

*EVOLUTION OF VESICULAR TRANSPORT IN  
KINETOPLASTIDS: DYNAMICS AND NOVEL  
GENE PRODUCTS*



**Divya Venkatesh**

**Queens' College**

**Department of Pathology**

**University of Cambridge**

**This dissertation is submitted for the degree of Doctor of Philosophy**

**February 2016**



## DECLARATION

This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

In accordance with the requirements for the degree committee for Biology, this thesis does not exceed 60,000 words, and it contains less than 150 figures.

Signed: \_\_\_\_\_

Date: \_\_\_\_\_

Divya Venkatesh

Bangalore, India

## SUMMARY

The membrane trafficking system mediates delivery of macromolecules and metabolites to discrete intracellular compartments from their site of uptake or synthesis. For many pathogens the trafficking system has a special relevance as it is responsible for maintaining the host-pathogen interface, i.e., the cell surface. Both the surface and the underlying trafficking apparatus are intimately connected with immune evasion in many parasites including those belonging to the highly divergent order Kinetoplastida. Kinetoplastid parasites are etiological agents of several neglected tropical diseases such as African sleeping sickness, Chagas disease, and Leishmaniasis. Newly available sequences of many kinetoplastid genomes were used to reconstruct evolution of trafficking across this lineage, using three central paralogous trafficking families: Rabs, SNAREs and Rab-GAPs, which have defined roles in specific trafficking events. Further, proteomics was used to analyse a representative SNARE complex to explore compositional conservation between kinetoplastids and Opisthokonts.

Overall there is little evidence for large scale expansions or contractions of these protein families, excluding a direct association with parasitism or changes to host range, host immunosophistication or transmission mechanisms. The data indicate a stepwise sculpting of the trafficking system where the large repertoire of the basal bodonids is mainly retained by the cruzi group, while extensive losses characterise other lineages, particularly the African trypanosomes and phytomonads. Kinetoplastids possess several lineage-specific Rabs but all retain a core canonical Rab set; by contrast there is little novelty within the SNARE family even though certain canonical endosomal SNAREs appear to show a considerable degree of sequence divergence. Proteomics suggests that SNARE complex composition is largely conserved. The major changes in Rab and SNARE repertoires are associated with endosomal and late exocytic pathways, which is consistent with the considerable evolution of surface proteomes. Therefore, despite the absence of a transition *per se* associated with parasitism, adaptation of membrane trafficking is likely under active selection where it meets the host environment.

## ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Prof. Mark Field for his continual guidance and support. I would also like to thank all the past and present members of the Field Lab for an enjoyable and supportive working environment. In particular, Dr KaFai Leung, for the help and advice, Dr Paul Manna for the TbVAMP7B BSF cell line, Dr Amanda O'Reilly and Dr Luděk Kořený for help with the evolutionary analyses, Dr Cordula Boehm for help with the cryoimmunolisation experiments, Mrs Karen Wolfreys, for making our lab functional, and Dr Luke Maishman for always insisting that 'it's going to be fine'. I would like to thank Dr Steve Kelly for access to the unpublished kinetoplastid genome sequences used in this study. Thank you also to Mrs Jane Duncan, who continues to patiently help with endless administrative requests. Dr Barry Kingston, for support as part of my degree committee, and for being in the building on a fine Saturday morning and releasing me from the basement cold room.

I would like to thank my parents, who made everything possible and supported me all through, as well as David Antoš for being a consistently encouraging and interested non-biologist. I'm grateful to Dr Kavita Atre and (soon-to-be) Dr Goylette Chami for being a source of help and optimism during thesis writing.

This research was funded by the Wellcome trust along with a scholarship from the Gates Cambridge Trust. It was conducted mainly in the Department of Pathology, University of Cambridge, and partially at the Biological Chemistry and Drug Discovery Unit at the University of Dundee. The Lundgren Research Award and the Cambridge Philosophical Society Research Studentship provided financial assistance in my final year.

Finally, thank you to Kayla Friedman and Malcolm Morgan of the Centre for Sustainable Development, University of Cambridge, UK for producing the Microsoft Word thesis template used to produce this document.

# CONTENTS

<b>1</b>	<b>INTRODUCTION .....</b>	<b>1</b>
1.1	MEMBRANE TRAFFICKING .....	1
1.1.1	<i>Key players of the trafficking system .....</i>	<i>2</i>
1.1.2	<i>Evolution of the trafficking system.....</i>	<i>5</i>
1.2	RAB GTPASE, TBC RABGAP AND SNARE PROTEIN FAMILIES.....	9
1.2.1	<i>Mechanism of Rab, TBC and SNARE action .....</i>	<i>10</i>
1.2.2	<i>Evolution of Rabs, TBCs and SNAREs .....</i>	<i>15</i>
1.3	KINETOPLASTIDS .....	22
1.3.1	<i>The complex life cycles of parasitic trypanosomatids .....</i>	<i>25</i>
1.3.2	<i>Strategies of immune evasion.....</i>	<i>28</i>
1.4	CELL AND MOLECULAR BIOLOGY OF KINETOPLASTIDS.....	30
1.4.1	<i>Genome .....</i>	<i>30</i>
1.4.2	<i>Membrane trafficking.....</i>	<i>32</i>
1.5	OBJECTIVES OF THIS STUDY .....	39
<b>2</b>	<b>MATERIALS AND METHODS .....</b>	<b>41</b>
2.1	BIOINFORMATIC APPROACHES .....	41
2.1.1	<i>Sequence data collection.....</i>	<i>41</i>
2.1.2	<i>Phylogenetic reconstruction .....</i>	<i>42</i>
2.2	EXPERIMENTAL APPROACHES .....	44
2.2.1	<i>Trypanosome cell culture.....</i>	<i>44</i>
2.2.2	<i>Nucleic acids and Recombinant DNA methods .....</i>	<i>44</i>
2.2.3	<i>DNA isolation .....</i>	<i>45</i>
2.2.4	<i>RNA isolation.....</i>	<i>45</i>
2.2.5	<i>Polymerase Chain Reaction.....</i>	<i>45</i>
2.2.6	<i>Quantitative reverse transcription PCR.....</i>	<i>45</i>
2.2.7	<i>Preparation of expression constructs .....</i>	<i>48</i>
2.2.8	<i>Transfection of PCF T. brucei .....</i>	<i>50</i>
2.2.9	<i>Immunofluorescence microscopy.....</i>	<i>50</i>
2.2.10	<i>Western immunoblotting.....</i>	<i>51</i>
2.2.11	<i>Proteomics .....</i>	<i>51</i>
<b>3</b>	<b>EVOLUTION OF KINETOPLASTID RABS AND RABGAPS.....</b>	<b>55</b>
3.1	INTRODUCTION .....	55
3.2	ADAPTIVE SHAPING OF TRAFFICKING REPERTOIRES .....	57

3.2.1	<i>Coding content and trafficking repertoires</i> .....	57
3.2.2	<i>Comparing Rabs vs TBCs in kinetoplastids</i> .....	58
3.3	PHYLOGENETIC ANALYSIS OF KINETOPLASTID RABS AND TBCs .....	61
3.3.1	<i>Kinetoplastid Rabs</i> .....	61
3.3.2	<i>Kinetoplastid TBCs</i> .....	70
3.4	PATTERNS IN THE EVOLUTIONARY HISTORY OF RABS AND RAB GAPS .....	77
3.4.1	<i>Secretory Rabs are well conserved</i> .....	80
3.4.2	<i>The GAPs for secretory Rabs</i> .....	84
3.4.3	<i>Primordial endocytic Rabs and their GAPs</i> .....	86
3.4.4	<i>Lysosome and flagella related Rabs and their GAPs</i> .....	89
3.4.5	<i>GAPs of unknown associations</i> .....	94
3.4.6	<i>Orphans and unassigned lineage-specific Rabs and RabGAPs</i> .....	94
3.5	CONCLUSIONS.....	97
<b>4</b>	<b>EVOLUTION OF KINETOPLASTID SNARES .....</b>	<b>99</b>
4.1	INTRODUCTION .....	99
4.2	THE PHYLOGENY OF KINETOPLASTID SNARES AND ASSIGNMENT OF SUB-TYPES 100	
4.2.1	<i>Qa SNAREs</i> .....	103
4.2.2	<i>Qb SNAREs</i> .....	108
4.2.3	<i>Qc and Qbc SNAREs</i> .....	111
4.2.4	<i>R SNAREs</i> .....	116
4.3	SEQUENCE-LEVEL COMPARISON OF SNARES .....	119
4.3.1	<i>SNARE domain</i> .....	119
4.3.2	<i>The C-terminal transmembrane domain</i> .....	121
4.4	PATTERNS IN THE EVOLUTIONARY HISTORY OF SNARES .....	125
4.4.1	<i>ER to Golgi and back</i> .....	127
4.4.2	<i>Exocytic SNAREs</i> .....	128
4.4.3	<i>Endocytic and lysosomal SNAREs</i> .....	132
4.4.4	<i>Longin-only proteins</i> .....	134
4.4.5	<i>SNAREs by sub-cellular localisation</i> .....	136
4.5	CONCLUSIONS.....	139
<b>5</b>	<b>INTERACTOME ANALYSIS OF SNARES IN <i>T. BRUCEI</i> .....</b>	<b>141</b>
5.1	INTRODUCTION .....	141
5.2	DEVELOPMENTAL REGULATION OF SNARES.....	142
5.3	TAGGING AND LOCALISATION OF SELECT SNARES .....	145

5.3.1	<i>Challenges in tagging T. brucei SNAREs</i> .....	146
5.3.2	<i>T. brucei R-SNARE localisations are conserved</i> .....	150
5.3.3	<i>VAMP7 is associated with Golgi duplication</i> .....	159
5.3.4	<i>Localisation of Qa-SNARE TbSynPM</i> .....	160
5.4	PROTEOMICS .....	163
5.4.1	<i>SNARE interactions of TbVAMP7C</i> .....	168
5.4.2	<i>Interactions of TbVAMP7B</i> .....	179
5.4.3	<i>Others</i> .....	181
5.5	CONCLUSIONS .....	184
<b>6</b>	<b>CONCLUSIONS</b> .....	<b>187</b>
<b>7</b>	<b>REFERENCES</b> .....	<b>193</b>



## LIST OF TABLES

<b>Table 1.1 Summary of the diversity of lifestyles of kinetoplastids and relatives ....</b>	<b>24</b>
<b>Table 3.1 Lineage-specific Rabs.....</b>	<b>69</b>
<b>Table 3.2 Comparison of kinetoplastid and LECA Rab and TBC repertoires.....</b>	<b>77</b>
<b>Table 4.1 Similarity of duplicated Qa-SNARE sequences .....</b>	<b>107</b>
<b>Table 4.2 Comparison of sequence identity between Npsn paralogues .....</b>	<b>111</b>
<b>Table 4.3 Comparison of kinetoplastid and LECA SNARE repertoires .....</b>	<b>125</b>
<b>Table 5.1 Summary attempts to tag TbSNAREs <i>in-situ</i>.....</b>	<b>147</b>
<b>Table 5.2 Methodological differences between <i>T. brucei</i> life-cycle stages.....</b>	<b>149</b>
<b>Table 5.3 Tagging of SNAREs in procyclic form trypanosomes .....</b>	<b>150</b>
<b>Table 5.4 List of putative interacting proteins - negative control .....</b>	<b>170</b>
<b>Table 5.5 List of putative interacting proteins – TbVAMP7C::HA.....</b>	<b>171</b>
<b>Table 5.6 List of putative interacting proteins (TbVAMP7C minus wild-type) ...</b>	<b>172</b>
<b>Table 5.7 Proteins identified as potential interactors for TbVAMP7C .....</b>	<b>174</b>
<b>Table 5.8 Proteins identified as potential interactors for TbVAMP7B .....</b>	<b>181</b>
<b>Table 5.9 Proteins identified as potential interactors of TbVAMP7C .....</b>	<b>182</b>

## LIST OF FIGURES

<b>Figure 1.1 Schematic of the endomembrane system</b> .....	2
<b>Figure 1.2 Steps involved in vesicle transport</b> .....	4
<b>Figure 1.3 Evolution of molecular machinery of trafficking</b> .....	8
<b>Figure 1.4 Structure and mechanism of Rab GTPase and TBC Rab GAP action</b> .	11
<b>Figure 1.5 Mechanism of SNARE-mediated fusion</b> .....	14
<b>Figure 1.6 Evolutionary relationships among kinetoplastids</b> .....	23
<b>Figure 1.7 Life cycle of trypanosomatids</b> .....	26
<b>Figure 1.8 Schematic of trypanosomatid ultrastructure</b> .....	33
<b>Figure 3.1 Representation of SNARE, Rab and TBC coding sequences in selected eukaryotic genomes and kinetoplastids.</b> .....	58
<b>Figure 3.2 Ratio of the number of Rabs:TBC in kinetoplastids and select eukaryotes</b> .....	59
<b>Figure 3.3 Topology of the kinetoplastid Rab family</b> .....	62
<b>Figure 3.4 Kinetoplastid Rab phylogeny – Part1</b> .....	63
<b>Figure 3.5 Kinetoplastid Rab phylogeny – Part2</b> .....	64
<b>Figure 3.6 Kinetoplastid Rab phylogeny – Part3</b> .....	65
<b>Figure 3.7 Kinetoplastid Rab phylogeny – Part4</b> .....	66
<b>Figure 3.8 Assignment of kinetoplastid Rabs</b> .....	68
<b>Figure 3.9 Topology of kinetoplastid TBC family</b> .....	71
<b>Figure 3.10 Kinetoplastid TBC phylogeny – Part1</b> .....	72
<b>Figure 3.11 Kinetoplastid TBC phylogeny – Part2</b> .....	73
<b>Figure 3.12 Kinetoplastid TBC phylogeny – Part3</b> .....	74
<b>Figure 3.13 Assignment of kinetoplastid TBCs</b> .....	75
<b>Figure 3.14 Representation of Rabs in kinetoplastids</b> .....	78
<b>Figure 3.15 Representation of TBCs in kinetoplastids</b> .....	79
<b>Figure 3.16 Phylogenetic assignment of Rab1/18-related sequences</b> .....	82

<b>Figure 3.17 Evolutionary history of kinetoplastid Rab5 .....</b>	<b>87</b>
<b>Figure 3.18 Evolutionary history of kinetoplastid Rab32 .....</b>	<b>92</b>
<b>Figure 3.19 Phylogenetic analysis of TBC-ExA subfamily .....</b>	<b>96</b>
<b>Figure 4.1 Topology of kinetoplastid SNARE family .....</b>	<b>100</b>
<b>Figure 4.2 Topology of human, yeast and trypanosome SNARE families.....</b>	<b>102</b>
<b>Figure 4.3 Assignment of SNAREs to intracellular pathways.....</b>	<b>103</b>
<b>Figure 4.4 Assignment of kinetoplastid Qa-SNAREs .....</b>	<b>105</b>
<b>Figure 4.5 Phylogeny of kinetoplastid Qa SNAREs.....</b>	<b>106</b>
<b>Figure 4.6 Assignment of kinetoplastid Qb-SNAREs.....</b>	<b>109</b>
<b>Figure 4.7 Kinetoplastid Qb-SNARE phylogeny .....</b>	<b>110</b>
<b>Figure 4.8 Assignment of kinetoplastid Qc- and Qbc-SNAREs .....</b>	<b>113</b>
<b>Figure 4.9 Kinetoplastid Qc- and Qbc-SNARE phylogeny.....</b>	<b>114</b>
<b>Figure 4.10 Kinetoplastid Qbc-SNAREs .....</b>	<b>115</b>
<b>Figure 4.11 Assignment of kinetoplastid R-SNAREs .....</b>	<b>117</b>
<b>Figure 4.12 Kinetoplastid R-SNARE phylogeny.....</b>	<b>118</b>
<b>Figure 4.13 Variability of the kinetoplastid SNARE domain .....</b>	<b>120</b>
<b>Figure 4.14 Pattern of TM domain loss in SynPM and Syx16.....</b>	<b>122</b>
<b>Figure 4.15 Comparing L. major SynPM1 sequences .....</b>	<b>123</b>
<b>Figure 4.16 Representation of SNAREs in kinetoplastids.....</b>	<b>126</b>
<b>Figure 4.17 Phylogenetic analysis of eukaryotic Qbc SNAP-25 like SNAREs .....</b>	<b>129</b>
<b>Figure 4.18 Phylogenetic analysis of eukaryotic ‘synaptobrevin’ domain containing sequences.....</b>	<b>131</b>
<b>Figure 4.19 Phylogenetic analysis of non-SNARE longins.....</b>	<b>136</b>
<b>Figure 4.20 SNAREs by predicted sub-cellular location.....</b>	<b>137</b>
<b>Figure 5.1 Developmental regulation of T. brucei SNAREs .....</b>	<b>144</b>
<b>Figure 5.2 Validation of tagged VAMP7A, B, C and Ykt6 in PCF .....</b>	<b>153</b>
<b>Figure 5.3 Sub-cellular localisation of PCF R-SNARE TbVAMP7A .....</b>	<b>154</b>

<b>Figure 5.4 Sub-cellular localisation of PCF R-SNARE TbVAMP7B.....</b>	<b>155</b>
<b>Figure 5.5 Sub-cellular localisation of PCF R-SNARE TbVAMP7C .....</b>	<b>156</b>
<b>Figure 5.6 Sub-cellular localisation of PCF R-SNARE TbYkt6.....</b>	<b>157</b>
<b>Figure 5.7 Validation of tagged R-SNARE TbVAMP7B in BSF .....</b>	<b>158</b>
<b>Figure 5.8 Validation of tagged Qa-SNARE TbSynPM in PCF.....</b>	<b>162</b>
<b>Figure 5.9 Optimising immunoisolation of tagged SNAREs.....</b>	<b>164</b>
<b>Figure 5.10 Testing buffer conditions for extraction of TbVAMP7C::HA .....</b>	<b>165</b>
<b>Figure 5.11 Testing buffer conditions for extraction of 2HA::TbYkt6.....</b>	<b>166</b>
<b>Figure 5.12 Validation of tagging of putative SNARE interactors of TbVAMP7C .....</b>	<b>177</b>
<b>Figure 5.13 Validation of putative non-SNARE interactors of TbVAMP7C.....</b>	<b>179</b>
<b>Figure 6.1 Reconstruction of birth and death of SNARE, Rab, and TBC subfamilies in kinetoplastids .....</b>	<b>189</b>

## LIST OF ABBREVIATIONS AND ACRONYMS

AP	adaptor protein complexes
Arf	ADP ribosylation factor
BSF	bloodstream form
CCV	clathrin coated vesicles
CNS	central nervous system
EE	early endosomes
ER	endoplasmic reticulum
ENTH	Epsin N-terminal homology domain
FP	flagellar pocket
ILV	intra-luminal vesicles
ISG	invariant surface glycoprotein
LE	late endosomes
LECA	last eukaryotic common ancestor
LRO	lysosome related organelle
MVB	multi-vesicular bodies
OPH	organelle paralogy hypothesis
PM	plasma membrane
PTU	poly-cistronic transcription unit
Rab	Ras-related in brain
RabGAPs	Rab-GTPase-activating proteins
SNAREs	Soluble N-ethylmaleimide-sensitive factor attachment protein receptors
SS	short stumpy
TBC	Tre-2/Bub2/Cdc16
TGN	<i>trans</i> -Golgi network
TM	transmembrane
VSG	variant surface glycoprotein



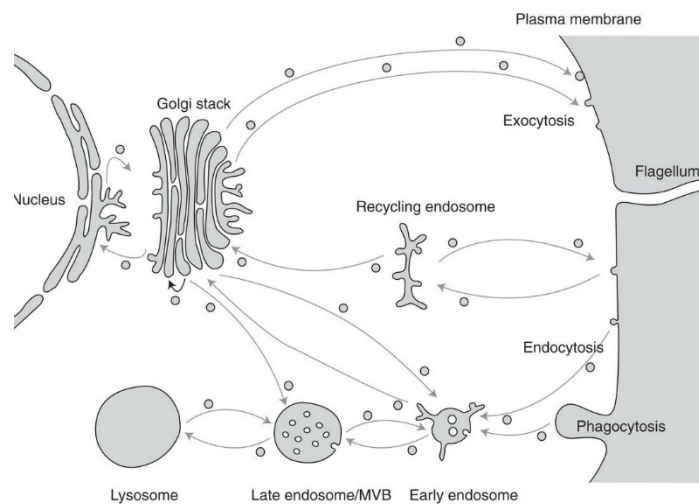
# 1 INTRODUCTION

## 1.1 Membrane Trafficking

The membrane trafficking system mediates delivery of macromolecules and metabolites to discrete intracellular compartments from their site of uptake or synthesis. Trafficking is essential to nearly all eukaryotic cells, contributing towards nutrient acquisition, protein processing and turnover, and compartmental division of labour. In multicellular organisms, trafficking participates in higher order cellular organisation and communication. This importance is reflected in the many diseases associated with trafficking, including diabetes, Alzheimer's and cystic fibrosis (Olkkonen and Ikonen, 2006; Rajendran and Annaert, 2012; Birault *et al.*, 2013; Seixas *et al.*, 2013). Moreover, the development of membrane trafficking is a major evolutionary feature that in part enabled the transition from prokaryotic to eukaryotic cells (de Duve, 2007; Schlacht *et al.*, 2014). For many pathogens, the trafficking system has a special relevance, as it is responsible for maintaining the host-parasite interface, i.e., the cell surface. Both the surface and the underlying trafficking apparatus are intimately connected with immune evasion, pathogenesis, and life cycle progression.

Extracellular substances too large to pass through the cell membrane via channels, components of the cell surface, and ligands attached to receptors on cell surface are internalised via membrane-bound vesicles in a process known as endo- or phagocytosis. Eukaryotic cells contain several internal compartments or 'organelles', schematically presented in **Figure 1.1** below. Newly synthesized proteins are transported through the endoplasmic reticulum (ER) where they are folded, modified

(glycosylation, cleavage), and assembled (Lodish *et al.*, 2000b). From the ER, they are carried away as cargo by vesicles which may fuse with other vesicles and the ERGIC (ER-Golgi intermediate compartment) before they reach the *cis*-face of the Golgi complex. As the proteins pass through the stacks of the Golgi they undergo further post-translational modifications such as glycosylation and proteolytic cleavage (Lodish *et al.*, 2000a, 2000c). At the *trans*-face, they bud off as vesicles and are sorted for further transport to various organelles via an elaborate network of vesicles called the *trans*-Golgi network (TGN). From this point, the cargo may be transported in several directions: anterograde to the plasma membrane (PM) for secretion (exocytosis), retrograde back through the Golgi or towards the ER, merge with the early endosomes (EE) from the PM, or late endosomes (LE) heading for the lysosome. These organelles are largely conserved throughout the eukaryotes although their morphology may show differences that are not necessarily correlated with phylogeny.



**Figure 1.1 Schematic of the endomembrane system**

Various organelles of the eukaryotic endomembrane system are presented in a generalised eukaryotic cell.

### 1.1.1 Key players of the trafficking system

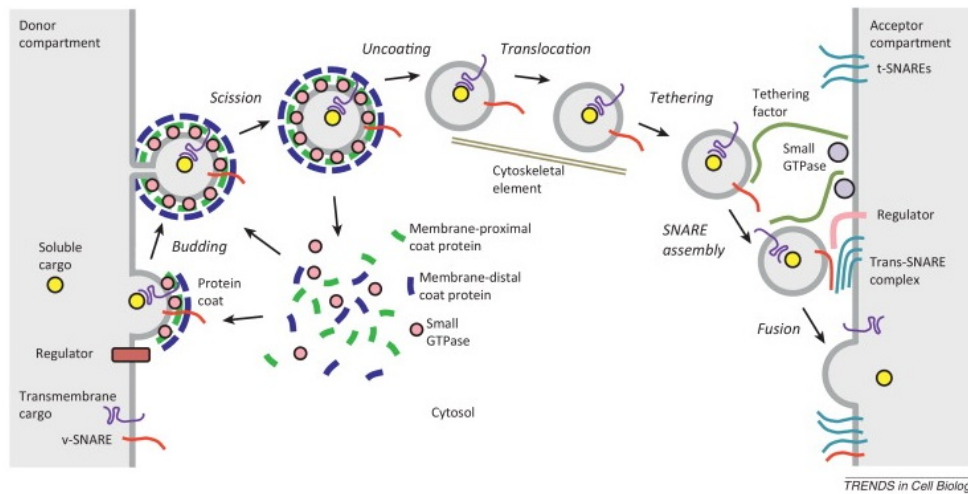
In 1975, George Palade and colleagues first proposed that newly synthesized proteins pass through a series of membrane enclosed organelles - the ER, Golgi complex and secretory granules, on their way out of the cell; and that these proteins



are found enclosed in small vesicles interspersed among the organelles along the pathway (Palade, 1975). This gave rise to the vesicle transport hypothesis, an updated version of which still holds. Thus, transport is executed in several steps, specifically vesicle formation, translocation and tethering at the destination membrane, docking, and finally fusion to release cargo (See **Figure 1.2** below). This requires the coordinated action of many protein families, including Rab (Ras-related in brain) and Arf (ADP ribosylation factor) GTPases, coat complexes, tethers, and SNAREs (soluble N-ethylmaleimide-sensitive factor attachment protein receptors). Members of these paralogous families encode combinatorial specificity for individual transport events and define organelles (Dacks *et al.*, 2009).

**Vesicle formation:** Specific phosphoinositides or membrane GTPases recruit coat proteins to the nascent vesicle or bud site. Cargo proteins may enter the vesicle by bulk flow, or in interaction with coat proteins or adaptor proteins (Dacks and Field, 2004). These are likely to include trafficking proteins required for mediating subsequent steps of the process such as SNAREs. Polymerisation of the coat proteins deforms the membrane to form a bulge, which eventually buds off. The mechanism of budding is largely the same in all parts of the pathway, but the individual protein components vary. Vesicles budding off the ER towards the Golgi are formed by the Sar1 GTPase and COPII coat complex, while those undergoing retrograde transport within the Golgi and from the Golgi towards ER are formed by the Arf GTPase and COPI coat complex (Bednarek *et al.*, 1995). Vesicles mediating post-Golgi transport use the clathrin coat, which is also recruited by the Arf GTPase or specific phosphoinositides. Clathrin adaptors include monomeric proteins such as GGAs (Golgi-localised, gamma ear-containing, Arf-binding proteins) and Epsin1 (an Epsin N-terminal homology (ENTH) domain containing protein), as well as multimeric proteins such as the adaptor protein complexes AP1-5 (Bonifacino and Lippincott-Schwartz, 2003). Several kinases and phosphatases regulate clathrin and the process requires accessory factors such as the dynamin GTPase for scission of the budding vesicle. Scission likely involves constriction of dynamin's ring like structure around the membranes, although the exact mechanism is still debated (Roux, 2014). Clathrin also uses chaperone-mediated uncoating involving the ATPase Hsc70 and its co-factor auxilin (Sousa and Lafer, 2015). Thus, clathrin-mediated transport is more

complex than the COP coats, reflecting the greater diversity of post-Golgi trafficking routes in eukaryotic cells.



**Figure 1.2 Steps involved in vesicle transport**

A generic representation of the steps involved in vesicle transport. *Vesicle budding.* Membrane-proximal coat components recruited to the donor compartment by small GTPase or phosphoinositide enable membrane deformation and cargo begins to assemble at the site. Membrane-distal coat components assemble and polymerise, increasing the curvature of the nascent vesicle, where cargo is concentrated. *Scission.* Coat or accessory proteins cause severance of the ‘neck’ of the vesicle from the donor compartment. *Uncoating.* Inactivation of small GTPase, hydrolysis of phosphoinositide, or uncoating enzymes act to remove the coat, whose components are recycled for vesicle formation. *Translocation.* ‘Naked’ vesicle is guided to destination compartment by cytoskeleton and attaches to it by action of Rabs and tethering factors. (6) The v/R and t/Q SNAREs assemble into a four-helix bundle. (7) The ‘trans-SNARE complex’ enables fusion of vesicle and target membrane, releasing the cargo in to acceptor compartment and SNAREs are recycled for further rounds of fusion. Figure is taken from (Bonifacino, 2014).

**Vesicle translocation, tethering, and fusion:** Newly formed vesicles are transported by diffusion or motor mediated transport via the cytoskeleton. Implicated molecular motors include kinesin, dynein and myosin (Cramer, 2008). Rab GTPases aid in the assembly of tethers, which are usually multi-subunit protein complexes or long coiled-coil proteins that mediate the initial association of the vesicle and target membrane. Like coat proteins, tethers also show specificity in their localisation and function. For instance, the TRAPP-I complex acts on the ER to Golgi pathway, while TRAPP-II acts on endosome to Golgi and events within the Golgi, and the HOPS complex functions between the endosome and vacuole (Whyte and Munro, 2002). Uncoating of the vesicle is associated with the Arf or Rab-GAP (GTPase activating protein) mediated hydrolysis of GTP in the GTPase to GDP, although a causal relationship has not been demonstrated (Tanigawa, 1993). Thus vesicles reach their target organelle membrane, and once apposed to each other, their fusion is mediated

by SNARE proteins, thus releasing vesicle contents into the target organelle (Bonifacino, 2004).

### 1.1.2 Evolution of the trafficking system

Most of what we know about the trafficking system as described above comes from studies in yeast and mammalian systems, which dominate our understanding of the biology of eukaryotes. However yeast and mammals belong to only one of five eukaryotic supergroups, which are listed below (*sensu* Adl *et al.* (Adl *et al.*, 2005) except for SAR-CCTH grouping which is according to (Burki *et al.*, 2009)):

1. Opisthokonta (animals, fungi and their single cell relatives)
2. Amoebozoa (amoebae and slime moulds)
3. Archeplastida (land plants and algae)
4. Excavata
5. SAR – CCTH
  - Stramenophiles (brown algae, oomycetes),
  - Alveolata (ciliates, dinoflagellates and apicomplexans)
  - Rhizaria (diverse - Radiolaria, Foraminifera, and Cercozoa)
  - CCTH (cryptomonads, centrohelids, telonemids, and haptophytes)

Our understanding of the ‘general’ principles of cell biology would therefore be incomplete without considering the breadth of eukaryotic diversity. Post-completion of the human genome project and especially with the development of next-generation sequencing, a great number of genomes from diverse organisms have been sequenced. These resources along with improved phylogenetic tools have made possible a much more detailed analysis of the diversity of extant eukaryotes, the origins of eukaryogenesis, and the evolution of the trafficking system.

Compartmentalisation exists in all domains of life, but it is distinctly complex in eukaryotes which all appear to be derived from a highly compartmentalised cell, having retained the basic cell plan in all lineages (Diekmann and Pereira-Leal, 2013). Even though simple internal membrane systems and endosymbionts have been described in a few prokaryotes, they appear to be phylogenetically scattered, arising

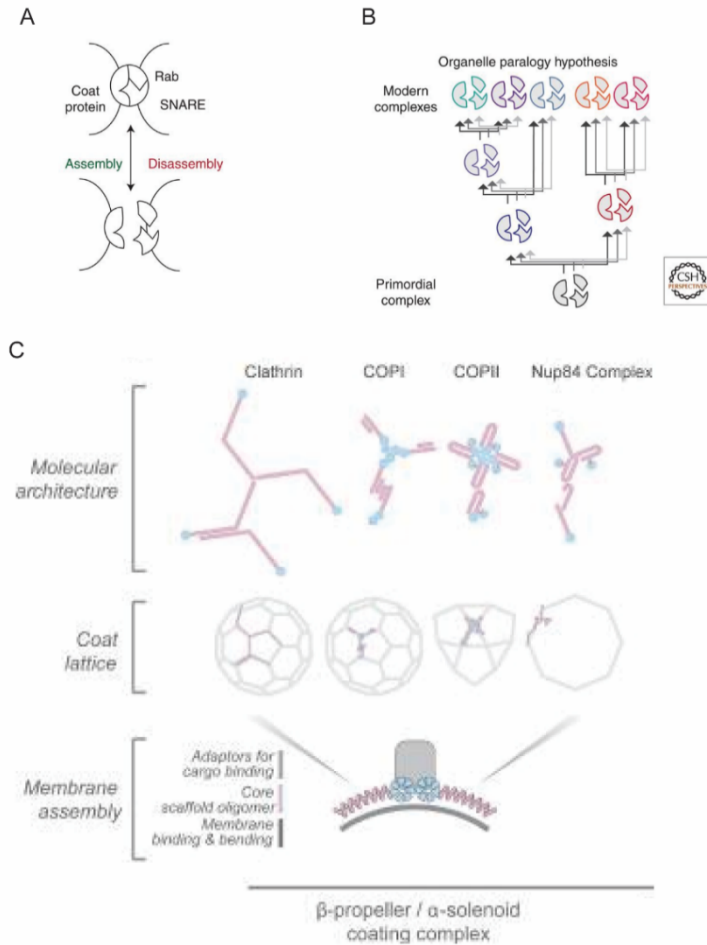
independently in the different lineages. Besides, prokaryotes usually possess only one type of internal compartment and lack vesicular trafficking *sensu stricto* (Diekmann and Pereira-Leal, 2013). Nevertheless, candidate prokaryotic homologues of some trafficking proteins have been identified. These proteins have little or no sequence identity with their eukaryotic counterparts but show parallels in terms of their functions and 3D structures. For example there are structural parallels between bacterial V4R and the Bet3 subunit of the TRAPPI complex (Podar *et al.*, 2008), and MPT63 of *Mycobacterium tuberculosis* and eukaryotic adaptins (Goulding *et al.*, 2002). A Ras-like GTPase (Mg1A) and its cognate GAP (Mg1B) have also been described in several bacterial and archaeal lineages (Hartzell, 1997; Koonin and Aravind, 2000). Unlike mitochondria and plastid organelles with clear parallels to  $\alpha$ -proteobacteria and cyanobacteria respectively, unequivocal prokaryotic homologues for organelles or proteins involved in membrane trafficking are limited. In addition, these organelles lack the characteristic traits of endosymbionts such as addressed double membranes, associated genomes, and autonomous replication. These observations together suggest autogenous, rather than endosymbiotic origins for the eukaryotic membrane trafficking system (Dacks and Field, 2007).

Comparative genomics, which can identify genes common between different genomes, has facilitated the reconstruction of the evolutionary history of eukaryotes. Studies indicate that molecular determinants of the endomembrane and other cellular systems are conserved in organisms of all of the main eukaryotic lineages suggesting the existence of a 'last eukaryotic common ancestor' LECA that was remarkably complex and possessed all the major cellular systems from the endomembrane system to meiosis (Koumandou *et al.*, 2013). At the organelle level, there is evidence that LECA possessed the endoplasmic reticulum, plasma membrane, multi-vesicular bodies (MVB), stacked Golgi apparatus as well as recycling and degradative endocytic routes of the trafficking system (Roger, 1999; Dacks *et al.*, 2003; Leung *et al.*, 2008). At the molecular level, near-complete complements of all major trafficking protein families such as GTPases of the Rab and Arf families, vesicle coats, adaptors, tethers and SNAREs appear to be present (Dacks and Field, 2007).

Such a deep level of conservation, together with the observation that in many trafficking protein families, each member performs functions similar to other members, but at a specific organelle or trafficking pathway, have led to a theory

explaining the autogenous evolution of the endomembrane system. Called the organelle paralogy hypothesis (OPH, **Figure 1.3**), it postulates that novel autogenous organelles arose as a result of gene duplication and neo-functionalisation of existing trafficking machinery (Field *et al.*, 2007a; Dacks *et al.*, 2008). Computer simulations based on theoretical calculations of protein-protein interactions and evolution of specificity among paralogues have also confirmed that such a mechanism could generate new organelles (Ramadas and Thattai, 2013). This model is further supported by the protocoatmer hypothesis (DeGrasse *et al.*, 2009). This hypothesis recognises shared architectures amongst proteins constituting multiple vesicle coat complexes (clathrin/adaptin, COPI, COPII), the nuclear pore complex, and intraflagellar transport, which share  $\beta$ -propeller or  $\alpha$ -solenoid building blocks (or a combination of both), and postulates that these different components arose from a common ancestral ‘protocoatmer’ (See **Figure 1.3** below for a visual representation).

Thus, paralogous expansion and diversification would explain how the complex configuration of LECA was achieved. An example of how this works can be seen in the emergence of a differentiated Golgi complex with *cis* and *trans* compartments (Dacks and Field, 2007). The coats servicing the *cis* and *trans* Golgi compartments, namely the hetero-tetrameric F-COP coatmer subcomplex and adaptins respectively, were found to be products of a set of gene duplications (Schledzewski *et al.*, 1999). Thus, the development of *cis* and *trans* Golgi subcompartments from an undifferentiated organelle was likely concurrent with or causal to the duplications that produced the two coat complexes.



**Figure 1.3 Evolution of molecular machinery of trafficking**

(A) Simplified depiction of components of vesicle formation and fusion machinery that act cooperatively and encode organelle specificity. (B) A hypothetical phylogeny shows rise of complexity via duplication and divergence of identity encoding genes, the organelle paralogy hypothesis (OPH). Taken from (Schlacht *et al.*, 2014) (C) Schematic representation of shared architectures of membrane deforming complexes composed of  $\beta$ -propeller (blue) or  $\alpha$ -solenoid (pink) building blocks, leading to the postulation of a common ancestor, the protocoatomer hypothesis. Taken from (Field *et al.*, 2011).

While paralogous expansion appears to be the dominant mode of evolution in the trafficking system, there are exceptions. Among the ESCRT sub-complexes in LECA, components of ESCRT I and II subunits show no homology amongst each other, even though subunits of ESCRT III and III-like appear to be derived from paralogous expansion (Dacks *et al.*, 2009). Multi-subunit tethering complexes (MTCs), most of whose subunits are predicted to have already existed in LECA, show no homology between the complexes. Thus, accretion of non-homologous components is also a mechanism for the evolution of the trafficking system (Dacks *et al.*, 2009).

Post-LECA paralogous expansion and diversification continued to be one of the dominant modes of evolution. There is evidence for both acquisition of significant complexity (e.g., in vascular plants, metazoan lineages, several protists) as well as decrease in complexity (many fungi, some algae, most kinetoplastids, apicomplexans and others) (Dacks and Field, 2004). The decrease is potentially either because of parasitism, where some functionality can be ‘offloaded’ into the host or because of extreme adaptation to specific environments. This finding overturned inferences of the initial rRNA phylogenetic studies (Sogin, 1991; Van de Peer and De Wachter, 1997), which seemed to indicate an ever-increasing complexity as eukaryotes evolved into a ‘crown’ group of multicellular organisms.

Molecular examples of lineage-specific innovation include components of clathrin-mediated endocytosis and the ESCRT late endosomal complex, with many gene products restricted to metazoan taxa (Field *et al.*, 2007a; Leung *et al.*, 2008). Others, such as sortilin paralogues in *Tetrahymena thermophila* and Rab5 in trypanosomatids, that, despite retaining similar functions to their metazoan counterparts, are non-orthologous but have evolved convergently *via* paralogous expansion from their respective ancestor proteins (Field *et al.*, 1998; Briguglio *et al.*, 2013). Arf GTPases also exhibit lineage-specific expansions (Rojas *et al.*, 2012). Finally, adaptin complexes, important cargo selectors, despite being widely conserved are frequently subject to secondary loss (Nevin and Dacks, 2009; Manna *et al.*, 2013). In the case of African trypanosomes loss of AP-2 is coupled to the origins of antigenic variation and the need for a highly rapid endocytic system (Manna *et al.*, 2014). Moreover, many trafficking proteins demonstrate patchy distribution, indicating multiple secondary losses during evolutionary history (Schlacht *et al.*, 2014). These expansions and losses may be viewed as probable adaptations to specific evolutionary pressure.

## 1.2 Rab GTPase, TBC RabGAP and SNARE protein families

The Rab, TBC RabGAP and SNARE protein families are three of the central protein families whose members function at most of the different trafficking pathways in a cell. These families are the subject of analyses in this study. Hence, this section

elaborates on the mechanism of their action and their evolutionary history across eukaryotes.

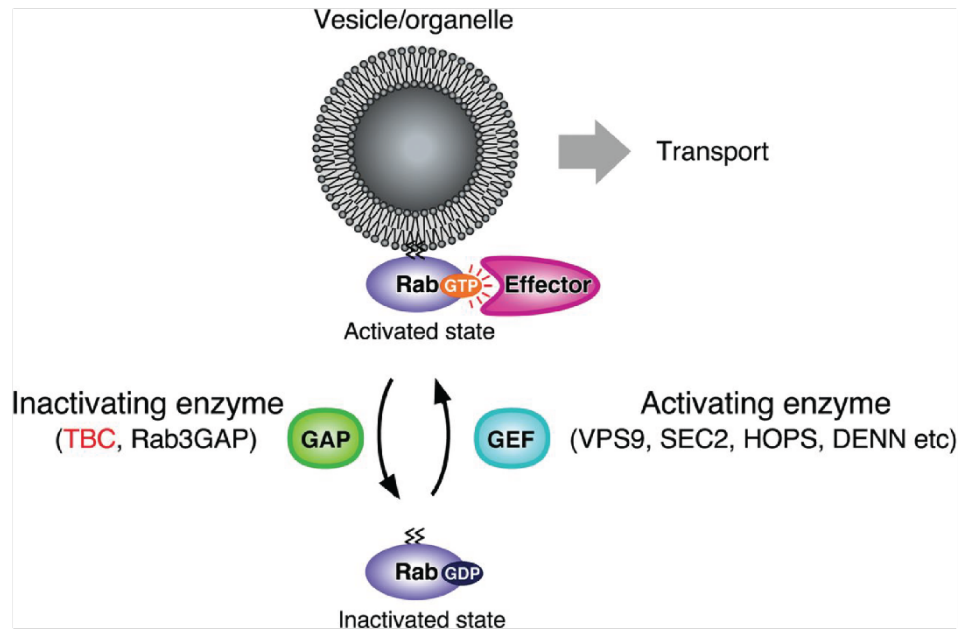
### 1.2.1 Mechanism of Rab, TBC and SNARE action

Rab GTPases belong to the Ras-superfamily of GTP hydrolases whose activity is regulated by the GTP/GDP binding states. They function as molecular switches, cycling between GTP bound active states and GDP bound inactive states. Active Rabs can recruit specific binding partners such as adaptor proteins, tethers, SM proteins, kinases, phosphatases and motor proteins which then execute vesicle formation, transport, tethering and fusion (Stenmark, 2009). Rabs are small proteins (20-25kDa) whose core structure comprises of the six-stranded  $\beta$ -sheet and five  $\alpha$ -helices, and they possess several interaction surfaces particularly the surface loops, with which they associate with regulatory molecules and downstream effectors to exert their functions (Nikolova *et al.*, 1998) (See **Figure 1.4** below). Nucleotide binding sites G1-3,  $Mg^{2+}$  binding sites PM1-3 are conserved at the Ras superfamily level. Five short regions each with 5-6 residues named RabF1-5 are unique to Rabs. They form the switch regions to which regulators and effectors bind, distinguishing between GTP/GDP bound conformations. RabSF1-4 are short regions conserved within Rab subfamilies rather than between different subfamilies. They form two surfaces for different interactions providing flexibility and complexity to Rab interactions (Pereira-Leal and Seabra, 2000).

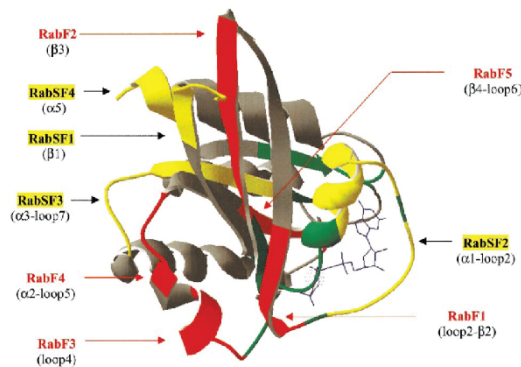
Rab-GTPase-activating proteins (RabGAPs) regulate Rab activity. Most RabGAPs have a Tre-2/Bub2/Cdc16 (TBC) Rab-binding domain (Richardson and Zon, 1995; Neuwald, 1997), usually at the C-terminal end of the protein. RabGAPs that do not possess this domain are rare and ~90% of known RabGAPs are of the TBC-containing canonical structure (Gabetnet-Castello *et al.*, 2013).



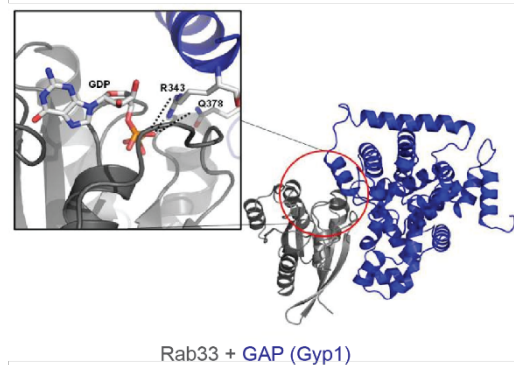
A



B



C



**Figure 1.4 Structure and mechanism of Rab GTPase and TBC Rab GAP action**

(A) Schematic overview of Rab cycle in membrane trafficking. Cycling between inactive GDP bound state and active GTP bound state is regulated by activating GEF and deactivating GAP proteins. Activated Rab is recruited to specific organelle/vesicle where it promotes transport via specific effectors. Taken from (Fukuda, 2011). (B) Representative Rab 3-D structure of Rab3A-GTP. PM/G-motifs (green) are involved in phosphate/Mg<sup>2+</sup> or guanidine nucleotide binding, RabF (red) are Rab-family specific regions, Rab-SF (yellow) are Rab sub-family specific regions, Mg<sup>2+</sup> and nucleotide are in blue. Taken from (Pereira-Leal and Seabra, 2000). (C) Representative Rab/GAP complex structure. Close-up view shows the involvement of arginine finger and glutamine finger in the interaction with GDP. Taken from (Park, 2013).

Signature sequences in TBC proteins include a catalytic arginine (R) finger that inserts into the nucleotide site as found in the Ras and Rho GAPs (Albert *et al.*, 1999) (See **Figure 1.4** above). However, the determinants of Rab recognition are not yet known. TBC RabGAPs may also contain other domains such as the PTB domain

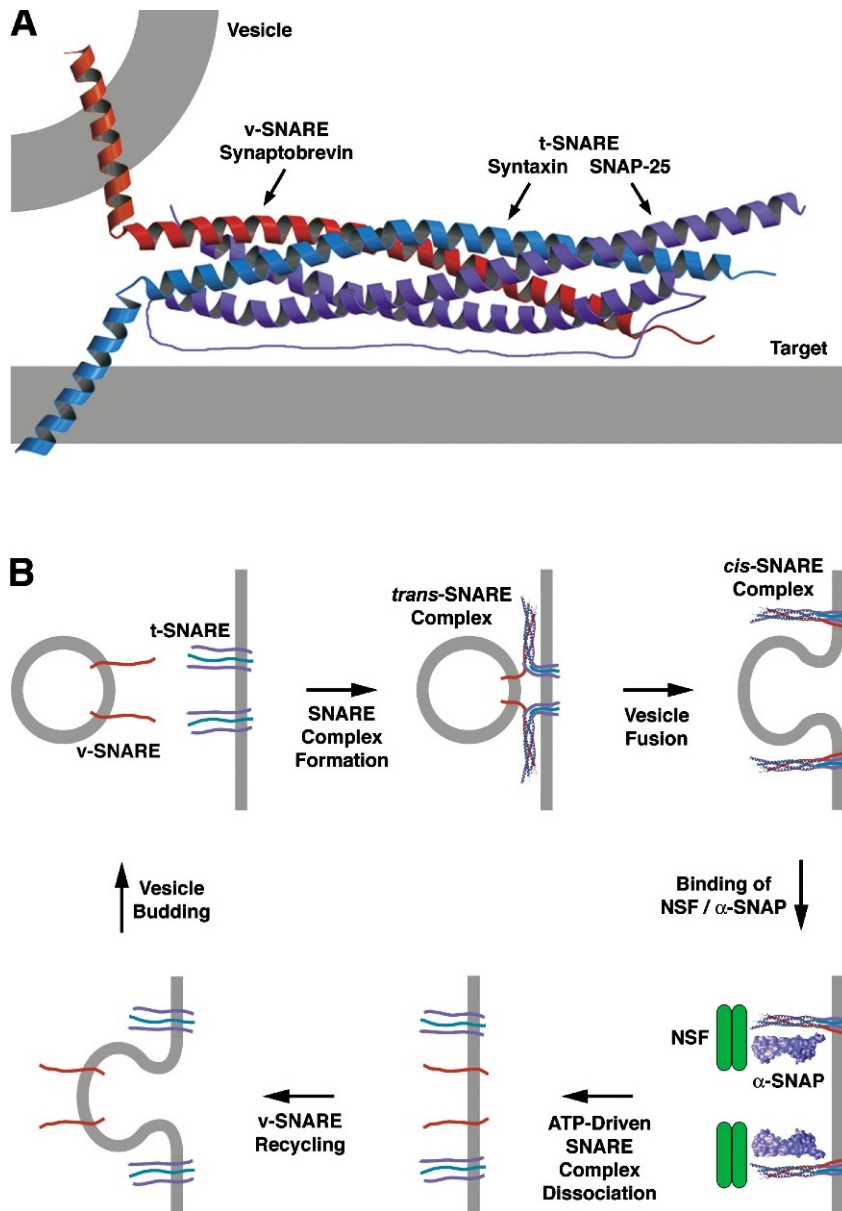
(phosphotyrosine-binding domain), PH (pleckstrin homology) domain, GRAM (glucosyl-transferase, Rab-likeGTPase activator and myotubularin) domain, RUN domain and/or CC (coiled-coil) domain. Some of these (RUN, GRAM) are associated with other small G protein signalling molecules such as Rap (Fukuda, 2011). RabGAPs are known to also bind non-substrate Rabs for which they may use these domains. For example, TBC1D11 directly binds Rab36 via its N-terminal PTB domain (Kanno *et al.*, 2010). Rab-deactivating RabGAPs are thought to act in Rab cascades in a manner similar to Rab-activating Rab guanine nucleotide exchange factor (RabGEF). GEFs are recruited by the GTP bound Rab and they activate the next Rab along the pathway. GAPs are hypothesised to be recruited by the newly active Rab at a given membrane to deactivate the previous Rab, thus restricting the spatial and temporal overlap of two Rabs and maintaining directionality of action (Rivera-Molina and Novick, 2009). An example of this cascade system is in yeast, where Ypt6 (yeast Rab6) and Ypt31/32 (yeast Rab11) occupy different regions of the Golgi and active Ypt31/32 can bind to the Rab-GAP Gyp6 which can then act as a GAP for Ypt6, thus deactivating it (Suda *et al.*, 2013).

SNAREs are small 25-35 kDA proteins, having a characteristic evolutionarily conserved 'SNARE' motif, which is 60-70 amino acid long and consists of largely hydrophobic heptad repeats. The original SNARE hypothesis (Söllner *et al.*, 1993a) proposed that recruitment of a distinct 'v-SNARE' on each vesicle and cognate 't-SNARE' on the target membrane enabled close apposition of the two membranes; the subsequent dissociation of the SNARE complex by the ATPase activity of NSF (N-ethyl maleimide sensitive factor) triggers fusion (Chen and Scheller, 2001).

Further advances have refined this hypothesis. In the SNARE complex, the unstructured SNARE motifs of single proteins formed elongated coiled coils of four intertwined parallel  $\alpha$ -helices, in which each helix is contributed by a different SNARE motif (Antonin *et al.*, 2002). Closer examination showed that at the centre of this bundle were 16 stacked layers of interacting side chains that were largely hydrophobic, except for the central '0' layer which was ionic. This layer was highly conserved and almost invariably constituted three glutamine (Q) and one arginine (R) residues, one from each contributing SNARE motif. SNAREs were therefore reclassified as Q or R-SNAREs according to the contribution of the SNARE motif to the '0' layer (Sutton *et al.*, 1998; Antonin *et al.*, 2002). Each of the contributing

motifs were named Qa, Qb, Qc and R and it is likely that functional complexes required each one of these motifs (in a three Qabc plus one R configuration) to form a hetero-oligomeric parallel four-helix bundles (Bock *et al.*, 2001; Jahn *et al.*, 2003) (See **Figure 1.5A** below). The assembly proceeds from the N-terminal end of the motif towards the C-terminus in a ‘zippering’ fashion (Li *et al.*, 2014) and the resultant release of energy is used to overcome the energy barrier for closer membrane apposition and fusion (as opposed to subsequent ATPase mediated dissociation, as the original hypothesis proposed). Once fusion is complete, the vesicle membrane is incorporated into the target membrane and the trans-SNARE complexes thus acquire a cis configuration. They are then dismantled into individual SNARE proteins by the action of  $\alpha$ SNAP and NSF ATPase proteins by ATP hydrolysis (Jahn and Scheller, 2006) (See **Figure 1.5B** below).

