__Syb4.1_
Ptet_XP_001462132.1
Ptet_CAH69609.2__Syb5.1_
Vamp7
Ptet_CAD97455.1__Syb1_
Ptet_XP_001459953.1
AT_VAMP724_sp|O23429|
Crei_XP_042925318.1
Crei_XP_001692324.1
Crei_XP_001692312.1
Crei_XP_001692216.1
Ddis_XP_635770.1
Ddis_XP_641249.1
Ptet_CAH69605.1__Syb8.1_
Ptet_CAK73918.1___Vamp741__
Ptet_CAH69604.1__Syb9.1_CV_spo
Ptet_XP_001441393.1
Ptet_CAK90226.1__Syb10.1_
Ptet_CAK75272.1__Syb11.1_
Ptet_CAK87252.1__Syb12.1_
Synaptobrevin
HS_SC222B
PSj_Sec22
AT_SEC22_gi|15221123|ref|NP_17
Crei_XP_001694037.1
16|cl|Naegr1_44614_unnamed_pro
Rame_52261
Ddis_XP_646421.1
Ptet_XP_001440492.1
KS_R_LonginOnly_Tb927_10_12450
Tcru_XP_816253.1
SEC22_YEAST
LmjF.19.0400
KS_R_Sec22_1383Tb927_10_15130
Tcru_XP_806440.1
Sec22
LmjF.35.2120
KS_R_Ykt6_1411Tb927_9_14080
Tcru_XP_817229.1
PSj_Ykt6
YKT6_YEAST
Crei_XP_001697175.1
AT_YKT6|gi|15242933|ref|NP_2006
HS_YKT6
Ddis_XP_635162.1
Rame_46666__1678...2220_
14|cl|Naegr1_36593_unnamed_pro
Ptet_CAH69608.1__Syb6.1_CV_spo
Ptet_XP_001442950.1
Ddis_XP_629024.1
Ptet_CAH69607.1__Syb7.1_
Ptet_XP_001455010.1
Ykt6
0.5


B
KS_R_LonginOnly_Tb927_10_12450
Tcru_XP_816253.1
HS_SC222B
PSj_Sec22
AT_SEC22_gil15221123lreflNP_17
Crei_XP_001694037.1
16lclINaegr1_44614_unnamed_pro
Rame_52261
Ddis_XP_646421.1
Ptet_XP_001440492.1
SEC22_YEAST
LmjF.19.0400
KS_R_Sec22_1383Tb927_10_15130
Tcru_XP_806440.1
Sec22
KS_Vamp7C_R_1388Tb927_10_790
LmjF.21.1290
Tcru_XP_808020.1
Tcru_XP_820574.1
"Vamp7"
HS_VAMP7
Rame_52847
AT_VAMP711_gil15236687lreflNP_
Rame_07684
Crei_XP_042923816.1
PSj_Vamp7
PSj_Vamp73
Ddis_DDB0231542__VAMP7B_
KS_Vamp7B_R_1403Tb927_5_3560
LmjF.08.0030
Tcru_XP_815061.1
4lclINaegr1_68250_unnamed_prot
Rame_26660
Rame_10363
Tcru_XP_810999.1
Tcru_XP_805874.1
Ddis_XP_639950.1
Ptet_CAD97455.1__Syb1_
Ptet_XP_001459953.1
5lclINaegr1_32174_unnamed_prot
Ptet_CAD97456.2__Syb2.1_CV_spo
Ptet_CAD97457.1__Syb2.2_CV_spo
KS_Vamp7A_R_1398Tb927_2_5120
LmjF.27.2350
Tcru_TcCLB.511627.60__VAMP_CV_
Tcru_XP_816323.1
AT_VAMP724_spIO23429l
Crei_XP_042925318.1
Crei_XP_001692324.1
Crei_XP_001692312.1
Crei_XP_001692216.1
Vamp7
Ptet_CAD97458.1__Syb3_
Ptet_CAH69611.1__Syb4.1_
Ptet_XP_001462132.1
Ptet_CAH69609.2__Syb5.1_
Ddis_XP_635770.1
Ddis_XP_641249.1
Ptet_CAH69605.1__Syb8.1_
Ptet_CAK73918.1___Vamp741__
Ptet_CAH69604.1__Syb9.1_CV_spo
Ptet_XP_001441393.1
Ptet_CAK90226.1__Syb10.1_
Ptet_CAK75272.1__Syb11.1_
Ptet_CAK87252.1__Syb12.1_
Synaptobrevin
0.4



Supplementary Figure 6 Ultrafast bootstrap maximum likelihood (ML) phylogenies for classification of R SNAREs in contractile vacuole (CV) model organisms and *Reclinomonas americana*. From the initial phylogeny (**A**), maximally supported clades were iteratively removed (**B**, **C**) until Vamp7 + synaptobrevin remained (Figure 19). *R. americana* sequences are in red, other eukaryotes are in purple. Initial alignment from Venkatesh et al. (2017).