All Qa, as well as some Qb and Qc SNAREs called Syntaxins, have an N-terminal anti-parallel 3-helix bundle ( $H_{abc}$ ). R-SNAREs may have profilin-like folds called the longin domain at the N-terminal. These domains are involved in correct localisation and regulation of SNARE activity by interaction with other trafficking proteins such as SM proteins and tethers. Most SNAREs have a C-terminal transmembrane (TM) domain connected to the SNARE domain by a short linker sequence. Exceptions include SNAREs that have two SNARE motifs joined by a linker with no other N or C-terminal domains. These SNAPs or synaptosome-associated proteins (named so as they were first discovered in neuronal exocytosis) and other SNAREs such as the highly conserved R-SNARE Ykt-6, have post-translational modifications that mediate membrane anchorage. The linker between the SNARE motifs of SNAPs is palmitoylated and Ykt6 has a CAAX box that is farnesylated (Gonzalo and Linder, 1998; Fukasawa *et al.*, 2004). Some SNAREs with TM domains may also be palmitoylated for example yeast SNARE Tlg1. It has been proposed that this modification protects against ubiquitylation and subsequent degradation (Valdez-Taubas and Pelham, 2005).



**Figure 1.5 Mechanism of SNARE-mediated fusion**

(A) Crystal structure of the synaptic SNARE complex with Q/t-SNAREs Syntaxin and SNAP-25 and R/v-SNARE synaptobrevin. (B) SNARE cycle. A single R-SNARE on the vesicle interacts with trimeric Q-SNAREs on the target membrane forming a trans-SNARE complex. Forming of this stable four-helix bundle results in fusion.  $\alpha$ -SNAP binds to the resultant cis-SNARE complexes on the fused membrane and recruits NSF, which hydrolyses ATP to dissociate the component SNAREs. These can be packaged into vesicles and reused for fusion. Taken from (Bonifacino, 2004).

*In-vitro*, SNAREs can form promiscuous complexes of varying stability and non-physiological complexes may also drive liposome fusion. However, out of the 300 different possible combinations of yeast SNAREs, only nine were found to be fusogenic *in-vitro* (Malsam *et al.*, 2008). *In-vivo*, this is further restricted by coordination of the localisation of SNARE proteins. Non-fusogenic complexes of

SNAREs, as in the binding of inhibitory iSNAREs has been shown to have a regulatory effect (Varlamov *et al.*, 2004). Physiologically specific sets of SNAREs can be assigned to most fusion steps. However, there may be a common SNARE which contributes to formation of different SNARE complexes: e.g., yeast Qa SNARE Sed5p (=Syx5) has been found to act in two different SNARE complexes (Sed5p/ Bos1p/ Sec22p /Bet1p and Sed5p/ Gos1p/ Ykt6/ Sft1p) mediating two distinct transport steps between the ER-Golgi and within the Golgi complex respectively (Parlati *et al.*, 2002).

### 1.2.2 Evolution of Rabs, TBCs and SNAREs

The evolution Rab GTPases across eukaryotes has been extensively studied and their roles in trafficking have been experimentally investigated in a variety of taxa. In comparison, less is known about SNAREs, especially in the non-opisthokont lineages, where both evolutionary and functional studies are fewer. Even less is known about the Rab activating TBC domain proteins. Known aspects of the evolution of these families is discussed in this section.

#### 1.2.2.1 LECA repertoires

Pan-eukaryotic analyses have revealed 23 Rabs (Elias *et al.*, 2012; Klöpper *et al.*, 2012), 10 TBC subtypes (Gabernet-Castello *et al.*, 2013), and 20 conserved SNARE functional groups (Klopper *et al.*, 2007) that likely represent the repertoire of LECA. Based on experimental data from mainly yeast and mammalian cells, SNAREs were classified according to the part of the trafficking pathway they are likely to be associated with: ER-I, Golgi-II, trans-Golgi network (TGN) – IIIa, digestive endosomal compartments – IIIb and plasma membrane – IV. Therefore, a QaIV SNARE would refer to the class of Qa SNAREs localised to the plasma membrane. As SNAREs cycle to and from their target compartment and can form part of multiple complexes, this classification is only indicative and not rigid. It is, however, a useful framework to compare SNAREs from different lineages. The 20 SNARE functional groups can be divided into four subfamilies: Qa, Qb, Qc and R by phylogeny, indicating that these represent the ancestral forms that subsequently diversified (Bock *et al.*, 2001). Given the requirement of one SNARE of each kind in the formation of a functional complex, the phylogeny also bolsters the idea that “QabcR” represents an ancient principal structural arrangement of SNARE function

(Kloepper *et al.*, 2007). Rab phylogenies also reveal higher order clades. In particular, the primordial endocytic Rabs (5, 20, 24, 21, 22 and 50) and primordial exocytic Rabs (1, 8, 18, 2, 4, 14 and 11) can be discerned (Elias *et al.*, 2012). Higher order groupings of TBC sub-types do not have very high statistical support but are consistently found to divide into super-clades including TBC-G and M; TBC-B, D, E, and F; and TBC-A, K, N, Q and RootA (Gabernet-Castello *et al.*, 2013). The biological significance of such groupings remains to be elucidated.

The LECA cohort is large, and within these families, members share both sequence similarity and structural motifs. Therefore, they are each likely to have arisen by duplication and divergence from a single prototypic unit (Kloepper *et al.*, 2007). The higher order groupings may be indicators of the intermediate steps involved in the evolution of the LECA repertoire. In the case of Rabs, it is likely that there were primordial Rabs that established the endocytic and exocytic pathways in the proto-eukaryote, from which the more complex system arose in LECA. The domain structure may also provide hints about the evolutionary history: As Q and R SNAREs have different N-terminal domains when present (Habc Syntaxin and Longin respectively), it has been hypothesised that the SNARE complement has arisen from a primordial Q and R SNARE. However, the phylogeny indicates a split between R+Qb and Qa+Qc subfamilies (Kloepper *et al.*, 2007). How the LECA SNARE repertoire of 20 conserved subgroups emerged is therefore, not yet determined.

At least 13 Ras superfamily (to which the Rab GTPase family belongs, along with 6 other GTPase families) proteins with all five characteristic motifs of GDP/GTP-binding domain have been identified in prokaryotic lineages (Dong *et al.*, 2007). A computational method that can trace gene footprints through a gene functional network has uncovered four prokaryotic SNARE-like proteins, two of which have been localised to the plasma membrane (Chen *et al.*, 2013). Thus, although these prokaryotic proteins have not yet been functionally analysed, it is possible that they present a starting point for the evolution of Rab and SNARE proteins in eukaryotes.

### 1.2.2.2 Expansions and losses

Both Rab and SNARE repertoires are found greatly expanded in Metazoa and Embryophyta (land plants), and are correlated with the rise of multicellularity in both these lineages (Kloepper *et al.*, 2007; Kienle *et al.*, 2009a; Diekmann *et al.*, 2011). In comparison to *Monosiga brevicollis*, the closest unicellular relative of Metazoa, 14 Rab subfamilies were estimated to have arisen at the base of Metazoa, with likely roles in regulated secretion and Golgi-related trafficking (Diekmann *et al.*, 2011). There is a distinct increase in SNARE numbers between unicellular and multicellular metazoans and embryophytes, even with the basal multicellular forms of former showing relatively larger repertoires than their unicellular cousins (Kienle *et al.*, 2009a). In general, while the early secretory pathways are conserved, expansions are seen in predicted post-Golgi SNAREs (Kloepper *et al.*, 2007). All multicellular metazoans and embryophytes (but not uni- and multicellular chlorophytes) show expanded endosomal SNAREs (such as Qa-Syx7, R-VAMP7) and some secretory SNAREs (Qbc SNAREs). In vertebrates and embryophytes in particular, the exocytic/secretory SNAREs (namely Qa-Syx1, Qbc SNAREs and R-VAMP7 subtypes) have greatly expanded (Kienle *et al.*, 2009a). Therefore Rab and SNARE expansions are most distinct in the complex rather than the simpler multicellular organisms of these lineages. In fact, *P. yezoensis*, a multicellular red alga, has only five Rabs, the smallest number of Rab paralogues recorded in eukaryotes (Petrželková and Eliáš, 2014). Similarly, fungi, despite major diversity in morphology and transitions between uni- and multicellular forms, have a basic set of SNAREs largely unchanged from the predicted LECA repertoire (Kienle *et al.*, 2009b) and a much reduced cohort of Rabs (Pereira-Leal, 2008). Their numbers vary little between species of different lineages and lifestyles. Even where duplicated Rabs such as Ypt31 and 32 are present, they appear to have redundant functions (Pereira-Leal, 2008). Such reduction has been attributed to their loss of the phagocytic mechanism and mainly saprophytic lifestyle as compared to the internal digestion seen in Metazoa (Kienle *et al.*, 2009a) and hence, can be viewed as a probable lineage-specific adaptation.

Rab and SNARE repertoires are rather heterogenous across eukaryotes and the differences can only partially be explained by differences in genome size. Even though closely related taxa have similar Rab repertoires, the number of Rabs can vary

greatly between phyla within the same group; in the Alveolata, there is a large difference between the apicomplexans (*Plasmodium falciparum*, 11 Rabs) and ciliates (*Tetrahymena thermophila*, ~70 Rabs) (Brighthouse *et al.*, 2010). This can be explained by the parasitic nature of the apicomplexans, which have undergone extensive secondary losses. However, not all parasitism results in small Rab repertoire as seen in amoebozoan parasite *Entamoeba histolytica*, which has ~90 Rabs and the excavate parasite *Trichomonas vaginalis*, whose cohort exceeds 300 Rabs (Brighthouse *et al.*, 2010). These are likely to be taxon-specific expansions, but the adaptive advantage, if any, of these repertoires, remains elusive.

The TBCs are found expanded in metazoans, vascular plants and certain unicellular organisms, and generally correlate with the Rab complement of a given genome (Gabernet-Castello *et al.*, 2013). As the number of Rabs increase, the number of TBCs also tends to increase, indicating some maintenance of specificity level. However, this relationship tends to break down with large increases in Rab repertoire size, where TBC numbers tend to lag behind the Rab repertoire (Gabernet-Castello *et al.*, 2013). Accordingly, TBCs show little substrate specificity *in vitro*, although localisation and regulatory elements may control specificity *in vivo*. For example, both yeast TBCs Gyp8 and Gyp1 act as a GAP for Rab1 *in vitro* (De Antoni *et al.*, 2002), but while the former is retained at the ER and acts on the ER pool of Rab1, the latter is localised to the Golgi complex.

Rab and SNARE expansions appear only loosely correlated, if at all. While several unicellular organisms show expanded SNARE repertoires, their numbers have a smaller dynamic range between which they vary, compared to the Rabs. For example, while *Paramoecium tetraurelia* has 70 SNAREs, it is predicted to have 229 Rabs (<http://www.rabdb.org/>), while by contrast the expanded cohort of *T. vaginalis* is composed of 40 SNAREs (and 300 Rabs) (Kloepper *et al.*, 2007; Brighthouse *et al.*, 2010).

Furthermore different lineages appear to have expanded specific Rab subfamilies, such as Rab7 (late endosomes, autophagosome) in Amoebozoa, Rab8 (exocytosis) in Vertebrata, Rab2 (secretion) in Ciliophora, and Rab11 (recycling endosomes) in Archeplastida. These suggest independent expansions as seen in the case of Rab5 (early endosomes) which shows one of the most common independent expansions seen in Rabs. Usually, duplicated products have spatially and functionally



related roles. Some exceptions include the Golgi Rab1 duplication where duplicate Rab35 acts at endocytosis at the plasma membrane rather than at the Golgi (Kouranti *et al.*, 2006; Sato *et al.*, 2008). Also, even though LECA Rab2 is found in a clade with the endosomal Rabs 4, 11, 14, it is localised at the Golgi (Short *et al.*, 2001; Klöpper *et al.*, 2012). Where there are losses, it may implicate loss of particular pathways or structures during specialization. For example, loss of flagella/cilia associated Rab23, IFT27 and RTW is concomitant with the loss of these structures (Klöpper *et al.*, 2012). Certain Rabs that show frequent losses such as Rab4, Rab21 and 22, appear to share some effectors with more well conserved Rabs such as Rab11 and 5 respectively. As with the Rabs, the SNARE subtypes found expanded in different lineages are endosomal (Syx7, Vti1, VAMP7) or exocytic (Syx1, Qbc, VAMP7/Syb1). This is true of expansions associated with multicellularity in plants and metazoans, as well as those seen in complex unicellular organisms such as *P. tetraurelia*.

### 1.2.2.3 Conservation of subfamilies

The functions and localisations of Rab orthologs are exquisitely conserved across eukaryotes, such that they are used as markers of organelle identity (Woollard and Moore, 2008; Stenmark, 2009; Brighthouse *et al.*, 2010). Among the three families, the TBCs GAPs starting with the smallest repertoire (10) in LECA (Gabernet-Castello *et al.*, 2013), appear to have dynamically evolved in different lineages, particularly due to the number of accessory domains they feature, and which exhibit extensive swapping, insertions and deletions (Brighthouse *et al.*, 2010). Yeast two-hybrid methods could reveal a limited number of interactions of *T. brucei* GAPs and Rabs; many interactions were inconsistent with orthologous pairs between known interacting Rabs and GAPs (Gabernet-Castello *et al.*, 2013). There was also sparse co-expression for detected pairs of Rabs and TBCs indicating that further analysis was necessary (Gabernet-Castello *et al.*, 2013). When phylogenetic data of TBCs was compared to experimental data available in other organisms, both conserved and divergent features were revealed. The location and Rab specificity of TBC-M and B are conserved between yeast and humans – while the former acts on Rab1 at ER exit sites, the latter acts on Rab7 at the vacuole in both organisms. However, the Rab5 GAP, RNTre, which belongs to TBC-O, is only found in Opisthokonts and Amoebozoa while Rab5 is found across all eukaryotic groups and would require a

different GAP in lineages lacking TBC-O (Gabernet-Castello *et al.*, 2013). Rab associated proteins such as the RabGEFs and other effectors are notoriously divergent, even between yeast and humans, which belong to the same supergroup. In comparison, Rab-RabGAP interactions appear to be better conserved (Brighthouse *et al.*, 2010). Further phylogenetic and experimental studies in divergent lineages will reveal the extent and nature of conservation and divergence of these proteins.

So far, experimental studies in fungi, mammals, and embryophytes reflect largely conserved SNARE complexes and localisations. Some individual subtypes of plant SNAREs do show unconventional localisations such as the Qc SNARE Syp71, with dual localisation at ER and plasma membrane (Suwastika *et al.*, 2008), but this is rare. Outside of these three groups, experimental studies on SNAREs are rare, so it is not yet confirmed whether the composition of SNARE complexes and their phylogenetically assigned localisations are in fact conserved in the other eukaryotic lineages. A few cases of novel SNAREs with changed domain combinations have been identified. These include: (i) Vam7 SNARE in fungi, which have an N-terminal Phox homology domain (PX), (ii) Syx17 in metazoans with two adjacent C-terminal transmembrane (TM) domains, (iii) Sec22-like in metazoans, lacking the SNARE domain but retaining the N-terminal longin domain and having three consecutive TM domains at the C-terminus (Kienle *et al.*, 2009a), and (iv) phytolongins in plants which are derived from VAMP72 and lack the SNARE domain while retaining the longin domain (Vedovato *et al.*, 2009). These are all lineage-specific innovations that likely enable SNAREs to interact with novel factors and perform modified functions. Usually, the core SNAREs that they are derived from remain conserved within the genomes.

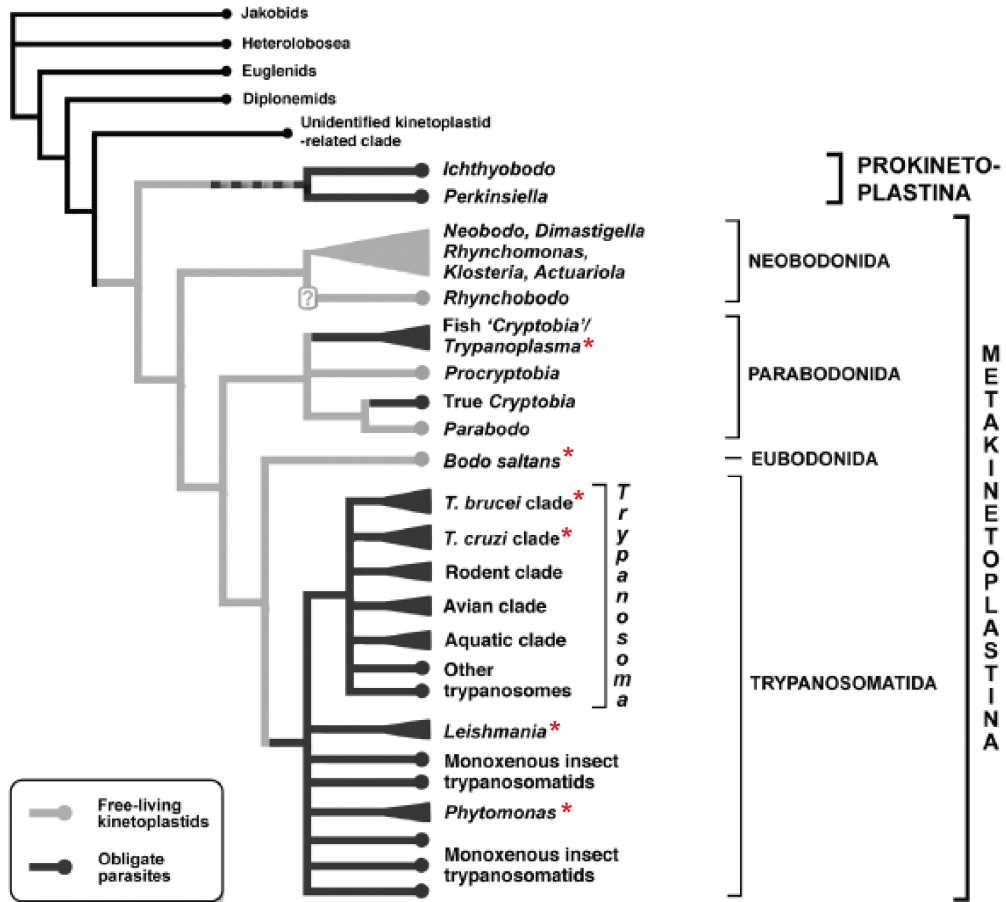
Therefore, while a subset of Rabs and SNAREs appear to be universally conserved, there is a great deal of dynamic shaping of repertoires particularly at the taxon specific level, with differing independent paths to complexity, multicellularity, and specialisation. Novel subclasses of TBCs are found in a wide range of lineages, but the core cohort is largely conserved and secondary losses are also very common (Gabernet-Castello *et al.*, 2013). Rabs appear to have evolved by a combination of secondary loss, and less frequently emergence of novel paralogues (Elias *et al.*, 2012). SNARE repertoires tend to remain relatively more stable and do not show the same

extents of losses and gains as seen in the Rabs, indicating the requirement of a minimal set for survival.

### 1.3 Kinetoplastids

Kinetoplastids are unicellular flagellated protists belonging to phylum Euglenozoa of the supergroup Excavata, and exhibit varied life-styles, host range, geographical isolation, and specialisation over long evolutionary periods. They are characterised by the kinetoplast, a network of circular mitochondrial DNA (called kDNA), physically attached to the flagellum basal body (Lukes *et al.*, 2002). Phylogenetic studies place kinetoplastids (or Kinetoplastea) as a branching lineage within Euglenozoa, with sister groups Euglenida (e.g., *Euglena gracilis*), Symbiontida as well as Diplonemea (Moreira *et al.*, 2004). Euglenozoa, along with Heterolobosea (e.g., *Naegleria gruberi*) belongs to Discicristata, a group of protists characterized by a unique feature, discoidal mitochondrial cristae (Adl *et al.*, 2012). Within kinetoplastids, phylogenetic studies reveal an early branching Prokinetoplastea with only two known representatives: *Ichthyobodo*, a biflagellate fish ectoparasite, and *Perkinsella*, which resides like an organelle in the cytoplasm of certain amoebae that parasitize the gills of fish. The rest of the known kinetoplastids have been assigned to Metakinetoplastea, which is further subdivided as Neobodonida, Parabodonida, Eubodonida and Trypanosomatida (Moreira *et al.*, 2004; Simpson *et al.*, 2006). (See **Figure 1.6** below for phylogeny)

Kinetoplastids include both free-living as well as parasitic species, but all known members of the group Trypanosomatida are parasitic and include several human pathogens. These include *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania* species, which cause African sleeping sickness, Chagas disease and Leishmaniasis respectively (Stuart *et al.*, 2008). Half a billion people, mainly in tropical and subtropical regions, are at risk for contracting these diseases; 20 million people are estimated to be affected, causing much suffering and up to 100,000 deaths per year (Stuart *et al.*, 2008). The parasites also infect livestock, which are sources of nutrition and farm labour, raising the economic toll of these diseases. Other members of the group, the Phytomonads, are responsible for plant disease, while still others are parasites of a wide range of animals (Camargo, 1999; Jaskowska *et al.*, 2015).



**Figure 1.6 Evolutionary relationships among kinetoplastids**

Phylogeny based on SSU rRNA gene trees and protein phylogenies. Circles denote single or few known representatives in a given clade, while triangles denote several known members; light grey represents free-living kinetoplastids, while dark grey represents obligate parasites; red asterisks indicate clades whose members are analysed in this study, and for which completed genomes are available, Adapted from (Stevens, 2008), originally published in (Simpson *et al.*, 2006).

Organism	Host	Vector	Flagella	Major cell surface proteins*	Infection type
<i>N. gruberi</i>	heterotroph	n/a	Multiple	unknown	Free living
<i>E. gracilis</i>	auto+heterotroph	n/a	2	unknown	Free living
<i>B. saltans</i>	heterotroph	n/a	2	unknown	Free living
<i>T. borreli</i>	fish	leech	2	unknown	extracellular, bloodstream
<i>P. serpens</i>	plant	hemipteran insects	1 free	gp63	extracellular, phloem
<i>P. EMI</i>	plant, symbiotic	hemipteran insects	1 free	gp63	extracellular, latex tubes
<i>P. HARTI</i>	plant	hemipteran insects	1 free	gp63	extracellular, phloem
<i>L. braziliensis</i>	vertebrates	phlebotomine sandfly	1 free	gp63, LPG; amastin	intracellular, macrophages
<i>L. major</i>	vertebrates	phlebotomine sandfly	1 free	gp63, LPG; amastin	intracellular, macrophages
<i>L. infantum</i>	vertebrates	phlebotomine sandfly	1 free	gp63, LPG; amastin	intracellular, macrophages
<i>L. mexicana</i>	vertebrates	phlebotomine sandfly	1 free	gp63, LPG, amastin	intracellular, macrophages
<i>L. donovani</i>	vertebrates	phlebotomine sandfly	1 free	gp63, LPG; amastin	intracellular, macrophages
<i>T. cruzi</i>	vertebrates	triatomine bugs	1 attached	gp63, mucins, trans-salidases; amastin	intracellular, many cell types
<i>T. grayi</i>	crocodiles	tse-tse fly	1 attached	trans-salidases	extracellular, bloodstream
<i>T. theileri</i>	mammals	ticks	1 attached	unknown	extracellular, bloodstream
<i>T. carassii</i>	fish	leech	1 attached	mucins	extracellular, bloodstream
<i>T. vivax</i>	mammals	tse-tse fly	1 attached	VSG; Procyclins	extracellular, bloodstream
<i>T. congolense</i>	mammals	tse-tse fly	1 attached	VSG; Procyclins	extracellular, bloodstream
<i>T. brucei brucei</i>	mammals	tse-tse fly	1 attached	VSG; Procyclins	extracellular, bloodstream
<i>T. brucei gambiense</i>	mammals	tse-tse fly	1 attached	VSG; Procyclins	extracellular, bloodstream

**Table 1.1 Summary of the diversity of lifestyles of kinetoplastids and relatives**

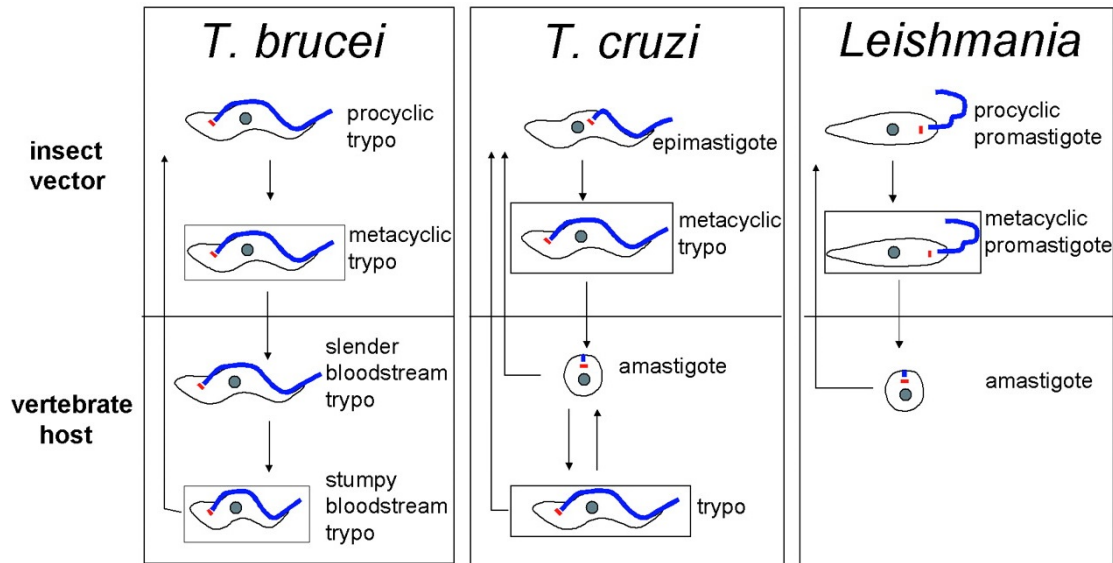
Kinetoplastids are grouped and coloured according to phylogeny as in **Figure 3.1**. LPG, lipoproteoglycan; VSG, variant surface glycoprotein.\*Major cell surface proteins of many newly sequenced kinetoplastids, especially the bodonids and some cruzi group members, phytomonads, as well as outgroup species *N. gruberi* and *E. gracilis* are as yet not very well described.

Kinetoplastids show several peculiar features that are not found in other eukaryotes: compartmentalisation of glycolytic enzymes in ‘glycosomes’, unconventional kinetochores, poly-cistronic gene clusters and trans-splicing of all mRNA transcripts with the absence of regulation of transcription initiation, and the complex and energy consuming editing of mitochondrial RNA (Ginger, 2005). In fact, Kinetoplastids are hypothesised to have diverged rather early from the eukaryotic lineage and lie close to the eukaryotic root (Cavalier-Smith, 2010; He *et al.*, 2014),

although this is a matter of much debate. Further, kinetoplastid surface components are also highly divergent, from one another and from other eukaryotes (Gurkan *et al.*, 2007; Dacks *et al.*, 2008). Extant trypanosomatids represent taxa that likely diverged 1000-1250 mya (Parfrey *et al.*, 2011) and are adapted to different insect vectors, show different host and geographical ranges, life-styles and immune evasion strategies; some of which may be reflected in specialisations within membrane trafficking. (See **Table 1.1** above for a summary of these differences)

### 1.3.1 The complex life cycles of parasitic trypanosomatids

The best-studied kinetoplastids are disease-causing trypanosomatids, which are usually digenetic, i.e., cycle between two hosts: vertebrate or plant, as well as an insect or leech vector. Developmental progression is accompanied by regulated changes in morphology, gene expression, metabolism, surface composition, and membrane transport. Different morphological forms, mainly discerned by the position of the kinetoplastid (*kt*), are associated with different life cycle stages. In trypomastigotes, *kt* is found at the posterior end of the cell, thus the flagellum folds back along the parasite's body and forms an undulating membrane (*um*) as it remains attached to the cell body and emerges at the anterior end. In epimastigotes, the *kt* is found anterior to the nucleus and the flagellum emerges from the centre of the cell, forming an *um* along half of the cell. In promastigotes, the flagellum emerges from the anterior of the cell and remains free. Finally in amastigotes, which are the non-motile intracellular forms, there is no flagellum extending out from the spherical cell (Clayton *et al.*, 1995). **Figure 1.7** below summarises the life cycles of three trypanosomatid representatives.



**Figure 1.7 Life cycle of trypanosomatids**

Stages in insect vector and vertebrate host are depicted for *T. brucei*, *T. cruzi* and *Leishmania*. Proliferative and non-proliferative (boxed) stages, with the position of the kinetoplast (red) and flagellum (blue) with respect to the nucleus (gray) are shown. (trypo = trypomastigote). Taken from (McConville *et al.*, 2002).

*T. brucei* ‘metacyclic’ trypomastigotes found in the saliva of the tse-tse fly are transferred to the mammalian host when the fly bites and takes a blood meal. Here, called the long slender (LS) forms or bloodstream forms (BSF), the parasite remains extracellularly in the blood and lymph, divides by binary longitudinal fission, and undergoes antigenic variation to evade the host immune response. This causes the first stage of trypanosomiasis, which in humans is characterized by generic symptoms: fever, headaches, and joint pains. BSFs eventually cross the central nervous system (CNS) barrier causing severe mental, sensory, and sleep anomalies which give the disease its name (sleeping sickness). Some of the parasites in the bloodstream however, undergo cell-cycle arrest turning into short stumpy (SS) forms, which are pre-adapted for uptake by the insect vector. SS forms that are taken up by the tse-tse fly in its blood meal differentiate into ‘procyclic’ trypomastigotes in the insect gut where they undergo several rounds of replication. They then travel to the salivary glands of the cell, where they transform into epimastigotes and then into metacyclic trypomastigotes, ready to be injected into the next mammalian host (Matthews, 2005).

*T. cruzi* are found as epimastigotes in the gut of their vectors, the triatomine bugs. As they reach the rectum, they transform into cell-cycle arrested metacyclic



trypomastigotes, which then enter their hosts via stercorarian transfer. Once in the vertebrate host, this form, incapable of extracellular replication, is able to invade the cells of a variety of tissues including the heart, gut, CNS, smooth muscle, and adipose tissues. The trypomastigotes at the cell membrane recruit the host cell lysosome to the plasma membrane likely by using the  $\text{Ca}^{2+}$ -dependent lysosome exocytosis pathway that is usually involved in repair of PM lesions in mammalian cells. Thus, they enter the cell via the parasitophorous vacuole forming at the PM. They then disrupt the vacuole membrane and enter the cytoplasm where they transform into non-motile amastigotes and undergo several rounds of binary fission. The amastigotes then differentiate into trypomastigotes that are released from the cell. These flagellated forms can then invade other host cells and be taken up by triatomine bugs during blood feeding. Continuous immune assault on persistent parasites is the primary cause of tissue damage in chronic disease (Stuart *et al.*, 2008).

Insect vectors, in this case, the sand fly, also transmit *Leishmania*. Over 21 species of this genus are known to cause diseases in humans – these include self-healing cutaneous leishmaniasis (CL), which may occasionally metastasize into mucosal membranes of the nose and mouth causing disfiguration and which can be difficult to cure (mucosal leishmaniasis, ML). The most serious form of the disease is visceral leishmaniasis (VL) which can affect several internal organs and is fatal if untreated. Metacyclic promastigotes in the anterior gut and mouth parts of the sand fly are transferred to the host by a regurgitation-like activity resulting in the expulsion of the parasites into the bite wound. Unlike *T. cruzi*, *Leishmania* are not able to actively invade cells, but are instead taken up by host macrophages by phagocytosis or receptor-mediated endocytosis, thus limiting the number of host cell types they are found in. The parasite-containing phagosome/endosome fuses with the lysosome to form the phagolysosome where the parasites are able to resist the low pH and hydrolytic enzymes. Within the phagolysosome, they transform into amastigotes, which then undergo binary fission. The amastigotes may be released by bursting of the host cell or by slow lysosomal exocytosis. When sand flies ingest infected macrophages during a blood meal, the amastigotes turn into procyclic trypomastigotes, which are capable of replication. When the replication stops, they turn into metacyclic trypomastigotes thus completing the cycle (Stuart *et al.*, 2008).

### 1.3.2 Strategies of immune evasion

The life cycle of the three representative trypanosomatids are largely similar, but they differ in several aspects such as the transmitting vector, mechanism of transmission into host, and the strategy of immune evasion in the host. This is reflected in differences seen in the surface composition, the underlying membrane trafficking mechanisms, and metabolism. Trypanosomatids show various polymorphic proteins with diverse glycolipid conjugates on their surface. The abundant cell surface glycoproteins are, however, largely mutually exclusive between *T. brucei*, *Leishmania* and *T. cruzi* (El-Sayed *et al.*, 2005a; Jackson, 2015).

*T. brucei* metacyclic trypomastigotes in the insect salivary glands are pre-adapted to the mammalian host environment by the acquisition of a dense immunogenic coat of 'variant surface glycoprotein' (VSG). In the mammalian host, VSG protects the parasite from complement mediated lysis, and even though it draws a powerful immune response from the host, the parasite is able to periodically switch to a new antigenically-distinct VSG (by a process called antigenic variation), thus requiring a whole new response from the host. The active recycling of the surface coat, together with the surface drag created by hydrodynamic forces generated by trypanosome motility, results in clearance of host antibodies bound to the parasite surface VSG. Once endocytosed, the antibodies are targeted to the lysosome for degradation while the VSGs are recycled back to the surface (Gadelha *et al.*, 2011).

*T. cruzi* trypomastigote surfaces have dispersed gene family -1 (DGF-1), mucins TcMUC and GP63, mucin associated surface protein family (MASP), and the trypomastigote alanine serine valine-rich proteins (TcASVs) as well as trans-sialidases. Amastin is the major surface protein in intracellular amastigotes (Jackson, 2015) *T. cruzi* survival in the mammalian host mainly relies on subverting the complement system, and on inhibitory effects on the mononuclear phagocytic system (Norris *et al.*, 1991; Flávia Nardy *et al.*, 2015). The parasite trans-sialidases transfer host sialic acid molecules to the terminal  $\beta$ -galactosyl residues of the cell surface mucin-like molecules. Such sialylated mucins protect the antigenic determinants of the surface from host attack of anti-galactosyl antibodies and complement factor B. Parasite-derived mucins can also inhibit the transcription of IL-2 in T cells, which leads to impairment of dendritic cell function. Sialyl-glycoproteins also inhibit early events in T-cell activation (Flávia Nardy *et al.*, 2015).

Non-protein lipophosphoglycan (LPG), glycoinositol phospholipid (GIPL), metallo-proteases GP63 (amastin, and GP63 in the amastigote stage) dominate the *Leishmania* cell surface. Its surface molecules protect the parasite from the complement system. LPG prevents the attachment of the final complement C5b-C9 complex to the surface, while GP63 protease can inactivate C3b and prevent formation of C5 convertase (Brittingham *et al.*, 1995; Yao *et al.*, 2013). LPG and GP63 also bind to host cell receptors such as fibronectin and mannose/fucose receptor respectively, and aid in parasite phagocytosis (Wilson and Pearson, 1986; Brittingham *et al.*, 1999). Inside the cells, GP63 can provide defence against the host anti-microbial peptides. Furthermore, GP63 released in exosomes are able to activate host tyrosine phosphatase SHP1 by cleavage, which results in attenuation of IFN- $\gamma$  induced immune activation (Cecílio *et al.*, 2014). They can also reduce the amount of IL-12 released by macrophages (Marth and Kelsall, 1997; Grazia Cappiello *et al.*, 2001), and increase the amount of TGF- $\beta$  and IL-10 released by macrophages and T-cells respectively (Ghalib *et al.*, 1995), effectively leading to immunosuppression and leaving the host susceptible to other infections.

## 1.4 Cell and molecular biology of kinetoplastids

The kinetoplastid cell body can adopt different morphologies at different life stages. It usually is in the form of an elongated spindle whose shape is maintained by an array of microtubules underneath the cell membrane. There is usually one (trypanosomatids) or two (bodonids) flagella, each of which emerges from a basal body which is found attached to the kinetoplast (in mitochondria) through a tripartite attachment complex (Ogbadoyi *et al.*, 2003). The flagellum emerges from the cell through a distinct invaginated domain in the membrane called the flagellar pocket (FP). As it folds back over the cell body, it can be attached to the cell (as in African trypanosomatids) or remain free (*Leishmania*) (Field *et al.*, 2007b).

### 1.4.1 Genome

In trypanosomatids, genes are arranged in poly-cistronic transcription units (PTUs), often containing over 100 genes, whose functions are not related (Martínez-Calvillo *et al.*, 2010). The gene order or synteny is greatly conserved between different species (El-Sayed *et al.*, 2005a). Even though the arrangement of coding regions along the chromosomes is similar in *Bodo saltans*, the gene order is not well conserved, perhaps due to the presence of a large number of Bodo-specific genes (Jackson *et al.*, 2008). In contrast with the synteny, the amino acid identity of syntenic genes (between *T. brucei*, *T. cruzi* and *L. major*) is only 40-55% which is in agreement with the proposed evolutionary distances between them (Ghedini *et al.*, 2004). Genes within a PTU are transcribed from the same strand, but adjacent PTUs can be on different strands. Therefore, regions that separate PTUs are called strand-switch regions (SSRs). Even though gene regulation is not specifically controlled at the level of transcription, at least a ten-fold higher initiation rate of transcription is found in SSRs of *L. major* (Martínez-Calvillo *et al.*, 2003, 2004). Once transcribed, individual coding regions are excised and capped via trans-splicing and polyadenylated to produce mRNA (Clayton, 2002). This type of genome organization is reflected in the reduced composition of RNA polymerase specific subunits and basal transcription factors as well as the underrepresentation of transcriptional regulators (Ivens *et al.*, 2005; Schimanski *et al.*, 2005)

Beyond these broadly conserved core chromosomal regions, many species-specific features are seen in subtelomeric regions. These are mainly involved in disease mechanisms and contribute to distinctive cell surface architectures (Handman *et al.*, 1995; Almeida *et al.*, 2007). In *T. brucei* bloodstream forms VSG genes are found in subtelomeric polycistronic expression sites as the terminal gene, before the start of the telomeric repeats region. VSG expression is specifically regulated such that only one of the 20-30 available expression sites is transcribed, by requiring its presence at a discrete nuclear site that has all the transcription enabling machinery (Navarro and Gull, 2001). Subtelomeric locations also contain VSG genes in long tandem arrays of repeated pseudogenes. Telomeres of mini-chromosomes also harbour up to 200 VSG genes. With recombinational rearrangements, these arrays essentially provide limitless opportunities for stochastic variation of the repertoire (Barry and McCulloch, 2001). The greatly expanded repertoires of the *T. cruzi* transsialidase (41400) and mucin (4860) genes, which are important parasite cell surface proteins when in the mammalian host, are also encoded within subtelomeres (El-Sayed *et al.*, 2005b). It has been proposed that recombinogenic tendencies of telomeres (Barry *et al.*, 2003) may enable diversification of these gene families too (Ginger, 2005). Unlike the VSG (which is expressed one at a time), more than one of each kind of these genes is expressed at the same time. *Leishmania* subtelomeres on the other hand, lack dead retro-elements, expansive gene families and specialized transcription units (El-Sayed *et al.*, 2005a). However, its subtelomeres do encode enzymes required for the assembly of the many types of glycoconjugates found at the cell surface, raising the possibility that they could be (however modestly) diversified too (Ginger, 2005).

The extra-nuclear genome i.e., the mitochondrial DNA is organized as an array of concatenated circular molecules known as the kinetoplast (kDNA), which gives the kinetoplastids their name. The kDNA circles are of two types, maxicircles and minicircles; these have been mainly studied in the trypanosomatids. A few dozen identical copies of maxicircles, which can be 20-40 kb each, depending on the species, are present. Thousands of minicircle copies are present, usually 0.5 to 10kb in size and of heterogenous sequence. Maxicircles encode typical mitochondrial gene products (e.g., respiratory chain complex subunits) and some proteins (Lukes *et al.*, 2002). These transcripts are usually cryptic and they undergo RNA editing, which is

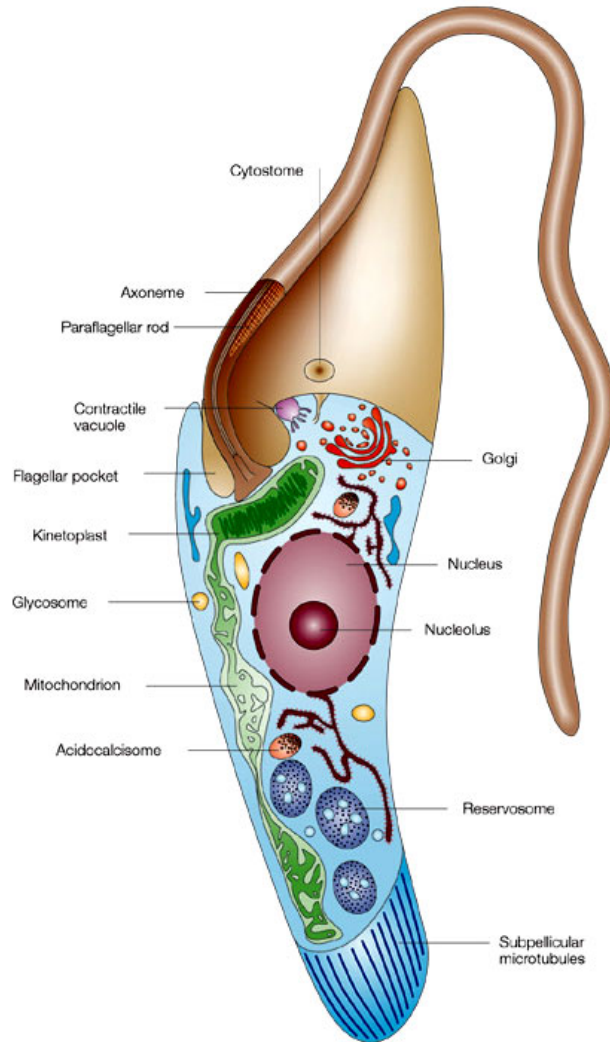
another peculiar feature of the kinetoplastids. RNA editing involves the insertion and deletion of uridine residues at specific sites in the transcripts. Guide RNAs (gRNAs) provide the genetic information for guiding and are largely encoded by minicircles; this appears to be the main purpose of minicircles (Lukes *et al.*, 2002). Studies beyond the trypanosomatids have revealed a great diversity of kDNA structures (Blom *et al.*, 1998). For instance, pro-kDNA, found in bodonids such as *Bodo saltans*, which are closely related to trypanosomatids, is composed not of networks but of individual 1.4-kb minicircles, with only a few very small catenanes (Blom *et al.*, 2000).

### 1.4.2 Membrane trafficking

Kinetoplastids have complex membrane trafficking systems that largely recapitulate the molecular machinery seen in the opisthokont model systems (See **Figure 1.8** below). However, as in other aspects of cell and molecular biology discussed above, their membrane trafficking systems also show distinct features. Classical organelles associated with membrane trafficking i.e., endoplasmic reticulum (ER), a single Golgi complex, a single lysosome, Multi-vesicular bodies (MVBs), early and late endosomal compartments (EE and LE respectively), have all been described in trypanosomatids (Clayton *et al.*, 1995). Further, organelles resembling lysosome related organelles (LROs) such as acidocalcisomes, which appear to be conserved in both eukaryotes and prokaryotes are also found (Docampo, 2015). Others, such as reservosomes (Souza *et al.*, 2000) are restricted to a single sub-genus *Schizotrypanosoma* (represented in this study by *T. cruzi*). Another organelle called cytosome-cytopharynx complex is also restricted to the schizotrypanum among trypanosomatids studied so far, but is present in the free-living bodonids as well (Attias *et al.*, 1996; de Souza *et al.*, 2009).

The *T. brucei* ER is a distinct sheet-like system of interconnected cisternal membranes with an elongated polarized structure. In the posterior half of the cell, the ER features an exit site intimately connected to the Golgi. It is characterized by an external tubular network and many vesicles of variable size (Engstler *et al.*, 2007). The BSF Golgi typically consists of 5–8 discrete stacked cisternae (Grab *et al.*, 1987), while it is double this size with many more cisternae in the PCF cell. The replication of the Golgi complex is coordinated with the cell cycle; it shows early division, is

connected to the formation of a new ER exit site, and involves transfer of materials from the old stack (He *et al.*, 2004).



**Figure 1.8 Schematic of trypanosomatid ultrastructure**

A schematic representation of internal architecture of *T. cruzi* epimastigote cell as a representative trypanosomatid. The cytosome, contractile vacuole and reservosomes are absent from other trypanosomatids. Picture taken from <http://www.nature.com/scitable/topicpage/kinetoplastids-and-their-networks-of-interlocked-dna-14368046#>

Both *T. brucei* and *Leishmania* utilize endocytosis to obtain nutrition, evade the host immune response, and survive in the host. In contrast, only insect-stage epimastigotes of *T. cruzi* show endocytosis; the process is low or absent in metacyclic trypomastigote and amastigote forms (de Souza *et al.*, 2009). Endocytosis (and exocytosis) is conducted through the flagellar pocket (Field and Carrington, 2009)

domain, which lacks the subpellicular microtubules. In *T. cruzi* epimastigotes, endocytosis also occurs through another specialized domain, the cyto-pharyngeal complex. The cytostome is an invagination of the plasma membrane (0.3  $\mu\text{m}$  diameter) coupled to a few special microtubules. As the cytostome penetrates deep into cell to the nuclear region, it becomes progressively narrower and called the cytopharynx (de Souza *et al.*, 2009).

The organelles of the membrane trafficking system are found to be highly polarized in these cells, and occupy conserved positions in the cytoplasm (Field *et al.*, 2007b). In *T. brucei*, all the endo- and exocytic structures including the lysosome and Golgi are found in the posterior of the cell body. This facilitates rapid recycling of surface components, which is required to evade the host immune response (Gadelha *et al.*, 2011). *Leishmania* are characterized by long endo/lysosomal multi-vesicular tubules (MVTs), spanning both the anterior and posterior of the cell (Vidugiriene *et al.*, 1999; de Souza *et al.*, 2009). In *T. cruzi* epimastigotes, reservosomes are found towards the posterior of the cell near the nucleus and the cytostome is found close to the flagellar pocket (de Souza *et al.*, 2009).

Endocytosis is exclusively clathrin-mediated (Allen *et al.*, 2003), unlike in opisthokonts and some other eukaryotes where up to five modes are recognized (Kirkham and Parton, 2005). No caveolins or flotillins have been identified in kinetoplastids. Therefore, the sorting of cargo does not happen at the surface but internally post-uptake. Three arms of the endocytic pathway are described in *T. brucei*: lysosomal targeting, Rab4-mediated, and Rab11-mediated recycling to the cell surface (Field *et al.*, 2007b). In the mammalian host, *T. brucei* features rapid endocytic recycling used to maintain the VSG coat, internalize the antibodies bound to it (Barry, 1979), and target the antibodies for lysosomal degradation (Pal *et al.*, 2003). *Leishmania* amastigotes endocytose and degrade MHC II molecules of the macrophages in which they are resident, thus limiting antigen presentation (Fruth *et al.*, 1993; Kima *et al.*, 1996).

Two classes of clathrin coated vesicles (CCVs) are described in *T. brucei* (Grünfelder *et al.*, 2003). CCV I are large (135nm) vesicles that originated from the clathrin coated pits of the flagellar pocket and are rich in surface macromolecules, mainly the VSG. These vesicles are not seen in the insect-stage procyclic forms (PCF). CCV II are small (50-60nm) vesicles that bud from the endosomal



compartments and contain concentrated fluid phase tracers but specifically exclude VSG. CCV I shed the coat and fuse with early endosomes (EE), which are perinuclear, large circular cisternae marked by the presence of Rab5. It is found juxtaposed to the late endosomes (LE) which are irregularly shaped (Engstler *et al.*, 2007). Two isoforms of Rab5 are found, A and B; they both colocalise in PCF but mark different compartments in the BSF and are likely involved in transport of different cargo sets (Pal *et al.*, 2002). EE localisation of Rab5 is largely conserved in *Leishmania* (Singh *et al.*, 2003) and *T. cruzi* (Araripe *et al.*, 2005). In the former, tubular structures (TS) 60 nm in size and located next to the Golgi and FP have been proposed to act as EE (Weise *et al.*, 2000). In the latter, endocytosed vesicles which fuse with an acidic tubular vesicular network spread from perinuclear to posterior of cell, are proposed to act as EE (Porto-Carreiro *et al.*, 2000).

Recycling endosomes (RE) continuously exchange membrane with the late endosomes (LE), PM and the trans-Golgi network (TGN). Unlike mammalian microtubule-associated tubular structures, the *T. brucei* RE is a giant, flattened, fenestrated structure that extends from the FP throughout the posterior of the cell (de Souza *et al.*, 2009). Here, ligands uncouple from their receptors and are sorted into CCV II along with fluid-phase cargo, which bud off from the rims of RE towards the LE/lysosome. Thus, the luminal VSG, transferrin receptor, and transmembrane (TM) proteins such as invariant surface glycoproteins (ISGs), remain in the RE by default (Grünfelder *et al.*, 2003). The recycling of VSG uses two pathways. It can be directly transported from RE by disc-shaped exocytic carriers (Rab11 positive) that fuse with the FP (Pal *et al.*, 2003). VSG can also be transported to the LE (Rab7 positive) from where it undergoes slow transfer to RE and eventually the PM. In *T. cruzi* epimastigotes, all endocytosed material is transported to the reservosomes which act as a store for macromolecules as well as lysosomal hydrolases and is the main site of protein degradation in the cell (Porto-Carreiro *et al.*, 2000).

MVBs are structures where further sorting takes place before lysosomal degradation in *T. brucei*. MVBs possess intra-luminal vesicles (ILVs) formed by inward budding of the membrane. Internalized ubiquitylated proteins are targeted to the MVB and their ILVs. *T. brucei* lacks the typical E3 ligases that ubiquitylate such cargo in mammalian cells, but possesses a majority of the downstream systems, i.e., the ESCRT complexes which recognize ubiquitylated cargo and regulate their

delivery to the MVBs. Thus, kinetoplastids likely use a different mechanism for the early targeting step. MVBs eventually fuse with the lysosome. Bloodstream form *T. brucei* shows increased lysosomal activity, in keeping with the increased endocytic activity (Engstler *et al.*, 2007).

*T. brucei* shows significant developmental changes that reflect the rapid endocytosis and recycling that is found in the BSFs. Clathrin and Rab11 are significantly upregulated in BSFs (Jeffries *et al.*, 2001; Morgan *et al.*, 2001). Unlike the PCF where the two Rab5 isoforms co-localise, they show distinct localisations and functions in BSF (Hall *et al.*, 2004). The lysosomal protein p67 is more elaborately glycosylated, and turned over more rapidly in BSFs, even though the level of proteolytic processing does not differ significantly between the two stages (Alexander *et al.*, 2002). In *Leishmania*, the dominant LPG antigen is downregulated in mammalian stages, and high proteolytic activity in amastigotes correlates with changes to the morphology of the lysosome and appearance of multivesicular megasomes (Courret *et al.*, 2001). AP-1 and clathrin appear to be constitutively expressed (Denny *et al.*, 2005) but there is potential developmental regulation of Qa SNAREs (Besteiro *et al.*, 2006).

The *T. brucei* endocytic system and cargo sorting has implications for resistance to the trypanosome lytic factor (TLF) found in human serum. TLF is composed of high-density lipoprotein (HDL) containing apolipoprotein LI (apoL-I), which is trypanolytic, and the haptoglobin-related protein (Hpr), which binds to haemoglobin (Vanhamme *et al.*, 2003). TLF is thus taken up through the trypanosome's haptoglobin-haemoglobin receptor (HpHbR). After endocytosis, when TLF reaches the late endosomes, the apoL-I undergoes a conformational change, leading to pore formation in the lysosomal membrane. This results in the influx of Chloride ions, osmotic swelling and eventually, cell lysis (Pérez-Morga *et al.*, 2005). *T. brucei brucei* is sensitive to TLF and hence cannot establish infections in humans (Vanhollebeke *et al.*, 2007). *T. brucei rhodisiense*, which causes the acute form of sleeping sickness in humans encodes a serum resistance antigen (SRA) which interacts with apoL-I and neutralizes it, preventing its transport to the lysosome (Van Xong *et al.*, 1998). *T. brucei gambiense*, does not possess SRA but has mutations in the HpHbR resulting in defective TLF uptake, at a fitness cost because haem is important for optimal growth (Ortiz-Ordóñez and Seed, 1995; Kieft *et al.*, 2010). This

is possibly the reason it causes a slower chronic version of the disease. Interestingly, in non-human primates, the apoL-I is mutated such that SRA cannot bind and neutralize it, and hence they are resistant to *T. b. rhodisiense* (Thomson *et al.*, 2009).

Exocytosis in *T. brucei* is geared towards the production of GPI-anchored molecules in both BSF (VSG) and PCF (procyclin) trypanosomes. While VSG has a globular and  $\alpha$ -helical structure that is N-glycosylated, procyclin is rather unstructured with proline repeat helix forming 50% of the mature polypeptide and only one isoform is glycosylated. Thus, chaperone and glycosylation requirements are higher in the BSF than PCF (Field *et al.*, 2007b). While key regulators of ER and Golgi transport processes, Rab1 and 2 are constitutively expressed (Dhir *et al.*, 2004), there is a marginal increase in the expression of ER-BiP (Bangs *et al.*, 1993) and Golgi Rab18 (Jeffries *et al.*, 2002) in BSF, which also shows a more elaborate Golgi than does the PCF. *Leishmania* and *T. cruzi* also mainly synthesise and transport GPI-anchored molecules to the cell surface. While the former produces complex glycolipids and glycoconjugates, the latter produces heavily glycosylated mucin molecules. Unlike *T. brucei*, these organisms have an established role for secretion of factors from the cell in disease progression (Field *et al.*, 2007b). These can take the form of extracellular vesicles (such as exosomes) that can alter signaling pathways involved in immune response to pathogens (Mantel and Marti, 2014).

The molecular machinery involved in all these different trafficking steps are largely conserved in kinetoplastids, even though significant absences exist along with a number of 'hypothetical' trypanosome-specific proteins of unknown functions. The two main coatomer COPI and II fully represented (Maier *et al.*, 2001), as are adaptor protein (AP) complexes except that AP-2 is missing in *T. brucei* and AP-3 is missing in *L. major* likely due to secondary losses, as they are otherwise retained across kinetoplastids (Manna *et al.*, 2013). Together with the absence of non-clathrin coat proteins at the cell surface, the loss of AP-2 may be viewed as a specialization. Given the high concentration of VSG at the surface and similar concentrations at the flagellar pocket during internalisation (without need for concentration), the sorting of internalized cargo takes place internally. Coats such as GGA and Stonins are not present in kinetoplastids, and are indeed found to be specific to opisthokonts indicating that they are relatively novel innovation of the lineage rather than ancient proteins. While adaptin interactors such as epsinR, AAK-1, RME-8, and synaptojanin have

been identified, a large number of them are missing in the genomes of kinetoplastids. EpsinR also lacks the ubiquitin interacting motif (UIM) which again has been shown to be opisthokont-specific (Field *et al.*, 2007a). Therefore, there must be an alternative mechanism for cargo selection in kinetoplastids. All tethering complexes apart from the TRAPPII are represented in kinetoplastids, even though not all subunits described in other eukaryotic lineages have been identified (Koumandou *et al.*, 2007).

The Rab GTPase family have been extensively studied in *T. brucei*, and they show conserved localisations and functions (Field and Carrington, 2004). Rab GAPs, which regulate Rab activity, have been identified *in-silico* but preliminary studies are inconclusive about the extent of functional conservation of these proteins (Gabernet-Castello *et al.*, 2013). SNAREs of the Qa-type, have been experimentally studied in *Leishmania major*, and also show largely conserved localisations as their opisthokont counterparts (Besteiro *et al.*, 2006). Kinetoplastids also possess the full complement of the SM proteins that aid and regulate SNARE actions (Koumandou *et al.*, 2007).

## 1.5 Objectives of this study

The sequences of many kinetoplastid genomes are now available and these resources have been used to reconstruct evolution of trafficking across this lineage, using Rab, SNARE and Rab-GAP paralogous families as proxies for defining intracellular compartment and transport pathway complexity. Specifically, the genomes of 18 kinetoplastids have been used in this analysis, encompassing basal bodonids and trypanosomatids. *Bodo saltans*, a free-living organism, and *Trypanoplasma borreli*, a digenetic parasite transmitted by leeches and infecting fish, represent the bodonids. Trypanosomatids are composed of *Phytomonas spp.*, *Leishmania spp.*, and *Trypanosoma spp.* *Phytomonas spp.* (*P. serpens*, *P. EMI* and *P. HARTI*) infect plants while *Leishmania spp.* (*L. braziliensis*, *L. major*, *L. infantum*, *L. mexicana* and *L. donovani*) infect many species of mammals and reptiles and form a distinct clade. Another clade is formed by the *Trypanosoma spp.*, which includes the monophyletic salivarian trypanosomes (*T. brucei brucei*, *T. brucei gambiense*, *T. congolense*, and *T. vivax*), henceforth referred to as the brucei group. They also include non-salivarian trypanosomes (*T. cruzi*, *T. grayi*, *T. theileri* and *T. carassii*), which will be referred to as the cruzi group. **Table 1.1** above details a brief comparison of the host, vector, morphology, and lifestyle aspects of kinetoplastids. Given the range of life-styles that is exhibited by different species across kinetoplastids, the correlation between changes in the trafficking system and attributes such as parasitism, monogenous or digenous life cycle, antigenic variation or intracellular modes of host immune evasion were explored. The possibility of coevolution between the three trafficking families was also investigated.

In addition, proteomics was used as an orthogonal approach to validate the phylogenetic reconstruction and determine whether the conservation of SNARE complexes described in opisthokont lineages extends to a highly divergent organism such as *T. brucei*. As there is a paucity of experimental studies of SNAREs in lineages outside of Opisthokonta and land plants, the results of this study will be an important step to determine the functional validity of SNARE phylogenetic assignments in general, i.e., the extent to which orthologous SNAREs from different eukaryotic lineages maintain their localisations and interactions.



# 2 MATERIALS AND METHODS

## 2.1 Bioinformatic approaches

### 2.1.1 Sequence data collection

A validated dataset of 26 predicted SNARE protein sequences, derived from *T. brucei* (Murungi et al. 2014) and 27 sequences from *L. major* (Besteiro *et al.*, 2006) were used as a query set to search the published proteomes of predicted kinetoplastid genomes ) for homologs by BLAST. Genomes are: *Trypanosoma brucei brucei* 927, *Trypanosoma brucei gambiense*, *Trypanosoma congolense*, *Trypanosoma vivax*, *Trypanosoma cruzi*, *Trypanosoma grayi*, *Leishmania major*, *Leishmania mexicana*, *Leishmania braziliensis*, *Leishmania infantum*, *Bodo saltans* (obtained from <http://www.GeneDB.org>), and *Phytomonas serpens*, *Phytomonas. EMI*, *Phytomonas. HART 1* (Porcel et al. 2014) and the heterolobosid *Naegleria gruberi* (Fritz-Laylin *et al.*, 2010) (obtained from <http://www.genome.jgi.doe.gov/Naegr1>). Further, tBLASTp was used to search unannotated genome data from *Trypanosoma theileri*, *Trypanosoma carassii*, *Trypanoplasma borreli*, and *Euglena gracilis* (unpublished data, kind access provided by Steve Kelly).

Two methods of collecting the BLAST results were employed in order to collect all possible SNARE sequences. In the first method, top three hits from each genome were collected [969 sequences] and in the second, all hits  $e \leq 10$  were collected [828 sequences]. ClustalW (Clustalw et al. 2003) was used to align the sequences from each dataset and to generate neighbour-joining trees (Saitou & Nei

1987). Sequences, which were outliers in the NJ tree or weakly clustering, were validated for the presence of the SNARE domain via Interpro (Hunter et al. 2009) and SNARE-DB (Kloepper et al. 2007). Sequences which contained domains associated with SNAREs (i.e. SNARE, t/v-SNARE, sec20, syntaxin, longin and synaptobrevin) were retained; those sequences in which no domain was detected but were between 70 and 500 amino acids long were also retained as possible SNAREs. Sequences which contained domains unrelated to SNAREs were discarded. The two datasets were then merged (removing duplicates) to obtain a dataset of 924 putative SNAREs. A number of duplicates were found in the unannotated datasets of *T. theileri*, *T. grayi*, *T. carassi*, and *E. gracilis*, which were removed using 98% sequence identity as a criterion. The duplications in the *T. cruzi* genome due to inclusion of Esmeraldo and non-Esmeraldo type were also removed, using TritypDB as a source to confirm equivalence. *N. gruberi* and *E. gracilis* sequences were also set aside as they were numerous and difficult to assign. The final non-redundant SNARE dataset had 518 sequences.

Predicted protein sequences obtained from a pfam Ras domain HMMSCAN against 50 eukaryotic proteomes (including the 19 genomes listed above) with e-values better than  $10^{-3}$  were used to generate a neighbour-joining tree. Kinetoplastid Rab candidates were identified and tentatively assigned based on clustering with known Rabs in the NJ tree. One sequence from each subfamily cluster was used as a query to define the cluster by reciprocal best hit (rbh) BLAST and another round of rbhBLAST was performed using results from the first rbh BLAST as queries. The tree was annotated accordingly to verify initial assignments. Additionally, these assignments were checked using Rabifier (Diekmann *et al.*, 2011). All collected kinetoplastid sequences were classed either as a tentative Rab subfamily member or as stray, to be further analysed by phylogeny. Similar procedure was undertaken with the TBC domain HMMSCAN to recover putative TBC domain containing proteins in kinetoplastids.

### 2.1.2 Phylogenetic reconstruction

The SNARE, Rab and Rab-GAP/TBC datasets were aligned using MAFFT (Katoh *et al.*, 2005) with the E-INS-i strategy. The alignments were then manually edited in Jalview (Waterhouse et al. 2009) and the edited alignments were used to generate maximum-likelihood trees in PhyML v3.0 (Guindon *et al.*, 2010). The



parameters used were: LG model of amino acid substitution, number of substitution rates – 4/6, starting tree – BioNJ, Tree topology search - NNI moves and statistics – aLRT SH-like and Bootstrap (100/0 replicates). Bayesian inference was implemented in MrBayes v3.2 (Huelsenbeck and Ronquist, 2001), generally with  $12 \times 10^6$  MCMC generations where convergence was achieved, as measured by a splits frequency below 0.01 being reached. Substitution models employed for inferring the trees were selected using ProtTest v3 (Abascal *et al.*, 2005) for PhyML and the mixed model was used for MrBayes. Iterative tree building was used to refine the datasets.

One representative from each well supported clade ( $>0.9$ , 90, otherwise mentioned in the Results) along with a panel of known eukaryotic SNAREs, Rabs and TBC Rab-GAPs respectively, were analysed with PhyML and MrBayes as described above to determine orthology, if any to known SNAREs, Rabs and TBC RabGAPs respectively. Rabs and TBC RabGAP panels were obtained from previous datasets (Elias *et al.*, 2012; Gabernet-Castello *et al.*, 2013). A SNARE panel was created from SNARE sequences from *Homo sapiens*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Phytomonas sojae* and *Entamoeba histolytica/ Dictyostelium discoideum* obtained from the SNARE-DB. *N. gruberi* sequences that were initially set aside were also included after a preliminary analysis of the sequences with human and yeast SNAREs to divide them into Qa, Qb, Qc and R subgroups. The assignment was confirmed using motif classification on SNARE-DB (Kloepper *et al.*, 2007) and NCBI-conserved domain database search. When sub-types were not assigned to known SNAREs, especially with the Qc SNAREs, the panel was expanded to include representatives from non-Opisthokont lineages in order to check if undetermined SNAREs could be better assigned to these lineages. Qc SNAREs were analysed with a wider set of sequences from *Guillardia theta*, *Aphanomyces astacii*, *Saprolegnia parasitica*, *Chlamydomonas reinhardtii*, and *Oxytricha trifallax*, which are represented in the final tree. In the final trees for Qa, Qb and R, only sequences of the original set of eukaryotic representatives was included. The trees obtained were annotated in Fig Tree v1.4 (Rambaut, 2009) and Adobe Illustrator v6. Coulson plot generator (Field *et al.*, 2013) was used to create pie chart graphics (<https://sourceforge.net/projects/coulson/>).

## 2.2 Experimental approaches

### 2.2.1 Trypanosome cell culture

Procyclic culture form (PCF) *T. b. brucei* MIT at 1.2 (Lister 427) was grown as previously described (Brun *et al.*, 1979; Hirumi and Hirumi, 1994). Cells were grown in SDM-79 medium (supplemented with 15% heat-inactivated fetal bovine serum, 100U/ml penicillin, 100U/ml streptomycin, and 2mM L-glutamine) at 28°C in non-adherent culture flasks with non-vented caps. They were maintained at  $1 \times 10^5$  to  $2 \times 10^7$  cells/ml. Bloodstream form (BSF) cells from the same cell line were grown in HMI-9 medium (supplemented as above) at 37°C with 5% CO<sub>2</sub> in non-adherent culture flasks with vented lids, and maintained at  $2 \times 10^6$  cells/ml.

To count cell density, 10µl (PCF) or 100µl (BSF) aliquots were withdrawn from cultures and diluted with 10ml isoton II (Beckman Coulter). Cell number was determined with a Z2 Coulter Counter (Beckman Coulter), averaging three measurements. A haemocytometer was also used for counting cells. Expression of plasmid constructs was maintained using antibiotic selection at the following concentrations in PCF: G418 and hygromycin B at 25µg/ml, blasticidin at 10µg/ml and puromycin at 2µg/ml. In BSF cell cultures, 2.5µg/ml G418 and 0.2µg/ml puromycin was used.

### 2.2.2 Nucleic acids and Recombinant DNA methods

All plasmids were grown in *Escherichia coli*, following transformation by heat shock at 42°C. PCR products and gel-embedded DNA were purified using QIAquick gel extraction kit (Qiagen) or Macherey & Nagel gel extraction kits according to instructions from the manufacturer. Plasmids were isolated using QIAprep® (Qiagen) or Macherey & Nagel Spin miniprep kit and QIAfilter™ Plasmid midi kit (Qiagen). Restriction and Antarctic phosphatase enzymes were obtained from New England Biolabs and T4 DNA ligase was obtained from Invitrogen and used according to the manufacturer's instructions. DNA was separated on 1% agarose gels using TAE buffer (40mM Tris acetate, 1mM EDTA) and visualised using 0.5µg/ml Ethidium bromide. Gel images were obtained using G:BOX imaging system and analysed using GeneTools software from Syngene, Cambridge, UK. DNA sequencing was done by

Geneservice, SourceBiosciences Cambridge, UK, and DNaseq, an in-house DNA sequencing facility at the University of Dundee, UK.

### 2.2.3 DNA isolation

For genomic DNA isolation from PCF cell,  $2 \times 10^7$  cells were harvested by centrifugation at 800g for 10min at 4°C, followed by washing with ice-cold PBS. The cell pellet was processed using the Qiagen Blood and Tissue kit according to the manufacturer's instructions.

### 2.2.4 RNA isolation

$2 \times 10^7$  BSF and PCF cells were harvested by centrifugation at 800g for 10 min at 4°C and washed once with PBS and snap frozen in dry ice. Total RNA was extracted using RNeasy Mini kit (Qiagen) and quantified using a ND-1000 spectrophotometer and Nanodrop software (Nanodrop Technologies).

### 2.2.5 Polymerase Chain Reaction

PCR reactions were carried out using Hercules II Fusion DNA polymerase from Aligent Technologies. One unit of polymerase was used along with 1µM final concentration of each primers, 200µM dNTPs, varying concentrations of genomic or plasmid DNA template and milliQ water (Millipore) were used in 50µl reactions. Oligonucleotide primers were obtained from Sigma-Aldrich and Invitrogen. Thermal cycling was programmed as follows – 95°C initial denaturation for 5 minutes, followed by 30 cycles of amplification (95°C for 30 sec, 52-63°C for 45 sec and 72°C for 45 sec) and one cycle of final extension at 72°C for 5 minutes.

### 2.2.6 Quantitative reverse transcription PCR

First strand cDNA was synthesised using SuperScript<sup>TM</sup> III Reverse Transcriptase from Invitrogen, following the manufacturer's instructions. qRT-PCR was performed using iQ<sup>TM</sup>-SYBRGreen Supermix on a MiniOpticon Real-Time PCR Detection System (Bio-Rad). Quantification was done using Opticon3 software (Bio-Rad). 12.5µl of 2X iQ<sup>TM</sup>-SYBRGreen Supermix (2X reaction buffer with dNTPs, iTaq DNA polymerase, 6mM MgCl<sub>2</sub>, SYBR Green I, fluorescein and stabilisers), 0.4µM forward and reverse primers and cDNA equivalent to 100ng total RNA was mixed and final volume was made up to 25µl with nuclease-free water. Replicate

samples were assembled as a master-mix, to which different templates were added. The reaction mix was transferred into Multiplate PCR plates set on a MiniOpticon Real-time PCR system, both from Bio-Rad. The amplification profile was as follows: 95°C initial denaturation for 5 minutes, followed by 35 cycles of amplification (95°C for 30 sec, 58°C for 30 sec and 72°C for 30 sec). A melting curve from 60-95°C was obtained to ascertain the specificity of the PCR amplification. Normalised expression  $\Delta\Delta CT$  of mRNA was determined based on PFR and TERT as reference genes (Brenndörfer and Boshart, 2010) using the BioRad CFX manager software. cDNA from PCF and BSF was analysed with primers for 26 SNARE genes (see table below), and samples with no Reverse Transcriptase treatment were used as control.

AGTGCAACGACACGTGTGAGGA	Tb927.3.5570 qf
GCAGATCCCGCAGTGGCTCC	Tb927.3.5570 qr
GCACACAGGCGGATGGAGCA	Tb927.10.11980 qf
ACGCGCTGCTCACGCTCAT	Tb927.10.11980 qr
GCACACAGGCGGATGGAGCA	Tb927.9.13030 qf
ACGCGCTGCTCACGCTCAT	Tb927.9.13030 qr
CGCACACGTCTGGCACGAAC	Tb927.10.14200 qf
TGTGGTCGATCCCCCGTGGT	Tb927.10.14200 qr
GGAAGGCCTGCAAAGCGCC	Tb927.11.8790 qf
TGCATGGCCAGTTGGTGGGC	Tb927.11.8790 qr
CCACGAGAACTGGACCTGCCA	Tb927.10.9950 qf
GGTGCCACCGTCACCCTCCT	Tb927.10.9950 qr
ACAGGTGGCGTTGCTAAAAC	Tb927.8.1120 qf
CTCTTGCATGGTGGCATATC	Tb927.8.1120 qr
ATGATGAGACGAGCGGTGTT	Tb11.01.2030 qf
GCGGTATTGCCGAATGTAGT	Tb11.01.2030 qr
CAAGACTTATTGCGCAGACG	Tb927.7.6440 qf
TTGCCATTTCGGTGGATAAAT	Tb927.7.6440 qr

GAAACCCTTCAGGCTCTGC	Tb927.8.1320 qf
CCAACTGCCTGAAAAACCAT	Tb927.8.1320 qr
TCACTGATGCCGTCTCTGAA	Tb10.70.5385 qf
CAGGTCATCGAGCATTGTGT	Tb10.70.5385 qr
GCGCTTAAACCGTACGATGA	Tb927.3.3720 qf
CACCGGGACTTCAGCATTAT	Tb927.3.3720 qr
GTGCATCGAGTGGTGTCAAT	Tb11.03.0965 qf
TATCGCAGTAGAACGCACCA	Tb11.03.0965 qr
TTTTACAACAACCCGCAAAA	Tb927.5.3560 qf
GGGCCAGTACCTTGTCCATA	Tb927.5.3560 qr
TGGTGATCTCGTTCAGCAAT	Tb09.211.4610 qf
TTCTCCAAAGTTTCCGTCAA	Tb09.211.4610 qr
ACTTGGAAGACCTTATGGTGCGTGG	Tb10.61.1380 qf
GCATCCTGTTCATGCGAGCCGT	Tb10.61.1380 qr
AAAATATATGGCGGGGTCGT	Tb927.10.790 qf
TGAGTTCCGGAATGGAAAAG	Tb927.10.790 qr
GGATCTCTTCTGCGTTGTCC	Tb11.01.7050 qf
AACCGTACCATCCTCACAGC	Tb11.01.7050 qr
CCCTTTCAGTCGTGAGGGGACA	Tb927.8.3470 qf
CTGCTGCATGCGGTGGACGA	Tb927.8.3470 qr
GGCCCAACAACCTTCTCTCTG	Tb10.61.0870 qf
GGAGCCCATGAGCTGATCTA	Tb10.61.0870 qr
TAGGGAAACGGGTATCAACG	Tb10.70.6010 qf
CATGTCTGCGAGCATCTTGT	Tb10.70.6010 qr

CGACGATACGGAAGAAAAGG	Tb09.160.2420 qf
GGAGAAGCAAACAGCAAAGG	Tb09.160.2420 qr
TTCAGTCCAGTGCTAGGTCGT	Tb927.2.5120 qf
TAGTGCCGCAAATGAAAATG	Tb927.2.5120 qr
ACGTGTCATGCAACGTGGACAA	Tb10.70.7410 qf
TCCACCACATTTGTTGGCGGAGAT	Tb10.70.7410 qr
TGCGGGAGACAATAAAGAGG	Tb927.10.14690 qf
CGCTAACGTCAAATCACGAA	Tb927.10.14690 qr

### 2.2.7 Preparation of expression constructs

Putative trypanosome SNAREs TbVAMP7C (Tb427.10.790), TbVAMP7A (Tb427.2.5120), TbVAMP7B (Tb427.5.3560), TbYkt6 (Tb927.9.14080) were amplified from *T. b. brucei* 427 genomic DNA using Hercules DNA polymerase (Aligent Technologies).

For hemagglutinin (HA)-tag fusion, the PCR products containing sequence for a C-terminal HA-epitope were cloned into the PCF expression vector plew79 (Wirtz *et al.*, 1999) using AvrII and BamHI (TbVAMP7A,B, and C) using the following primers (written 5' to 3'):

TTGTGTCCTAGGATGCTTATATCTGCCTCCTT	plewVAMP7A_F
ACTCAAGGATCCTTAAGCGTAATCTGGAACATC GTATGGGTACTTTTTGCACTTGAGGTTAG	plewVAMP7A_R
TTGTGTCCTAGGATGCCCATTAATATAGTTG	plewVAMP7B_F
ACTCAAGGATCCTTAAGCGTAATCTGGAACATC GTATGGGTATGACTTGCACTTGGAAGT	plewVAMP7B_R
TTGTGTCCTAGGATGCAGGGAGGAACAAAAT	plewVAMP7C_F
ACTCAAGGATCCTTAAGCGTAATCTGGAACATC GTATGGGTACTTCTTTTCTCTTTTTTAC	plewVAMP7C_R

The PCR product of TbYkt6 was cloned into pHD1034 containing an N-terminal HA-epitope using HindIII and AflI using the following primers:

GGCCAAGCTTTATACTCCCTGGCAAT	pHD1034Ykt6F
CCGTCTTAAGTCACATGACGGTGCAACA.	pHD1034Ykt6R

Putative SNARE interactors TbSyx16B (Tb427tmp.211.3920), TbSynE (Tb427.03.5570), TbVti-likeA (Tb427.8.3470), TbVti-likeB (Tb427.08.1120), TbSyx6-like1 (Tb427.10.1830) and TbSyx8-like (Tb427.10.2340) were also similarly amplified. For 6x cMyc PCR, the product of each gene was cloned into pRPΔOP (with thanks from Dr. Lucy Glover, Dr. David Horn's Lab, University of Dundee) containing 6x cmyc-epitope using HindIII and XbaI using the following primers:

GCGCGCAAGCTTATGAGCGGGGACGGCGTTGG	pRPCTb427.8.1120F
GCGCGCTCTAGAACTTTCCCCAGAACTTCC	pRPCTb427.8.1120R
GCGCGCAAGCTTATGGACGATCCAAGTTGGCA	pRPCTb427.3.5570F
GCGCGCTCTAGATACTTTATGGTACGCAACGA	pRPCTb427.3.5570R
GCGCGCAAGCTTATGTCGTCTCTGCAAGATCC	pRPCTb427.10.1830F
GCGCGCTCTAGAACTAAAGACACAATAGAAGA	pRPCTb427.10.1830R
GCGCGCAAGCTTATGTCTAAACAAGAA	2F_PRP_Tb427.10.2340_C
GCGCGCTCTAGAAAGTATTTAAAAGCAC	2R_PRP_Tb427.10.2340_C
GCGCGCAAGCTTATGTCATCTGATCTT	3F_PRP_Tb427.08.3470_C
GCGCGCTCTAGACTTCCAAAATACAAT	3R_PRP_Tb427.08.3470_C
GCGCGCAAGCTTATGGCGACCCGTGAC	4F_PRP_Tb427.211.3920_C
GCGCGCTCTAGAAGACAGCATCTTTTG	4R_PRP_Tb427.211.3920_C

Putative interactors VPS45 (Tb427.10.6780) and Sly1 (Tb427tmp.160.0680) were cloned into pMOT vector (Oberholzer *et al.*, 2006) with 3xV5 tag using the following primers:

AGGTCCTGTGCACGCCTGCATCGGTGGGACTGGAGTC CTTAACAGTGAAACCTTCCTGAGCCTGCTAGCAGCGC ACGCAGGTACCGGGCCCCCCTCGAG	VPS45pMOT_F
GTATTTTGGTTTCGTTTATTCATACCACCATGCGGAGG CGCAATGTCCCCGCCAAAACAGGCGAGGGCGGCACA TGGCGGCCGCTCTAGAACTAGTGGAT	VPS45_pMOT_R
GGTTAGTTATGGCTGTACCGCAATGCTGACGGGGAAT GAAGCACTGCGCCAGCTTACTGTTCTTGGTGAAGGAA TATCAGGTACCGGGCCCCCCTCGAG	Sly1_pMOT_F
AAAGCACGTTAGGATAGTATCTGAAAGTGGGAAAAC GCCAAATGGCACAAAGACCAAACGGCCGGGCCGGT GCTGGCGGCCGCTCTAGAACTAGTGGAT	Sly1_pMOT_R

All constructs were verified by sequencing and linearized with NotI prior to transfection into cells. Clonal transformants were selected by resistance to antibiotics as relevant to each vector and cell line.

### 2.2.8 Transfection of PCF *T. brucei*

1.6 x 10<sup>7</sup> cells per transfection were harvested at 4°C, washed in cytomix (2mM EGTA, 120mM KCl, 0.15mM CaCl<sub>2</sub>, 10mM KPO<sub>4</sub>, 25mM HEPES, 5mM MgCl<sub>2</sub>, 0.5% glucose, 100µg/ml BSA, 1mM hypoxanthine, pH 7.6) and resuspended in 500ul cytomix. Electroporation was performed with 5-15 µg of DNA using a Bio-Rad Gene Pulser II (1.5 kV and 25 µF).

### 2.2.9 Immunofluorescence microscopy

Cells were prepared as previously described (Leung *et al.*, 2008). Antibodies were used at the following concentrations: rat anti-HA (Roche), 1:1000; mouse anti-c-myc (SantaCruz Biotech 9E10), 1:500; mouse anti-p67 (from J. Bangs, University of Wisconsin-Madison) 1:1000; rabbit anti-GRASP (from Graham Warren, Vienna, Austria) 1:500 in 20% FCS in PBS (v/v). Wide-field epifluorescence images were acquired using a Nikon Eclipse E600 epifluorescence microscope equipped with a Hamamatsu ORCA charge-coupled device camera, and data were captured using



MetaMorph (Universal Imaging, Marlow, UK). Quantitation was performed on raw images gathered under non-saturating conditions using ImageJ (Rasband, 2012)

### 2.2.10 Western immunoblotting

Whole cell lysates in standard SDS-PAGE sample buffer containing  $10^7$  cells/lane were resolved by SDS-PAGE. Proteins were transferred to polyvinylidene fluoride membranes (Millipore) and blocked using 5% semi-skimmed milk in PBST (3.2 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 mM KH<sub>2</sub>PO<sub>4</sub>, 1.3 mM KCl, 135 mM NaCl, 0.05% Tween® 20, pH 7.4) for 1 hour. Membranes were then incubated for 1 hour at room temperature or 4°C overnight with primary antibody at appropriate dilution in blocking solution of 1% milk. Excess antibodies were removed by washing membrane thrice in PBST before incubating with secondary antibody. Primary antibodies were used at the following concentrations: rat anti-HA (Roche) at 1:10000 and mouse anti-c-myc 1:5000 in PBST (Phosphate-buffered saline + 0.1% Tween 20). Primary antibody binding was detected using secondary anti-IgG horseradish peroxidase (HRP) conjugates (Sigma-Aldrich) at 1:10000 in PBST. After incubation in secondary antibody, blots were washed again in PBST, thrice. Detection of HRP-conjugated secondary antibodies was by chemiluminescence using luminol (Sigma-Aldrich) and x-ray film (Kodak) or the G:BOX chemiluminescence imaging system from SynGene.

### 2.2.11 Proteomics

Interactions between selected tagged SNAREs and other trypanosome proteins were analyzed by immunoisolation after cryomilling of parasites.  $5 \times 10^{10}$  procyclic cells or  $1.2 \times 10^{10}$  bloodstream form cells harbouring genetically tagged HA epitope were lysed by mechanical milling in a Retsch Planetary Ball Mill PM200 using liquid nitrogen cooling (Retsch, United Kingdom). Aliquots of powder were thawed in the presence of solubilization buffer. Several buffer conditions were tested. The final conditions used for the various samples are listed in the table below.

<b>Sample</b>	<b>Buffer #</b>	<b>Buffer composition</b>	<b># of mass spec runs</b>
<b>TbVAMP7C::HA (PCF)</b>	Buffer1	1.0% CHAPS, 50mM Tris pH7.5, 150mM NaCl, 1mM NEM	1
	Buffer2	1.0% Triton-X-100, 50mM Tris pH7.5, 100mM NaCl, 1mM NEM	3
	Buffer3	0.1% Triton-X-100, 50mM Tris pH7.5, 100mM NaCl, 1mM NEM	2
	Buffer4	0.1% Triton-X-100, 0.5% Triton-X-114, 50mM Tris pH7.5, 100mM NaCl, 1mM NEM	2
<b>TbVAMP7B::HA (PCF)</b>	Buffer2	1.0% Triton-X-100, 50mM Tris pH7.5, 100mM NaCl, 1mM NEM	1
<b>HA::TbYkt6 (PCF)</b>	Buffer2	1.0% Triton-X-100, 50mM Tris pH7.5, 100mM NaCl, 1mM NEM	1
	Buffer5	0.5% Brij, 50mM Tris pH7.5, 50mM NaCl, 1mM MgCl <sub>2</sub> , 10μM CaCl <sub>2</sub>	1
<b>TbVAMP7B::HA (BSF)</b>	Buffer2	1.0% Triton-X-100, 50mM Tris pH7.5, 100mM NaCl, 1mM NEM	1
	Buffer6	1.0% Triton-X-100, 0.5% Triton-X-114, 50mM Tris pH7.5, 100mM NaCl, 5mM EDTA	2

HA-tagged SNAREs were isolated using Pierce anti-HA magnetic beads. All washes were in the same buffer without NEM. Following analysis of an aliquot by SDS-PAGE, affinity-isolated proteins were precipitated in 90% ethanol and sent for mass spectrometry. The number of replicates analysed for each buffer condition on

each sample is shown in the table above. Peak lists were submitted to ProFound and searched against an in-house *T. brucei* database using data from GeneDB (<http://www.genedb.org>).



# 3 EVOLUTION OF KINETOPLASTID RABS AND RABGAPS

## 3.1 Introduction

RabGTPases are master regulators of trafficking in the eukaryotic cell and key mediators of specificity in the endomembrane system. They have been used as markers in comparative cell biology due to the high degree of specificity of their sub-cellular localisation, which is conserved even in very distantly related orthologues. They have thus been used as predictors of the existence and complexity of membrane trafficking pathways. The evolutionary history of Rabs across eukaryotes has been reconstructed by several groups (Elias *et al.*, 2012; Klöpper *et al.*, 2012) and they have also been explored within lineages such as the Fungi (Pereira-Leal, 2008) and Metazoa (Diekmann *et al.*, 2011). Rabs are in turn regulated largely by RabGAP proteins. Studies of RabGAPs have mainly been conducted with in yeast and Metazoa (Itoh *et al.*, 2006; Fukuda, 2011; Frasa *et al.*, 2012). There has been a pan-eukaryotic analysis of these proteins which has allowed identification of orthologues across a

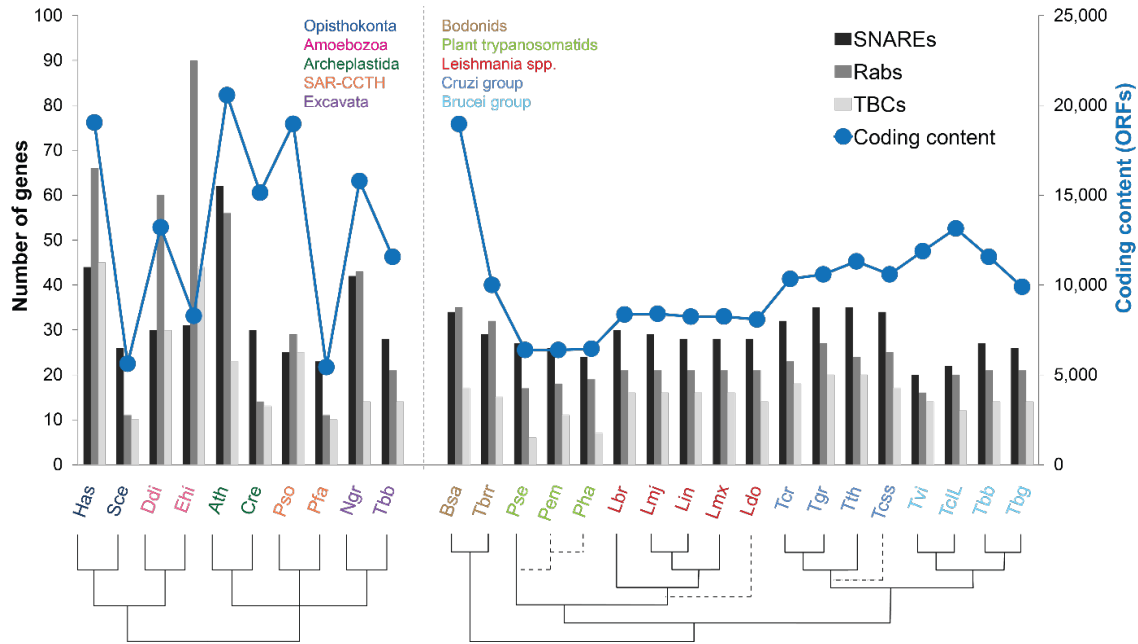
great diversity of eukaryotes, but experimental studies are still rare outside of the established yeast and mammalian systems (Gabet-Castello *et al.*, 2013). In this study, newly sequenced genomes were used to reconstruct the evolutionary history of the Rabs and RabGAPs in a deeply divergent class of organisms, Kinetoplastida. This has been used to assess the variation of trafficking complexity within this group, and the relationship between the Rabs and RabGAPs. In this chapter, the overall comparison of trafficking repertoires across kinetoplastids is presented, followed by the detailed phylogenetic analyses of kinetoplastid Rabs and RabGAPs. SNAREs are addressed in the next chapter.

## 3.2 Adaptive shaping of trafficking repertoires

### 3.2.1 Coding content and trafficking repertoires

Previous analysis has suggested that there is a loose correspondence between genome size and the repertoire of Rab genes (Gabernet-Castello *et al.*, 2013). While this may reflect correspondence between genome size and compartmental complexity in some manner, the huge diversity across eukaryotes makes evaluation of complexity a difficult parameter to define. Hence, this analysis was performed in a group of organisms where basic cellular bauplan is well conserved, both to re-evaluate this concept and to provide an overview of gene family size.

**Figure 3.1** below summarises the total number of protein coding genes (coding content) and total number of Rabs, TBC GAPs, and SNAREs found in each kinetoplastid genome as well as two representatives from each eukaryotic supergroup. At 19,000 genes, the coding content of the free-living bodonid *Bodo saltans* is considerably larger than that of the parasitic species. The number of genes in *Trypanoplasma borreli* is yet unknown and the figure only reflects an estimate. *Naegleria gruberi* and *Euglena gracilis* also have rather large genomes (Fritz-Laylin *et al.*, 2010) (MCF, Kelly, S., and Ebenezer, TG, unpublished data). While the phytomonads and *Leishmania spp.* have a particularly small number of protein coding genes (under 9000), this is not the case in the brucei group trypanosomes, which have ~10,000 genes (numbers according to latest <http://www.tritrypdb.org> gene metrics data). In keeping with genome size, *B. saltans*, *N. gruberi* and *E. gracilis* have larger numbers of SNAREs, Rabs and TBCs compared with the remaining kinetoplastids, and the core trypanosomatid repertoire is further reduced. However, in the brucei group, repertoires are smaller even though their coding content is comparable with that of the cruzi group, suggesting further secondary losses in the former. Even *Leishmania spp.* have slightly larger numbers of these proteins despite smaller coding content. Thus, overall, this suggests an adaptive shaping of trafficking system gene repertoires, which does not simply reflect genome size.



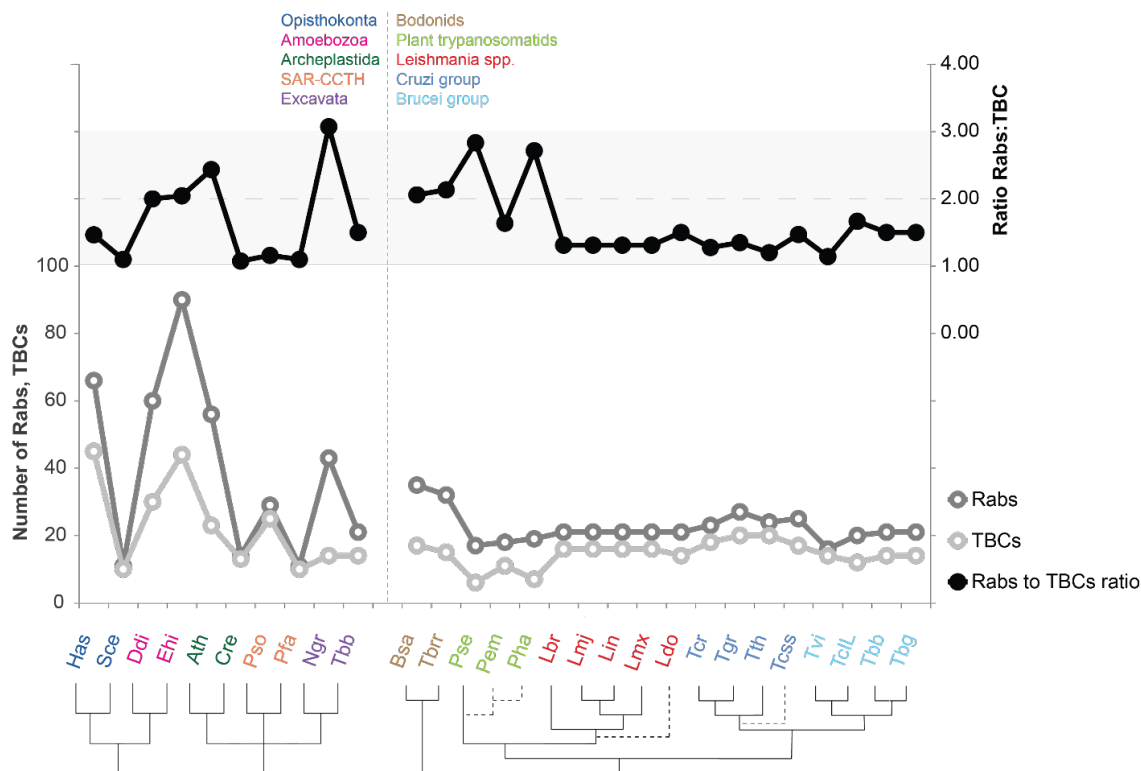
**Figure 3.1 Representation of SNARE, Rab and TBC coding sequences in selected eukaryotic genomes and kinetoplastids.**

Genomes are arranged by phylogenetic relationships. The five classically recognised, *sensu* Adl 2004, eukaryotic super groups and each sub-group of kinetoplastida *sensu* Manna 2013 are colour-coded according to the colour key on either side of the dividing dashed line respectively. Blue symbols and solid line represent the total coding content of the respective organism by total number of predicted ORFs (reads are shown on the y-axis, right). Numbers of SNARE, Rab and TBC ORFs are represented by dark, medium and light grey bars respectively (y-axis, left). (Hsa – *Homo sapiens*, Scv – *Saccharomyces cerevisiae*, Ddi – *Dictyostelium discoideum*, Ehi – *Entamoeba histolytica*, Ath – *Arabidopsis thaliana*, Cre – *Chlamydomonas reinhardtii*, Pso – *Phytophthora sojae*, Pfa – *Plasmodium falciparum*, Ng – *Naegleria gruberi*, Tbb – *Trypanosoma brucei*; Bsa – *Bodo saltans*, Tbr – *Trypanoplasma borreli*, Pse – *Phytomonas serpens*, Pem – *Phytomonas* EM1, Pha – *Phytomonas* HART1, Lbr – *Leishmania braziliensis*, Lmj – *Leishmania major*, Lin – *Leishmania infantum*, Lmx – *Leishmania mexicana*, Ldo – *Leishmania donovani*, Tcr – *Trypanosoma cruzi*, Tgr – *Trypanosoma grayi*, Tth – *Trypanosoma theileri*, Tcss – *Trypanosoma carassi*, Tvi – *Trypanosoma vivax*, TcIL – *Trypanosoma congolense*, Tbg – *Trypanosoma brucei gambiense*)

### 3.2.2 Comparing Rabs vs TBCs in kinetoplastids

A previous study comparing the number of Rabs and TBCs in a broad range of eukaryotic genomes showed that even though there is a general correlation between Rab and TBC cohorts, there is some variation in the total number of TBCs, as well as the Rab:TBC ratio within supergroups (Gabernet-Castello *et al.*, 2013). The authors concluded that the evolution of the TBC family is highly dynamic. This study compared the Rab and TBC cohorts of eighteen kinetoplastid genomes to assess the level of variation in greater resolution within a single taxonomic order.





**Figure 3.2** Ratio of the number of Rabs:TBC in kinetoplastids and select eukaryotes

Genomes are arranged by phylogenetic relationships. The five classically recognised, *sensu Adl 2004*, eukaryotic super groups and each sub-group of kinetoplastida *sensu Manna 2013* are colour-coded according to the colour key on either side of the dividing dashed line respectively. Number of Rab, TBC genes are plotted on the left Y-axis, and the Rab:TBC ratio is plotted on the right Y-axis. Empty dark grey circles represent Rabs, empty light grey circles represent TBCs, and full black circles represent Rab:TBC Ratio. Shaded area corresponds to ratio of 1 through to 3, and the dashed line represents a ratio of 2. (Hsa – *Homo sapiens*, Sce – *Saccharomyces cerevisiae*, Ddi – *Dictyostelium discoideum*, Ehi – *Entamoeba histolytica*, Ath – *Arabidopsis thaliana*, Cre – *Chlamydomonas reinhardtii*, Pso – *Phytophthora sojae*, Pfa – *Plasmodium falciparum*, Ng – *Naegleria gruberi*, Tbb – *Trypanosoma brucei*; Bsa – *Bodo saltans*, Tbr – *Trypanoplasma borreli*, Pse – *Phytomonas serpens*, Pem – *Phytomonas EM1*, Pha – *Phytomonas HART1*, Lbr – *Leishmania braziliensis*, Lmj – *Leishmania major*, Lin – *Leishmania infantum*, Lmx – *Leishmania mexicana*, Ldo – *Leishmania donovani*, Tcr – *Trypanosoma cruzi*, Tgr – *Trypanosoma grayi*, Tth – *Trypanosoma theileri*, Tcss – *Trypanosoma carassi*, Tvi – *Trypanosoma vivax*, TcIL – *Trypanosoma congolense*, Tbg – *Trypanosoma brucei gamiense*)

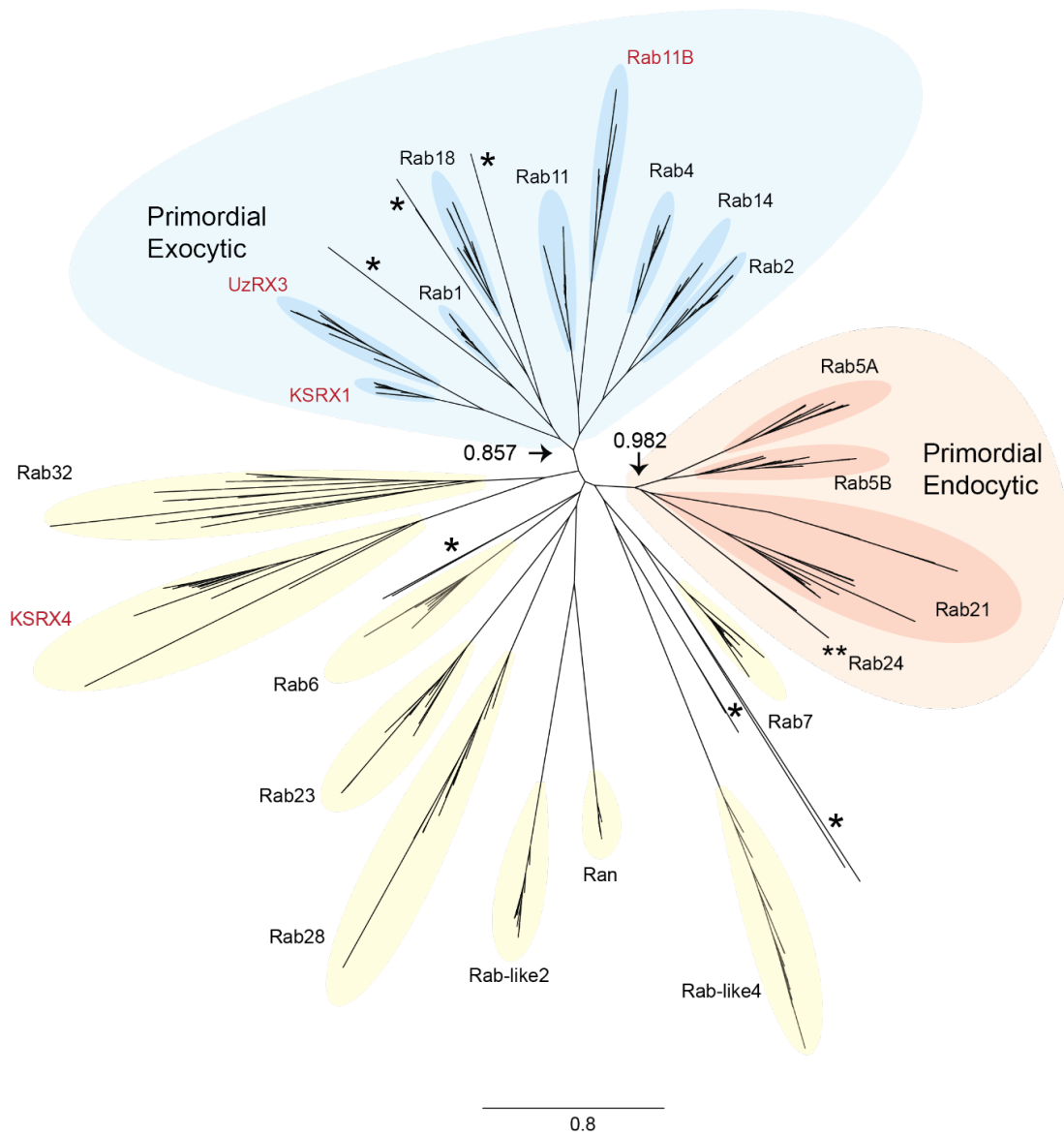
**Figure 3.2** above represents the number of Rabs, TBCs (left Y-axis) and the ratio of the number of Rabs:TBCs (right Y-axis) in each organism across kinetoplastids and reference eukaryotic representatives. As seen in other eukaryotes, the number of TBCs is lower than the number of Rabs in kinetoplastids, and the Rab:TBC ratio generally lies between 1 and 2. Ratios >2.0 (dotted horizontal line) were observed in lineages such as *N. gruberi*, land plants and *D. discoideum*, which possess a highly expanded set of Rabs (Gabernet-Castello *et al.*, 2013). In kinetoplastids, bodonids that show Rab:TBC just slightly above 2 also have similarly expanded Rab cohorts. Two of the three phytomonads also have a ratio >2, but it

appears to be not due to Rab expansion but a contraction of the TBC complement as the clear decrease in TBC numbers shows. In contrast, in *Trypanosoma vivax* there is a decrease in the number of Rabs compared to closely related species such as *Trypanosoma brucei* and *Trypanosoma congolense*, leaving it with nearly equal Rab and TBC repertoires (Rab:TBC ratio  $\approx 1$ ). Thus, there is a moderate level of variation in the Rab:TBC ratio even among genomes within the kinetoplastids, confirming the dynamic shaping of these repertoires. These data also suggest that the specificity of Rab and TBC interactions is complex. Indeed, multiple Rabs act as substrates for GAP activity of a single TBC and vice versa. For example, Gyp1, the *Saccharomyces cerevisiae* TBC-D protein is able to act as a GAP for yeast Rab1, 7 and 5-like (Du, Collins, and Novick 1998).

### 3.3 Phylogenetic analysis of kinetoplastid Rabs and TBCs

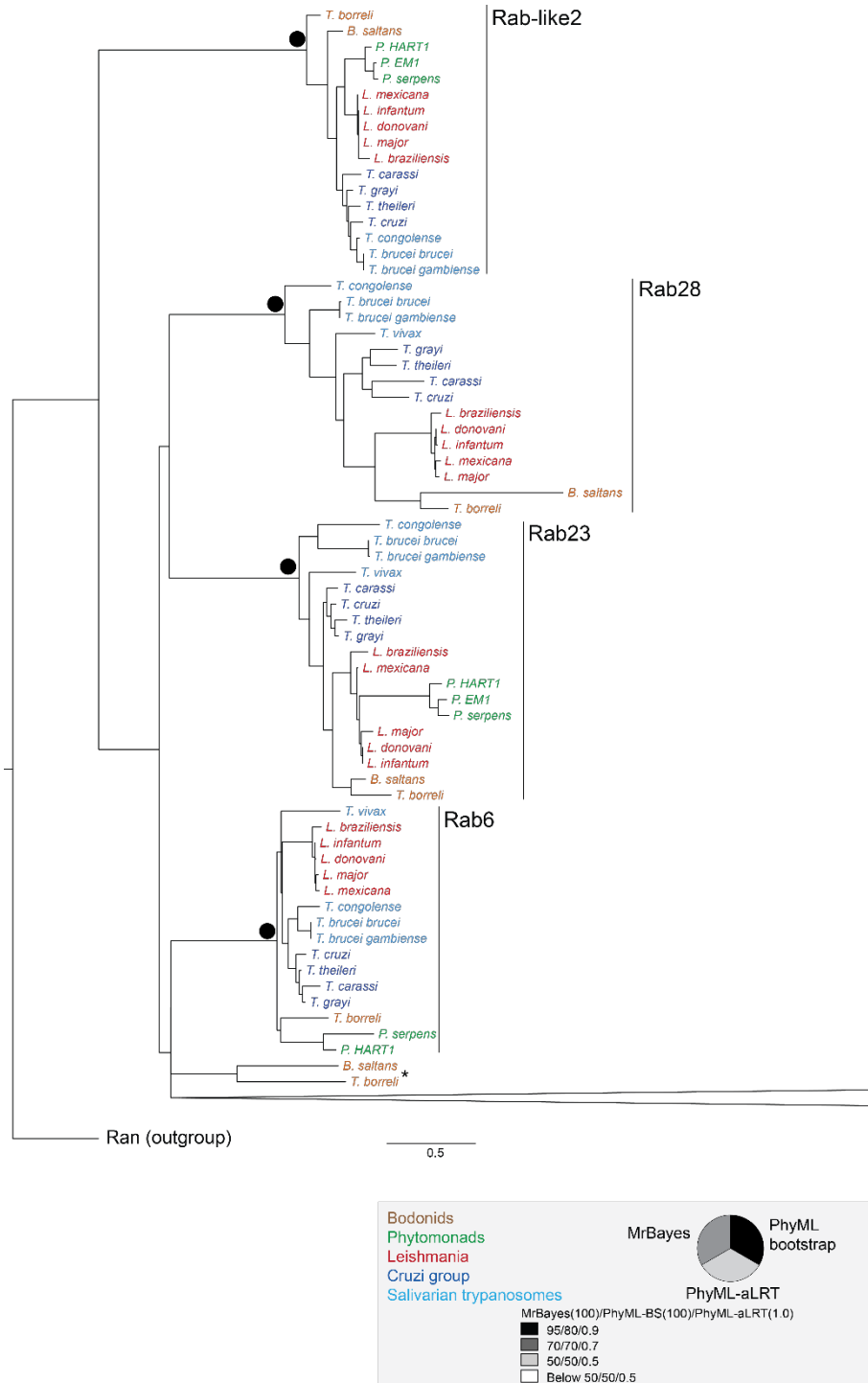
#### 3.3.1 Kinetoplastid Rabs

The non-redundant dataset of kinetoplastid Rabs comprised 382 sequences from 18 genomes. 21 well-defined clusters of Rab-subtypes were obtained but some sequences, mainly from *B. saltans* and *Trypanoplasma borreli* remained ‘orphan’. Twenty of these are presented in **Figure 3.3** below, which shows the topology of the kinetoplastid Rab family, with putative ‘orphans’ marked with a single asterisk. The cluster that was eventually assigned to Rab-like5 was omitted due to formation of long branches. Only PhyML aLRT (approximate likelihood ratio test) analysis is presented as higher order relationships between clades in PhyML bootstrap analysis was not well resolved and achieved low support. In contrast, individual clades formed by kinetoplastid Rabs were particularly well conserved with statistical support from PhyML aLRT (1.0), PhyML bootstrap (100) and MrBayes (%) always exceeding 0.9/90/90 and often close to full support. Due to the large size of the tree of all kinetoplastids taken together, it has been presented in four parts as **Figure 3.4**, **Figure 3.5**, **Figure 3.6**, and **Figure 3.7** below. *B. saltans* and *T. borreli* have an expanded set of Rabs, which appear much diverged and form several independent branches. Often the *B. saltans* and *T. borreli* sequences appear in pairs, indicating a common bodonid origin. It is unclear whether they emerged after the divergence of the trypanosomatid lineage but before the species radiation within the bodonids, or were secondarily lost in trypanosomatids. Sometimes however, single or two *B. saltans* sequences are found, which indicate species-specific innovations.



**Figure 3.3 Topology of the kinetoplastid Rab family**

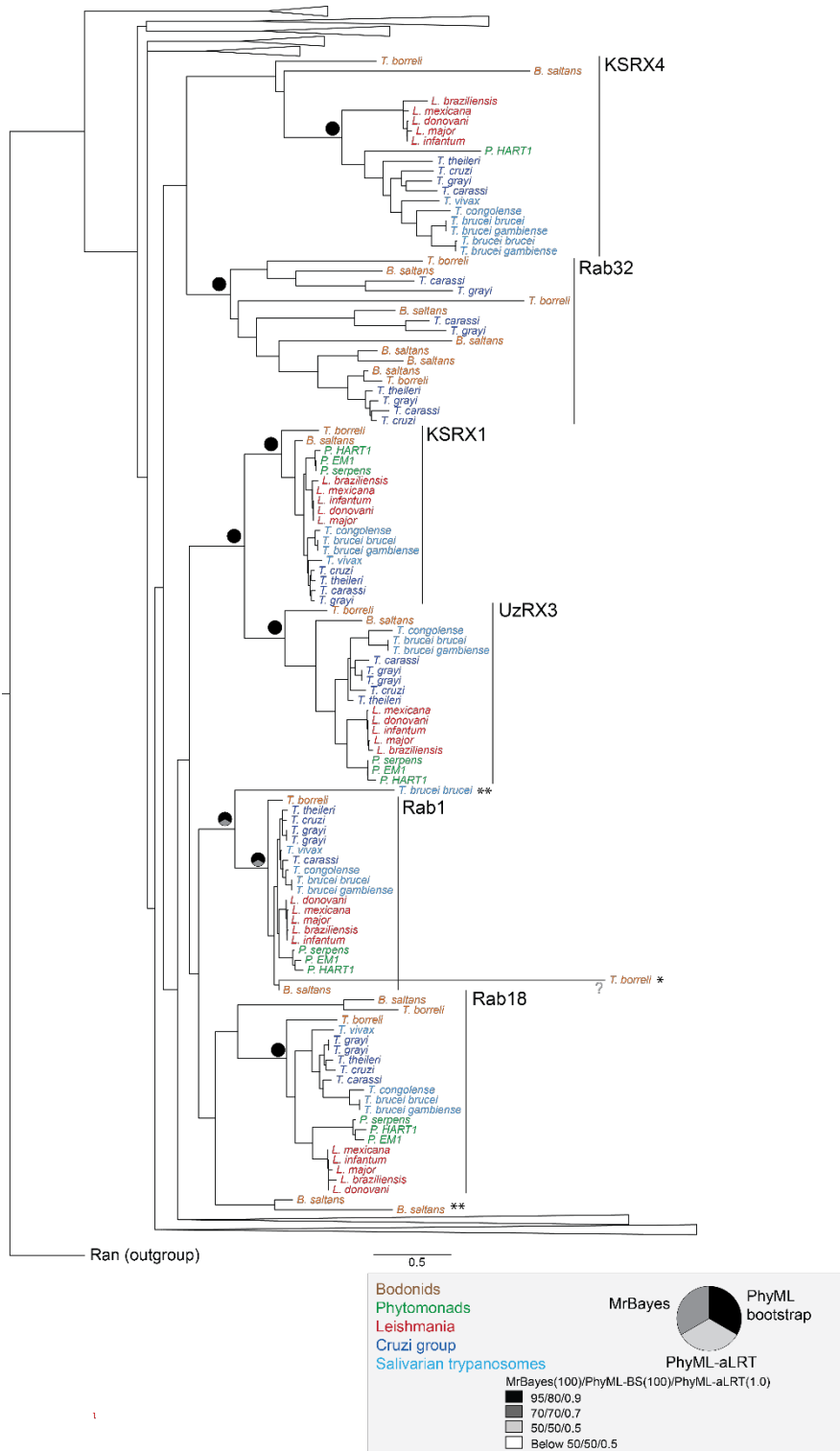
Best unrooted PhyML topology of all putative kinetoplastid Rabs is shown. Lighter pink and blue shades represent the two previously described primordial endocytic and exocytic super-clusters respectively (arrows indicate PhyML aLRT support values). Individual Rab subfamily clusters are also shaded in dark blue/dark pink/yellow. Names of lineage-specific clades are in red, names of LECA Rabs are in black, single asterisks indicate ‘orphan’ sequences that failed to cluster with kinetoplastid or LECA Rabs, double asterisks indicate orphan sequences assigned as Rab24.



**Figure 3.4 Kinetoplastid Rab phylogeny – Part1**

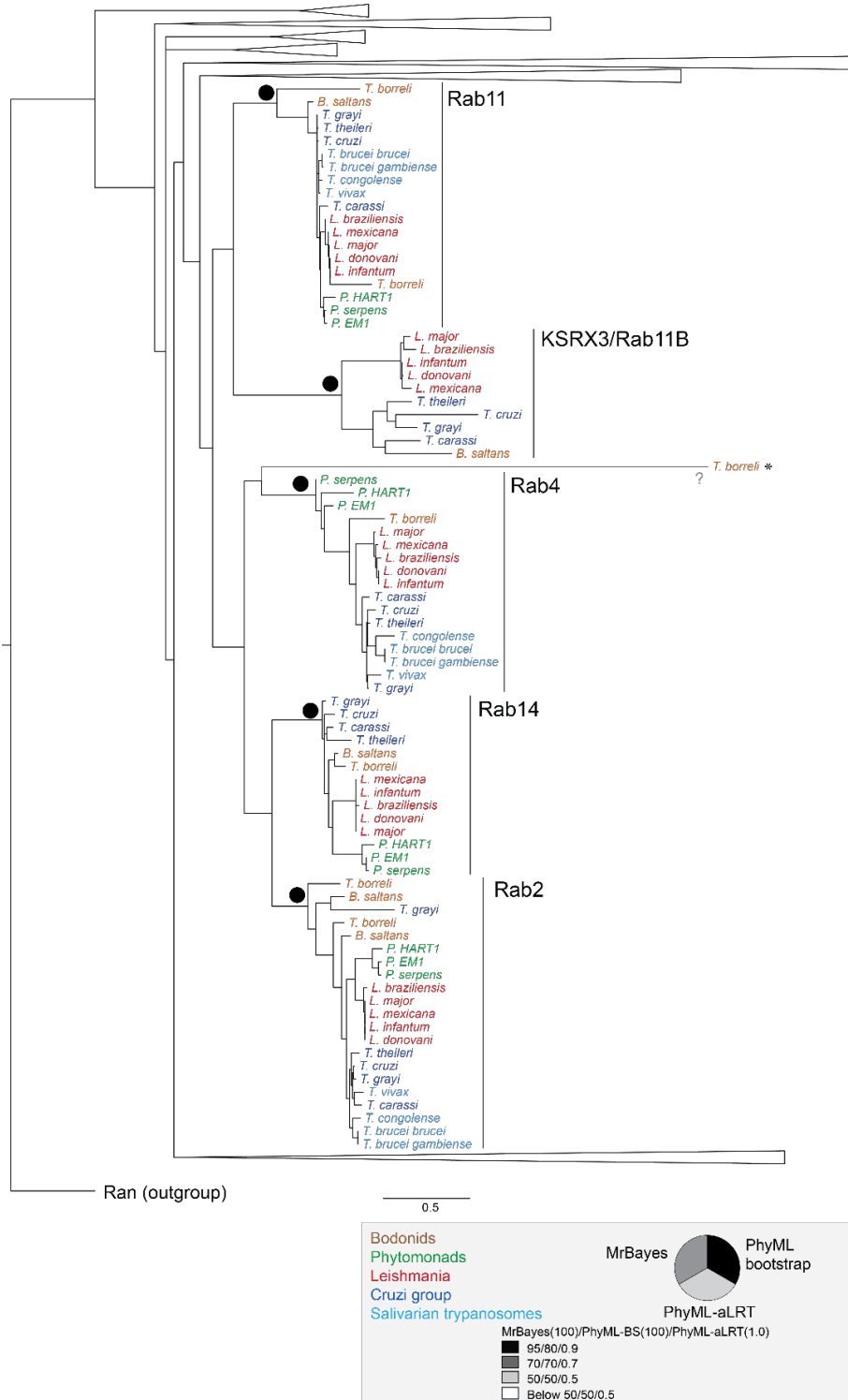
Best PhyML topology is presented in four parts. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 100) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Taxa are coloured by phylogenetic relationships as shown in the key. Each cluster is marked with a vertical line and named according to the assignment of the representative of the said cluster according to **Figure 3.8** below.

Chapter 3: Evolution of Kinetoplastid Rabs and RabGAPs



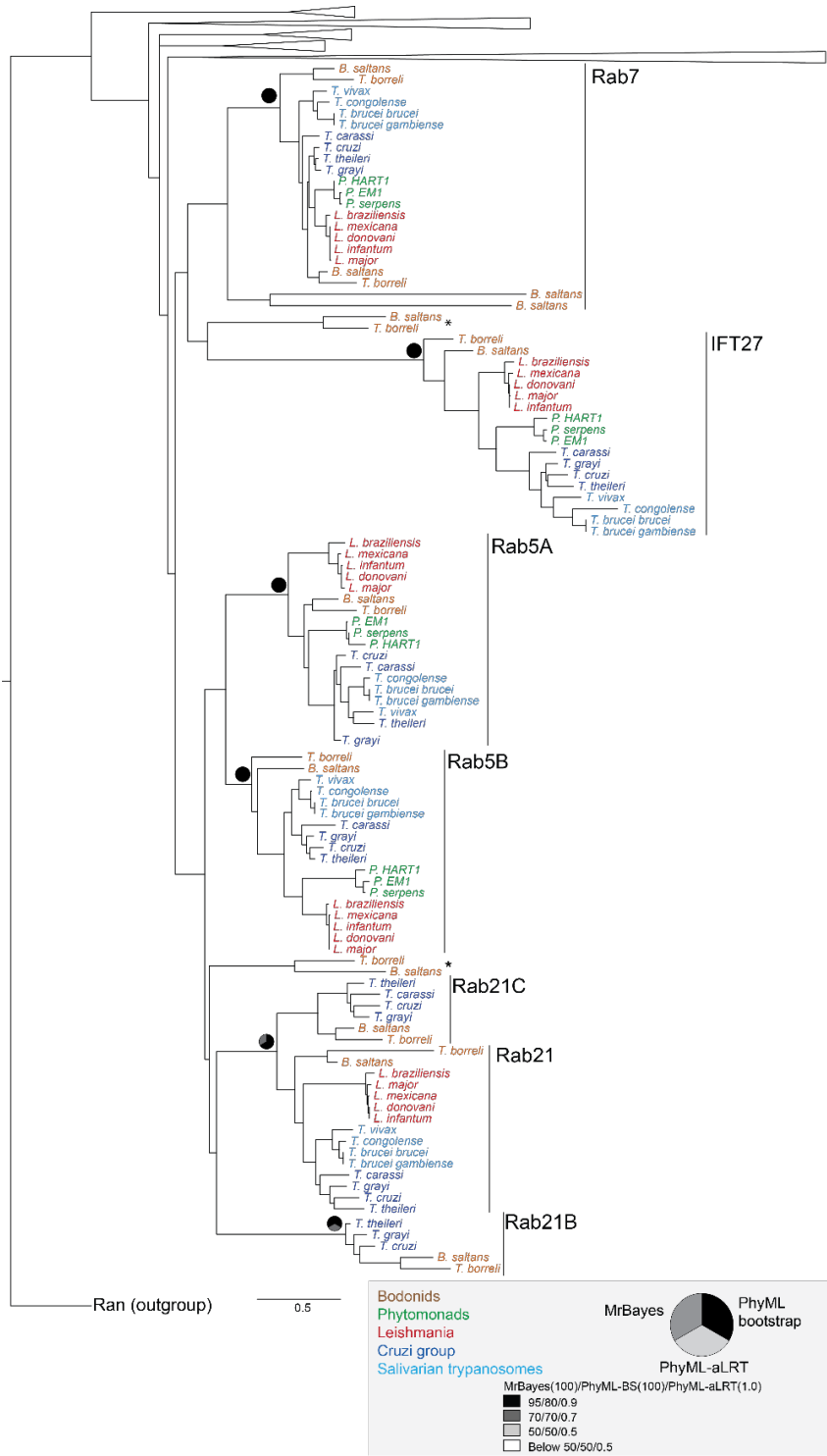
**Figure 3.5 Kinetoplastid Rab phylogeny – Part2**

Best PhyML topology is presented in four parts. See legend for Figure 3.4 above. Single asterisk denotes unassigned orphan Rabs, double asterisks denote tentatively assigned Rabs (based on clustering in at least 2 out of 3 analyses with >70% support).



**Figure 3.6 Kinetoplastid Rab phylogeny – Part3**

Best PhyML topology is presented in four parts. See legend for Figure 3.4 above. Single asterisk denotes unassigned orphan Rabs, double asterisks denote tentatively assigned Rabs (based on clustering in at least 2 out of 3 analyses with >70% support).



**Figure 3.7 Kinetoplastid Rab phylogeny – Part4**

Best PhyML topology is presented in four parts. See legend for Figure 3.4 above. Single asterisk denotes unassigned orphan Rabs, double asterisks denote tentatively assigned Rabs (based on clustering in at least 2 out of 3 analyses with >70% support).

Two representative sequences from each cluster along with all the orphans were added to a curated dataset of eukaryotic Rabs obtained from (Elias *et al.*, 2012).



This new dataset was analysed using MrBayes and PhyML to infer putative orthology of the kinetoplastid sequences. The resulting Mr Bayes tree, along with added support values from PhyML aLRT, is presented in **Figure 3.8** below. Even though the PhyML bootstrap analysis largely recapitulated in the PhyML aLRT and MrBayes topologies, no significant support values were obtained. The dataset was tested with RaxML bootstrap analysis (data not presented), which yielded comparable results to MrBayes and PhyML aLRT. The individual sub-clusters representing each landmark Rab-subtype was replicated with robust support in both PhyML aLRT and MrBayes analyses. Fifteen of 16 such sub-clusters were supported by  $>0.9$  likelihood ratio in PhyML aLRT and  $>90\%$  posterior probability in MrBayes (0.9/90)). It was therefore possible to assign Rab 1, 2, 4, 5, 6, 7, 11, 14, 18, 21, 23, 24, 28, 32, RTW and IFT27 representing a considerable proportion of the 23 Rabs that make up the LECA complement. Only the Rab14 sub-cluster nested in the Rab2, 4 and 14 cluster, had a lower support value ( $>0.7/70$ ). These assignments were mapped on to **Figure 3.3** above.

It was not possible to assign the representatives of several clusters to known eukaryotic subtypes. Some of these are listed on the Rab database (Diekmann *et al.*, 2011) as putative kinetoplastid specific Rabs and while others are unidentified “RabX”s or have very low confidence ( $\ll 0.25$ ) to known Rabs. They are summarised in **Table 3.1** below. After subsequent separate analysis with closely related clusters, we were able to assign some of these Rabs, for examples those closely related to and possibly originating from a duplication of Rab1/18 (KSRabX1, UzRabX3) and Rab11 (KSRabX3); this is discussed below. While the putative KSRabX4 was found closely associated with the Rab32 cluster in our kinetoplastid tree (see **Figure 3.5** above), it was not possible to assign it as such using representative eukaryotic sequences, suggesting a divergence of this sequence in kinetoplastids. From the overall topology of kinetoplastid Rabs, they appear to form two main super-clades representing the primordial exocytic and endocytic Rabs as described in (Elias *et al.*, 2012).

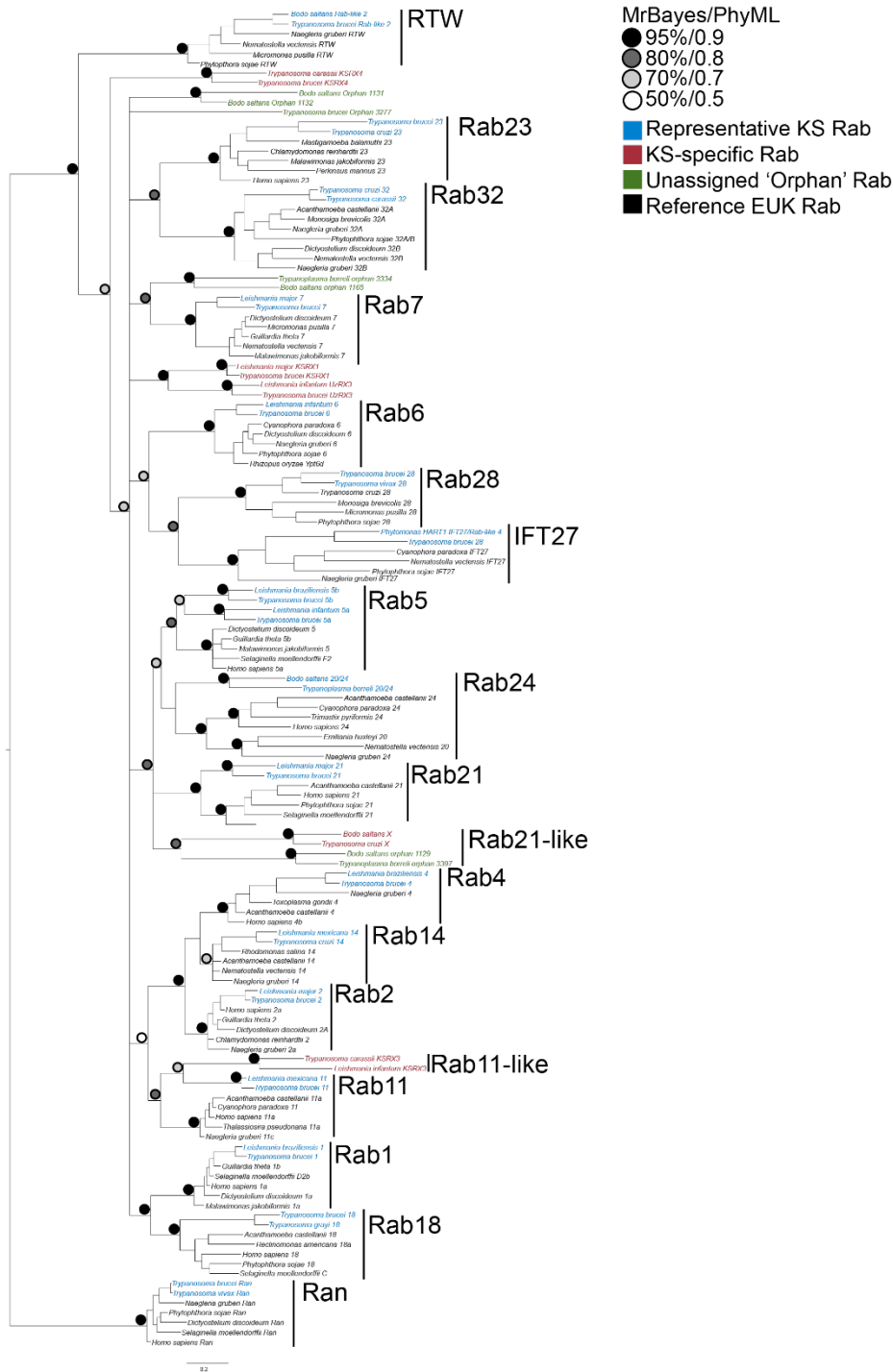


Figure 3.8 Assignment of kinetoplastid Rabs

Best PhyML topology is presented. Node values are iconised as circles for two support values each representing PhyML approximate likelihood ratio test (1.0) and MrBayes posterior probabilities (%) and colour-coded in grayscale as shown in the key. Two representative kinetoplastid Rabs from each sub-type cluster (blue), kinetoplastid/lineage-specific Rabs (red), and unassigned 'orphan' Rabs (green) are presented along with eukaryotic representative Rabs (black). KS – kinetoplastid specific, EUK – eukaryotic.

	<b>Rab</b>	<b>Previous name</b>	<b>Rabifier</b>	<b>Accession</b>	<b>Assigned as</b>
Representatives of lineage-specific clusters	TbKSRabX1	RabX1, Trab1	KSRabX1	Tb927.8.4610	Rab1/18 related
	TbUzRabX3	RabX2, Trab7	RabX	Tb927.8.4620	Rab1/18 related
	TbKSRabX4	TbRX2	RabX	Tb927.4.4220	Unassigned
	TbKSRabX4	duplicate TbRX2	RabX	Tb927.8.8140	Unassigned
	TcKSRabX3	not described	KSRabX3	TcCLB.511245.180	TcRab11B
	3470	not described	RabX (0.25)	TcCLB.511277.190	TbRab21C
Orphans	3277	TbRab1B	AtRabD1 (<0.25)	Tb927.9.15930	TbRab1-like
	1131, 1132	not described	RabX (0.5), Rab10 (<<0.25)	BSA07915.1, BSA07920.1	Rab1/18-related
	1156, 3336	not described	Rab13 (<<0.25), RabX (0.5)	BSA27275.1, Tbr_Locus_1334	Rab1/18-related
	1146, 3387	not described	RabX (0.5), RabX (0.5)	BSA14955.1, Tbr_Locus_7838	Rab24
	1165, 3334	not described	RabX (0.5), RabX (0.5)	BSA35135.1, Tbr_Locus_11581	putative Rab7-like
	1129, 3397	not described	Rab17 (<<0.25), Rab11 (<<0.25)	BSA04415.1, Tbr_Locus_9980	unassigned
	1149, 3324	Not described	Rab21 (<<0.25), Rab5 (<<0.25)	BSA18640.1, Tbr_Locus_10849	Rab21D

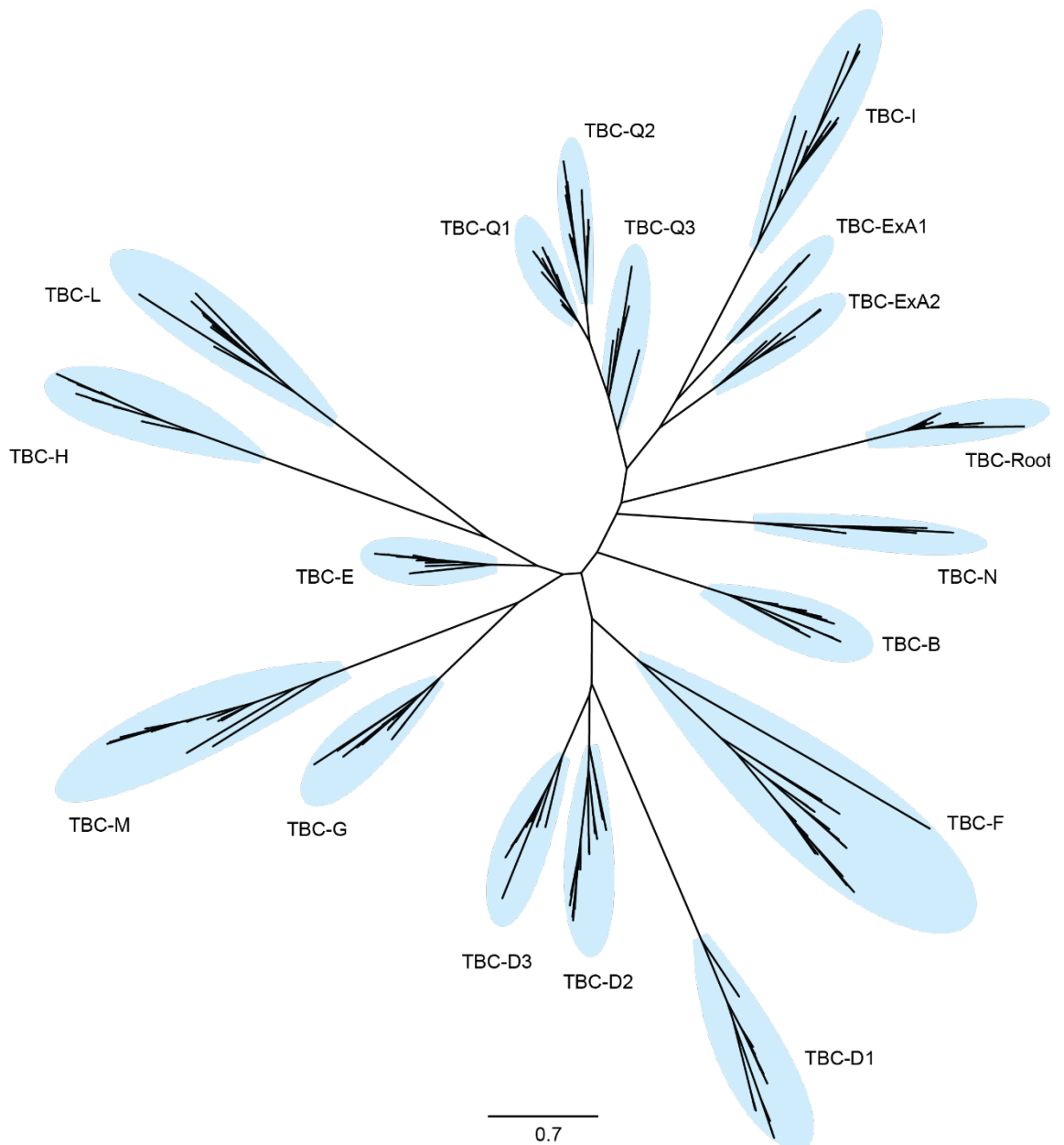
**Table 3.1 Lineage-specific Rabs**

Representative sequences from unassigned clusters as well as ‘orphan’ sequences are presented here. They are named according to previous classification, or a unique 4-digit ID code from this study. Previous descriptions in the literature if any are in column 2, assignment from ‘Rabifier’ (with confidence if support is 0.5 or below) is in column 3, Gene ID, in column 4, and tentative new assignments if any according to the phylogenetic analyses are in column 5. Criteria for assignment is clustering in at least two out of three types of phylogenetic analyses with support >70% (Rab11B, 21C, 1-like, 7-like) or >60% (Rab1/18-related).

### 3.3.2 Kinetoplastid TBCs

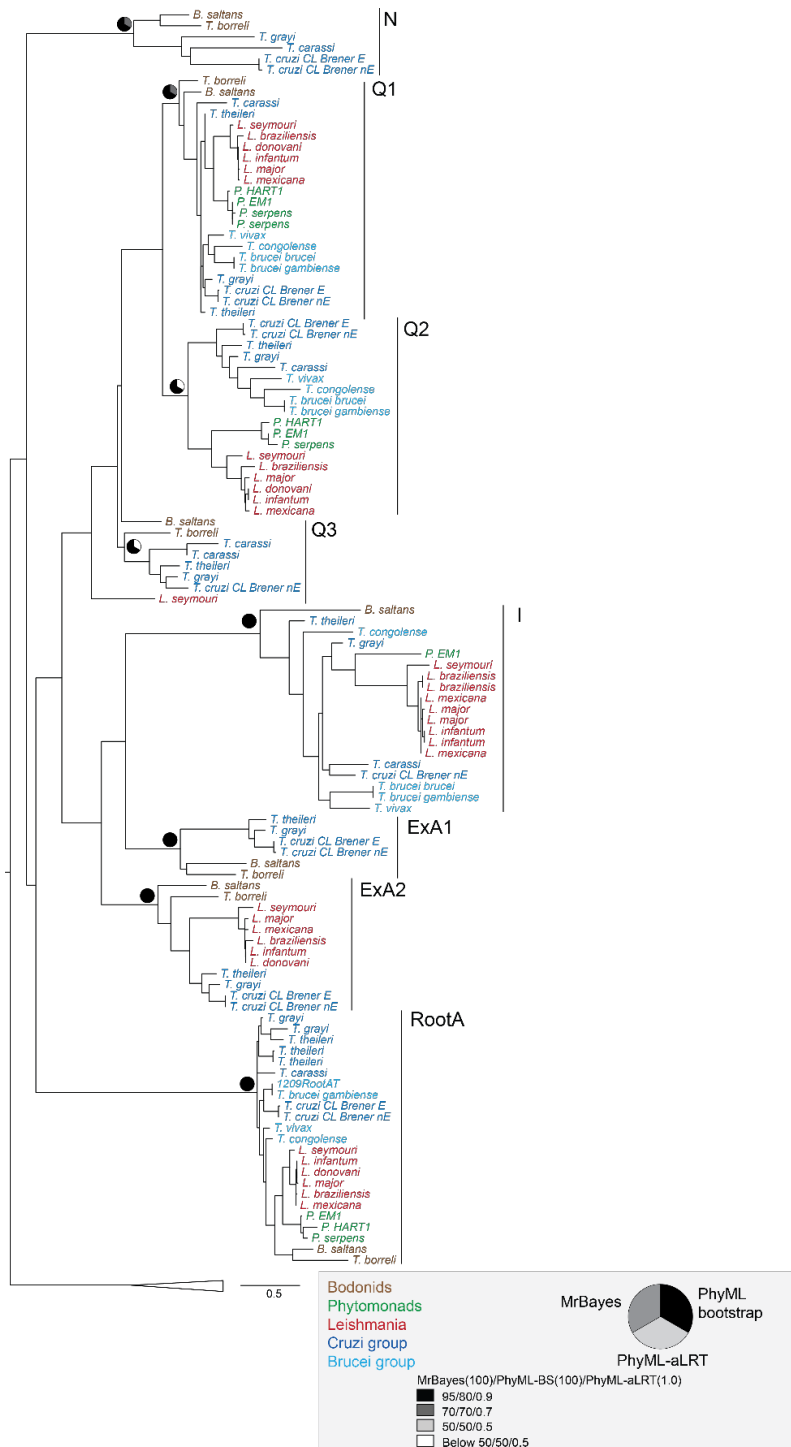
The non-redundant dataset of kinetoplastid TBC RabGAPs comprised 307 sequences from 18 genomes. 21 well-defined clusters of TBC-subtypes were obtained, which are presented in **Figure 3.9** below. Unlike the Rabs no ‘orphan’ sequences from the bodonids or any other kinetoplastid were found. As in the Rabs, the individual clades formed by kinetoplastid TBCs are particularly well conserved with statistical support from PhyML aLRT (1.0), PhyML bootstrap (100) and Mr Bayes (%) always exceeding 0.9/90/90 and often close to full support. Some clusters (Qs, F, D3, G and E) have lower support, particularly from PhyML bootstrap but all clusters are replicable across all three analyses as well as RaxML bootstrap analysis (data not presented). Due to the large size of the tree of all kinetoplastids taken together, it is presented in three parts as **Figure 3.10**, **Figure 3.11**, and **Figure 3.12**.

One representative sequence from each cluster was taken together with a curated dataset of eukaryotic TBCs obtained from (Gabernet-Castello *et al.*, 2013) and analysed in a manner similar to the Rab analysis. The resultant tree is presented in **Figure 3.13** below. Of the 18 putative subtypes, all were assigned to TBC D, L, B, I, Q, RootA, N, E, G, H, F and M. TBC-D and Q have three further subtypes as shown in Figure 3.9 below. Two of the clusters corresponded to the excavate-specific TBC ExA as previously defined in *T. cruzi* sequences (Gabernet-Castello *et al.*, 2013). As in the Rab analysis, all TBC sub-clades, with a few exceptions (TBC-Q, ExA and I) were supported robustly, i.e., above 0.8/800/95 (PhyML aLRT (1.0)/PhyML bootstrap (1000)/ MrBayes (%)). For these exceptions, even though the support values were lower, their assignments were replicable across all three analyses. These assignments are mapped on to **Figure 3.9**. Two kinetoplastid sequences, *T. brucei* D3 and *T. brucei* E were assigned as such in a previous study (Gabernet-Castello *et al.*, 2013), but showed an uncertain position in these phylogenies, often branching just outside the assigned clusters (indicated by a question mark in **Figure 3.13** below), thus suggesting divergent sequence.



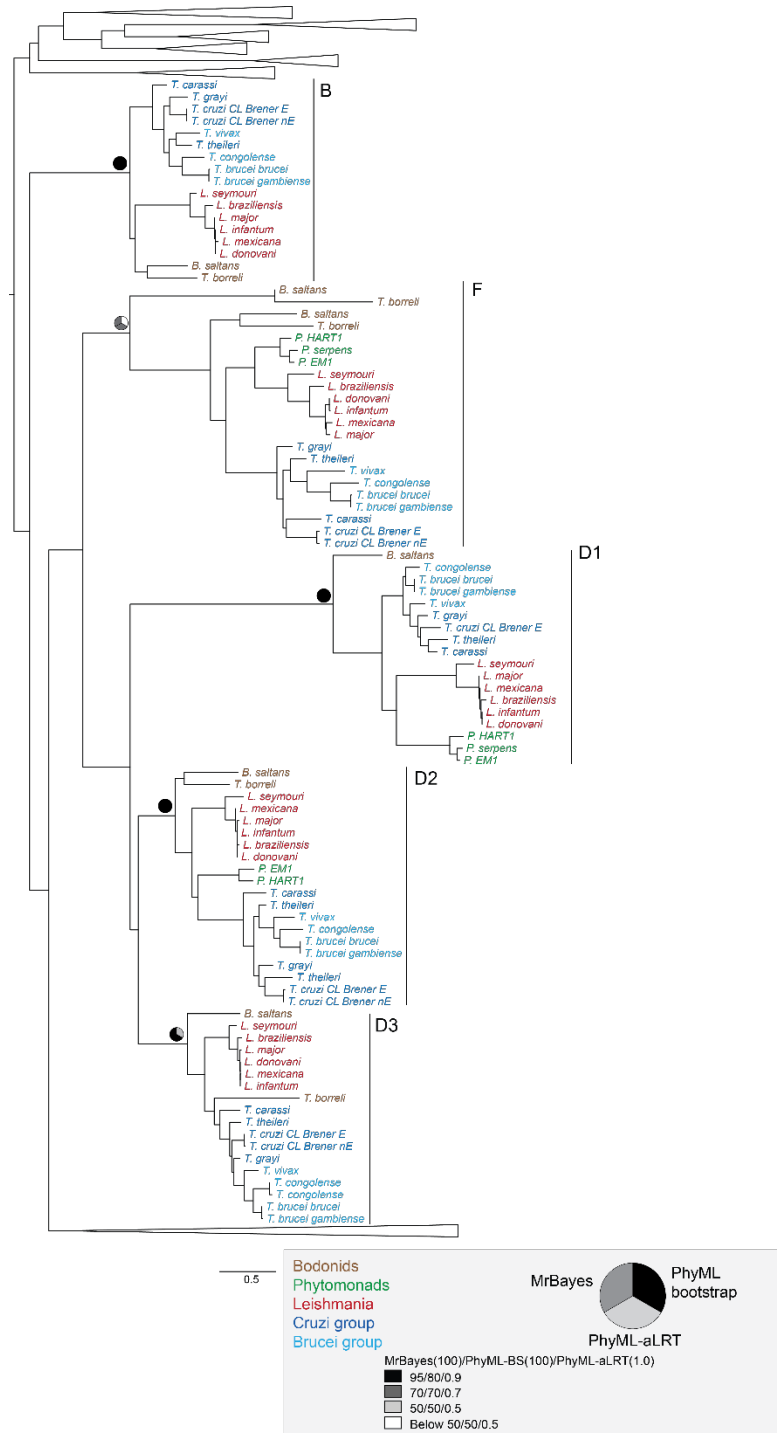
**Figure 3.9** Topology of kinetoplastid TBC family

Best unrooted PhyML topology of all putative kinetoplastid TBCs is shown. Individual TBC subfamily clusters are shaded in blue.



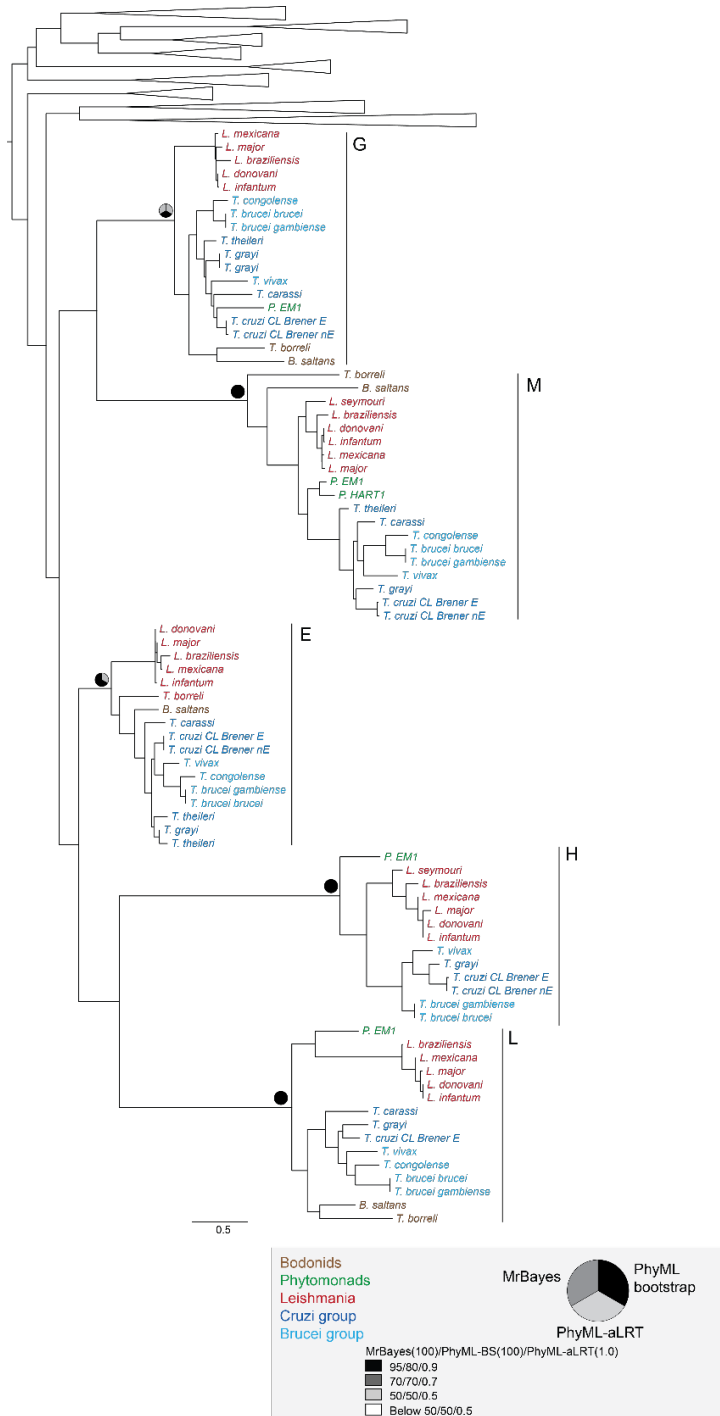
**Figure 3.10 Kinetoplastid TBC phylogeny – Part1**

Best PhyML topology is presented in three parts. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 100) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Taxa are coloured by phylogenetic relationships as shown in the key. Each cluster is marked with a vertical line and named according to the assignment of the representative of the said cluster according to **Figure 3.13** below.



**Figure 3.11 Kinetoplastid TBC phylogeny – Part2**

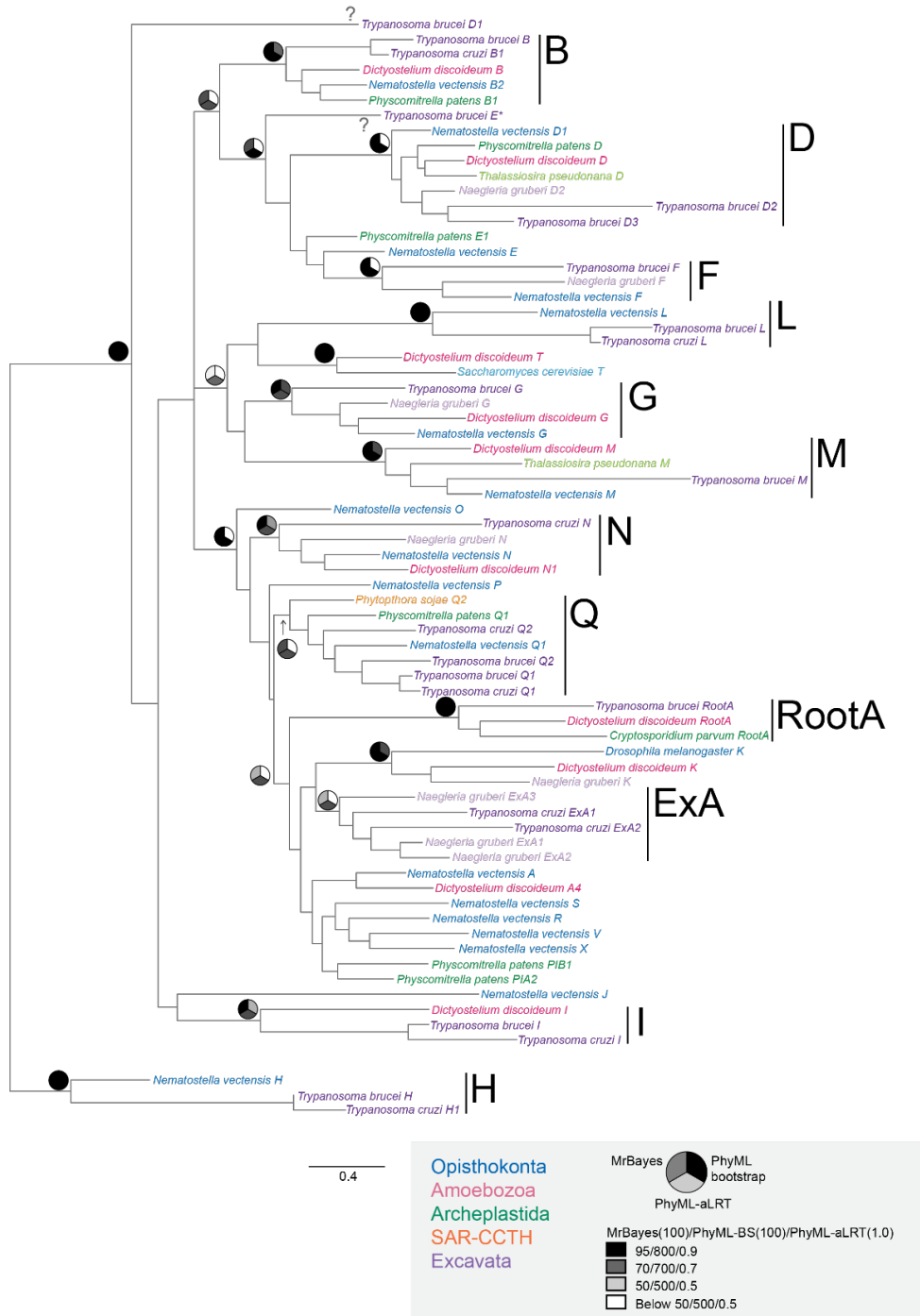
Best PhyML topology is presented in three parts. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 100) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Taxa are coloured by phylogenetic relationships as shown in the key. Each cluster is marked with a vertical line and named according to the assignment of the representative of the said cluster according to **Figure 3.13** below.



**Figure 3.12 Kinetoplastid TBC phylogeny – Part3**

Best PhyML topology is presented in three parts. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 100) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Taxa are coloured by phylogenetic relationships as shown in the key. Each cluster is marked with a vertical line and named according to the assignment of the representative of the said cluster according to **Figure 3.13** below.





**Figure 3.13** Assignment of kinetoplastid TBCs

Best PhyML topology is presented. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 100/0) and MrBayes posterior probabilities (%) and colour-coded in grayscale as shown in the key. Each phylogeny shows one representative kinetoplastid TBC from each sub-type cluster along with eukaryotic representative TBCs from five supergroups, colour-coded as in the key. Vertical lines mark the TBC-subtype clusters, ‘?’ indicates uncertain placement.

Higher order relationships among the kinetoplastid TBCs were not very well resolved. Kinetoplastid TBC tree shows a slightly altered topology from the previously described eukaryotic TBC phylogeny (Gabernet-Castello *et al.*, 2013). This may be because not all the TBC diversity of the eukaryotes, which likely arose after some of the major lineages diverged, (and hence restricted to specific groups of organisms) were represented in this tree. TBC-G and M are still clustered together, however there was no TBC-B, D, E and F cluster – just TBC-D and F were found on the same branch. The excavate-specific ExA sequences consistently clustered with TBC-I in the kinetoplastid tree but not in the pan-eukaryotic tree. In the kinetoplastid analysis, this relationship is well supported by the maximum likelihood analysis (0.951), less well by Bayesian posterior probabilities (66%) and not at all by the bootstrap analysis (10). Therefore, it is not clear from which TBC gene the excavate-specific sequences arose.

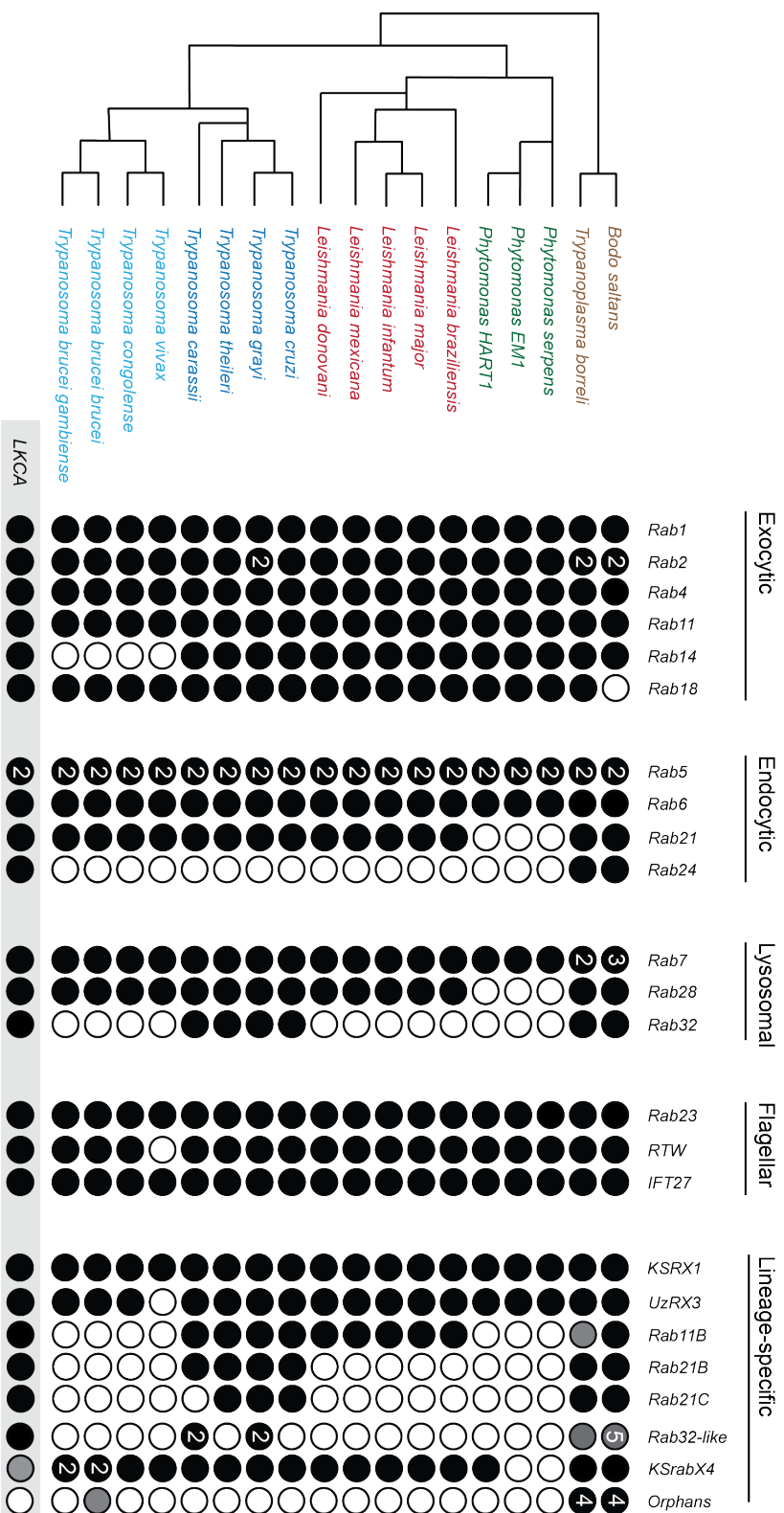
### 3.4 Patterns in the evolutionary history of Rabs and Rab GAPs

Overall, as can be seen in **Table 3.2** below, 72% of the Rabs and 100% of the TBCs found in LECA are nominally represented in the kinetoplastid genomes suggesting a well conserved cohort. On detailed examination of Rab and TBC-clusters, many instances of loss as well as pan-kinetoplastid and lineage restricted expansions were found. These patterns are represented in **Figure 3.14** (Rabs) and **Figure 3.15** (TBCs) below, and their implications are discussed in this section.

Trafficking family	Putative LECA repertoire	Number in LECA	Absent in kinetoplastids	Conservation	Lineage-restricted novel sub-types
<b>Rab GTPases</b>	1, 2, 4, 5, 6, 7, 8, 11, 14, 18, 20/24, 21, 22, 23, 28, 32A/B, 34, 50, RTQ, IFT27, Titan	22	8, 22, 32B, 34, 50, Titan	16/22 (72%)	KSRX1, UzRX3; Rab11B; Rab21B,C; Rab32-like; KSRabX4; Orphans
<b>TBC RabGAPs</b>	B, D, E, F, I, L, M, N, Q, RootA	10	none	10/10 (100%)	TBC ExA

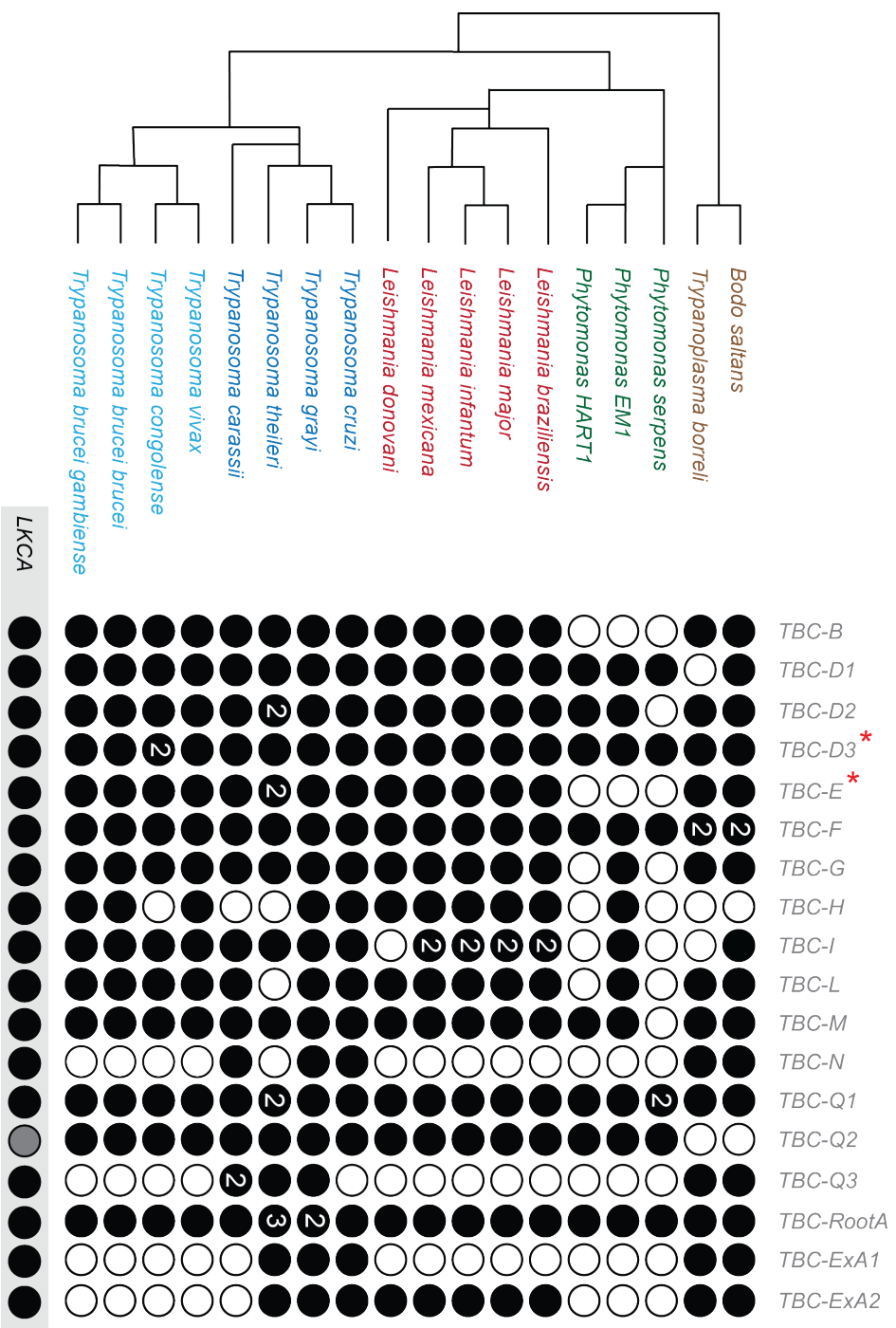
**Table 3.2 Comparison of kinetoplastid and LECA Rab and TBC repertoires**

Putative LECA repertoires of Rabs and TBCs are presented, along with subfamilies absent from kinetoplastids as well as novel lineage-specific subtypes.



**Figure 3.14 Representation of Rabs in kinetoplastids**

Individual subtype clades as found in kinetoplastids (assigned to known eukaryotic sub-types or lineage-specific) are shown by columns, with taxa shown as rows, with the hypothetical last common kinetoplastid ancestor as the lowest row highlighted by a grey box. A schematic phylogeny for the taxa is drawn on the left, derived from (Manna *et al.*, 2013). Black circles indicate at least one member of the clade has been found with phylogenetic support of 80/0.5/50 (MrBayes/PhyMLal.RT/PhyML bootstrap) or more; grey circles indicate lower support values but above 50 (MrBayes) or in the case of the *LKCA*, indeterminate presence from given data.



**Figure 3.15 Representation of TBCs in kinetoplastids**

Individual subtype clades as found in kinetoplastids (assigned to known eukaryotic sub-types or lineage-specific) are shown by columns, with taxa shown as rows, with the hypothetical last common kinetoplastid ancestor as the lowest row highlighted by a grey box. A schematic phylogeny for the taxa is drawn on the left, derived from (Manna et al. 2013). Black circles indicate at least one member of the clade has been found with phylogenetic support of 80/0.5/50 (MrBayes/PhyML/RT/PhyML bootstrap) or more; grey circles indicate lower support values but above 50 (MrBayes) or in the case of the LKCA, indeterminate presence from given data. Red asterisks indicate uncertain assignment according to Figure 3.13 above.

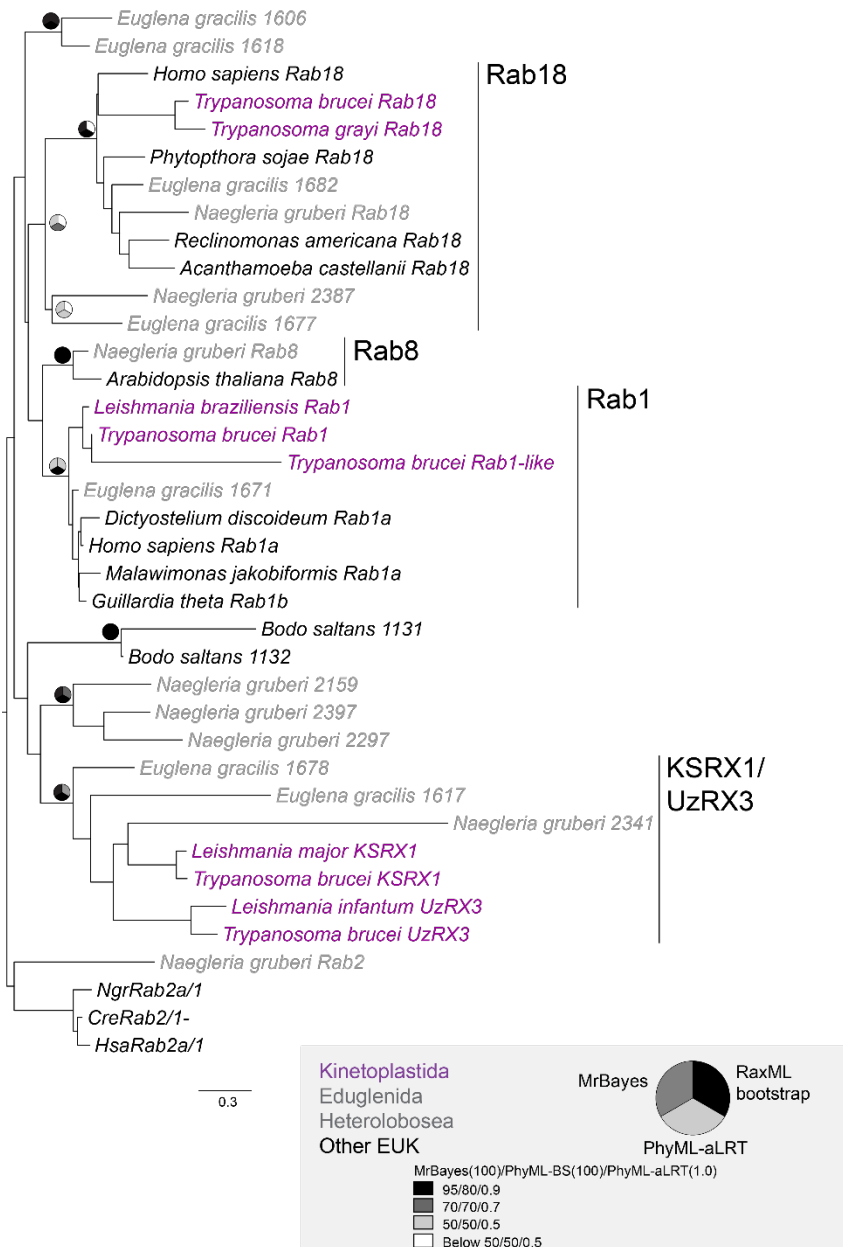
### 3.4.1 Secretory Rabs are well conserved

Rab1, 8 and 18, associated with anterograde pathways, are a monophyletic group of ancient Rab family paralogs that have been shown to be part of the higher order ancestral clades of Rabs that are primarily linked with exocytic processes, as described in (Elias *et al.*, 2012). Rab8, which showed a complex pattern of losses in non-metazoan eukaryotes, was lost from all kinetoplastids while canonical Rab1 and 18 were present. Rab8 was present in the heteroloebosid *N. gruberi* but was not identified in *E. gracilis*, indicating that this may be a loss from the whole euglenid lineage rather than just kinetoplastids. Rab1 was found duplicated several times in Opisthokonta, mainly in Metazoa, giving rise to Rab19, 30, 33, 35, and X6. Independent duplications of Rab1 were found in many other phyla such as apicomplexans, heterokonts and land plants (Klöpffer *et al.*, 2012). On the other hand, Rab18 showed a simpler evolutionary history, being sporadically duplicated in metazoan lineages or lost in several phyla but largely stably retained.

Rab1/18 shows a complex pattern of expansion in kinetoplastids. It is likely that paralogous expansion of Rab1 or 18 may have given rise to two lineage-specific Rabs KSRX1 and UzRX3. It was not possible to phylogenetically resolve the identity of the Rab that gave rise to these proteins, which are found across kinetoplastids and appear to have related paralogs in both *N. gruberi* and *E. gracilis*. In order to clarify the relationships, putative Rab1, 8, 18 and KSRX1 and UzRX3 representatives from kinetoplastids, and the curated eukaryotic set, along with the *N.gruberi* and *E.gracilis* sequences that clustered with them were analysed (see Figure 3.16 below). KSRX1 and UzRX3 have a sequence similarity of 60.7% and 44% identical (according to pairwise sequence alignment of *T. brucei* sequences by EMBOSS Matcher) and are very closely related to each other 1.0/100/100 (PhyML aLRT (1.0)/PhyML bootstrap (100) /MrBayes (%), see **Figure 3.5** above). They thus appear to have arisen from a duplication event and lie next to each other in the genome (gene IDs Tb927.8.4610 (TbKSRX1) and Tb927.8.4620 (TbUzRX3)). The highly conserved nucleotide binding sequence ‘WDTAGQ’ was retained as such in KSRX1 but changed to ‘WDTSGQ’ in UzRX3, which was shown to be a GTP-locked form of the protein (Field *et al.*, 2000). Their relationship with the Rab1/18 clade was well supported in the maximum-likelihood tree (0.925 PhyML aLRT) but not in the Bayesian analysis

(59%). Regardless of their relationship to canonical Rabs however, they likely arose before the formation of the kinetoplastid lineage due to their clustering with sequences from *N. gruberi* and *E. gracilis*, which are called as bonafide Rabs by Rabifier, but their assignation to known sub-types has generally very low confidence ( $\ll 0.25$ , see **Table 3.1** above for details). These proteins were previously described as RabX1 and 2 respectively and studied in *T. brucei*. It is interesting to note that they were found to have a role not in trafficking but in regulating infectivity in the fly midgut (Natesan et al. 2009). While the localisation of UzRabX3 was comparable to Rab1 (Golgi, early secretory pathway), KSRX1 showed a diffuse localisation spread throughout the cell. This is reminiscent but perhaps an even more divergent example of the Rab1-related Rab35 in opisthokonts, which has a role in endocytosis at the plasma membrane rather than at the ER-ERGIC-Golgi where Rab1 is usually active (Kouranti et al. 2006; Sato et al. 2008).

*N. gruberi* and *E. gracilis* possess further Rab1/18 related sequences, as does *B. saltans*, which appears to have two additional proteins related to the clade (called *Bodo saltans* 1131 and 1132 in **Figure 3.16** below). These two proteins share little sequence similarity (38.4%, EMBOSS Matcher) but appear to have arisen from the same ancestral gene in a species-specific manner (phyML aLRT grouping support 0.99, MrBayes 100%). Further, a lone *T. brucei* protein, previously described as Rab1B (Ackers *et al.*, 2005) and which Rabifier (Diekmann *et al.*, 2011) describes as AthRabD1 (a plant paralog of Rab1) *albeit* with very low confidence, is also found to group consistently within the Rab1 clade, albeit with a relatively long branch (see **Figure 3.16** below). These results indicate an expansion of the Rab1/18 clade at different points of the discicristate lineage while the ancestral forms of Rab1 and 18 remain conserved. At least in *T. brucei*, this neo-functionalisation is likely related to non-transport functions; experimental studies in other organisms in the lineage may reveal if this is the case elsewhere. These results will have important implications for our understanding of the role of G protein signalling in protozoa.



**Figure 3.16 Phylogenetic assignment of Rab1/18-related sequences**

Best PhyML topology rooted on Rab2 sequences is presented. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 100/0) and MrBayes posterior probabilities (%) and colour-coded in grayscale as shown in the key. PhyML bootstrap values were too low to be meaningful and hence not shown here for clarity. Taxa are coloured according to the key, EUK, eukaryotes.

Rabs 2, 4 and 14 form another monophyletic group of primordial exocytic Rabs. Rab2, which mediates ER to Golgi trafficking, was independently duplicated several times in lineages such as Metazoa, Heterokonta and Angiosperma. In kinetoplastids, Rab2 was found duplicated in the bodonids as well as in *T. grayi* and the three sequences were clustered together on the outer branch of the main Rab2



clade (See Figure 3.6 above). This indicates an origin prior to the establishment of the trypanosomatid lineages, most of which appear to have lost this duplicated Rab2 gene.

Rab4, which is involved in endosome to plasma membrane recycling, shows no major expansions in eukaryotes and is only sporadically lost, was found singly across all kinetoplastids. Rab14, associated with lysosomal delivery of phagosomes, was lost in the brucei group trypanosomes (see Figure 3.14 above), possibly aiding in further streamlining of the rapid endocytosis which is their main mechanism of immune evasion in mammalian host bloodstream. Like AP-2 however (Manna *et al.*, 2013), Rab14 was not lost in the other extracellular trypanosomes of the cruzi group. Not much is known about how these organisms, which lack VSG, survive in the host bloodstream but it appears that the mechanisms they employ are divergent rather than parallel to those seen in the brucei group.

Rab11 is closely related to the Rab2, 4, 14 group and is involved in the recycling of endosomes from the trans-Golgi network to the surface and it is implicated in the phagophore formation step of autophagy in mammals. It appears to be utilized in the recycling of different cargoes in *T. cruzi* and *T. brucei* respectively, both of which are necessary for immune evasion in the host. *T. cruzi* Rab11 localizes to the contractive vacuole to regulate trafficking of trans-sialidase to the plasma membrane (Niyogi *et al.*, 2014). *T. brucei* Rab11 is implicated in recycling of variant surface glycoprotein (VSG), which takes up a dominant share of the endocytic activity while in the host. Rab11 was stably retained across kinetoplastids and found duplicated in *T. borreli*. A Rab11-like protein (named KSRab11B) was also identified, which likely arose from duplication of Rab11 after divergence of the kinetoplastid lineage as none of the *E. gracilis* or *N. gruberi* sequences cluster with them. It appears to be retained only in *B. saltans*, *Leishmania spp.* and the cruzi group. In the mammalian lineage, analogous lineage specific duplication products Rab11a and b have distinct functions and localise to distinct vesicular compartments despite high sequence homology (Lapierre *et al.*, 2003). Rab11 phylogeny in Archaeplastida also exhibits a pattern of less conserved clusters suggesting neo-functionalisation of the duplicate genes (Petrželková and Eliáš, 2014). It is therefore likely that the kinetoplastid Rab11B has also acquired a different function: it may add nuance to the surface remodelling that intra-cellular parasites such as *T. cruzi* and *Leishmania* may require for invasion and survival in host cells, while not being necessary in African

trypanosomes. Again, not enough is known about immune evasion and host interactions of the other putative extracellular trypanosomes of the cruzi group, which also retain KSRab11B, to comment on a possible role.

Rab6 branches close to the primordial endocytic group but is not a member. It localises mainly to the Golgi where it mediates retrograde transport within the Golgi and to the ER (Luo and Gallwitz, 2003). It is well conserved across eukaryotes and is expanded in vertebrates and duplicated in angiosperms and present singly across all kinetoplastids.

### 3.4.2 The GAPs for secretory Rabs

Among the TBC subfamilies that regulate these secretory/anteriograde Rabs is TBC-Q whose members were found to act as a GAP for Rabs 1, 4, 14, 11 and 6 (Fukuda, 2011). TBC-Q was found expanded independently in many eukaryotic lineages including in the kinetoplastids (Gabernet-Castello *et al.*, 2013). Three kinetoplastid subfamilies of TBC-Q were found: Q1, Q2, and Q3. Both the basal bodonids (but not *E. gracilis* and *N. gruberi*) were represented in each of these indicating that expansion took place likely at the base of the kinetoplastid lineage. Q2 is retained as a single copy while Q1 is found duplicated in *T. theileri* and *P. serpens*. Q3 appears to be lost in most trypanosomatids except in the cruzi group where *T. carassi* has two copies and *T. theileri* and *T. grayi* possess one copy each. The pattern of Q3 representation showed no correlation with Rab1, its putative duplicates, or Rab6, but was roughly similar to Rab32, and 32-like, Rab21B and C, which were retained only in the bodonids and the various cruzi group members (see Figure 3.14 and Figure 3.15 above). TBC-M (mammalian TBC1D20) is also found to be a GAP for Rab1 and 2 in mammals (Haas *et al.*, 2007). It shows no expansions in eukaryotes and is lost in several archeplastid, SAR and excavate species, but found singly across all kinetoplastids and is lost only from *P. serpens*.

The yeast homolog of TBC-D (Gyp1) was shown to act as a GAP on Ypt1 (yeast Rab1) and is required for the recycling of membrane material in Snc1 (yeast VAMP7) positive vesicles via the fusion of these endosomal vesicles with the Golgi (Lafourcade *et al.*, 2004). It also putatively affects Rab7, 8 and Rab-like5 (Du *et al.*, 1998). TBC-D is stably retained across all eukaryotes with occasional duplications. In kinetoplastids, three TBC-D clusters D1, 2, and 3 were found. All of them were well

represented across all taxa and appeared to have arisen before the radiation of kinetoplastid species. Exceptions include only sporadic losses and gains, which may suggest a degree of redundancy and perhaps some plasticity in TBC repertoires: *T. borreli*, doesn't possess TBC-D1, *P. serpens* does not have TBC-D2, and *T. theileri* has an extra copy of TBC-D3.

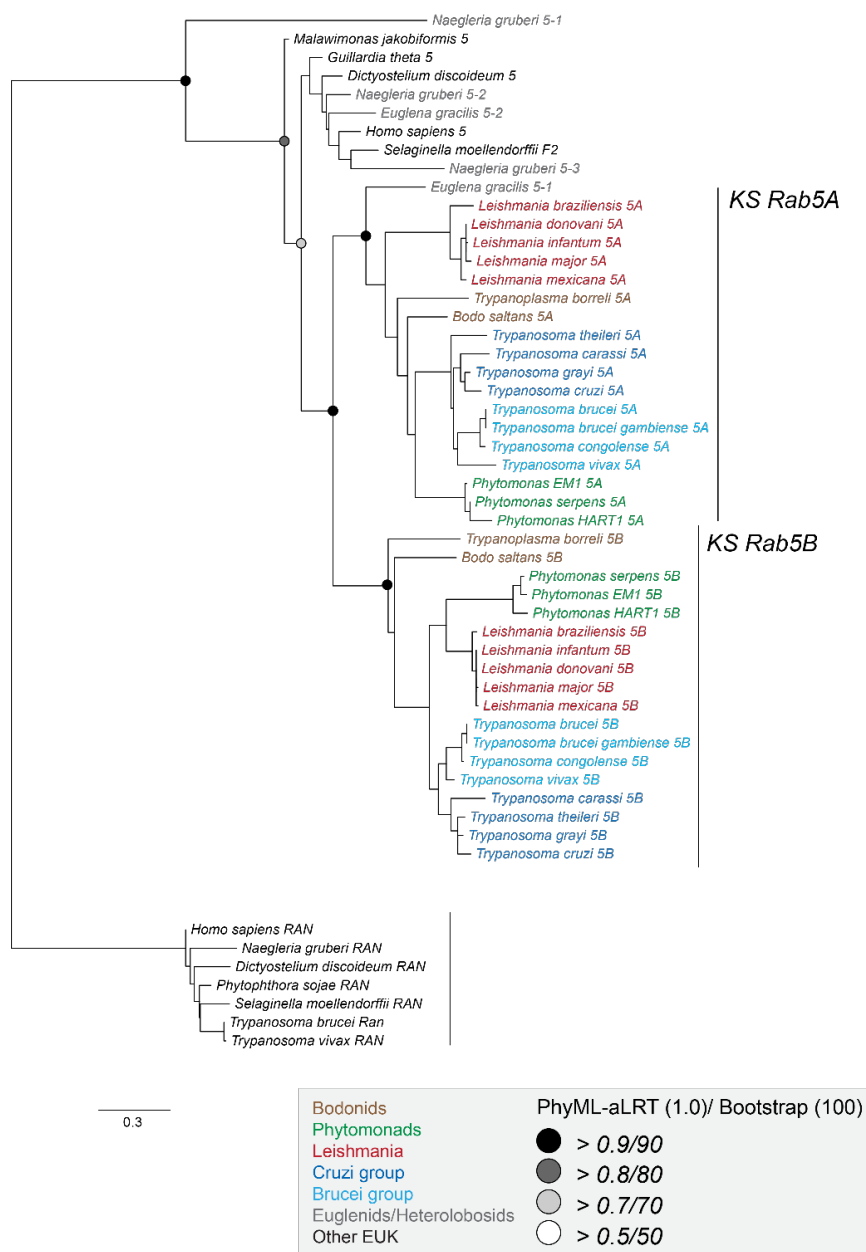
The mammalian TBC-E (TBC1D13) was reported to be a Rab35-specific GAP involved in mediating GLUT4 trafficking in adipocytes (Davey *et al.*, 2012). However, while Rab35 is derived from Rab1 duplication that is specific to Metazoa (Klöpffer *et al.*, 2012), TBC-E is found across eukaryotes, *albeit* patchily (Gabernet-Castello *et al.*, 2013). It is therefore possible that TBC-E has acquired a lineage-specific function in metazoans and their Rab substrate(s) in other eukaryotes remain to be identified. TBC1D13 is also binds Rab1 (found across eukaryotes) and Rab10 (Metazoa-restricted), but it does not have a GAP activity towards them. TBC-E is found singly in kinetoplastids, with the exception of the phytomonad lineage where it was not identified, and *T. theileri* where it was duplicated. This pattern does not correlate with Rab1 or its putative duplication products in kinetoplastids, which were generally found across all taxa (See **Figure 3.14**, Rab1, KSRX1, and UzRX3).

TBCs are not very well studied and the few studies available show evidence both for and against cross-species conservation of TBC function. Yeast two-hybrid studies in trypanosomes indicate that TBC-Rab interactions as found in yeast and mammals are only partially conserved in *T. brucei*. Rab1 GAP activity of TBC B, D, M, and Q members and the Rab11 GAP activity of TBC-Q1 is possibly retained. However, the other putative Rab substrates for TBC-Q1 (namely Rab 4, 18, 28, 5A) and TBC-Q2 (namely UzRabX3) largely differ from previously described interactions of TBC-Q members with Rabs 2A, 4, 6, 8A, 10, 14, 22A, 23, 35, 36 (Fukuda, 2011; Gabernet-Castello *et al.*, 2013). Moreover, studies in yeast showed that although the three TBCs D, M, and Q affect the same Rab1, the context and location of each activity differs. While Gyp8 (TBC-M) functions at the ER, Gyp1 (TBC-D) as well as Gyp5 and Gyl1 (TBC-Q) act as Rab1-GAPs at the Golgi complex (De Antoni *et al.*, 2002; Barr and Lambright, 2010). We do not know the effect of duplication of Rabs or TBCs on TBC activity. However, given the different putative Rab partners of TBC-Q1 (Rab11) and Q2 (UzRabX3), it is unsurprising that they have non-redundant

functions and show an additive effect when knocked down simultaneously in *T. brucei* (Gabernet-Castello *et al.*, 2013).

### 3.4.3 Primordial endocytic Rabs and their GAPs

Six ‘primordial endocytic’ Rabs have been described: Rab 5, 20/24, 21, 22 and 50 (Elias *et al.*, 2012). Of these, three are represented in kinetoplastids: Rab5, 21 and 24. Rab5 mediates early endocytic traffic and is found to be independently duplicated in many lineages including basal Fungi, Saccharomycotina, Kinetoplastida and Apicomplexa, and is further expanded in vertebrates and angiosperms. While yeast Rab5 paralogs were found to have overlapping functions (Singer-Kruger, 1994), mammalian paralogs have distinct functions in endocytosis of surface receptors and interaction with intracellular pathogens (Alvarcz-Dominguez and Stahl, 1998; Alvarez-Dominguez and Stahl, 1999; Barbieri *et al.*, 2000). Distinct functions were also found in *T. brucei*. While TbRab5A and B co-localise in procyclic cells, this is not the case in the bloodstream form found in the host bloodstream: here the TbRab5A endosomes were found to carry VSG and transferrin, which is endocytosed by the GPI-anchored transferrin receptor, and TbRab5B endosomes did not, and only carried ISG<sub>100</sub> which is essentially an unknown epitope (Pal *et al.*, 2002). Across kinetoplastids, the two Rab5 paralogs were stably represented in all taxa, indicating a basal origin of duplication (see **Figure 3.17** below).



**Figure 3.17 Evolutionary history of kinetoplastid Rab5**

Best PhyML topology is presented. Statistical support from PhyML aLRT and bootstrap analyses are presented as circles at key nodes, filled in in grayscale according to the key shown (bottom). PhyML bootstrap values were too low to be meaningful and hence not shown here for clarity. Taxa are coloured according to the key.

It is unclear if some kind of functional differentiation has taken place in kinetoplastids other than brucei group trypanosomes where many such GPI-anchored molecules are implicated in virulence systems (Ferguson 2000), and make such a bifurcation of uptake-function advantageous. Only a single Rab5 homologue was characterised in *T. cruzi* (Araripe *et al.*, 2005) but localisation and cargo association

were not studied. A single Rab5 has also been described in *L. donovani*, as a mediator of homotypic early endosome fusion of haemoglobin (Singh *et al.*, 2003). Thus, Rab5B remains unstudied in other kinetoplastids – it would be interesting to compare the possible differences in neo-functionalisation, or lack thereof in closely related organisms with distinct differences in the rates and requirements of endocytosis.

Rab21 is implicated in the trafficking of integrins, which are metazoan-specific receptors that mediate cell adhesion, migration, maintenance of cell polarity, and cytokinesis (Pellinen *et al.*, 2006, 2008; Mai *et al.*, 2011). They are particularly known for their function in polarised cells, where they are found in vesicles near the apical surface (Opdam *et al.*, 2000) and in non-metazoan *Dictyostelium discoideum*, it regulates phagocytosis (Khurana *et al.*, 2005). Rab21 shows no major expansions in any eukaryotic lineage and shows sporadic loss – it has been estimated to have been lost at up to six points during the evolution of Archaeplastida (Petrželková and Eliáš, 2014). Therefore, it was surprising to find three paralogs in kinetoplastids showing a complex pattern of descent (See **Figure 3.14** and Figure 3.7 above). The canonical Rab21 was present across the kinetoplastids except in the phytomonads, where it was not identified. This protein was also present in *N.gruberi* and *E. gracilis*. The second Rab21 cluster, which is closely related, was named Rab21B, and this protein was present in just the bodonids and the cruzi group and lost in the other trypanosomatids. Rab21C also shows a similar pattern but appears less closely related to the canonical Rab21. Further two of the ‘orphan’ Rabs found in *B. saltans* and *T. borreli*, which were closely related to each other, also clustered with the Rab21 group. Interestingly, this pair (putative Rab21D) clustered with another *E. gracilis* sequence with good support indicating that this may be an expansion dating back further than the root of the kinetoplastids. If this is the case, there were possibly up to two points of Rab21 expansion: Rab21B and C at the base of the kinetoplastids, and Rab21D in the common ancestor of kinetoplasts and euglenids. Its repertoire in kinetoplastids is mainly a result of loss. In *T. brucei*, Rab21 mediates intermediate endocytic steps. Here it acts downstream of Rab5A and in intimate connection with the trypanosome ESCRT (endosomal sorting complex required for transport) system to transport cargo from endosomes to the lysosomes. In doing this, it adds complexity to the degradative arm of the trypanosome endocytic system in which Rab4, 5A, 5B, 7, 21 and 28 are implicated (Ali *et al.*, 2014). The presence of up to four copies of this protein in the

bodonids and three in the cruzi group, just one in *Leishmania spp.* and the brucei group, and none in the phytomonads, possibly indicating a more elaborate and complex endocytic degradative system in the common ancestor of kinetoplastids, which has been sculpted to differential extents in the separate trypanosomatid lineages.

Rab5 and 21 were the only Rabs of the primordial endocytic group to be retained across kinetoplastids. A putative Rab24, which has a key role in autophagosome maturation, was found to be restricted to the bodonids. The bodonid Rab24 was so named because it was not possible to phylogenetically resolve whether it belonged to the Rab20 or 24 cluster, but it has been postulated that Rab20 may be a metazoan-specific paralog of Rab24 (Klöpper *et al.*, 2012). The Rab5-related Rab22 was found in *N. gruberi* and *E. gracilis* and so is a kinetoplastid lineage-specific loss. Rab50, on the other hand, was present in *N. gruberi* but absent from *E. gracilis* and all the kinetoplastids. This either indicates an early loss in Euglenozoa or that Rab50 may never have been present in the lineage, if the root of the eukaryotes, whose position is as yet unresolved, lies between Euglenozoa (a phylum that includes kinetoplastids) and the rest of eukaryotes as has been postulated in (Cavalier-Smith, 2010, 2013; Akiyoshi and Gull, 2014).

The putative RabGAP for Rab24 has not yet been found. A member of TBC-B was found to act as a GAP for Rab21, but this has not yet been well characterised; it will be discussed in the next section as a confirmed GAP for Rab7. TBC-Q members, which act as GAP for Rab6 (yeast Gyp6, mammalian TBC1D11, also act as a GAP of Rab1 and was discussed in the Section 3.4.2 above. In humans, RN-tre, a member of the TBC-O subtype acts as a GAP for Rab5 (Lanzetti *et al.*, 2004; Haas *et al.*, 2005). However, TBC-O is restricted to Opisthokonts so there must be another RabGAP that controls Rab5 in the remaining supergroups. This has not been identified so far.

#### 3.4.4 Lysosome and flagella related Rabs and their GAPs

In the phylogenetic analysis of eukaryotic Rabs, Rabs 7, 23 and 32 formed a likely monophyletic superclade. These proteins are usually associated with later endosomal pathways and organelles such as the lysosome.

Rab7 mediates late endosomal transport to the lysosome and is also reported to be involved in autophagosome maturation (Hyttinen *et al.*, 2013). It is one of the most

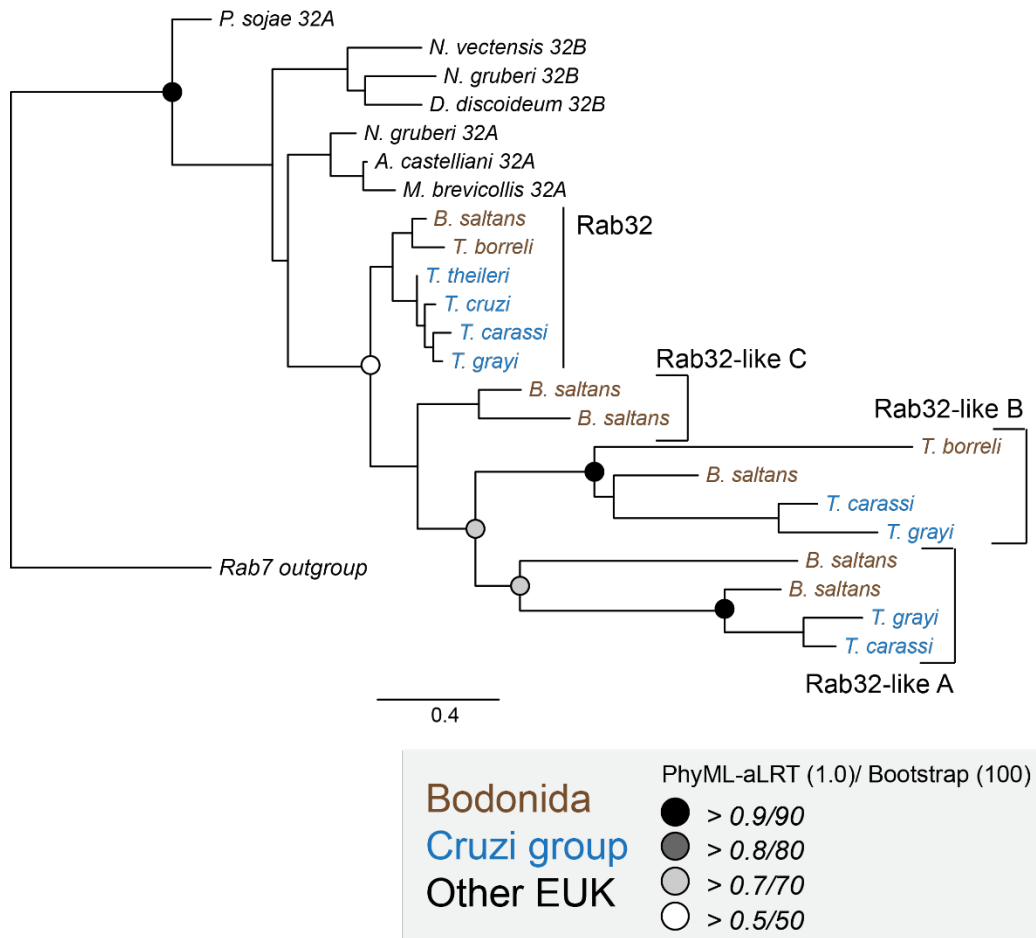
prolific Rabs in that it was found to be expanded in several lineages – in angiosperms and metazoa, but also in fungi (Pereira-Leal, 2008; Klöpper *et al.*, 2012). Again, there is evidence that each expanded set shows different localisations and functions. For example, in *E. histolytica*, while the Rab7A isoform is involved in fusion of cargo carrying post-Golgi compartments to late endosomes, Rab7B takes part in the fusion of late endosomes with lysosomes (Saito-Nakano *et al.*, 2007). In kinetoplastids, Rab7 was expanded in the bodonids, but represented by a single paralog in trypanosomatids. Unlike previously described Rabs (such as Rabs 11, 21), it appears that the duplication has occurred after the divergence of the trypanosomatid lineage from the basal bodonids. One pair of proteins from *B. saltans* and *T. borreli* form a closely associated pair attached to the main Rab7 cluster (Rab7B), and another single *B. saltans* protein forms an outer branch (BsRab7C). Another pair of *B. saltans* and *T. borreli* proteins, which were initially classified as ‘orphans’, were also closely associated with the Rab7 cluster (Rab7-like) in the pan-eukaryotic analyses with modest support (MrBayes 70%, RaxML bootstrap 48%, PhyML aLRT 0.853). This was not replicated in the pan-kinetoplastid analysis where this pair was not reliably associated with any cluster and RabDB (Diekmann *et al.*, 2011) classifies the sequences as RabX. This pair has a closely clustering *E. gracilis* sequence, indicating that the putative Rab7-like protein originated from before the kinetoplastid divergence and was lost from the trypanosomatids (possibly in a manner similar to Rab21D).

Rab32, which is involved in the biogenesis of lysosome-related organelles (LROs), is found at the ER-mitochondria intersection and is required for autophagosome formation in mammalian cells (Hirota and Tanaka, 2009). It was recently shown to be essential in the biogenesis and maintenance of acidocalcisomes (which resemble LROs) in *T. cruzi*, particularly in the exchange of material with the parasite’s contractile vacuole complex (Niyogi *et al.*, 2015). Rab32 presented a complex pattern of evolution in kinetoplastids (see **Figure 3.18** below). The canonical Rab32 was present in the bodonids and cruzi group only, which possess one copy each. It is possible that the loss in other trypanosomatids had to do with concurrent loss or decrease in the importance of the contractile vacuole system. There appears to have been further duplication and divergence giving rise to Rab32-like proteins, named A, B and C. Rab32-like A is found in the *B. saltans* (2 off), *T. grayi* and *T.*



*carassi*, while Rab32-like B is found singly in *B. saltans*, *T. borreli*, *T. grayi* and *T. carassi* and Rab32-like C is found only in *B. saltans* (2 off). Therefore, after the initial expansion and divergence, these proteins appear to have been lost on at least two occasions on the path to parasitism in *T. borreli*, the *Leishmania* plus *Phytomonas* super clade, the brucei group trypanosomes as well as two members of the cruzi group. This phylogeny could also be explained partially by lineage-specific duplications in *B. saltans* such as the emergence of two copies of Rab32-like C. The data indicate reduced or alternative pathways for autophagosome biogenesis, as is likely the case in several fungal, archeplastid and apicomplexan lineages lacking Rab32. In contrast, the emergence of Rab32-like proteins likely indicates development of novel or more complex LROs, particularly in *B. saltans*.

Rab23 is known to be involved in the formation of cilia/flagella and has been lost in organisms concomitantly with these structures. In *T. brucei*, it has been localised to the flagella, although its depletion does not result in any observable defect (Lumb and Field, 2011). It could function redundantly with other flagellar biogenesis Rab-like proteins such as IFT27 and RTW. In *T. brucei*, IFT27 has also been shown to be involved in both anterograde and retrograde flagellar transport (Huet *et al.*, 2014). Rab23 is retained singly across all kinetoplastids as is IFT27, and RTW was present in all but was not identified in *T. vivax*.



**Figure 3.18 Evolutionary history of kinetoplastid Rab32**

Best PhyML topology is presented. Statistical support from PhyML aLRT and bootstrap analyses are presented as circles at key nodes, filled in in grayscale according to the key shown (bottom). Taxa are coloured according to the key. EUK, eukaryotes.

Rab28 forms a sister clade to IFT27 and although it is classed as a ‘true Rab’ and not Rab-like protein, it is a divergent member of the family with low sequence homology to canonical Rabs (Pereira-Leal and Seabra, 2001). Indeed, the highly conserved G-protein binding sequence ‘WDTAGQ’ is changed to ‘WDIGGQ’ across all eukaryotes including kinetoplastids, indicating that the Rab28 sequence diverged early from the norm, possibly in LECA. It has been associated with flagellar transport due to coincident losses with flagella/cilia in eukaryotes (Klöpper *et al.*, 2012). However, experimental studies show no evidence for this. In rat endothelial cells, it is shown to have nuclear/cytoplasmic localisation in culture conditions, and putatively interact with NF- $\kappa$ B, perhaps aiding in transporting it to the nucleus (Jiang *et al.*, 2013). In *T. brucei* on the other hand, it appears to be endosomal and interacts with

the retormer and ESCRT complex and is required for turnover of invariant surface glycoproteins (Lumb *et al.*, 2011). Therefore, it appears that its function may not be conserved across eukaryotes even though Rab28s from distinct lineages such as humans and trypanosomes maintain up to 80% sequence similarity in the switch and interswitch regions which Rabs use to interact with effectors (Eathiraj *et al.*, 2005; Delprato and Lambright, 2007).

Rab34 and RabTitan are ancient Rabs that show a very patchy distribution across eukaryotes. The function of RabTitan is unknown, and its position as part of the Rab family has been questioned, and specifically citing the lack of C-terminal cysteines (Klöpffer *et al.*, 2012). Rab34 is reported to mediate a Rab7-independent pathway to phagolysosome biogenesis through size selective transfer of endo/lysosomal cargo into lysosomes in mammalian cells (Kasmapour *et al.*, 2012). Both these proteins were not found across kinetoplastids indicating an early loss, or absence from the lineage, if the eukaryotic root lies between Euglenozoa and the rest of eukaryotes.

Of the TBCs that affect these Rabs is the lysosomal TBC-B, various mammalian isoforms of which affect Rab7, 11, 21 and 2. TBC-B was found to be duplicated several times in opisthokont, archeplastid and amoebozoan lineages, but was absent from SAR (Gabernet-Castello *et al.*, 2013). There is conservation between yeast TBC-B Gyp7 and the mammalian homologue TBC1D15, both of which were found to act as a GAP for Rab7 (Peralta *et al.*, 2010; Zick and Wickner, 2012). In kinetoplastids, this protein was singly retained in all kinetoplastids but absent in the phytomonads that do not show a concomitant loss of Rab7, but have lost Rab21. It is possible that TBC-B may act on Rab21, with some other GAP controlling Rab7, but this needs to be experimentally tested.

TBC-N, whose target Rab is as yet unknown but also plays a role in autophagosome formation (Longatti and Tooze, 2012; Longatti *et al.*, 2012) shows a distribution similar to that of the classical Rab32 indicating that they may act in the same pathway. However, mammalian TBC-N, TBC1D14 was shown to bind directly to Rab11 to promote autophagosome formation but did not act as a Rab11-GAP (Longatti *et al.*, 2012).

### 3.4.5 GAPs of unknown associations

Other TBCs also showed lineage-specific duplications or losses but there is no obvious pattern, which suggests that the TBC family is subject to expansion and contraction through distinct selection mechanisms in different lineages.

TBC-F was found patchily distributed across eukaryotes, but was present across all taxa in kinetoplastids. It appears to have been separately duplicated in the bodonids and well as in *T. theileri*. The mammalian member of TBC-F, TBC1D5, regulates Rab7-dependent membrane recruitment of cargo-selective retromer subcomplex (Seaman *et al.*, 2009) and also Atg9 trafficking and initiation of autophagy. It mediates its interaction with Vps29 of the retromer complex and autophagosome via its two LC3 interacting regions (LIRs) (Popovic and Dikic, 2014). LIRs form the region that all known autophagy receptor and adaptor proteins bind to ATG8 modifiers. Although Vps29, Rab7 and Atg9 are found in kinetoplastids, the TbTBC-F lacks the LIRs and is thus unlikely to be involved in the manner described in mammals and its function needs to be experimentally determined.

TBC-I was also found patchily distributed across eukaryotes and found duplicated or expanded in half the organisms in which they occur (5 of 11 in the analysis of (Gabernet-Castello *et al.*, 2013)). In kinetoplastids, it is lost in two of three phytomonads, *T. borreli*, and *L. donovani*, and duplicated in the rest of the *Leishmania spp.* Not much is known about TBCs G, H, I, and L in terms of their Rab partners, although there is some information about functional relevance such as the inhibition of innate immunity signalling by mammalian TBC-I member TBC1D23 (De Arras *et al.*, 2012). The main pattern of these proteins is their absence from two *Phytomonas* species *P. serpens* and *P. HART*, perhaps reflective of their reduced genome size and smaller Rab complement.

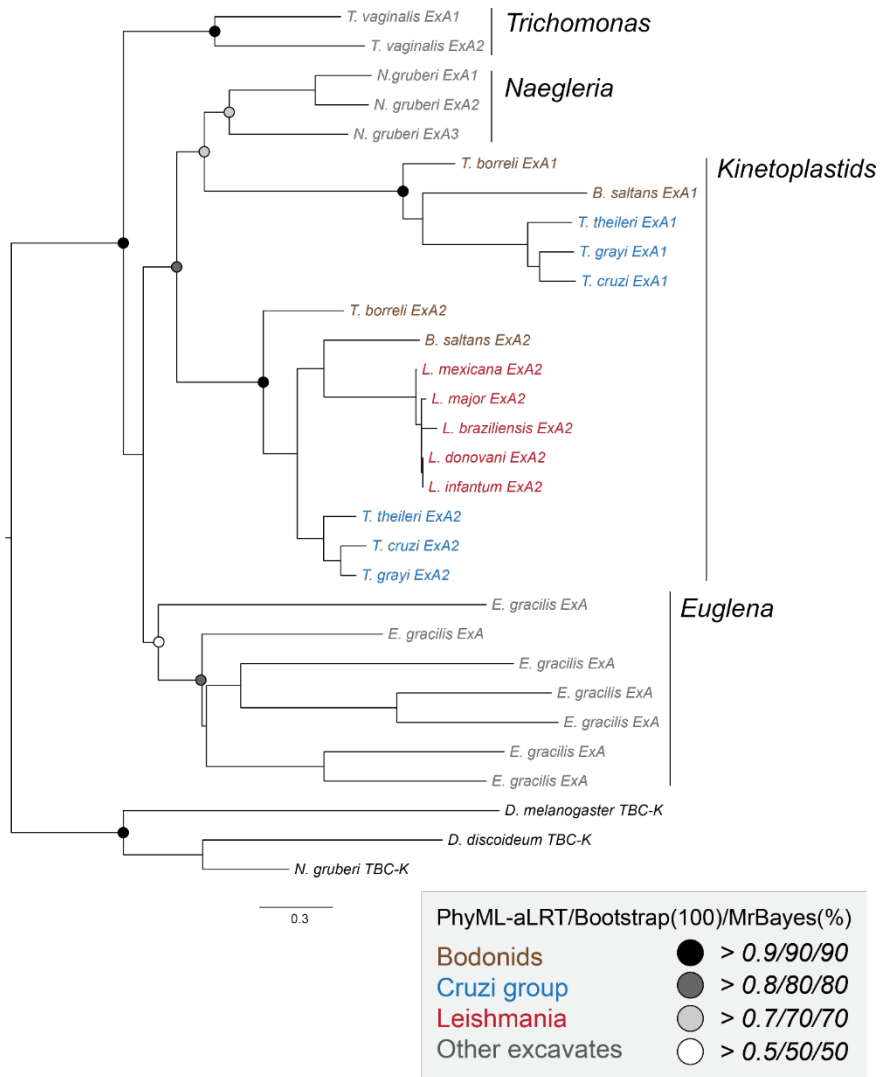
### 3.4.6 Orphans and unassigned lineage-specific Rabs and RabGAPs

Six pairs of ‘orphan’ Rabs were found in *B. saltans* and *T. borreli*. While one pair was assigned as the ancestral Rab24 that appears to be lost in all trypanosomatids, two others were eventually assigned as paralogs of Rab1/18, one pair as putative Rab7-like, and finally, another pair as putative Rab21D (see **Table 3.1** in Section 3.3.1 above). One pair remains unassigned. With the general trend towards loss of Rab paralogues in trypanosomatids, it is unclear if these orphan Rabs came

about in the bodonid lineage after radiation of trypanosomatids, or if they were lost in the latter, as seen in the case of Rab24. A single *T. brucei* sequence was also assigned as Rab1-like as it consistently grouped with the Rab1 clade. The caveat is that it had a relatively longer branch relative to the rest of the cluster (see **Figure 3.5**).

Other lineage-specific Rabs were better represented across kinetoplastids. Most of these were assigned as paralogs of Rab1/18 (KSRX1, UxRX3), Rab11, and Rab21. Only for one cluster, KSRX4, whose representatives Rabifier classifies as RabX, it was not possible to assign a close relationship to a core Rab paralog with confidence, most likely because the sequence is highly divergent. As seen in eukaryotic Rab32 and kinetoplastid UzRabX3, the usually highly conserved ‘WDTAGQ’ sequence is modified to ‘WDTAGL’ in KSRX4 sequences. Interestingly, this Q → L point mutation results in a dominant active form of Rab with reduced GTPase activity and high affinity to GTP. It has been used to study the function of Rabs in many eukaryotes including *T. brucei* Rabs 5 and 11 (Macara and Brondyk, 1995; Pal *et al.*, 2003). Thus, KSRX4 Rabs may be endogenously dominant active forms and their function needs to be elucidated. Furthermore, there appears to have been a *T. brucei* specific duplication event for this gene as only *T. b. brucei* and *T. b. gambiense* have two copies which branch together within the brucei group cluster, while the remaining species possess only a single paralog (see **Figure 3.5**).

The presence of up to three Excavate-specific TBCs named ExA were postulated (Gabernet-Castello *et al.* 2013). Examination with a wider selection of genomes within the supergroup indicated that these genes were unlikely to be monophyletic but probably arose at different points along different lineages within excavates (see **Figure 3.19** below). The expansions of these genes as seen in *T. vaginalis*, *N. gruberi* and *E. gracilis* were species-specific as their ‘ExA’ sequences clustered according to species. Expansion in kinetoplastids appears to be due to a duplication at the base of the lineage (rather than in individual species) which was asymmetrically retained in various taxa.



**Figure 3.19 Phylogenetic analysis of TBC-ExA subfamily**

Best PhyML topology rooted on TBC-K sequences is presented. Statistical support from PhyML aLRT, PhyML bootstrap, and MrBayes analyses are presented as circles at key nodes, filled in in grayscale according to the key shown (bottom right), taxa are coloured according to the key.

### 3.5 Conclusions

The analysis presented here indicates a highly dynamic evolutionary history of Rabs within the kinetoplastids. Seven novel Rab clades restricted to the Kinetoplastida/Euglenozoa lineage were found. All but one of these were traced back as paralogs of known LECA Rabs namely Rab1/18 (KSRabX1, UzRabX3), 11(KS RabX3), 21 (Rab21B and C) and 32 (Rab32-like). Only KSRX4 and one pair of the ‘orphans’ of *B. saltans* and *T. borreli* are much too diverged to be robustly placed. Based on previous observation in kinetoplastids (for example *T. brucei* Rab5A vs 5B), and other eukaryotes (for example *H. sapiens* Rab11a vs 11b) it can be predicted that these duplicated Rabs likely perform novel functions. Compared to the 90.8% sequence identity between human Rab11a and 11b, *L. major* Rab11A and 11B have only 41% and the corresponding *T. cruzi* pair has 31% identity (according to pairwise sequence alignment by EMBOSS Matcher). It is therefore likely that these duplications are evolutionarily less recent and/or that neo-functionalisation extremely likely. It would be very interesting to check for neo-functionalisation of these Rabs, and differences if any between trypanosomatid species that adopt different lifestyles. However, while the Rabs in kinetoplastids comprise a number of lineage-specific innovations, for the most part the canonical LECA Rabs remain present and identifiable by phylogeny, BLAST, and Rabifier. While the extent of conservation of putative TBC LECA subtypes in kinetoplastids is far greater than many extant multicellular organisms, there is also limited lineage-specific innovation, suggesting a complement that is rather similar to that predicted for the LECA.

Rabs of the early secretory pathway were particularly well conserved. There were no major expansions within genes associated with this part of the endomembrane system, with the exception of Rab1/18, of which there were excavate-specific duplications (namely KSRX1 and UzRX3) which appeared sufficiently diverged to have undergone neo-functionalization (see **Figure 3.16**) to a role apparently distinct from membrane transport (Natesan *et al.*, 2009). The endocytic and retrograde pathways represented the bulk of the lineage-specific losses and gains in kinetoplastids (see **Figure 3.14** and **Figure 3.15**). In particular, bodonids and the cruzi group retained an expanded set of Rab11, 21 and 32 which were absent from the other trypanosomatids.

There is no similar asymmetrical pattern seen among the TBC proteins; members of TBC-D and Q (act as GAPs for several post-Golgi Rabs such as 4, 7, 11 and 14) were expanded across all kinetoplastids. TBC-N, which affects Rab11 and plays a role in autophagy showed restricted distribution having been lost in all lineages but the bodonids and cruzi group. Canonical Rab32 showed a similar distribution as TBC-N but no functional association between the two have been reported so far. Whether TBC-N interacts with Rab11 duplicate Rab11B, which shows the same distribution as TBC-N, is a matter of speculation. It appears that the expanded TBC cohort present at the base of the kinetoplastids was more broadly retained than the Rabs. Exceptions include extensive losses in the phytomonad lineage, and to a lesser extent in the brucei group. Other losses and gains appeared largely sporadic and taxon-specific and showed no correlation to Rab losses.

Overall, the dominant theme, particularly in the trypanosomatids, was gene loss. This loss becomes even more obvious perhaps because of the much larger repertoire and genome sizes in the organisms from which the lineage likely diverged – *N. gruberi*, *E. gracilis* and *B. saltans* have a relatively expanded set of these trafficking proteins. The brucei group trypanosomes in particular were notable for losses, despite their larger genome size compared to *Leishmania spp.* and which suggests a streamlining of the trafficking pathways in these organisms.



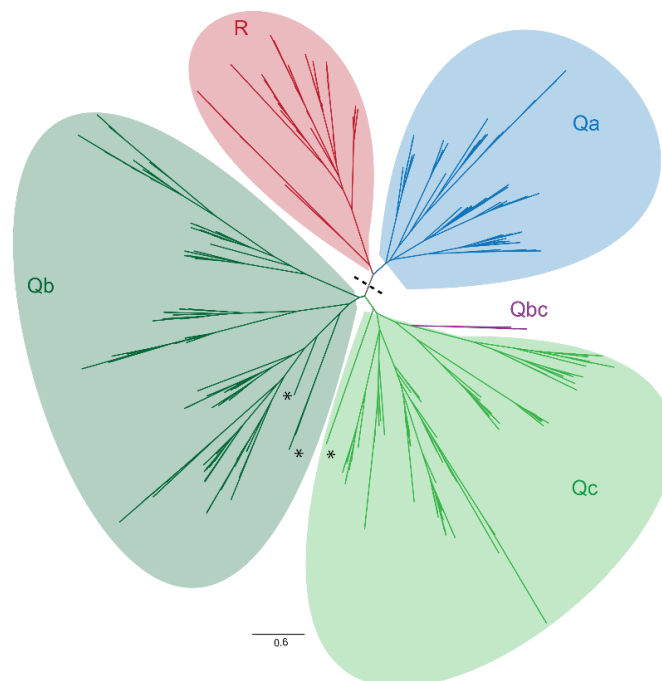
# 4 EVOLUTION OF KINETOPLASTID SNAREs

## 4.1 Introduction

SNAREs mediate the final step of membrane fusion in cellular trafficking. While they have been extensively studied in mammalian and yeast model systems, as well as in land plants, knowledge of their evolutionary history and functional roles outside of these models is once more, limited. There has been only one major pan-eukaryotic study of the SNAREs (Kloepper *et al.*, 2007). The authors provided a useful classification system of SNAREs into 20 sub-types that could be tentatively assigned to specific localisation in the cell. However, their dataset, by limitations on availability of sequence data, is skewed toward the animals, plants, and fungi. Thus, the total complexity of eukaryotes was not represented. There were also studies that reconstructed the evolutionary history of SNAREs within Metazoa and Embryophyta as well as fungi and these revealed the broadly conserved core set of SNAREs along with numerous lineage-specific innovations and expansions (Sanderfoot, 2007; Kloepper *et al.*, 2008; Kienle *et al.*, 2009a). In this study, newly sequenced genomes were used to reconstruct the evolutionary history of the SNAREs in a deeply divergent class of organisms, Kinetoplastida. It provides a counterpoint to compare the modes of evolution in vastly differing eukaryotic supergroups; it also assesses the variation of the SNARE repertoire, among the different lineages within Kinetoplastida.

## 4.2 The phylogeny of kinetoplastid SNAREs and assignment of sub-types

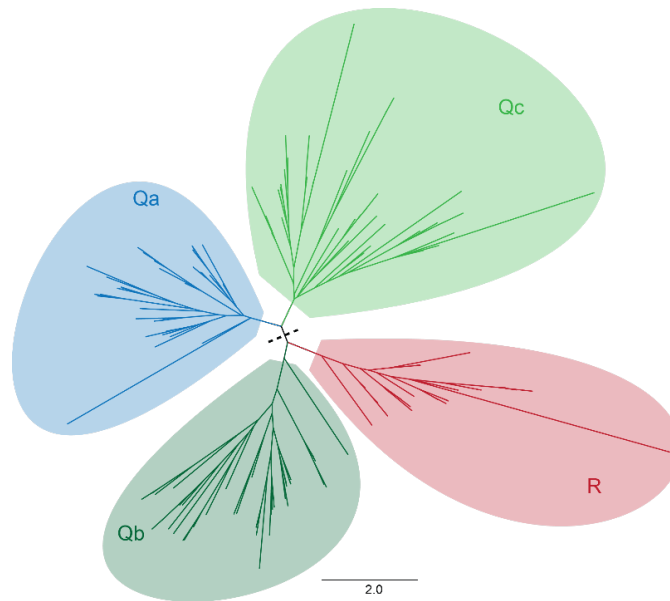
The non-redundant dataset of kinetoplastid SNAREs comprised 518 sequences from 18 genomes. Phylogenetic analysis indicated an overall division into the four classically defined subtypes: Qa, Qb, Qc, and R with reasonable support from PhyML aLRT (>0.7 for each subtype cluster) but not with bootstrapping or MrBayes (See unrooted tree presented in **Figure 4.1** below). At this point Qbc SNAREs were not distinctly discernible, perhaps because there were only three sequences of its kind. They formed a clade within the Qc group of the tree, indicated in purple. Some sequences from *B. saltans* and *T. borreli*, which were predicted to be SNAREs according to the SNARE-DB motif scan, remained outliers, are marked with asterisks. Like many of the outlier ‘orphan’ Rabs, these sequences were also tentatively linked to existing clusters as will be discussed below.



**Figure 4.1 Topology of kinetoplastid SNARE family**

Best unrooted PhyML topology of all putative kinetoplastid SNAREs is shown. Four clusters are identified Qa (blue), Qb (dark green), Qc (light green), and R (red). Putative Qbc SNAREs are shown in purple. Asterisks mark outlier sequences, short dotted line marks division of subfamilies into Qa+R and Qb+Qc groups.

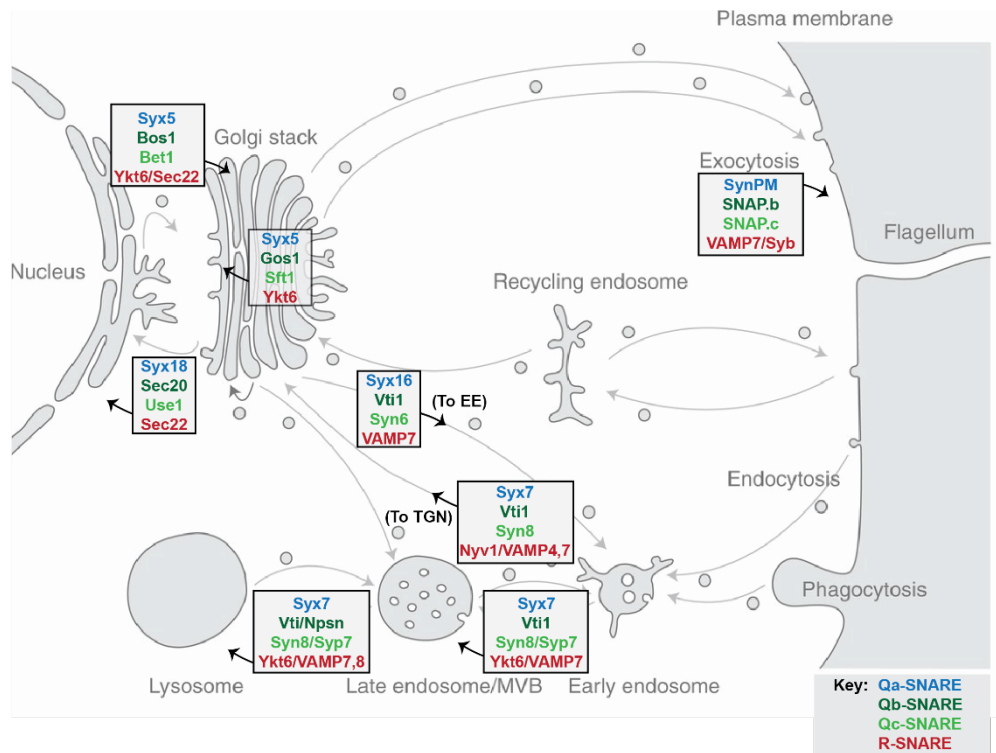
Kloepper *et al.* described the partitioning of the four basic subgroups into two elementary groups with one group containing R + Qb SNAREs and another containing Qa + Qc. This agrees with the knowledge of domain structures of SNAREs where both Qa and Qc possess N-terminal ‘syntaxin’ domains, while R-SNAREs possess longin domains and Qb SNAREs generally lack any structured N-terminal motifs. In the kinetoplastid tree, the divisions between families was however, distinct: Qa and R SNAREs formed one group (PhyML aLRT 0.93) while Qb and Qc SNAREs formed another group (PhyML aLRT 0.75). This topology changed when trees of representative kinetoplastids along with human and yeast SNAREs taken together were considered (see **Figure 4.2** below). Here, the previously described Qb + R (PhyML aLRT 0.72) and Qa + Qc (PhyML aLRT 0.71) pattern emerged. It is possible that the pattern in the kinetoplastid tree was an artefact – SNAREs are relatively small proteins, which does not allow for many residues of relevance to phylogenetic analysis, especially when a large dataset of 518 sequences are used. While individual clusters were very well supported in the large kinetoplastid tree, smaller datasets of less heterologous sequences were generally more consistent in terms of relationships between clusters. This is perhaps the reason for the low support values observed in the deeper nodes of the tree. Therefore, a combination of phylogenetics (analysis with yeast and human sequences) and online bioinformatics tools for sequence analysis was used to test representative sequences from each cluster in order to assign them to sub-groups (Qa, Qb, Qc or R) which could then be analysed separately to assign specific orthology to known proteins if any. Online tools included the SNARE-DB motif recognition tool (<http://bioinformatics.mpibpc.mpg.de/>), NCBI-conserved domain database search (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and Interpro domain search (<http://www.ebi.ac.uk/interpro/>). Thus, of the 518 sequences, 134 were assigned as Qa-SNAREs, 142 as Qb, 140 as Qc, and 102 as R-SNAREs.



**Figure 4.2 Topology of human, yeast and trypanosome SNARE families**

Best unrooted PhyML topology of *Trypanosoma brucei*, *Homo sapiens* and *Saccharomyces cerevisiae* SNAREs is shown. Four clusters are identified Qa (blue), Qb (dark green), Qc (light green), and R (red). Short dotted line marks division of subfamilies into Qa+Qc and Qb+R groups.

SNARE-mediated fusion in the cell involves the formation of cognate SNARE complexes where the SNARE domains of Qa, Qb, and Qc (or Qbc) SNAREs on the target membrane interact with the SNARE domain of an R-SNARE on the vesicle to enable fusion. The specificity of this reaction is maintained both by preference for interaction with cognate SNAREs and correct spatial segregation in the cell (Bethani *et al.*, 2007). A generalised overview of the specific SNARE complexes involved in the different steps of trafficking within the cell is presented in **Figure 4.3** below to provide context to the upcoming analysis. This overview was mainly based on knowledge of yeast and human SNARE complexes, and simplified for clarity. Components of complexes mediating fusion of vesicles from endoplasmic reticulum (ER) to Golgi and *vice versa*, from early endosomes (EE) to the trans-Golgi network (TGN) and *vice versa*, transport to the late endosomes and lysosomes, as well as exocytosis at the cell surface, are shown. For each sub-group, the phylogeny of all the kinetoplastid sequences is presented in parallel with the phylogeny of the representative kinetoplastids with select representative eukaryotic sub-group SNAREs, which was used to assign the clusters to known SNAREs.



**Figure 4.3 Assignment of SNAREs to intracellular pathways**

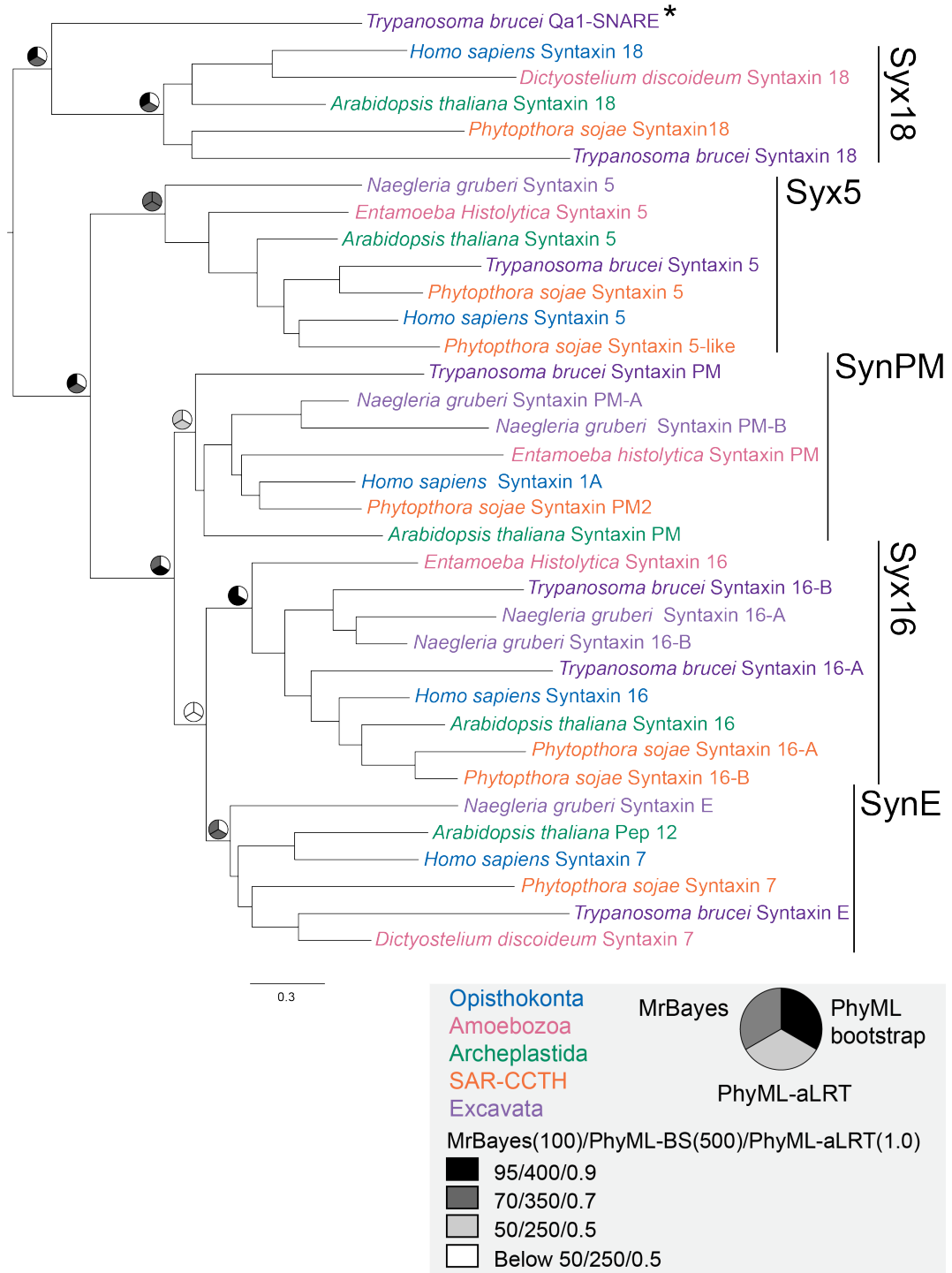
Components of SNARE complexes acting at various steps of membrane trafficking pathways of a generalised cell are presented in grey boxes, and coloured according to sub-family as shown in the key. Modelled on figure in (Jahn and Scheller, 2006).

#### 4.2.1 Qa SNAREs

The kinetoplastid Qa SNAREs separated into eight clusters: Qa1 (unassigned), SynE (=Syx7), Syx5, SynPM1, SynPM2, Syx16A, Syx16B and Syx18. **Figure 4.4** below shows representatives from each cluster analysed with Qa SNAREs of representative eukaryotic species. The three-variable pie-chart icons at nodes indicate the statistical support for the node in three analyses: PhyML aLRT (1.0), PhyML bootstrap (500) and MrBayes posterior probabilities (%). The strength of the statistical support is shown in gradation of colour from white to black as described in the key. In general, the support from PhyML bootstrap analysis was rather poor, only rising to > 350 in the case of Syntaxin5 and otherwise remaining below 250 i.e., 50% support. The Syntaxin 5 assignment was also well supported by the other two analyses. Syntaxin18 and 16 had >95% support from the Mr Bayes analysis, and over 0.7 and 0.9 respectively from PhyML aLRT, allowing them to be assigned with a higher degree of confidence. SyntaxinE and PM had lower EE support, but were consistently retrieved by

all three analyses. The lower support for SynE and SynPM suggests that they are perhaps more divergent in the kinetoplastids than Syx5, 18 and 16.

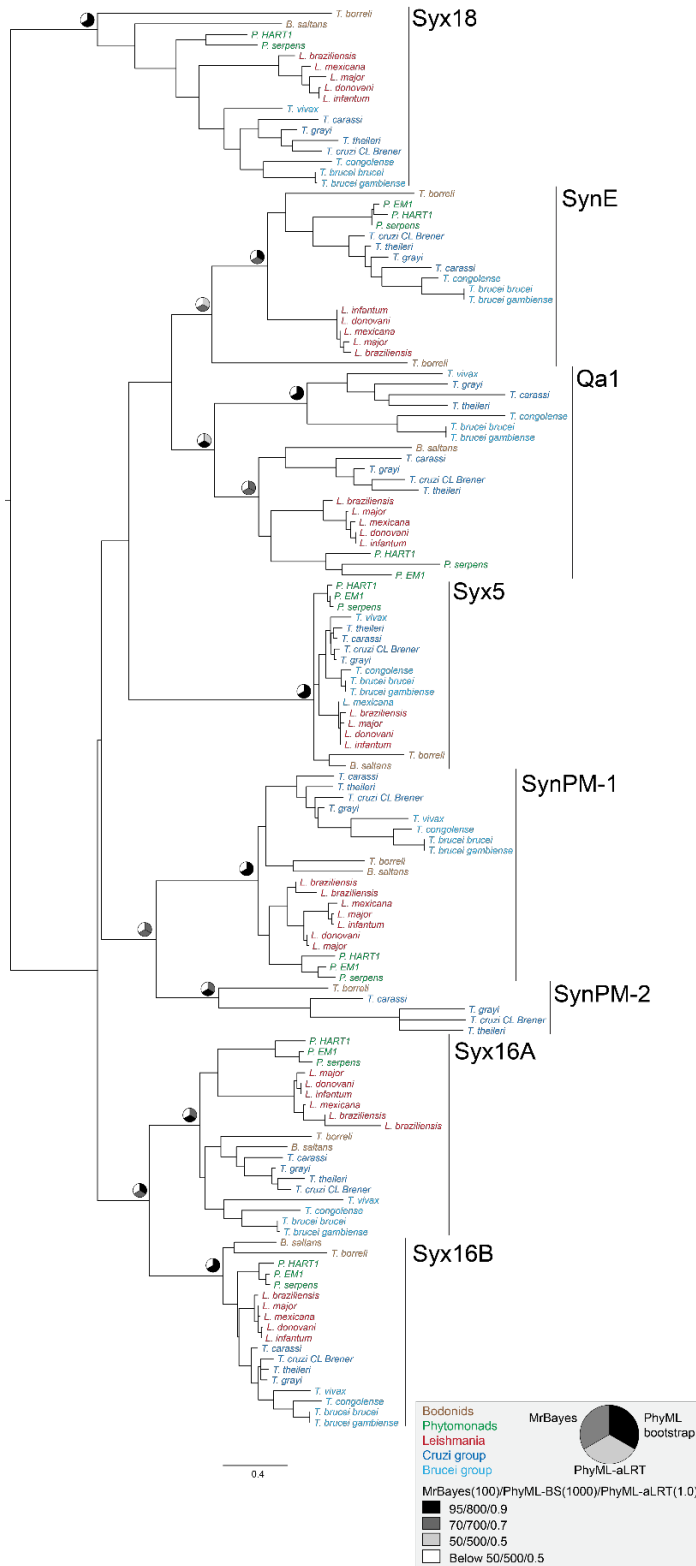
The SNARE designated as Qa1 remained unassigned – it was found to cluster with Syntaxin 18 in **Figure 4.4** below, but this clustering was not replicable in all trees. The pan-kinetoplastid analyses (**Figure 4.5** below) indicated a possible relationship between Qa1 and SynE as Qa1 consistently formed a sister cluster to SynE but it is likely that Qa1 is too divergent to reliably cluster with a pan-eukaryotic set of SNAREs. If that is indeed the case, a distinction is observed between Qa SNAREs of putative ER-Golgi function, which remain singular (Syx5 and Syx18) and post-Golgi Qa SNAREs all of which appear to have duplicated (SynE, Syx16, and SynPM). This expansion of post-Golgi Qa SNAREs was modest compared to the expansion seen in multicellular animals and plants. Animals, for instance, possess Syntaxins1 through 4, 7 (SynE equivalent), 13, 16, 17 and 20. However, analysis was in line with the general observation that the early secretory pathway proteins are more conserved than those involved in post-Golgi trafficking. Syntaxins16A and 16B were found throughout the kinetoplastids, indicating an origin early in the lineage before radiation of extant species. SynPM also had two clusters, one that was present across kinetoplastids (SynPM1) and another that was restricted to basal bodonid *T. borreli* and the cruzi group trypanosomes (SynPM2), again indicating an early emergence, but followed by losses in the many trypanosomatid lineages (and *B. saltans*).



**Figure 4.4 Assignment of kinetoplastid Qa-SNAREs**

Best PhyML topology is presented. Node values are iconised as circles for two support values each representing PhyML aLRT(1.0), PhyML bootstrap (500), and MrBayes posterior probabilities (%) and colour-coded in grayscale as shown in the key. One representative kinetoplastid Qa-SNARE from each sub-type cluster (dark purple), is presented along with eukaryotic representatives coloured according to supergroup as shown in the key. Each cluster is marked with a vertical line and named, asterisk indicates unassigned kinetoplastid QaSNARE.

## Chapter 4: Evolution of Kinetoplastid SNAREs



**Figure 4.5 Phylogeny of kinetoplastid Qa SNAREs**

Best PhyML topology is presented. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 1000) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Taxa are coloured by phylogenetic relationships as shown in the key. Each cluster is marked with a vertical line and named according to the assignment of the representative of the said cluster according to Figure 4.4 above.



The conservation of sequence identity between duplicated Qa-SNARE genes was rather poor: at the protein sequence level there is only approx. 24% identity between *T. brucei* Syx16 A vs B; *T. cruzi* SynPM1 vs SynPM2; and *T. grayi* SynE vs Qa1 (see **Table 4.1** below).

Sequences	% identity
<i>T. cruzi</i> SynPM1 + 2	26.00
<i>H. sapiens</i> Syx1 + 2	65.00
<i>T. brucei</i> Syx16A + B	24.20
<i>P. sojae</i> Syx16A + B	40.00
<i>A. thaliana</i> Syp41 + 42	65.50
<i>T. grayi</i> SynE + Qa1	22.60

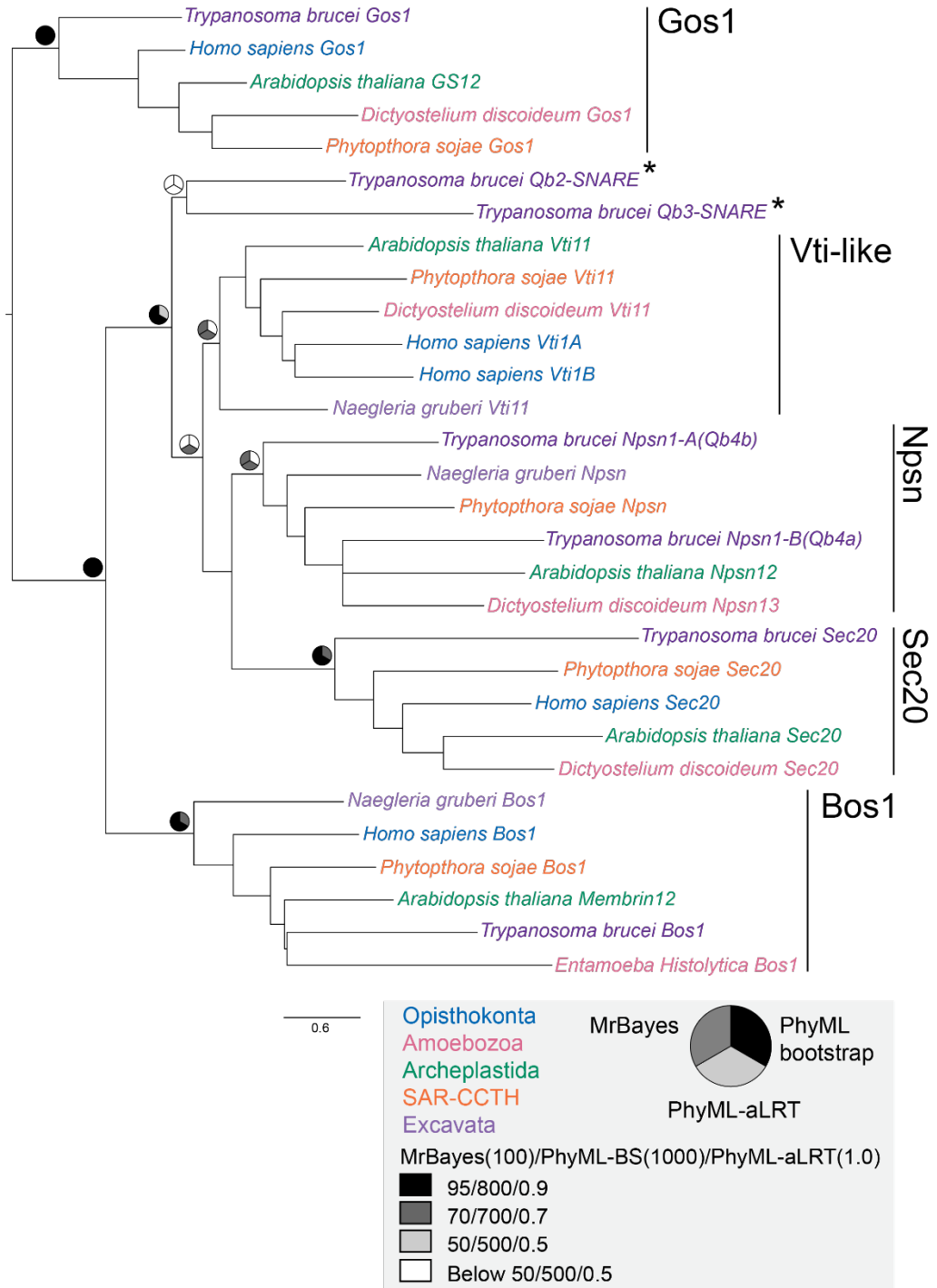
**Table 4.1 Similarity of duplicated Qa-SNARE sequences**

Percentage identity between paralogous pairs of Qa-SNARE sequences are presented. Identity was determined using EMBOSS matcher pairwise alignment (McWilliam *et al.*, 2013). *A. thaliana* Syp4 sequences are the equivalent of SynE in plants.

Therefore, it is interesting that while Syx16A and B, and SynPM1 and 2 were easily assigned to canonical eukaryotic SNAREs, Qa1 remained branching independently. The duplications of Syntaxin1(=SynPM) in animals (Syntaxin1-4) have developed multiple roles and have been correlated with the development of cell polarity in metazoans (Dacks and Doolittle, 2002). Considering that *H. sapiens* Syntaxin1 and 2 have different functions while being 65% identical (EMBOSS matcher pairwise alignment), it is likely that the duplicated Qa-SNARE genes with <30% identity will exhibit neo-functionalisation. Indeed, this study provides an indication of the different roles for *T. brucei* Syntaxins 16A and 16B, given that only Syx16B but not 16A interacted with the R-SNAREs VAMP7B and C. Given the conserved interaction between Syntaxin16 and VAMP7 in yeast and humans, it appears that Syntaxin16B is most probably the canonical gene while 16A may have assumed a new function and form a complex with a different set of SNAREs. Further functional studies will help to understand if this is the case.

### 4.2.2 Qb SNAREs

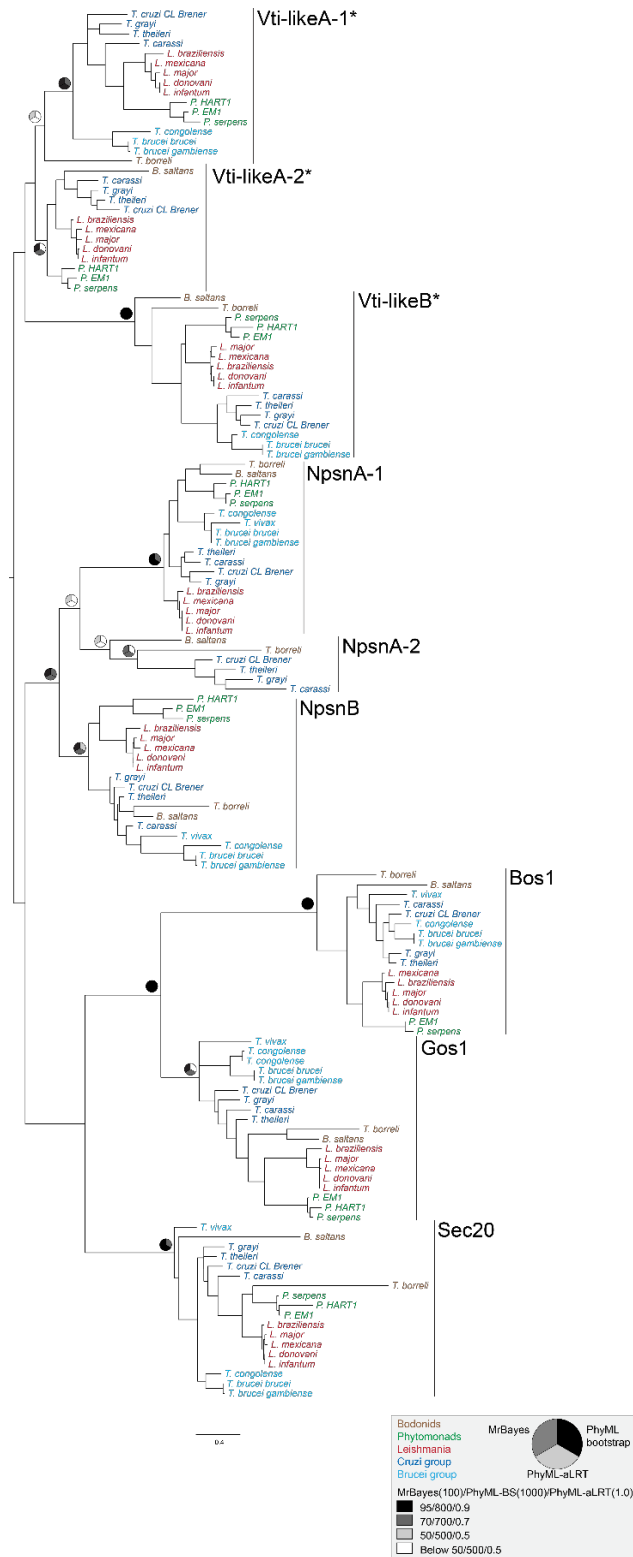
Of the nine clusters of Qb SNAREs, six were assigned as Bos1, Gos1, Sec20, Qb4A-1 and -2 (NpsnA), and Qb4b (NpsnB) by phylogenetic analysis. **Figure 4.6** below shows representatives from each cluster analysed with Qb SNAREs of representative eukaryotic species. Most clades were well supported by PhyML aLRT and MrBayes values. Three SNAREs called Qb2A, Qb2B, and Qb3, branched independently, and were tentatively assigned by BLAST. In pan-kinetoplastid Qb SNAREs analysis (**Figure 4.7** below), a majority of the nine clusters were well supported by PhyML aLRT and MrBayes analyses, but the bootstrap values were lower than those seen in kinetoplastid Qa SNARE analysis. Some bodonid sequences that remained as outliers in **Figure 4.1** remained unresolved in this analysis as well. This includes *T. borreli* Qb2 sequence, which branched on the Qb2A cluster with very low support. There appeared to be no corresponding *B. saltans* sequence in this cluster either but the core trypanosomatid sequences on their own were well supported (0.92/76/97, PhyML aLRT/PhyML bootstrap/ MrBayes). Representatives from both Qb2A and B returned Vti1 proteins when searched against both *S. cerevisiae* and *A. thaliana* and these were confirmed by reverse BLAST into *T. brucei* (Qb2A) and *T. cruzi* (Qb2B) respectively. They were therefore tentatively assigned as Vti-likeA1 and 2. Qb3 formed a well-defined cluster (0.99/100/100) whose origins nevertheless remained unresolved by phylogeny. It is an extremely short protein (123 amino acids) which on BLAST-searching identified Vti1B, Vti13 and Vti1p proteins in *H.sapiens*, *A. thaliana* and *S. cerevisiae* genomes respectively. However, the reverse BLAST invariably picked up one of the other Qb proteins. SNARE-DB classifies the SNARE motif in Qb3 as “Qb.III.d” which corresponded to Vti1. In addition, it clustered with the Qb2 SNAREs in all the analyses. Therefore, Qb3 was also assigned tentatively as Vti-like (Vti-likeB), albeit with less support than for the Qb2 proteins. Representative Vti sequences from all super-groups, including that of excavate *N. gruberi*, formed a reasonably well supported cluster (**Figure 4.6** below), so it appears that there was lineage-specific divergence of this protein in kinetoplastids.



**Figure 4.6 Assignment of kinetoplastid Qb-SNAREs**

Best PhyML topology is presented. Node values are iconised as circles for two support values each representing PhyML aLRT(1.0), PhyML bootstrap (1000), and MrBayes posterior probabilities (%) and colour-coded in grayscale as shown in the key. One representative kinetoplastid Qb-SNARE from each sub-type cluster (dark purple), is presented along with eukaryotic representatives coloured according to supergroup as shown in the key. Each cluster is marked with a vertical line and named, asterisk indicates

## Chapter 4: Evolution of Kinetoplastid SNAREs



**Figure 4.7 Kinetoplastid Qb-SNARE phylogeny**

Best PhyML topology is presented. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 1000) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Taxa are coloured by phylogenetic relationships as shown in the key. Each cluster is marked with a vertical line and named according to the assignment of the representative of the said cluster according to Figure 4.6 above. Names marked with an asterisk indicate clusters that were assigned by BLAST.

Bos1 and Gos1 proteins function in SNARE complexes that mediate anterograde transport from ER to Golgi and within Golgi trafficking. Sec20 mediates retrograde transport back from the Golgi to the ER. Kinetoplastid Qb-SNARE analyses (**Figure 4.7** above) showed all three to be singly conserved in kinetoplastids, as they are in most other eukaryotes. Even though only two types of post-Golgi trafficking Qb-SNAREs were found, these appeared to have undergone duplications to yield up to six subtypes of proteins Vti-like A1, A2, B, and Npsn A1, A2, B NpsnA2 with varying levels of support for each cluster (See **Figure 4.7** above). The level of sequence conservation between different forms of each subtype was low, e.g. NpsnA1, A2, and B sequences from *T. cruzi* and *B. saltans* were shown to be between 19 and 32% identical only, as presented in the percentage identity matrix in **Table 4.2** below. Such a considerable level of sequence divergence may indicate an early origin of these paralogues and a high likelihood of neo- or sub-functionalisation.

	TcNpsnA1	TcNpsnA2	TcNpsnB
TcNpsnA1	100	31.65	25.79
TcNpsnA2	31.65	100	22.83
TcNpsnB	25.79	22.83	100
	BsNpsnA1	BsNpsnA2	BsNpsnB
BsNpsnA1	100	19.65	22.94
BsNpsnA2	19.65	100	23.91
BsNpsnB	22.94	23.91	100

**Table 4.2 Comparison of sequence identity between Npsn paralogues**

Percentage identity between the three Npsn paralogues of *T. cruzi* (Tc) and *B. saltans* (Bs) are presented. Cells are shaded from light (low) to dark blue (high) to represent the level of identity.

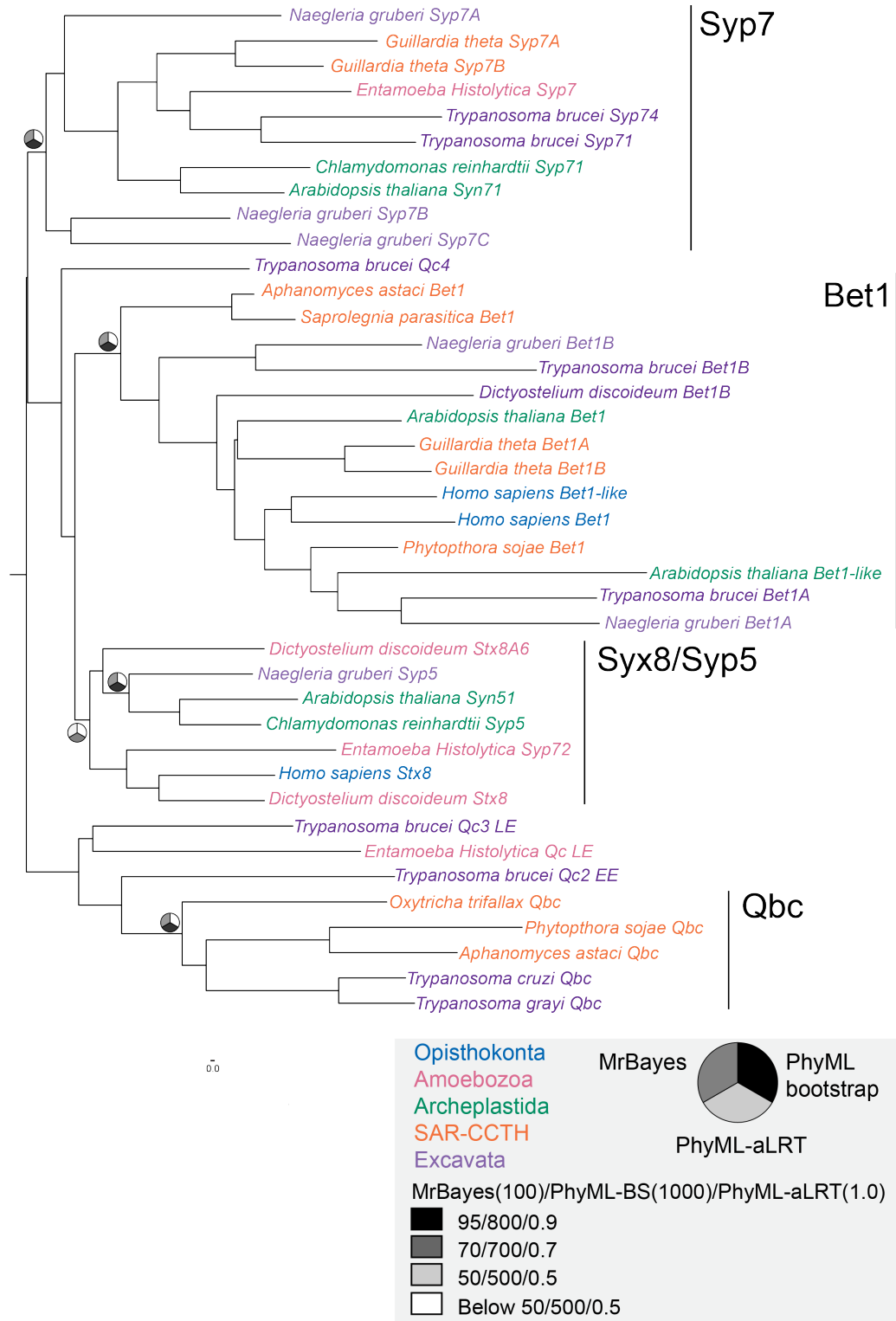
### 4.2.3 Qc and Qbc SNAREs

Eight major clusters found were assigned as Bet1A and B, Syp7A and B, Syx6-like 1 and 2, Syp5/Syx8-like and the Qbc-like proteins by phylogenetic analysis, or BLAST when phylogenetic support was low. **Figure 4.8** below shows representatives from each Qc-subtype analysed with Qc-SNAREs of representative eukaryotic species. In general, there was lower statistical support for Qc SNARE clusters than Qa and Qb-SNAREs, and only half of Qc-SNAREs were assigned by phylogeny because others tended to branch out individually. Two clusters were assigned to the Bet1 subtype, as they consistently formed a cluster in all three analysis

and have >60% MrBayes values, and PhyML aLRT support of 0.855. Two further clusters were assigned as Syp7 with similar supports. In both cases, the PhyML bootstrap values were negligible. Qbc SNAREs were also assigned, with only slightly better MrBayes and PhyML bootstrap values.

Unassigned Qc-SNAREs included clusters denoted Qc2, Qc3, and Qc4, whose representative sequences did consistently branch with eukaryotic representatives, and were tentatively assigned by BLAST. When *T. brucei* Qc3 sequence was searched against *H. sapiens* and *A. thaliana* proteomes, it yielded Syx8 and Syp5 respectively, and when these sequences were searched against *T. brucei* they brought up the original Qc3 sequence. The current nomenclature of Syx-8 type of SNAREs is somewhat confusing. In this study, opisthokont and amoebozoan Syx8 clustered with archeplastid Syp5 (and not Syp7 as has been previously described); and Syp7 appeared to be a plant/protist specific SNARE similar to Qb-SNARE Npsn. When representative Qc2 and Qc4 sequences (from *T. cruzi*, *L. major* and *L. Mexicana*) were searched against *H. sapiens*, *A. thaliana* and *P. sojae* genomes, they all yielded Syntaxin6, although this was not always confirmed by reverse BLAST. They were thus tentatively assigned as Syx6-like proteins.

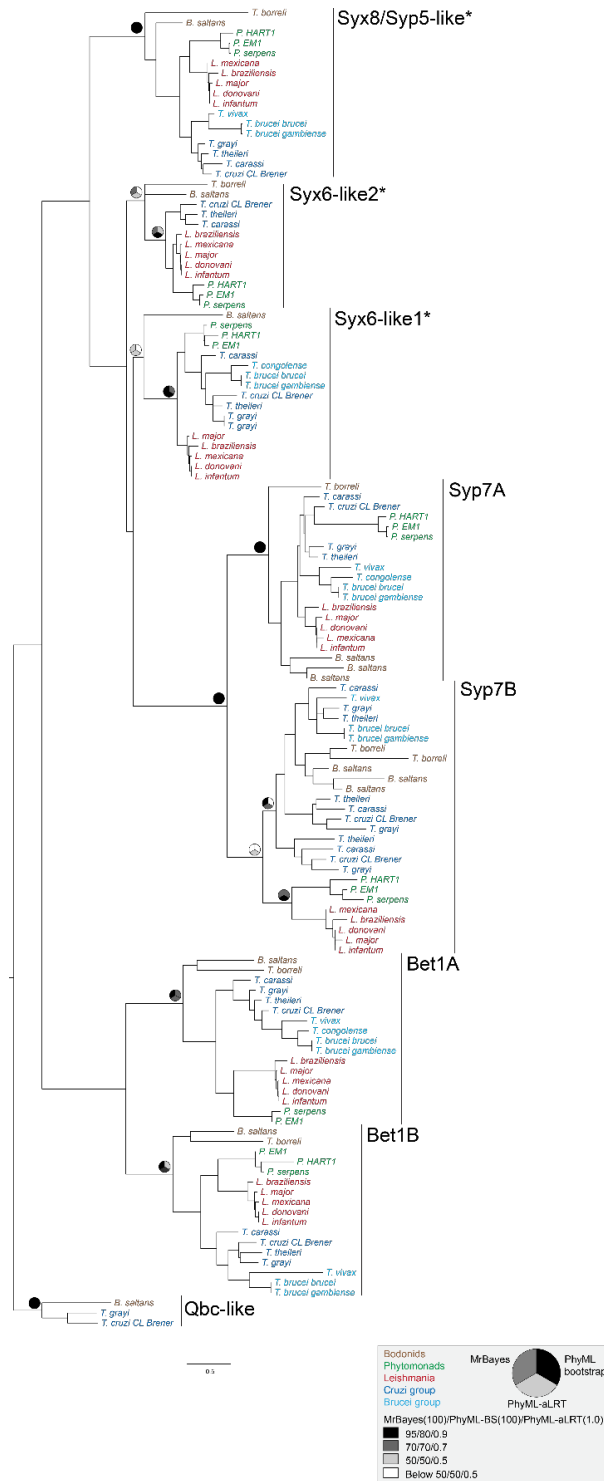
In the kinetoplastid Qc SNAREs phylogeny (see **Figure 4.9** below), two clusters of Bet1A and B were robustly clustered together, and the presence of all kinetoplastids in each indicated an early origin for this duplication, which was retained. The Syx8/Syp5-like cluster (originally named Qc3) was well conserved within the kinetoplastids even though their eukaryotic assignment was unsatisfactory and required the usage of BLAST to confirm the relationship. The relationship between the two clusters whose representatives were assigned as Syx6-like was unclear and likely unresolved. Two paralogs of Syp7 were identified. Syp7A was found across kinetoplastids singly, except in *B. saltans*, which had three paralogs. The three paralogs were clustered together within the Syp7A cluster indicating that this was likely a *B. saltans* specific expansion. Syp7B appeared to have a complex history in kinetoplastids. Multiple subgroup-specific clusters of bodonids, and the cruzi group were seen, suggesting that these were lineage-specific expansions rather than a result of early expansion followed by trypanosomatid loss which has been the predominant pattern seen so far.



**Figure 4.8 Assignment of kinetoplastid Qc- and Qbc-SNAREs**

Best PhyML topology is presented. Node values are iconised as circles for two support values each representing PhyML aLRT(1.0), PhyML bootstrap (1000), and MrBayes posterior probabilities (%) and colour-coded in grayscale as shown in the key. One representative kinetoplastid Qc-SNARE from each sub-type cluster, and two putative trypanosomatid Qbc-SNAREs (dark purple), are presented along with eukaryotic representatives coloured according to supergroup as shown in the key. Each cluster is marked with a vertical line and named. Note kinetoplastid Qc2, 3 and 4 sequences remain unassigned by phylogeny.

## Chapter 4: Evolution of Kinetoplastid SNAREs



**Figure 4.9 Kinetoplastid Qc- and Qbc-SNARE phylogeny**

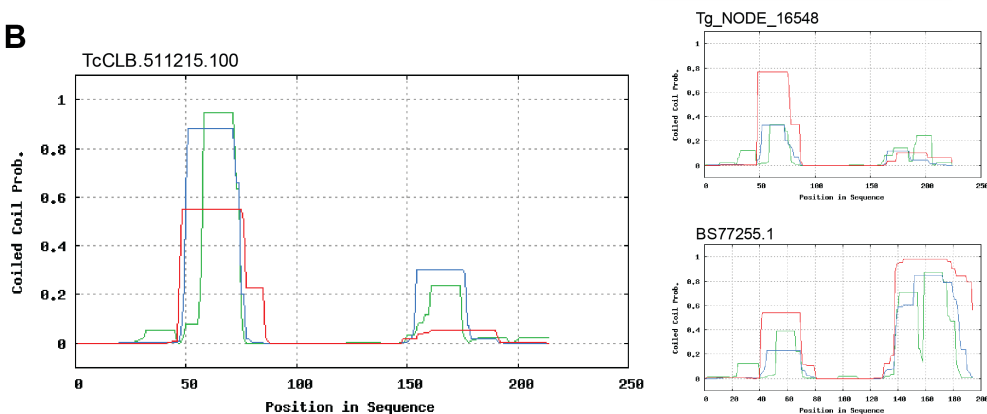
Best PhyML topology is presented. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 1000) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Taxa are coloured by phylogenetic relationships as shown in the key. Each cluster is marked with a vertical line and named according to the assignment of the representative of the said cluster according to Figure 4.8 above. Names marked with an asterisk indicate clusters that were assigned by BLAST.



Three putative Qbc proteins from *T. cruzi*, *T. grayi* and *B. saltans* were also included in this analysis. They formed a small well-supported cluster as an outgroup to the other Qc SNARE sequences. When scanned through the conserved domain database search at NCBI, all three sequences brought up a single Qbc-type domain (see **Figure 4.10A** below), and the *T. grayi* sequence had an additional “Syntaxin6 SNARE domain”. Top Human and Arabidopsis BLAST hits were all Qbc-type proteins. However, BLAST of *T. grayi* sequence into *S. cerevisiae* yielded Tlg1 (= yeast Syx6) consistent with the domain characterisation. The website <http://toolkit.tuebingen.mpg.de/pcoils> was used to scan for coiled-coil regions (Lupas *et al.*, 1991) and <http://www.cbs.dtu.dk/services/TMHMM-2.0/> to scan for *trans*-membrane domains. In keeping with the domain structure of the Qbc, two regions of coiled-coil domains (**Figure 4.10B**) were found and *trans*-membrane domains were absent. The presence in the basal bodonid lineage indicated that these proteins were lost in the other kinetoplastids. Thus, these proteins appeared particularly divergent, and further analysed as shown in Section 4.4.2 below.

**A**

Code	Putative Qbc	NCBI-CDD (non-specific hits)	<i>H. sapiens</i> BLAST	Reverse BLAST	<i>A. thaliana</i> BLAST	Reverse BLAST 2	<i>S. cerevisiae</i> BLAST	Reverse BLAST	COILS	TM
1552	TcCLB.511215.100	1x SNAP-47 (cd15888)	SNAP-25	yes	SNAP-33	yes	Sec9p	yes	2x	no
1656	Tg_NODE_16548	1x SNAP-25N (cd15861), 1x Stx6 SNARE domain (cd15851)	SNAP-29	yes	SNAP-30	no, Qc SNARE	Tlg1	no, Qc SNARE	2x	no
none	BS77255.1	1x SNAP-25N (cd15861)	SNAP-25	no significant hits	SNAP-30	no significant hits	Sec9	no significant hits	2x	no

**B**

**Figure 4.10 Kinetoplastid Qbc-SNAREs**

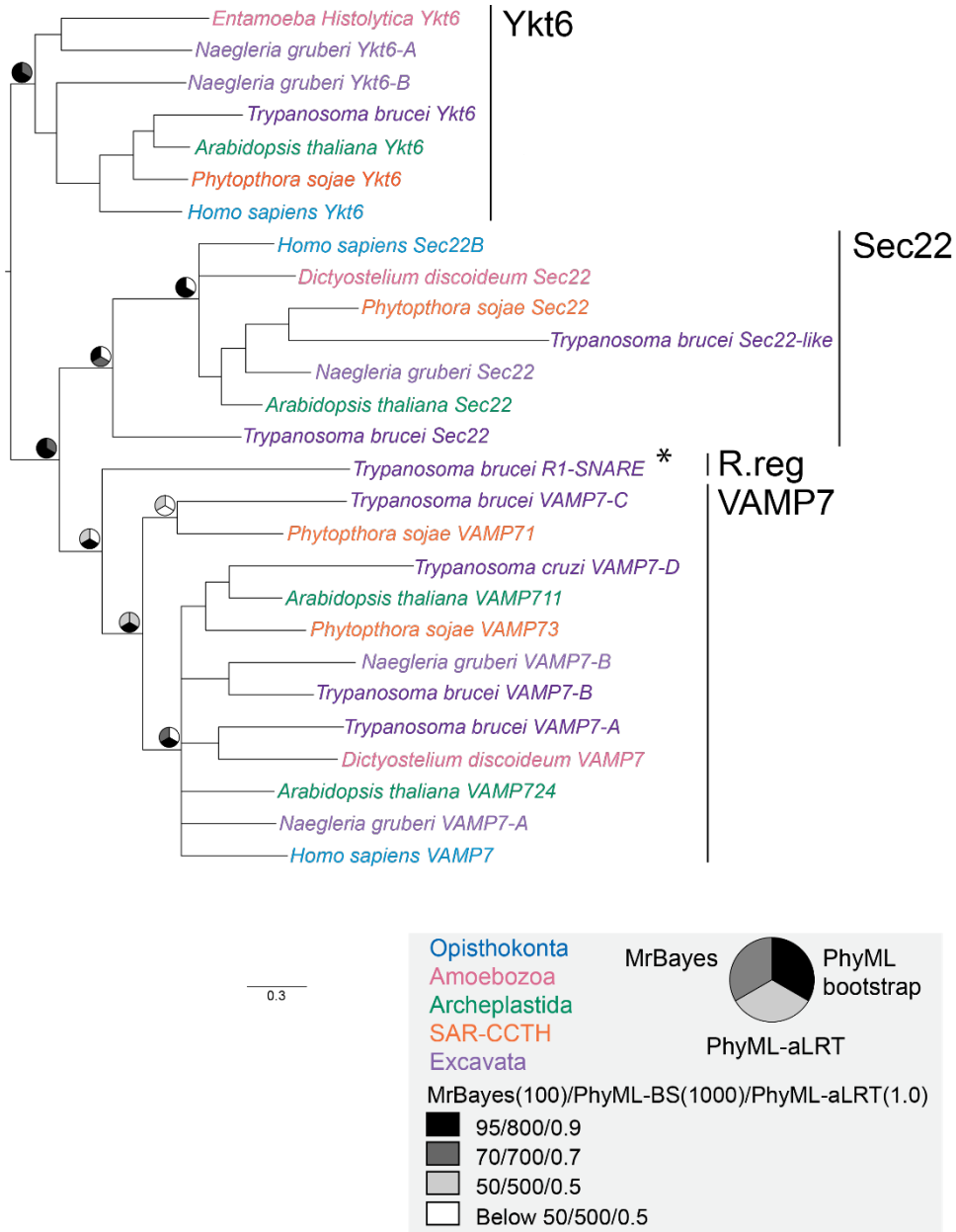
(A) Summary of BLAST and sequence analysis of putative kinetoplastid Qbc SNAREs. Hits for kinetoplastid sequences from NCBI-CDD, results of BLAST searching into genomes of *H. sapiens*, *A. thaliana* and *S. cerevisiae*, together with reverse BLAST into respective kinetoplastid genomes, the number of coiled coil domains, and the number of transmembrane domains are presented. (B) Graphical output of coiled-coils prediction from <http://toolkit.tuebingen.mpg.de/pcoils>.

#### 4.2.4 R SNAREs

Finally, the R-SNAREs in kinetoplastids could be divided into seven clusters, six of which were assigned as Ykt6, Sec22, VAMP7 A, B, C, and D by phylogeny. **Figure 4.11** below shows representatives from each cluster analysed with R-SNAREs of representative eukaryotic species. Ykt6 and Sec22 were assigned with very good support from PhyML aLRT and Mr Bayes analyses. Again, PhyML bootstrap values were less robust, but reached 733/1000 for Ykt6 indicating the well-conserved nature of the protein. A non-SNARE longin protein (with no SNARE domain but only the N-terminal longin domain) was found to consistently cluster with Sec22, and hence it was assigned as a Sec22-like protein. It was not possible to assign the representative of the seventh cluster of Tomosyn-like sequences due to long branching effects and the derived nature of kinetoplastid and eukaryotic representative sequences (such as yeast Sro7/Sro77). Tomosyn-like was however assigned based on domain structure (SNARE + WD-40 repeat regions) and sequence length. Also, as it consistently branched closely to the VAMP7 cluster, kinetoplastid Tomosyn-like is likely to have been derived from it. This was supported by the top BLAST hits for the *T. brucei* Tomosyn-like sequence: VAMP4 (derived from VAMP7) in *H. sapiens* and VAMP722 in *A. thaliana* and Snc1 (=yeast VAMP7) in *S. cerevisiae*. Within the VAMP7 cluster, no specific pattern relating the four kinetoplastid representative sequences that were used in the analysis was seen. It is likely that the duplications leading to several copies of the gene occurred at some point before the divergence of the kinetoplastid lineage.

The analysis of all kinetoplastid R-SNAREs is presented in **Figure 4.12** below. Single clusters of Ykt6, Sec22, and R.reg were found, and up to four VAMP7 clusters could be discerned. While three of these VAMP7A, B, and C were largely well represented across kinetoplastid taxa and formed well-defined robust clusters, VAMP7D was putatively present only in the bodonids and the cruzi group. This appeared to be classic case of duplication at the base of the lineage followed by a loss in most trypanosomatid lineages, but the origin of the duplication was not fully resolved in these analyses. The specific relationships between the different VAMP7 clusters also could not be satisfactorily discerned. Only VAMP7C lacked the longin domain, indicating it could be a brevin-type protein as described in animals (Synaptobrevins). However, even though it lacked the longin domain, it retained an

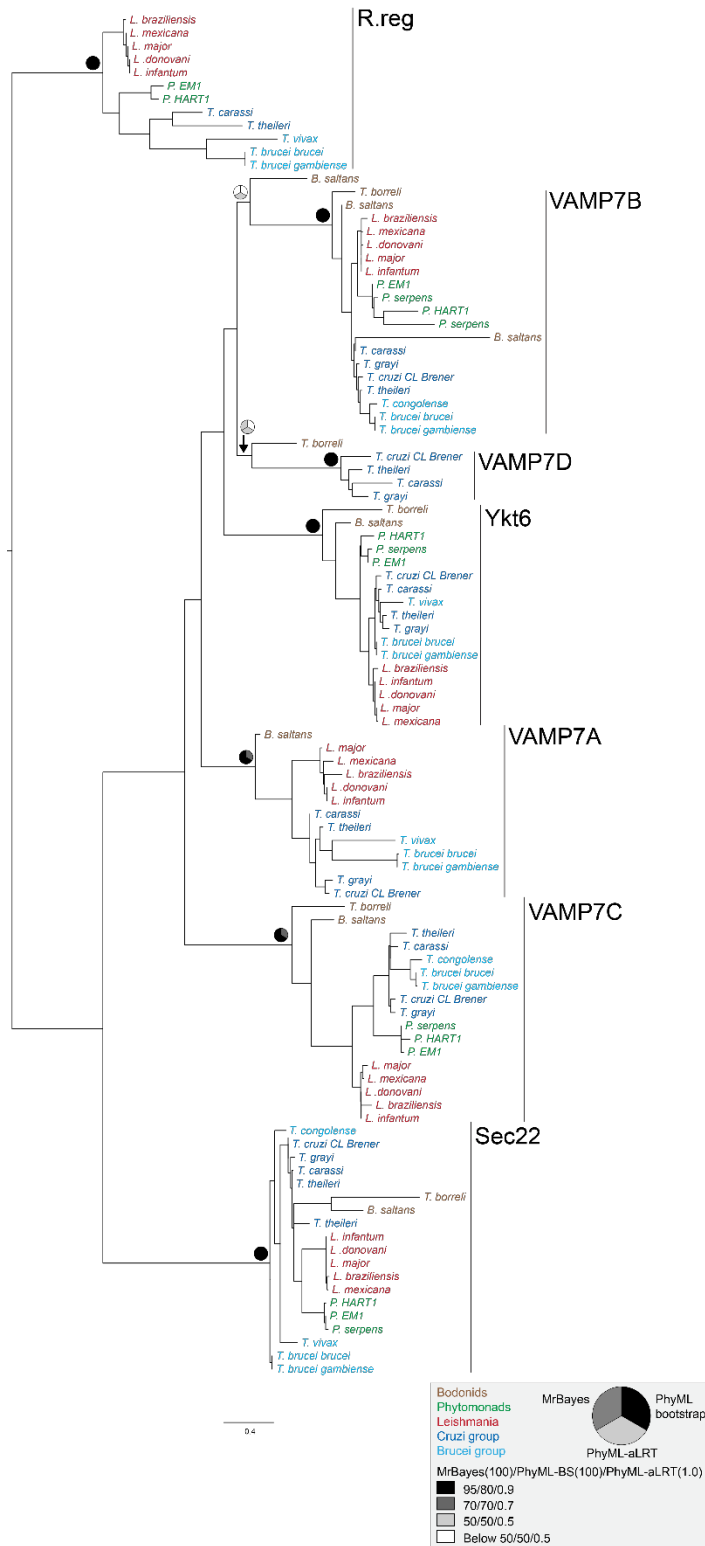
extended N-terminal sequence, which is likely unstructured. The assignment and eukaryotic context of this protein is discussed in Section 4.4.4 below.



**Figure 4.11 Assignment of kinetoplastid R-SNAREs**

Best PhyML topology is presented. Node values are iconised as circles for two support values each representing PhyML aLRT(1.0), PhyML bootstrap (1000), and MrBayes posterior probabilities (%) and colour-coded in grayscale as shown in the key. One representative kinetoplastid Qc-SNARE from each sub-type cluster, and two putative trypanosomatid Qbc-SNAREs (dark purple), are presented along with eukaryotic representatives coloured according to supergroup as shown in the key. Each cluster is marked with a vertical line and named. Note kinetoplastid Qc2, 3 and 4 sequences remain unassigned by phylogeny.

## Chapter 4: Evolution of Kinetoplastid SNAREs



**Figure 4.12 Kinetoplastid R-SNARE phylogeny**

Best PhyML topology is presented. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 1000) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Taxa are coloured by phylogenetic relationships as shown in the key. Each cluster is marked with a vertical line and named according to the assignment of the representative of the said cluster according to Figure 4.11 above.

## 4.3 Sequence-level comparison of SNAREs

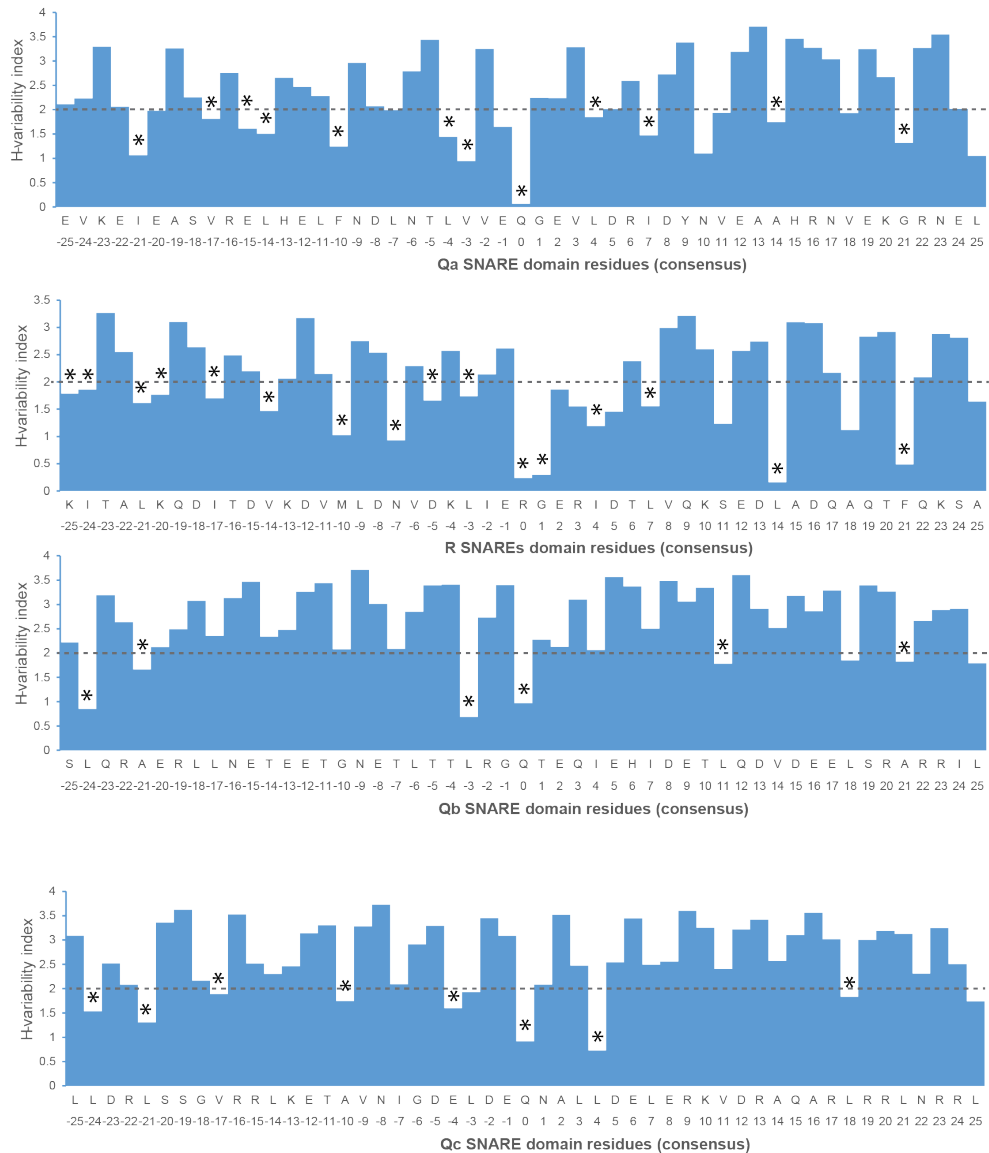
### 4.3.1 SNARE domain

SNAREs are relatively small proteins, usually 150-270 amino acids long. The subtype classification and therefore to some extent their specificity, are determined by the 60-70 amino-acid long SNARE domain. SNARE domains form a ternary complex forming coiled coils of four intertwined  $\alpha$  helices, one from each SNARE protein. This bundle contains 16 stacked layers of interacting side chains at its centre. Most of these are hydrophobic amino acids, but the central '0' layer, which is highly conserved with a Q (glutamine) or R (arginine) (Jahn and Scheller, 2006). In the 0-layer, the complementary interactions between the side chains of glutamine and the guanidino group of the arginine form strong hydrogen bonds whereas in the flanking regions, big bulky groups are complementarily packed together with smaller groups to enable correct orientation of the helices (Fasshauer, 1998).

The alignment of kinetoplastid sequences for each group Qa, Qb, Qc, and R was separately analysed. Each alignment was truncated to include 30 residues on either side of the central zero layer. This was then run through the Protein Variability Server at the website <http://imed.med.ucm.es/PVS/> (Garcia-Boronat *et al.*, 2008). Variability at each position was calculated according to the Shannon entropy method and plotted as a bar chart (**Figure 4.13** below).

The central 0-layer is indeed the best conserved residue overall. In kinetoplastids, the 0-layer was most conserved in Qa and R-SNAREs, where it is nearly universal, and less so in Qb and Qc SNAREs. A similar pattern was seen across the SNARE domain with Qa and R-SNAREs showing the lowest amount of variability in their residues. This pattern correlated with the greater dynamics and divergence of kinetoplastid Qb and Qc-SNAREs. Indeed, while most Qa and R-type SNAREs were readily assigned to eukaryotic sub-types, the assignment was less clear for Qb and Qc SNAREs - they formed less distinct clusters, with much lower support values than Qa and R-SNAREs, as seen in the sub-family-wise eukaryotic phylogenetic analyses of SNAREs.

## Chapter 4: Evolution of Kinetoplastid SNAREs



**Figure 4.13 Variability of the kinetoplastid SNARE domain**

Bar charts show variability in the residues of the SNARE domains of Qa, Qb, Qc and R-SNAREs (top to bottom). Variability is plotted on the Y-axis, and X-axis shows the consensus residues of the respective SNARE domains. 30 residues on either side of the 0-layer are plotted.

Parsing of the SNARE domains of representative sequences from a broad range of eukaryotes suggested that across eukaryotes, the 0-layer residues of Qa and R-SNAREs were more conserved than Qb and Qc-SNAREs. Thus, Qb and Qc-SNAREs maybe the main drivers of diversity in the SNARE family.

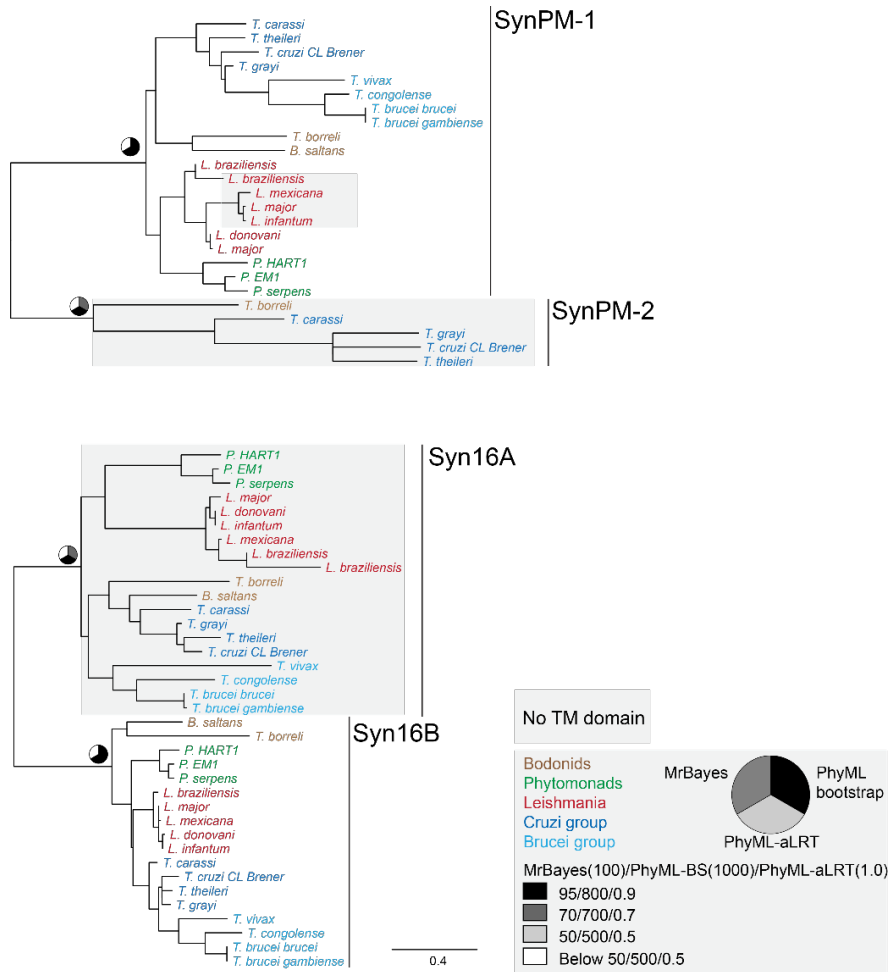
### 4.3.2 The C-terminal transmembrane domain

Almost all SNAREs contain a transmembrane domain at the C-terminal end of the protein, which helps to anchor them in the membrane. It is important for SNARE function: from correct localisation to actual execution of fusion, the presence and character (including length) of the TM domain has been shown to be necessary (Jahn and Scheller, 2006). Notable exceptions to this rule are two special proteins, the R-SNARE Ykt6, and the Qbc SNAREs with two SNARE domains, which are palmitoylated at the C-terminal end or between the SNARE domains respectively, in order to be anchored in lipid membrane.

Several kinetoplastid SNAREs that were not Ykt6 or Qbc appeared to lack the transmembrane region. These sequences were checked using the TMHMM server <http://www.cbs.dtu.dk/services/TMHMM/> to confirm the absence of the domain. Among the Qa SNAREs, two clusters in which all the members lacked the TM domain were found: Syx16A and SynPM-2. Further, some of the sequences from the *Leishmania* species in the SynPM1 cluster were also missing the TM domain. **Figure 4.14** below shows the phylogeny of SynPM and Syx16 SNAREs, with sequences lacking TM domain marked with a grey box.

Two copies of SynPM1 were found in both *L. braziliensis* and *L. major*. Both appeared to be duplications that took place at the same locus: LbrM28.1630 and LbrM28.1640, LmjF28.1470 and LmjF28.1480. These sets of sequences were identical and differed only in the last 40-50 residues of the proteins (See **Figure 4.15** below). In fact, the divergence began right at the start of the transmembrane domain (residue 225) and continued through to the end of the protein. While one sequence of each pair retained the transmembrane domain, the other had no detectable TM motif. The two paralogues were studied in *L. major*, where it was found that GFP-LmjF28.1470 (has TM) localised to a compartment close to but not quite at the flagellar pocket while GFP-LmjF28.1480 (no TM) was cytosolic (Besteiro *et al.*, 2006). It is therefore likely that the former retained its conserved function of fusion at the cell membrane while the latter may have lost the ability to form SNARE complexes. Furthermore, of the three *Leishmania* species that did not duplicate SynPM1, two (*L. infantum* and *L. mexicana*) had lost the TM domain, while one (*L. donovani*) retained it.

## Chapter 4: Evolution of Kinetoplastid SNAREs



**Figure 4.14** Pattern of TM domain loss in SynPM and Syx16

Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 1000) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Taxa are coloured by phylogenetic relationships as shown in the key. Each cluster is marked with a vertical line and named according to phylogenetic analysis. Grey boxes indicate sequences that do not possess C-terminal transmembrane (TM) domain; single asterisks indicate lineage-specific duplications.



A

```

1260LmjF.28.1470/1-287 1 MAELQQKHTENMQTVDEARSKALRGEIDEL SRETGNAAKAAKEKLDAMSK 50
1261LmjF.28.1480/1-266 1 MAELQQKHTENMQTVDEARSKALRGEIDEL SRETGNAAKAAKEKLDAMSK 50

1260LmjF.28.1470/1-287 51 NTAKLKSVPDSVQANSAIIRIEENQYTHLVLKLTMAMAEYQRQQSANEAY 100
1261LmjF.28.1480/1-266 51 NTAKLKSVPDSVQANSAIIRIEENQYTHLVLKLTMAMAEYQRQQSANEAY 100

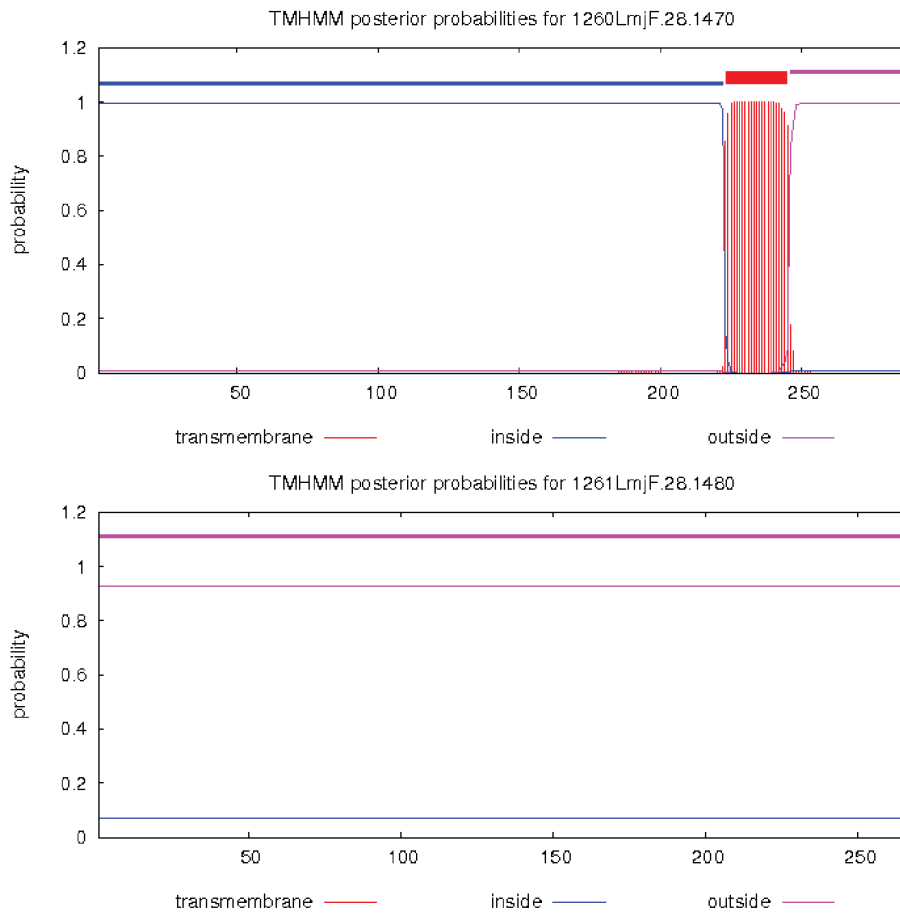
1260LmjF.28.1470/1-287 101 YKAQTQRQIKIKYTNPDGSAIDDSTAAQLAEQVMENDTSSYIFQQSKEVL 150
1261LmjF.28.1480/1-266 101 YKAQTQRQIKIKYTNPDGSAIDDSTAAQLAEQVMENDTSSYIFQQSKEVL 150

1260LmjF.28.1470/1-287 151 ASIIETRNDIYRIEQSMRDLNQLFNDLAFLVNEQGELMDVILANVQQSTR 200
1261LmjF.28.1480/1-266 151 ASIIETRNDIYRIEQSMRDLNQLFNDLAFLVNEQGELMDVILANVQQSTR 200

1260LmjF.28.1470/1-287 201 YVAKGRAALKKARRYQKKSRRKLCVLCVGMTIVALFVVVGVVLAATIKI 249
1261LmjF.28.1480/1-266 201 YVAKGRAALKKARRYQKKSRRKLRCSVLRKEYVEKTAAYSPWQEAASVQV 250

1260LmjF.28.1470/1-287 250 PRATPSPPAQVRGQGRGQVSDSHAGAAKAPPVRQFTVL 287
1261LmjF.28.1480/1-266 251 HR--HCCPEGEKW---AAWED----- 266
    
```

B



**Figure 4.15 Comparing *L. major* SynPM1 sequences**

(A) Pair-wise alignment of the two SynPM1 sequences of *L. major* show identical sequences but for the TM domain. (B) Analysis with TM-HMM predictor reveals lack of TM domain in one of the duplicated sequences.

Given that all *Syx16A* sequences lacked the TM domain, it is possible that this domain was rendered dysfunctional in the common ancestor of all kinetoplastids, i.e., before radiation of the species of this lineage. Interestingly even within this group,

there was a lineage specific duplication of the *L. braziliensis* Syx16A giving rise to two proteins both of which lacked a functional TM domain. A single Syx16B sequence, from *T. borreli* also lacked the TM domain. Thus, for a variation that was likely to render a SNARE non-functional, this is a surprising extent of conservation. It is possible that these proteins may be soluble and act as inhibitory SNAREs (iSNAREs), forming unproductive complexes with cognate SNAREs as a regulatory mechanism. It is worth testing this in tractable organisms like *T. brucei* and *L. major*. Paralogous Syx16 sequences from other organisms such as *N. gruberi*, *P. ramorum*, and *T. vaginalis* all retain their TM domain so this appears to be a kinetoplastid restricted phenomenon.

Some other Qa SNAREs have also lost the TM domain but showed a less dramatic distribution. These include three Syntaxin5 sequences from *T. theileri*, *T. congolense* and *T. borreli* as well as both the *T. theileri* sequences from Qa1. No cluster-wide losses of these domains was seen in Qb, Qc, and R SNAREs. The only major pattern discerned was with respect to *T. theileri* sequences. VAMP7A, Vti-likeA1, 2x Syp7B, Syx8/Syp5-like, and Syx6-like1 sequences from *T. theileri* have all lost a functional TM domain. This means that out of the 35 SNAREs found in this species, nine, i.e. 26% were missing TM domains. *T. borreli* has also lost the TM domain from both Npsn paralogues as well as Syx6-like2, further to the losses mentioned above (SynPM2, Syx16A and B, Syx5). It has therefore lost TM domains in 7 out of 28 SNAREs, which is also 25% of its repertoire. Again, it is possible that these proteins have a role in inhibitory regulation of cognate SNAREs. However, interestingly, these were not duplicates that diverged, but were part of the core SNARE repertoire. Except for Syntaxin5, all other SNAREs operate in post-Golgi pathways and appear to be largely sporadic losses, with one exception. Syx16A-Vti-likeA1-Syx6-like1-VAMP7A, all lost in *T. theileri* are putative orthologs of Syx16-Vti-Syx6-VAMP7 which form a SNARE complex mediating fusion in the *trans*-Golgi network. Experimental analysis may shed light on the functional significance of this significant TM domain loss.

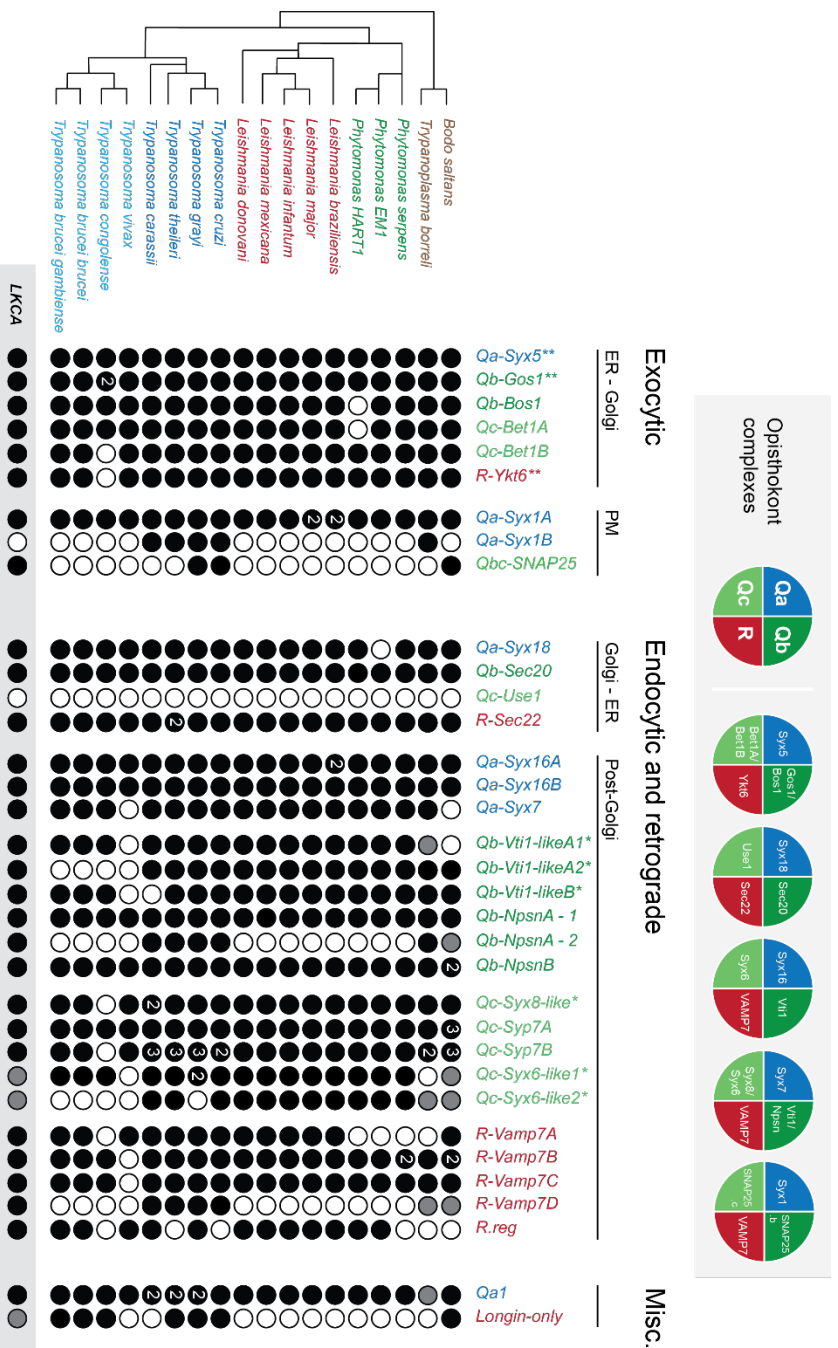
#### 4.4 Patterns in the evolutionary history of SNAREs

Overall, as can be seen in **Table 4.3** below, most SNAREs predicted to be present in LECA are nominally represented in the kinetoplastid genomes suggesting a well-conserved cohort. As will be discussed in Section 4.4.2, the ancient origin of the synaptobrevin is questioned in this analysis. Nevertheless, ~90% of LECA SNAREs are represented across the kinetoplastid lineage. Only the SNAP-25 shows a very restricted presence, all others are better represented. On detailed examination of SNARE clusters, many instances of loss as well as pan-kinetoplastid and lineage restricted expansions were found. These patterns are represented in **Figure 4.16** below (next page), and their implications are discussed in this section.

Family	Predicted LECA repertoire	LECA	Absent from kinetoplastids	Retention*	Lineage-specific paralogues
SNAREs		21		19/21 (90.4%)	
Qa	Syx18, Syx5, Syx16, Syx7, Syx1	5	none	5/5 (100%)	Qa1, Syx16 (2), Syx1(=SynPM)(2)
Qb	Sec20, Bos1, Gos1, Vti1, Npsn	5	none	5/5 (100%)	Vti1-like (3), Npsn (3)
Qc	Use1, Bet1, Syx6, Syp5/Syx8, Syp7	5	Use1	4/5 (80%)	Bet1 (2), Syp7 (2), Syx6-like (2)
Qbc	SNAP-25	1	Very restricted	1/1 (100%)	none
R	Sec22, Ykt6, VAMP7, R.reg, Syb-1**	5	Syb-1	4/5 (80%)	VAMP7 (4)

**Table 4.3 Comparison of kinetoplastid and LECA SNARE repertoires**

Putative LECA repertoires of SNAREs are presented; along with subfamilies absent from kinetoplastids as well as novel lineage-specific paralogues, total number of each paralogue putatively present at the base of the kinetoplastids is shown. Retention\* identifies SNAREs if present in at least one of the kinetoplastids studied. This study questions the ancient origin of the synaptobrevin Syb-1 but is included as per previous studies (Klopper *et al.*, 2007).



**Figure 4.16 Representation of SNAREs in kinetoplastids**

Columns, show individual subtype clades as found in kinetoplastids (assigned to known eukaryotic sub-types or lineage-specific) with taxa shown as rows, with the hypothetical last common kinetoplastid ancestor as the lowest row highlighted by a gray box. A schematic phylogeny for the taxa is drawn on the left, derived from (Manna et al. 2013). Black circles indicate at least one member of the clade has been found with phylogenetic support of 80/0.5/50 (MrBayes/PhyML-aLRT/PhyMLbootstrap) or more; gray circles indicate lower support but above 50 (MrBayes) or in the case of the LKCA, indeterminate presence from given data. SNARE subtypes are colour-coded as Qa – Blue; Qb – dark green, Qc – light green and Qbc – green, and R – red. Subtypes with single asterisks indicate that assignment was via BLAST only; double asterisks indicate SNARE is involved in retrograde pathway in addition to anterograde.

#### 4.4.1 ER to Golgi and back

Across eukaryotes, the SNARE complement that mediates the early secretory steps and retrograde transport between ER and Golgi were highly conserved, while the endocytic and exocytic pathways were expanded and diversified in many lineages. Qa-Syx5, Qb-Bos1, Qc-Bet1 and R-Sec22 have long been known to be mediators of transport from the ER to Golgi (Newman *et al.*, 1990; Hardwick and Pelham, 1992). See **Figure 4.3** in Section 4.2 above for a map of SNARE complexes within a generalised eukaryotic cell. This SNARE complex also mediates intra-Golgi transport and while the complex constituting Qa-Syx5, Qb-Gos1, Qc-Sft1 and R-Ykt6 is thought to exclusively mediate COPI transport within the Golgi (Ballensiefen *et al.*, 1998). Theoretically, the R-SNAREs Sec22 and Ykt6 should form the v-SNARE in these SNARE complexes. However, Q-SNAREs Sft1/Bet1 have also been described as v-SNAREs, while Sec22/Ykt6 as t-SNAREs in a three-part complex in each of these complexes (Parlati *et al.*, 2000).

While Sft1 and Bet1 were described as two different Qc SNAREs acting at these two stages of the pathway in fungi, mammals and archaeplastids in the SNARE database (Kloepper *et al.* 2007), they remained unresolved in this analysis. It is likely that Bet1 fulfils the requirements of the early secretory pathway – kinetoplastids possess two paralogs Bet1A and Bet1B, as do *H. sapiens* (Bet1 and Gs15 (Bet1-like)) while *A. thaliana* uniquely has an expanded set of four (described as Bet11, Bet12, Sft11 and Sft12) and *D. discoïdium* has just one. The other SNAREs in these complexes were all singletons in kinetoplastids bar some species-specific losses (see **Figure 4.16** above) and a duplication of Gos1 in *T. congolense*. Hence, these data suggest the possibility of some differentiation in this pathway consistent with known multiple budding pathways at the trypanosome ER (Sevova and Bangs, 2009).

The retrograde pathway from the Golgi to the ER is mediated by the SNARE complex Syx18-Sec20-Use1-Sec22 in yeast (Dilcher *et al.* 2003). As **Figure 4.16** shows, all but Use1 were conserved across kinetoplastids, except for single losses of Syx18 and Sec20 in *Phytomonas EMI* and *HART1* respectively, and duplication of Sec22 in *T. theileri*. Together with retention of the COPI coat this suggests that retrograde transport system are present, which is also consistent with the presence of

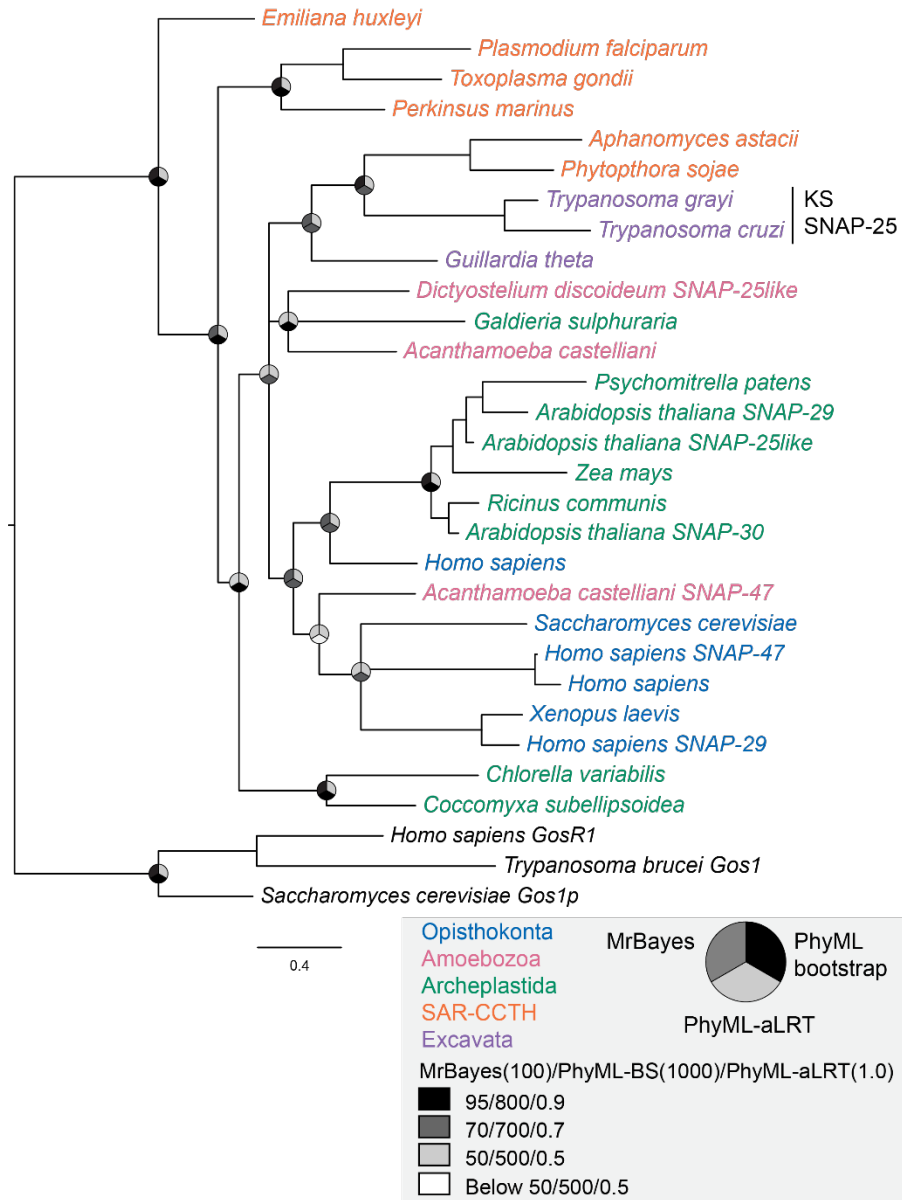
an ortholog of the KDEL receptor and ER-retrieval signals on major ER proxies such as BiP (Bangs *et al.*, 1996; Schwartz *et al.*, 2013).

#### 4.4.2 Exocytic SNAREs

The exocytic SNARE complex has mainly been studied in metazoan neurons and yeast. It is a special ternary complex composed of Syx1-SNAP25-Syb1 in which SNAP-25, a ‘Qbc’ SNARE, provides two SNARE domains and the R-SNARE is a ‘short’ VAMP or ‘brevin’ lacking the N-terminal longin domain. Syx1, representing plasma membrane syntaxins (SynPM) can be widely identified across eukaryotes and is likely an ancient SNARE protein present in LECA (Dacks and Doolittle 2002). SNAP-25 and Syb1-like proteins can be identified *in-silico* in several lineages outside of Opisthokonta, but there are very few experimental studies.

Kinetoplastids possess SynPM although it has a complex evolutionary history in kinetoplastids as described above with the lack of the C-terminal TM domains in certain species. The localization of TbSynPM1 was similar to the TM-domain containing SynPM1 from *L. major* which localises close to the flagellar pocket (Besteiro *et al.*, 2006). The putative Qbc-type SNAREs in kinetoplastids, present only in three species: *B. saltans*, *T. cruzi* and *T. grayi* could not be assigned with canonical yeast or human Qbc SNAREs. They were therefore analysed separately along with a wider diversity of putative Qbc SNAREs from across the eukaryotes. The result is displayed in **Figure 4.17** below.

Qbc SNAREs have expanded separately in the plant and metazoan lineages, similar to the pattern of Syntaxin1 (=SynPM) in these lineages (Dacks and Doolittle, 2002). Qbc-SNAREs formed two separate clusters that included two representative species in which their repertoire was expanded (*H. sapiens* and *A. thaliana*). The kinetoplastid sequences (marked by a vertical line) formed another separate cluster with other protists (stramenopiles and excavates). The higher order relationships between the sequences remained insufficiently resolved but the kinetoplastid and canonically described Qbc SNAREs appeared quite divergent. In contrast to the expansion in metazoan and land plants, Qbc SNAREs in diverse protist taxa show no expansion, even in species with very complex trafficking machinery such as *P. tetraurelia* or large genomes such as *N. gruberi*.



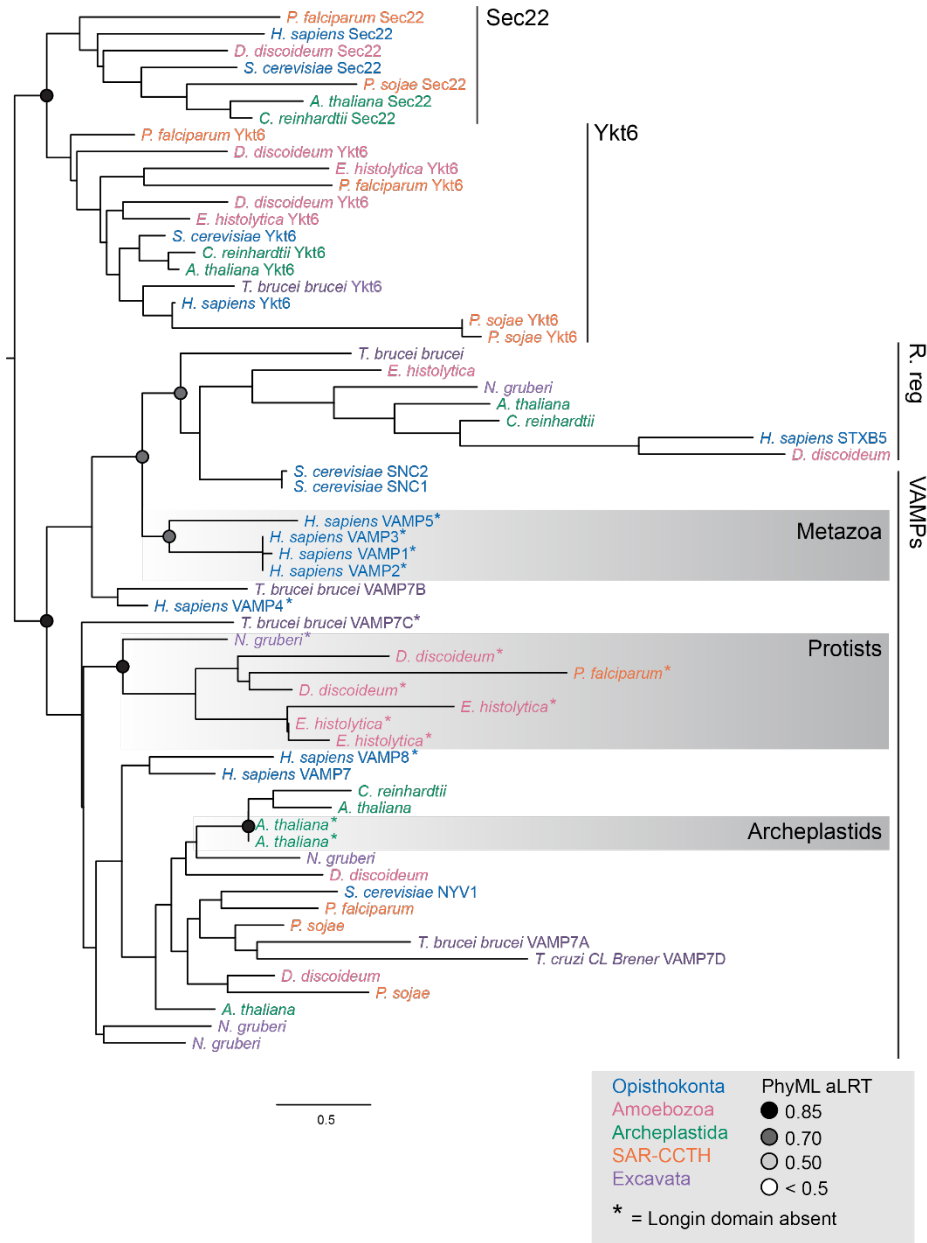
**Figure 4.17 Phylogenetic analysis of eukaryotic Qbc SNAP-25 like SNAREs**

Best PhyML topology rooted on Qb-Gos1 sequences is present. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 100/0) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Eukaryotic representative SNAP-25like SNAREs identified in Opisthokonta (Blue), Amoebozoa (Pink), Archeplastids (Green), SAR-CCTH (Orange) and Excavata (light Purple) are shown. Note expansions in archeplastids and opisthokonts are lineage specific. Higher-level relationships between clusters is not resolved, but presence of several separate clusters indicates divergence of sequences in different lineages. Kinetoplastid Qbc-like sequences are found to cluster with representative stramenopile sequences (*A. astacii* and *P. sojiae*), marked with a vertical line.

The eukaryotic phylogeny of synaptobrevins was also complex, and not easily resolved. **Figure 4.18** below represents a reconstruction of representative ‘synaptobrevin’ containing proteins (as determined by the presence of the interpro domain IPR001388) from select organisms across eukaryotes. These included Sec22, Ykt6, R.reg proteins, as well as VAMPs with and without the longin domain. Perhaps because of insufficient sequence information (as VAMPs are small proteins and the SNARE domain is relatively small), there was no robust support for the clustering in PhyML bootstrap trees and the MrBayes analyses failed to cluster sequences. Therefore, only the phyML aLRT tree is presented.

Here, SNAREs lacking the longin domain were found in different clusters of diverse taxa. However, they did not always resemble the Syb1-like protein (described in animals) where the sequence is truncated to a few residues at the N-terminus of the SNARE domain. Instead, extended N-termini were seen, but without an identifiable longin (or any other) domain. This indicates that such proteins likely emerged independently in several lineages at various points in evolution, likely from the loss of the longin domain from a ‘VAMP7’ and is not an ancient LECA protein asymmetrically retained in extant eukaryotes. A single independently evolved longin-lacking VAMP was present in kinetoplastids (VAMP7C) but in *T. brucei*, it showed an endosomal localization in juxtaposition to Golgi and lysosomal markers rather than cell membrane and did not form a complex with TbSynPM according to cryo-immunoprecipitation experiments in *T. brucei* (discussed in Chapter 5). However, TbVAMP7C was the only R-SNARE detected in the cell surface enriched proteome (from C. Gadhela), so it may be recycling to the surface undetected by immunofluorescence methods.





**Figure 4.18 Phylogenetic analysis of eukaryotic ‘synaptobrevin’ domain containing sequences.**

All R-SNARE synaptobrevin domain (IPR01388) containing sequences from selected eukaryotic representatives were analysed. PhyML bootstrap and Bayesian analyses were inconclusive due to very low supports and unresolved relationships respectively so only the PhyML aLRT analysis is shown. Statistical support at key nodes are presented as circles filled in in gray-scale according to the key shown. Eukaryotic representative sequence are colour-coded as Opisthokonta (Blue), Amoebozoa (Pink), Archeplastids (Green), SAR-CCTH (Orange) and Excavata (light Purple). Sec22 and Ykt6 are conserved compared to other VAMPs. R.reg forms long branches likely due to derived nature of the sequences. ‘Brevin’-like VAMPs, lacking the N-terminal longin domain are marked with an asterisk (\*).

SNAREs were experimentally studied in the protozoan ciliate *Paramecium tetraurelia*. The ciliate has a single Qbc SNARE PtSNAP-25 which unlike the cell-membrane restricted mammalian homologue, was found on the membranes of food

and contractile vacuoles, cell membrane and in soluble form in the cytoplasm, and may not have a role in regulated exocytosis (Schilde *et al.*, 2008). *In-silico* analysis failed to reveal Qbc sequences in other ciliates (Schilde *et al.*, 2008). R-SNAREs lacking the longin domain have also been found, one of which (named PtSyb10) localized to the cell membrane close to basal bodies, and oral cavity. It was proposed to have a role in vesicle trafficking by delivering components for the formation/maintenance/functioning of cilia or the cell membrane (Schilde *et al.*, 2010). It is unknown if the two SNAREs interact. As *P. tetraurelia* represents a particularly diverged and complex ciliate, experimental studies in other non-opisthokont lineages are needed before the conservation of the exocytic SNARE complex across eukaryotes can be determined. Thus SynPM, VAMP7 and Qbc SNAREs all show divergence in different lineages suggesting flexibility, likely owing to adaptive pressure, in the evolution of this complex.

#### 4.4.3 Endocytic and lysosomal SNAREs

The endocytic pathway had a larger cohort of Qb, Qc, and R SNARE proteins than the classical secretory pathway, reflecting the larger number of routes that are associated with endocytic trafficking, as the endosomal compartments may act as a sorting hub for incoming and outgoing traffic. This includes recycling to plasma membrane, retrograde transport to the Golgi/ER, transport to lysosomes and lysosomal exocytosis. The expansion of the SNARE set may allow for more varied and flexible routes.

Four Qb SNARE proteins (Vti-like A, B and NpsnA, B), four Qc SNAREs: Syp7A, B, Syx6-like and a Syx8-like protein, and four R-SNAREs VAMP7 A, B, C and D as well as the Tomosyn-like regulatory R-SNARE, were identified. For several of these proteins, multiple copies existed, particularly in the basal bodonids and the cruzi group (see **Figure 4.16** above). In contrast, a relatively small repertoire of two Qa-SNARE sub-types were found: SynE (=Syx7) and Syx16.

Several SNAREs were found to be present only in the bodonids and the cruzi group, having been selectively retained while other trypanosomatids lost them. These were SynPM2, NpsnA2, copies of Syp7B and VAMP7D. Considering that the canonical versions of all these SNAREs have been described to be involved in post-Golgi trafficking, it is possible that these four may constitute a novel SNARE

complex in the organisms in which they are retained. However, that SynPM2 may be a non-functional protein, lacking the TM domain, and the independent emergence of the Syp7B copies in bodonids and the cruzi group separately indicate that these proteins may not act together after all. Experimental evaluation of *T. cruzi* SNARE complexes could help to understand the role of these expanded SNAREs better.

Most SNARE losses were scattered across different lineages with no obvious pattern. There was however, an instance of putative co-evolutionary loss of the components of predicted SNARE complexes, at the base of the brucei group lineage. SNAREs Qa-SynPM2, Qbc-SNAP25 and R-VAMP7D, which could form a putative exocytic complex, were lost. Qb-Vti-like A2, Qb-Npsn A2, and Qc-Syx6-like2 that were predicted to form complexes with VAMP7 were also lost (see **Figure 4.16** above). Except SNAP-25 (which shows a broader pattern of loss in kinetoplastids), these were all SNAREs derived from kinetoplastid-specific expansions. Therefore, this possibly indicates loss of kinetoplastid-specific post-Golgi organelles/pathways.

In addition, the brucei group trypanosomes *T. vivax* and *T. congolense* showed further losses of SNARE subtypes. *T. vivax* has lost Qa-SynE, Qb-Vti-likeA1, Qb-Vti-likeB, Qc-Syx6-like1, R-VAMP7B, and R-VAMP7C. *T. congolense* has lost Qc-Bet1B, Qc-Syx8-like, Qc-Syp7B, R-Ykt6, R-VAMP7A, and R.reg Tomosyn-like; there was absolutely no overlap between the two sets of losses. The identification of these losses could be artefact of incomplete genome assembly although that is unlikely as no such extensive absences were seen in Rab and RabGAP proteins of these genomes. The respective SNARE repertoires of these two genomes were likely sculpted after their divergence into separate species. In *T. vivax*, these losses mean that they possess none of the Vti1-like or Syx6-like SNAREs and only one VAMP7 paralogue. Together with SynE, the four proteins form the endosomal SNARE complex. The complete loss of three of the four SNAREs that form this complex is a major loss indicating a much reduced endosomal system in *T. vivax*. In *T. congolense*, both early secretory and post-Golgi SNAREs were lost and they did not appear to be members that constitute a complete whole SNARE complex. They only formed parts of it: Bet1 and Ykt6 are in the same complex, which together with Syx5 and Bos1 mediate ER to Golgi transport; Syx8/Syp5 and VAMP7 mediate late endocytic traffic to the lysosome with Vti1 and SynE SNAREs.

In **Figure 4.16**, circles for some sequences of the post-Golgi pathway are coloured grey. These were SNAREs, mainly from the bodonid species, that were in the particular cluster consistently, but with low phylogenetic support, likely an indication of sequence divergence from the trypanosmatid consensus (of which there are 16 genomes represented, in contrast to two of the bodonids). Perhaps sequencing of other bodonid genomes (*B. saliens*, *B. curvifilus*, *B. caudatus* and other Bodo lineages, as well as related organisms from genus *Rynchomonas* and *Ichthyobodo*), would help place them better.

#### 4.4.4 Longin-only proteins

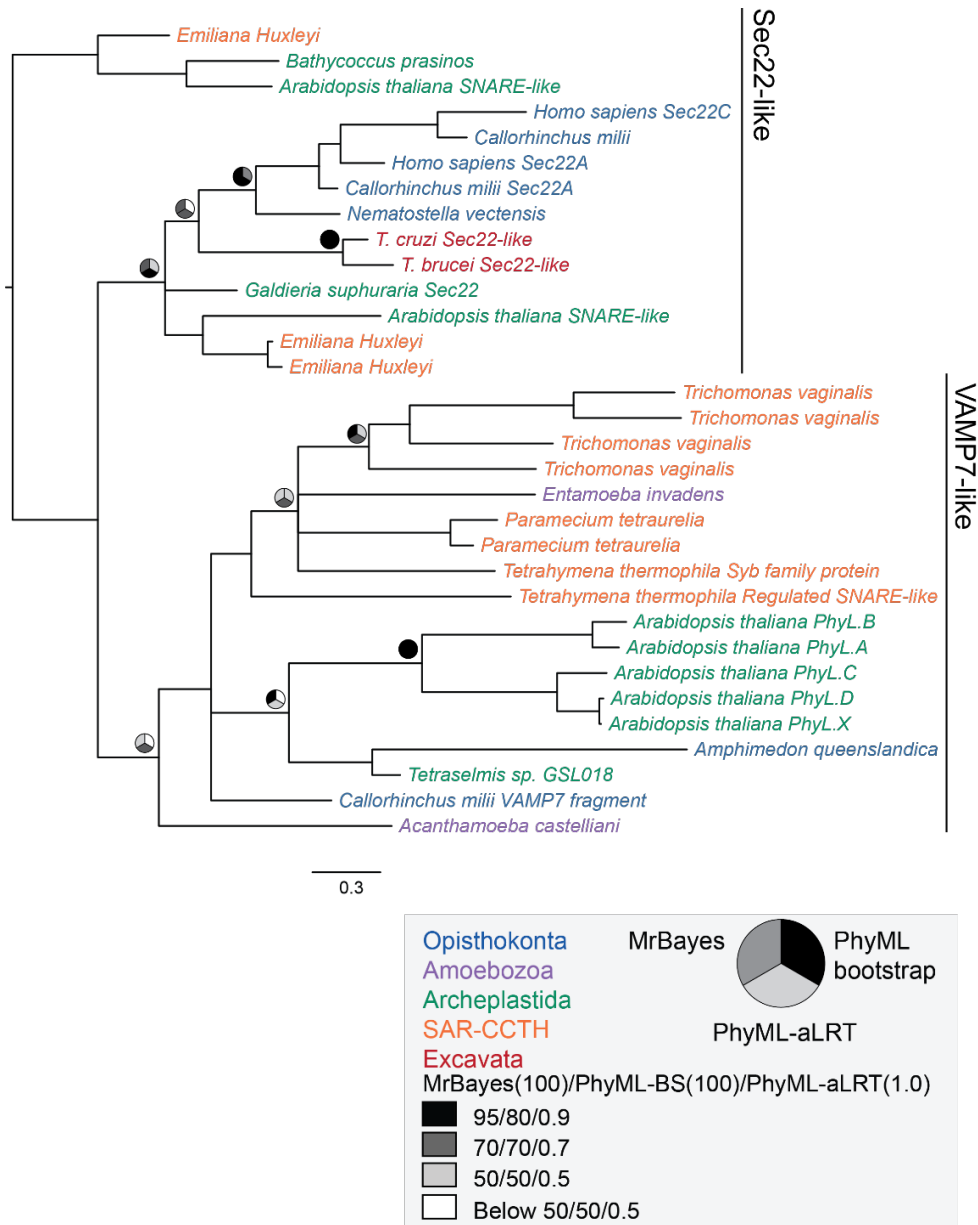
Longin-like folds are conserved in many trafficking proteins and are involved in their interactions with each other. These include tethers like Trs20 (Jang *et al.*, 2002); the  $\sigma$  and  $\mu$  subunits of clathrin adaptors (Collins *et al.*, 2002); and the SRX domain of the Sr $\alpha$  subunit of the signal recognition particle (SRP) (De Franceschi *et al.*, 2014). Most R-SNAREs, except the Tomosyn-like regulatory proteins, possess a longin domain through which SNAREs mediate interactions with other trafficking partners which helps regulate SNARE function. In kinetoplastids, proteins containing a longin-domain but not a SNARE domain were found. They formed a single cluster in the kinetoplastid tree indicating a common origin. They were however, only patchily retained in extant taxa, being found in *B. saltans*, the cruzi group (except *T. carassi*) and the brucei group (except *T. vivax*).

Such non-SNARE longin proteins have also been described in other organisms. In mammals, the R-SNARE Sec22b has two further isoforms lacking the SNARE domain, namely Sec22a and c which are reported to mediate ER-Golgi transport (Tang *et al.*, 1998). In addition, the alternative splicing of the SYBL1 gene results in three isoforms: one with both SNARE and longin domains (=VAMP7), one with just the SNARE, and finally, one with just the longin domain (Vacca *et al.*, 2011). A non-SNARE Sec22 protein encoded as such at the genome level (and not because of alternative splicing) was reported in the rice proteome (Sec22-like, Q6UU98). It was deduced to have arisen from the deletion of the SNARE domain from a Sec22 paralogue where the protein also lacks the C-terminal transmembrane (TM) domain due to a frameshift resulting from the new exon-intron boundary. Land plants also possess a unique type of non-SNARE longin, in which the SNARE domain is

replaced by PhyL (phytolongin domain) which is reportedly derived from the SNARE domain. Kinetoplastid non-SNARE longins cluster with the Sec22 group, and therefore were likely derived from them rather than VAMP7 as in the plants.

*T. brucei* and *T. cruzi* longin-only proteins were analysed with proteins with such domain structure from across eukaryotes. These were found by searching genomes of species with sequences containing the interpro longin domain signature: IPR010908. The longin domain is usually found along with the synaptobrevin domain, but only sequences lacking the latter were chosen. Such proteins were patchily distributed among different phyla. No such proteins were found in any available species' predicted proteome from Cryptophyta, Fornicata, Heterolobosea, Rhizaria and Stramenopiles. This may indicate a paucity in the number of sequenced genomes rather than real absence. In the Viridiplantae group, phytolongins as described before were found but there were also non-phytolongin sequences from both subgroup Streptophytes (to which phytolongins are meant to be restricted), as well as the subgroup Chlorophytes.

Phylogenetic analysis of all these proteins revealed two major subgroups into which the sequences clustered: one appears to be Sec22-derived the other, VAMP7-derived (see **Figure 4.19** below). The former cluster contains human and other metazoan Sec22a and c sequences (dark blue), as well as the kinetoplastid longin-only sequences (red). Also present were sequences from red algae *Galdieria sulphuraria* (light green), land plant *A. thaliana* (dark green), and the haptophyte *E. huxleyi* (brown). In the latter cluster, sequences from the alveolates *P. tetraurelia* and *Tetrahymena thermophile* (purple), and the diplomonid *Trichomonas vaginalis* were found. The four sequences from *T. vaginalis* appeared to be a species-specific expansion, which is unsurprising, given the large genome and expanded SNARE complement of the species. The classical phytolongins also form a separate cluster within this clade. Therefore, the phylogenetic analysis suggests a convergent evolutionary mode of emergence of such non-SNARE longin proteins. It would be interesting to see whether these proteins perform similar functions across different lineages. This study provides a starting point for comparison by the analysis of *T. brucei* VAMP7C.



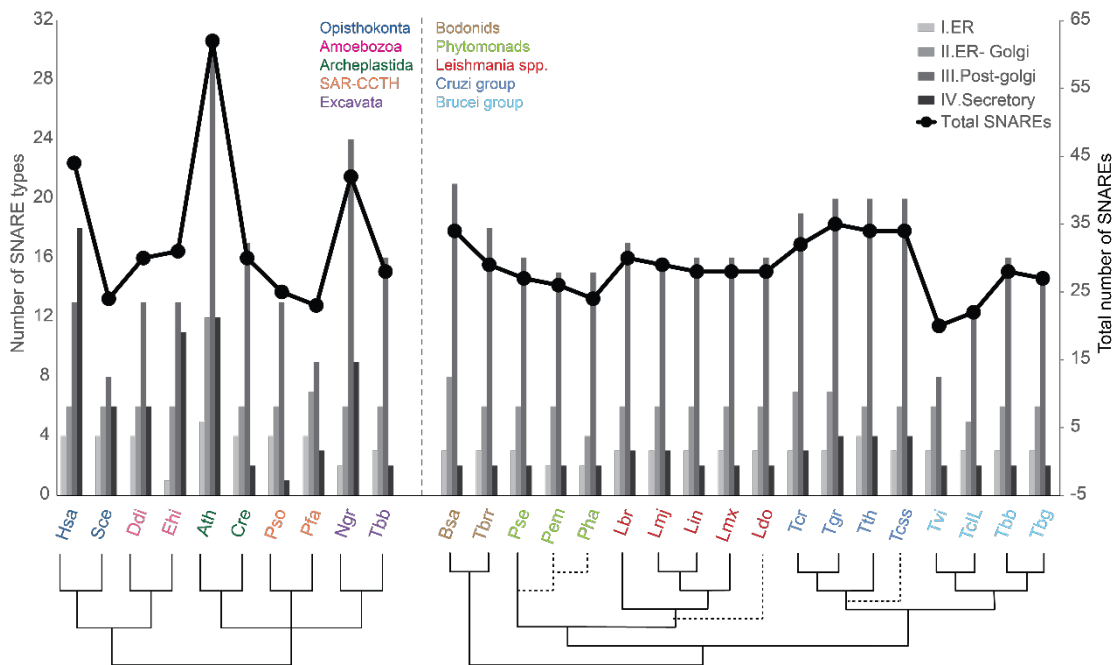
**Figure 4.19 Phylogenetic analysis of non-SNARE longins**

Best PhyML topology is presented. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test (1.0), PhyML Bootstrap (out of 100) and MrBayes posterior probabilities (%) and colour-coded as shown in the key. Eukaryotic representative sequence are colour-coded as Opisthokonta (Blue), Amoebozoa (Pink), Archeplastids (Green), SAR-CCTH (Orange) and Excavata (Red). All longin domain (IPR010908) containing proteins from selected eukaryotic species were collected and only those retaining the longin but no synaptobrevin (IPR001388) domain were used in this analysis. Two major clusters from which these proteins are derived can be discerned: Sec22-like and VAMP7-like.

#### 4.4.5 SNAREs by sub-cellular localisation

The kinetoplastid SNARE complement of the early secretory pathway and Golgi to ER retrograde trafficking were well conserved, in contrast to that associated with endocytic and post-Golgi trafficking, which is relatively diverged and expanded.

**Figure 4.20** below represents a quantification of SNAREs of different predicted sub-cellular locations across kinetoplastids and select eukaryotic representatives from all supergroups. SNAREs were classified according to the primary location they were associated with as Class I - ER, Class II - ER-Golgi, Class III - Post-Golgi/endosomal and Class IV - Secretory, as described in (Kloepper *et al.*, 2007). The total number of SNAREs (black symbols, y-axis right) across kinetoplastids was largely stable and comparable to the average eukaryotic repertoire. The smallest repertoires were found in the brucei group trypanosomes. The phytomonads did not show a drastically reduced SNARE repertoire in contrast to the TBC-RabGAP repertoires within these organisms.



**Figure 4.20** SNAREs by predicted sub-cellular location

A comparison of SNARE repertoires across kinetoplastids and representative eukaryotes by predicted sub-cellular localisation as per (Kloepper *et al.*, 2007). Genomes are arranged by phylogenetic relationships on X-axis. The five classically recognised, *sensu Adl 2004*, eukaryotic super groups and each sub-group of kinetoplastida are colour-coded according to the colour key on either side of the dividing dashed line respectively. Bars in grayscale represent number of individual SNARE subtypes, plotted on Y-axis (left). Colour of bars refers to predicted cellular location of action, coloured in according to the key. Black symbols and solid line represent the total number of SNAREs in each genome.

The total number of Class I and II SNAREs numbers were maintained across kinetoplastids (16/18 genomes) as well as representative eukaryotes (8/10 genomes). Most kinetoplastids possessed only a partial complement of the predicted Class IV secretory SNAREs. However, considering the patchy distribution of both Qbc and

canonical secretory brevin-like VAMPs across eukaryotes, there may be a need to reassess what SNAREs are labelled as ‘secretory’. While opisthokonts, amoebzoa and land plants have >five SNAREs (along with the occasional protist of particularly large genome size) in this category, algae, representatives of SAR-CCTH and kinetoplastids do not. Perhaps because the data set used to create this classification (as acknowledged) was biased against these less well-studied genomes (due to non-availability of sequenced genomes), it does not work when eukaryotes are considered as whole. For example, the R-SNARE IV brevin VAMPs found expanded in animals and land plants probably arose within those lineages, and hence longin-lacking VAMPs that arose elsewhere such as the kinetoplastid VAMP7C, were not phylogenetically classified as secretory SNAREs. Experimentally elucidating the mechanism of constitutive and stimulated exocytosis in broader eukaryotes is an important research area to pursue.

Class III SNAREs, which mediate post-Golgi trafficking, were expanded in all the genomes used in this analysis. Even those with only a modest expansion such as *S. cerevisiae* and *T. vivax* had a greater number of Class III SNAREs than any other. Only in *H. sapiens* (which have >12 Class III SNAREs) do the Class IV SNAREs (>16) outnumber Class III SNAREs. This large set reflects the complexity of these trafficking pathways in the ancient common ancestor of all extant eukaryotes.



## 4.5 Conclusions

Overall, the kinetoplastid phylogeny suggested a somewhat invariant set of SNAREs in this lineage. All SNARE sequences could be assigned to known individual eukaryotic orthologs. Some of these assignments were tentative and relied on evidence from BLAST (Syx8/Syp5, Syx6, and Vti proteins) but this is most likely a reflection of the divergence of the kinetoplastid lineage from other organisms used in this analysis. Furthermore, even though unusual domain variations were found, no truly novel kinetoplastid SNARE was found and distinct from Rabs and TBCs where at least one was found: Rab KSRX4 and TBC-ExA.

SNAREs mediating ER-Golgi and Golgi-ER were singly retained across kinetoplastids with few exceptions (loss of Use1, duplication of Gos1 in *T. congolense*). As in Rabs, majority of the losses, gains and divergence were seen in the post-Golgi trafficking pathways. Several post-Golgi SNARE genes likely expanded at the base of the kinetoplastid lineage were retained in a subset of extant taxa, the bodonids and the cruzi group (Qa SynPM2 and Qa1, Qb NpsnA2 and R VAMP7D), or more broadly across kinetoplastids (Qb-Vti-likeA/B and NpsnA/B, Qc-Bet1, Syp7A/B and Syx6-like1/2, and R-VAMP7A/B/C). On the other hand, within Syp7A and B there appeared to have been duplications at intermediate points in the evolution of kinetoplastids, within the bodonids and the cruzi group. The Qb, Qc, and R-SNAREs appeared to have expanded to a greater extent than the Qa-SNAREs at the base of the kinetoplastid lineage. It is therefore probable that kinetoplastid Qa SNAREs may be involved in the formation of multiple complexes at different stages of the pathway.

The level of sequence conservation of the SNARE domain, in particular the 0-layer residue, correlated with the level of conservation of each sub-family. Qa and R-SNAREs with near universal conservation of the 0-layer and the lowest level of variability of the SNARE domain residues show particularly well supported trees, whose representative sequences were assigned clearly to known eukaryotic subtypes. On the other hand, Qb and Qc SNAREs formed relatively less well-supported clusters, whose representative sequences appeared to be divergent from the landmark

eukaryotic sequences used in this analysis, and hence more difficult to assign confidently by phylogeny.

Several SNAREs with unconventional domain structures resulting from different combination of domains (domains N-terminal Habc or longin, the SNARE domain and the C-terminal *trans*-membrane domain) have been described across eukaryotes. Similar proteins appear to have emerged convergently in other lineages. Non-SNARE longins and VAMPs lacking N-terminal longin domain (VAMP7C) of independent origin were found in kinetoplastids as well as several other organisms (see **Figure 4.18** and **Figure 4.19**). Two Qa-SNARE sub-types (Syx16A and SynPM2) lacking the transmembrane domain have also been found. Phylogenetic analysis predicts that these proteins emerged before the radiation of the kinetoplastid lineages. They have not only been retained during the long evolutionary period up to the present, there have been further duplications and emergence of the TM-lacking SynPM proteins, particularly in the *Leishmania* genus. Whether these proteins form functional SNARE complexes has not yet been determined, but given their remarkable conservation, it would be interesting to investigate their role, if any, in experimentally tractable organisms like *Leishmania* and *T. brucei*.

In this study, the emphasis has been placed on kinetoplastid lineages in the comparative genomics survey of SNARE proteins along with representatives from several non-opisthokont taxa. This study provides further confirmatory evidence for the complex trafficking system in LECA, and the SNAREs associated with vesicles and organelles predicted to be present in it.

# 5 INTERACTOME ANALYSIS OF SNAREs IN *T. BRUCEI*

## 5.1 Introduction

SNARE proteins have been extensively studied in opisthokont model organisms and in some plant models but few others outside of these model systems. Some of the latter studies include localisation and functional studies in protists *P. tetraurelia* and *L. major*, which revealed a combination of conserved and divergent aspects. Qbc and Syb1-like proteins in *P. tetraurelia* show more extensive localisation in the cell rather than being restricted mainly to the cell membrane but may have a role in exocytosis (Schilde *et al.*, 2008, 2010), while Qa-SNAREs in *L. major* show both conserved and unusual localisations (Besteiro *et al.*, 2006). However, these studies do not test the identity of various binding partners and overall complex composition of these divergent SNAREs. In this chapter, after presenting a qPCR analysis of the expression of SNAREs in the two life cycle stages to confirm expression at mRNA level and consider stage-specific regulation, the results obtained from localisation and interactome analysis of SNARE proteins in *T. brucei* are presented.

## 5.2 Developmental regulation of SNAREs

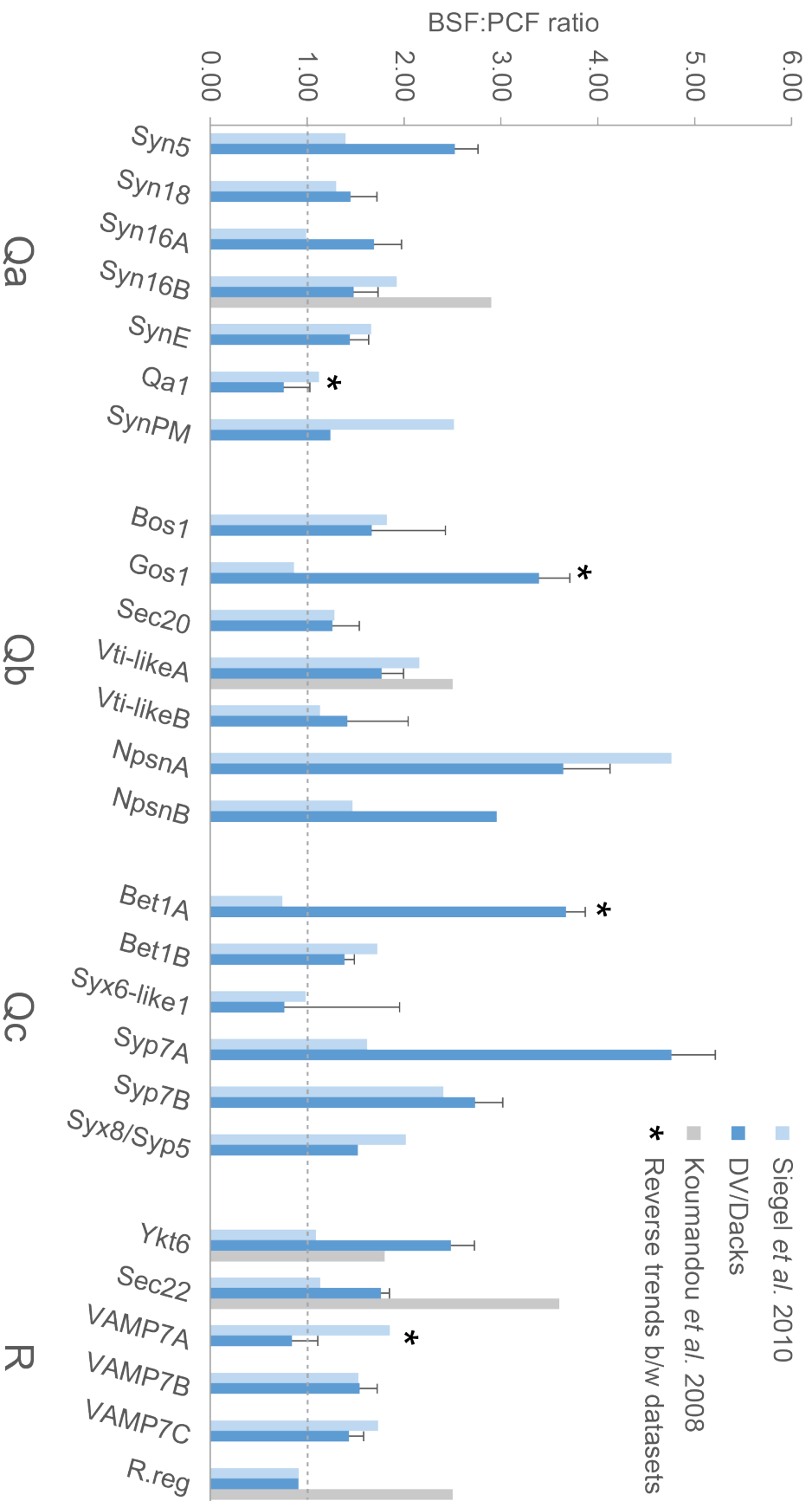
*T. brucei* has a complex life cycle. In the mammalian host, it invades multiple host tissue spaces including the bloodstream, lymph and the central nervous system. In the insect vector *T. brucei* also shows multiple stages of infection with clear changes in morphology and surface antigen expression. In the glucose and immune factors-rich bloodstream, the parasite is geared toward increased glycolytic activity, sequential expression of immunologically distinct VSGs, rapid endocytic recycling of cell surface proteins, manipulation of the host immune system and clearance of surface bound antibodies. In contrast, transition into the insect vector is accompanied by replacement of the VSG coat with procyclins, activation of the mitochondrion, alterations to cellular and organelle morphology and cell-cycle checkpoints, and a decrease in endocytic activity (Koumandou *et al.*, 2008). Such massive cellular remodelling is at least partially driven by changes to gene expression. Gene expression in *T. brucei* is poly-cistronic, with coding genes arranged head to tail in an array across the chromosome, all derived from a single transcriptional start site. Several mechanisms for the control of expression levels in response to differentiation signals have been postulated and even though all genes are constitutively transcribed, variations in mRNA levels are reported to accompany life-cycle changes (Jensen *et al.*, 2009).

The main mechanism of immune evasion in bloodstream form *T. brucei* (BSF) is antigenic variation combined with rapid endocytic recycling of surface proteins, mainly the VSG. Rapid endocytosis also aids in internalisation of host antibodies and targeting them to the lysosome to be degraded. In line with this, 6% of the trafficking-related transcriptome is found upregulated in the bloodstream form (Koumandou *et al.*, 2008). Previous high-throughput studies have indicated that some SNAREs maybe developmentally regulated; qRT-PCR was used to test these findings.

**Figure 5.1** below represents a chart of qPCR results along with data from previous studies (Koumandou *et al.*, 2008; Siegel *et al.*, 2010) where available. The ratio of BSF to PCF expression was plotted for each SNARE and SNAREs were arranged according to sub-groups Qa, Qb, Qc, and R. The overall picture in dataset from this study (dark blue) is consistent with what has been described before (light blue, gray): 22 of the 26 SNAREs that were analysed were found upregulated to

various degrees in the bloodstream form of the parasite. However, not all trends (i.e., up/down-regulation in a life cycle stage) were consistent between these datasets (marked with asterisk), which is likely due to variability in biological samples, and/or differences in normalisation bias between the two methods used. However, a large proportion (20/26, ~77%) showed similar trends. Of the 22 that were found upregulated in bloodstream form cells, eight showed >2 fold change. Considering the low abundance of SNARE proteins to start with, such a degree of change could have significant impact on protein levels.

There were no major differences in trends, i.e., whether they are up or down regulated in a certain life-stage between differently localised SNAREs (regardless of the extent of up or down regulation). In this dataset, most of the transcripts mediating ER-Golgi trafficking were found to be modestly upregulated in bloodstream forms (<2.5 fold) except for the Qb SNARE Gos1 and the Qc SNARE Bet1A (>3 fold). Surprisingly, while data from (Siegel *et al.*, 2010), who used high-throughput RNA sequencing to study stage-specific genome-wide steady-state mRNA levels, was largely consistent with this study, it differed particularly in these two genes Gos1 and Bet1A, which they found to be downregulated in bloodstream forms. It was not possible to reconcile these results. Similarly, most post-Golgi SNAREs were modestly upregulated in bloodstream form. However, it is again two Qb and two Qc SNAREs that show the highest levels of upregulation. These correspond to Qb- NpsnA, NpsnB and Qc- Syp7A, Syp7B. The plant counterparts of both these genes were shown to be involved in trafficking to the plasma membrane along with Qa SynPM and R VAMP7 (Zheng *et al.*, 2002; Suwastika *et al.*, 2008).



**Figure 5.1 Developmental regulation of *T. brucei* SNAREs**

SNAREs are arranged according to subfamily, Qa, Qb, Qc and R, on the X-axis, and ratio of BSF:PCF expression levels is plotted on the Y-axis. Dotted line (Ratio = 1.0) marks no change in expression level. Data from this study (dark blue), from Siegel *et. al.* 2010 (light blue), and Koumandou *et. al.* 2008 (grey) are presented. Single asterisks mark reverse trends between datasets.

Qa SynPM and R VAMP7s were also upregulated in bloodstream forms, but not to the degree as in the Qb and Qc SNAREs. All SNAREs were expressed but absolute quantification of the transcripts was not performed. Scaled counts of bloodstream form and procyclic form trypanosomes (PCF) were available from Siegel *et al.*. They revealed that the expression of SNARE proteins was relatively on the lower end of the scale when considering the proteome as a whole, ranging between 14.89 (Syp5/Syx8) to 154.28 (R.reg) in procyclic forms, and from 30 (Syp6/Syx8) to 217 (NpsnB) in bloodstream forms. In contrast, the scaled counts for histone protein was >4000 in both forms,  $\alpha$ -tubulin varied between 26000 (BSF) and 32000 (PCF) approximately, while VSG showed >51000 in BSF with negligible PCF expression (3.57). Within the bloodstream form cells the Golgi and post-Golgi SNAREs showed distinctly higher expression than SNAREs involved in the early secretory pathway, consistent with the high rates of endocytosis. This pattern was not seen in procyclic cells

### 5.3 Tagging and localisation of select SNAREs

SNAREs have been extensively studied in plant, metazoan and fungi model systems where their localisations and interactions were remarkably conserved despite low sequence similarity. There are considerably fewer experimental studies in other lineages. Kinetoplastids are evolutionarily distant from currently favoured model organisms in the study of SNAREs, and may possibly be close to the eukaryotic root and hence represent an independently evolving ancient lineage (Cavalier-Smith, 2010; Akiyoshi and Gull, 2014; He *et al.*, 2014). Phylogenetic analyses from Chapter 4 indicated that kinetoplastids possess a largely conserved set of SNARE proteins. However, certain sub-types such as the putative Syx6-like, Syx8-like, and Vti1-like were likely diverged; they tended to branch independently in eukaryotic SNARE phylogenies and could not be assigned based on them. Instead, they were assigned based on BLAST and reverse best hit BLAST. An analysis of selected SNARE proteins, determining location and interactions in *Trypanosoma brucei brucei*, was undertaken in an attempt to validate that predicted SNAREs in highly divergent taxa maintain their binding partners and overall complex composition. To achieve this, select SNAREs were first tagged and visualised in *T. brucei*.

### 5.3.1 Challenges in tagging *T. brucei* SNAREs

Several attempts were made to tag SNAREs *in-situ* using the pMOT tagging system for C-terminal tagging (Oberholzer *et al.*, 2006) and the pEnT6B system for N-terminal tagging (Kelly *et al.*, 2007). The pMOT system uses a one-step PCR amplification method to generate tagging constructs with ~80 nucleotide targeting sequences on each end for in-situ tagging. Five strategies were tested, as summarised in **Table 5.1**.

First, Qa-SNAREs were chosen because of the availability of localisation data for these SNAREs in *Leishmania major* and the extent of literature available for comparison, which indicated their importance in the regulation of SNARE-mediated fusion via their N-terminal domains. Both SMB and wild-type Lister 427 cell lines (Row 1 and 2 in **Table 5.1** below) were transfected with a 3xHA tag construct. SNAREs are small proteins so HA tags were chosen because of their small size, and their successful use in tagging SNAREs in mammalian systems (Gordon *et al.*, 2009). These however did not yield any transfectants.

Next, given the 1+3 configuration of trans-SNARE structures where R-SNAREs are found on one membrane and a trio of Q-SNAREs on the opposite membrane, R-SNAREs were prioritised for tagging instead of the Qa-SNAREs. This would enable a systematic and quicker elucidation of the specificity of SNARE interactions via pull downs further down the line. Transfections into the SMB cell line (by Dr Paul Manna, Row 3 in **Table 5.1**) yielded three transfectants for each of the R-SNAREs, which were confirmed by immunofluorescence. However, only one clone of VAMP7B-Tb427.05.3560 was detectable by western blotting.

Then, to simultaneously avoid the problems of obtaining a high efficiency of transfection in bloodstream forms and address the possibility that the C-terminal tag was interfering with the biology of the tagged protein, R-SNAREs were tagged at the N-terminal with a single HA tag using the pEnT6B construct (Row 4 in **Table 5.1**). Of the transfectants selected via antibiotic resistance, only one VAMP7C-Tb427.10.790 cell line was detectable by immunofluorescence and none was detectable by western blotting. Overall, these data suggested that there was a general issue with the addition of even a small epitope tag to trypanosome SNAREs, and an issue that was not encountered previously with several distinct classes of protein.



	SNAREs tagged	Cell line used	Tagging construct used	Number of Positives obtained	PCR	IFA evidence	Western blot evidence	Comments
1 (Qa)	Tb11.01.8000, Qa1	SMB	pMOT-3HA (C terminal)	0	n/a	n/a	n/a	Constructs were sequenced, all four experiments were repeated once
	Tb09.211.3920, Syx16B			0				
	Tb427.10.14200, Syx5			0				
	Tb11.12.0013, SynPM			0				
2 (Qa)	Tb11.01.8000, Qa1	BSF Lister 427	pMOT-3HA (C terminal)	0	n/a	-ve	n/a	Cell lines did not survive long enough to test for PCR and WB
	Tb09.211.3920, Syx16B			1				
	Tb427.10.14200, Syx5			0				
	Tb11.12.0013, SynPM			0				
3 (R)	Tb427.10.790, VAMP7C	SMB	pMOT-3HA (C terminal)	3	+ve for VAMP7B	+ve	+ve only for VAMP7B	Even the WB band for VAMP7B was very faint
	Tb427.5.3560, VAMP7B			3				
	Tb427.9.14080, Ykt6			3				
	Tb427.2.5120, VAMP7A			3				
4 (R)	Tb427.10.790, VAMP7C	PCF	pEnT6B-HA (N terminal)	3	+ve for VAMP7B	+ve only for VAMP7B	-ve for all	
	Tb427.5.3560, VAMP7B			3				
	Tb427.9.14080, Ykt6			3				
	Tb427.10.790, VAMP7C			4				
5 (R)	Tb427.5.3560, VAMP7B	PCF	pMOT-GFP (C-terminal)	2	not done	mislocalised - staining throughout cell	GFP-size bands (~27kDa)	Size of GFP is about the same as the size of SNAREs
	Tb427.2.5120, VAMP7A			1				
	Tb427tmp.01.7050, R.reg			1				

**Table 5.1 Summary attempts to tag TbsSNAREs *in-situ*.**

Details of cell line, tagging construct used for each experiment together with the results obtained are presented.

To overcome the problems of inefficient transfections in the bloodstream form trypanosomes and inconclusive immunofluorescence and blotting evidence, we then used procyclic form trypanosomes which reliably showed higher transfection efficiency with the pMOT system but with the GFP tag instead of the HA. Even though the GFP is a large tag, it has been used successfully to tag SNAREs in yeast (Gossing *et al.*, 2013). Furthermore, the cryoimmunoisolation experiments planned further along were optimised for the llama anti-GFP antibody so this could be useful. In this case (Row 5 of **Table 5.1** several positive transformants were obtained. However, when tested by IFA they did not show staining in specific locations in the cell, rather they stained punctate structures throughout the cytoplasm. When assessed by western blot, GFP sized bands were found but there were no bands at the expected size of SNARE-GFP fusion proteins. Once more, these data suggested a level of difficulty with tagging an endogenous copy of the SNARE proteins. Difficulty in endogenous tagging may be due to low abundance of SNARE proteins (see Section 5.2 above) making it difficult to visualise them once tagged. PCR amplification based pMOT construct, which uses 70-80 nucleotides for homologous recombination works well for many genes but not all, as observed for SNAREs and other genes in our laboratory. It is possible that longer homologous regions may be required for effective recombination *in-situ*. It is also possible that the tagged version of the gene is non-functional or even toxic even at endogenous levels of expression.

Therefore, over-expression constructs were then used to tag SNAREs as random/rRNA integration greatly increases chances of effective recombination. The caveat is that interpretation of data may be less reliable because of possible overloading of the trafficking pathway. It was also decided to use procyclic cells instead of bloodstream forms as this life form has higher transfection efficiencies. While the rate of endocytosis is significantly lower in procyclic form trypanosomes than the mammalian stage parasites, there is unlikely to be any major changes in the localisation and the specificity of SNARE interactions. As these are the aspects of SNARE biology of highest interest and because procyclics have considerable methodological advantages (see **Table 5.2** below), they were chosen instead of bloodstream forms for tagging. **Table 5.3** below shows a selected summary of attempts to tag SNAREs using overexpression vectors in procyclic cells. Repeated

attempts to tag R-SNAREs, particularly Sec22 and VAMP7A using the plew79 vector system that worked very well for VAMP7C were unsuccessful.

Therefore, there appears to be an intrinsic issue with the ability to tag this class of protein regardless of factors such as life-stage, *in-situ*, or ectopic expression, and endogenous or higher levels of expression.

	<b>Procyclic forms</b>	<b>Bloodstream forms</b>
<b>Doubling time</b>	12 hours	8 hours
<b>Tolerance to high cell numbers</b>	Yes (higher numbers can be grown in smaller volumes)	No
<b>Tolerance to low cell numbers</b>	No (too-sparse cultures will die out)	Yes (one cell can seed a culture)
<b>Transfection efficiency of pMOTtag vectors</b>	High	Low
<b>Suitability for cryo-pull down</b>	Yes	Difficult and untested
<b>Transfection time frame</b>	10 days-2 weeks	4 days-1 week

**Table 5.2 Methodological differences between *T. brucei* life-cycle stages**

Details of differences in the handling of the two proliferative life-stages of trypanosomes, namely procyclic forms and bloodstream forms, is presented.

SNARE	Cell line	Construct	Tag	Integration	Positives	IFA	WB
<b>VAMP7C</b>	<b>PTT</b>	<b>plew79</b>	<b>C-1xHA</b>	<b>rRNA</b>	<b>8</b>	<b>yes</b>	<b>yes</b>
VAMP7B	PTT	plew79	C-1xHA	rRNA	7	yes	yes
<b>VAMP7B</b>	<b>wt427PCF</b>	<b>plew79</b>	<b>C-1xHA</b>	<b>rRNA</b>	<b>1</b>	<b>yes</b>	<b>yes</b>
Sec22	wt427PCF/PTT	plew79	C-1xHA	rRNA	none	n/a	n/a
VAMP7A	wt427PCF/PTT	plew79	C-1xHA	rRNA	none	n/a	n/a
Tomosyn-like	wt427PCF/PTT	plew79	C-1xHA	rRNA	none	n/a	n/a
non-SNARE longin	n/a	plew79	C-1xHA	rRNA	n/a	n/a	n/a
<b>Ykt6</b>	<b>wt427PCF</b>	<b>pHD1034</b>	<b>N-2xHA</b>	<b>rRNA</b>	<b>10</b>	<b>yes</b>	<b>yes</b>
<b>VAMP7A</b>	<b>wt427PCF</b>	<b>pRPAOP</b>	<b>C-6xMyc</b>	<b>random</b>	<b>6</b>	<b>yes</b>	<b>yes</b>
<b>SynPM</b>	<b>wt427PCF</b>	<b>plew79</b>	<b>C-2xHA</b>	<b>rRNA</b>	<b>5</b>	<b>yes</b>	<b>yes</b>

**Table 5.3 Tagging of SNAREs in procyclic form trypanosomes**

Details of selected attempts to tag SNAREs using over-expression constructs are presented here. IFA – Immunofluorescence analysis; WB: Western blot. Successfully obtained tags are in bold.

### 5.3.2 *T. brucei* R-SNARE localisations are conserved

**Figure 5.2** below shows tagged VAMP7A, B, C, and Ykt6 cell lines by immunofluorescence and western blot. The molecular weight of SNAREs is generally around 25kDa. The estimated weight of TbVAMP7A is 25.3kDa, TbVAMP7B is 24.4kDa, TbVAMP7C is 26.5kDa and Ykt6 is 23.3kDa. As seen in Figure 5.2 B, the molecular weight of HA-tagged TbVAMP7B, C (1xHA, +1.2 kDa) and Ykt6 (2xHA, +2.4 kDa) are estimated by the western blot to be close to the 25kDa mark as expected. The 6xMyc tagged VAMP7A however, runs at a higher molecular weight (between 35 and 55 kDa) even though the expected weight of the tagged protein would be 32.5kDa [25.3 kDa (protein) + 1.2kDa(1xmyc) x 6]. It was not possible to determine the reason for this shift but it may include post-translational modification and incomplete denaturation. Importantly, probing an untagged wild-type sample with the same anti-myc antibody yielded a clear lane so this is likely to be correctly tagged.

Trypanosomes have a polarised trafficking route that is concentrated between the nucleus and the flagellar pocket and kinetoplast with all the major organelles lying in this region, with the exception of ER, which shows a reticular structure throughout

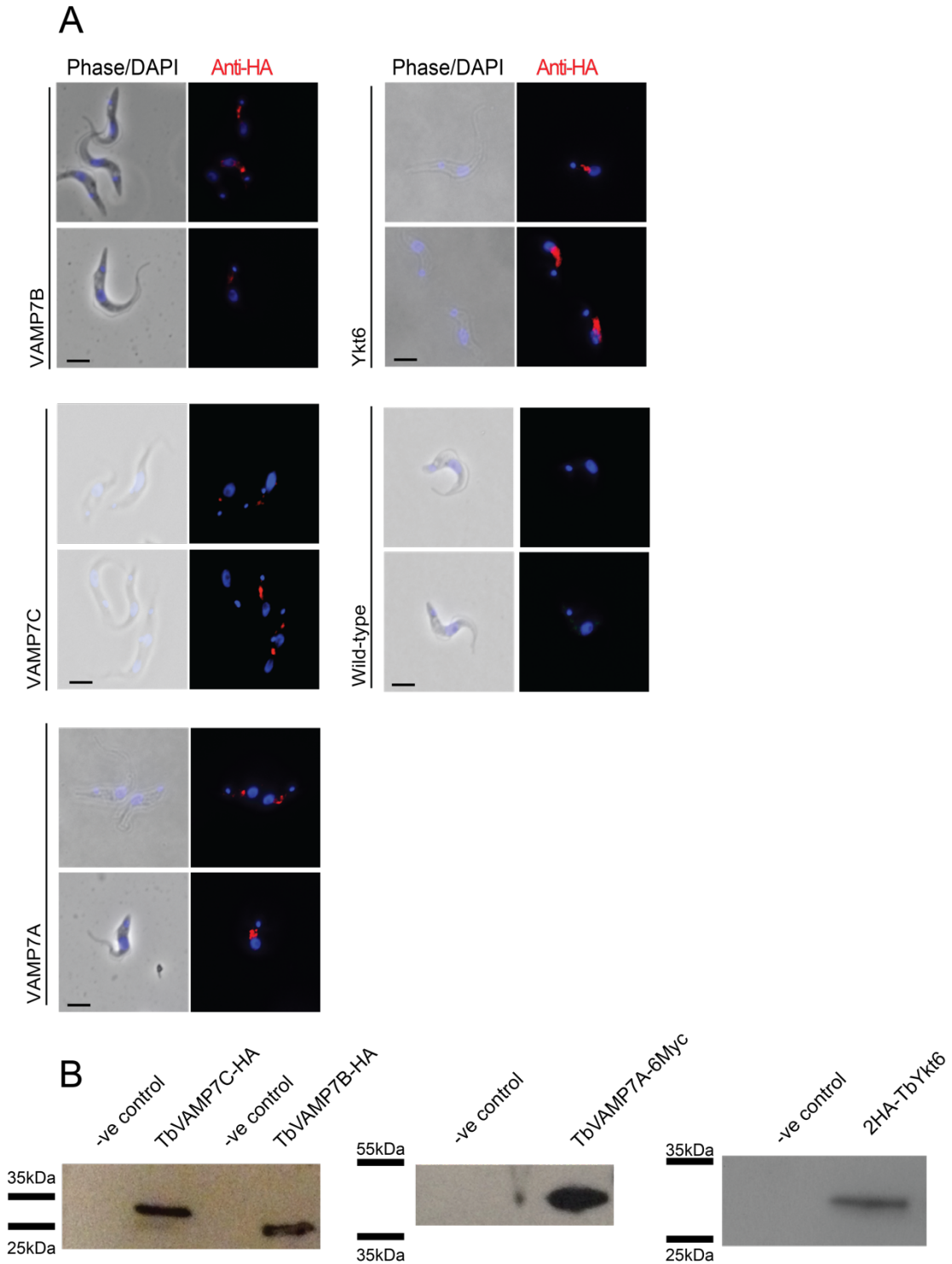
the cell body. Thus, the location of all the four R-SNAREs were within the general area where trafficking proteins are expected to be. Variation in the levels of expression were likely due to non-clonal starting population of selected cells. Further, since SNAREs are not only found at their organelle of action, but recycled post-fusion to enable repeated activity (Jahn and Scheller, 2006), there is likely to be some variation in their localisation, which is apparent in the figures that follow. Positive staining in tagged cell lines and absence of any signal in the wild-type negative control in both the IFA and western blots indicate that our proteins of interest are likely to be specifically tagged and expressed in the cell.

Localisation of tagged TbVAMP7A, B, C and Ykt6 (red) with reference to three cellular organelle markers: BiP for endoplasmic reticulum, GRASP for Golgi and p67 for the lysosome (green) are presented in **Figure 5.3**, **Figure 5.4**, **Figure 5.5**, and **Figure 5.6** below respectively. While VAMP7B and C showed no major overlap with the ER, VAMP7A showed partial co-localisation. VAMP7 co-localisation with ER has been reported in certain organisms such as *A. thaliana* (Uemura *et al.*, 2004) and animal intestinal cells (Siddiqi *et al.*, 2006) so TbVAMP7A may represent another instance of ER involvement of this protein. VAMP7A and C were juxtaposed to the lysosome with VAMP7A showing an occasional co-localisation, but with VAMP7B slightly more distant from the lysosome. Both VAMP7B and C were juxtaposed to the Golgi (colocalisation of VAMP7A with Golgi is not available) and this localisation will be discussed further in section 5.3.3. All three TbVAMP7s show slightly different localisations with respect to the lysosome – VAMP7A is found to be overlapping or at least partially overlapping with p67, VAMP7C is either juxtaposed or close to the lysosome while VAMP7B is also juxtaposed or completely distinct from the p67 staining. This partially differential localisations of the three proteins indicates that they may be involved in overlapping as well as distinct pathways.

Given the streamlined nature of the *T. brucei* trafficking system, it is likely that such juxtaposition with markers of the Golgi and the lysosome indicates that the VAMP7s may be endosomally located. Specific characterisation of the endosomal compartment was not possible due to unavailability of antibodies to endosomal marker proteins such as Rab5 (early endosome), Rab7 (late endosome) and Rab11 (recycling endosome). Endosomal localisation of VAMP7 would be consistent with previously reported data in yeast and metazoan lineages where they mediate transport

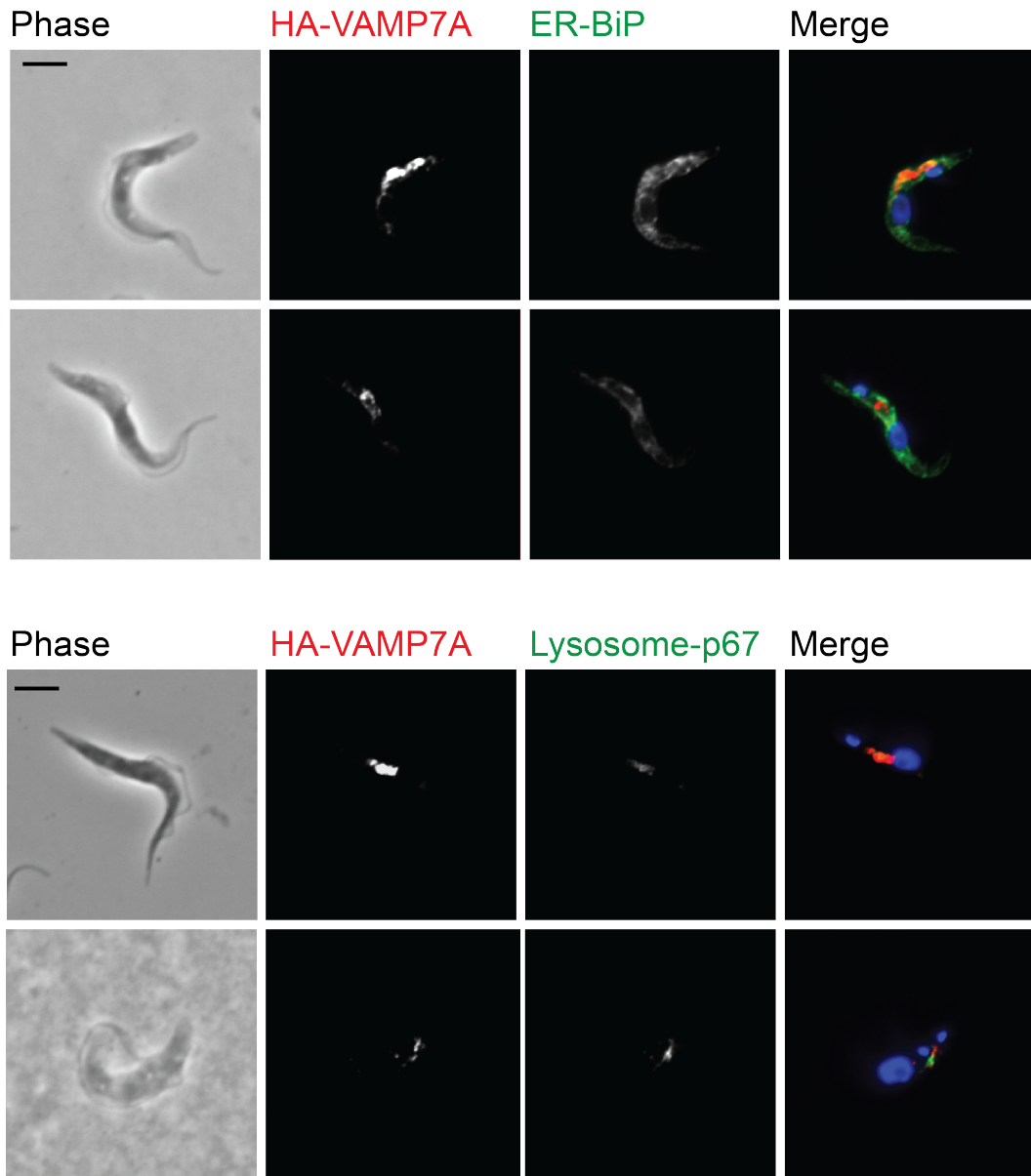
between the endosomes and lysosomes, and between endosomes and the plasma membrane (Burri and Lithgow, 2004; Jahn and Scheller, 2006). Colocalisation with above-mentioned Rabs would confirm specific endosomal location of these SNAREs.

The Golgi-SNARE Ykt6 showed a broader staining likely due to the higher level of expression of the N-terminal tagging vector used (pHD1034). The Ykt6 staining did overlap with the Golgi marker GRASP and also showed close association with the lysosome but less so with the BiP staining of the ER. This localisation is also consistent with the behaviour of the yeast and human orthologues of Ykt6. Further as mentioned above, the apparent differences



**Figure 5.2 Validation of tagged VAMP7A, B, C and Ykt6 in PCF**

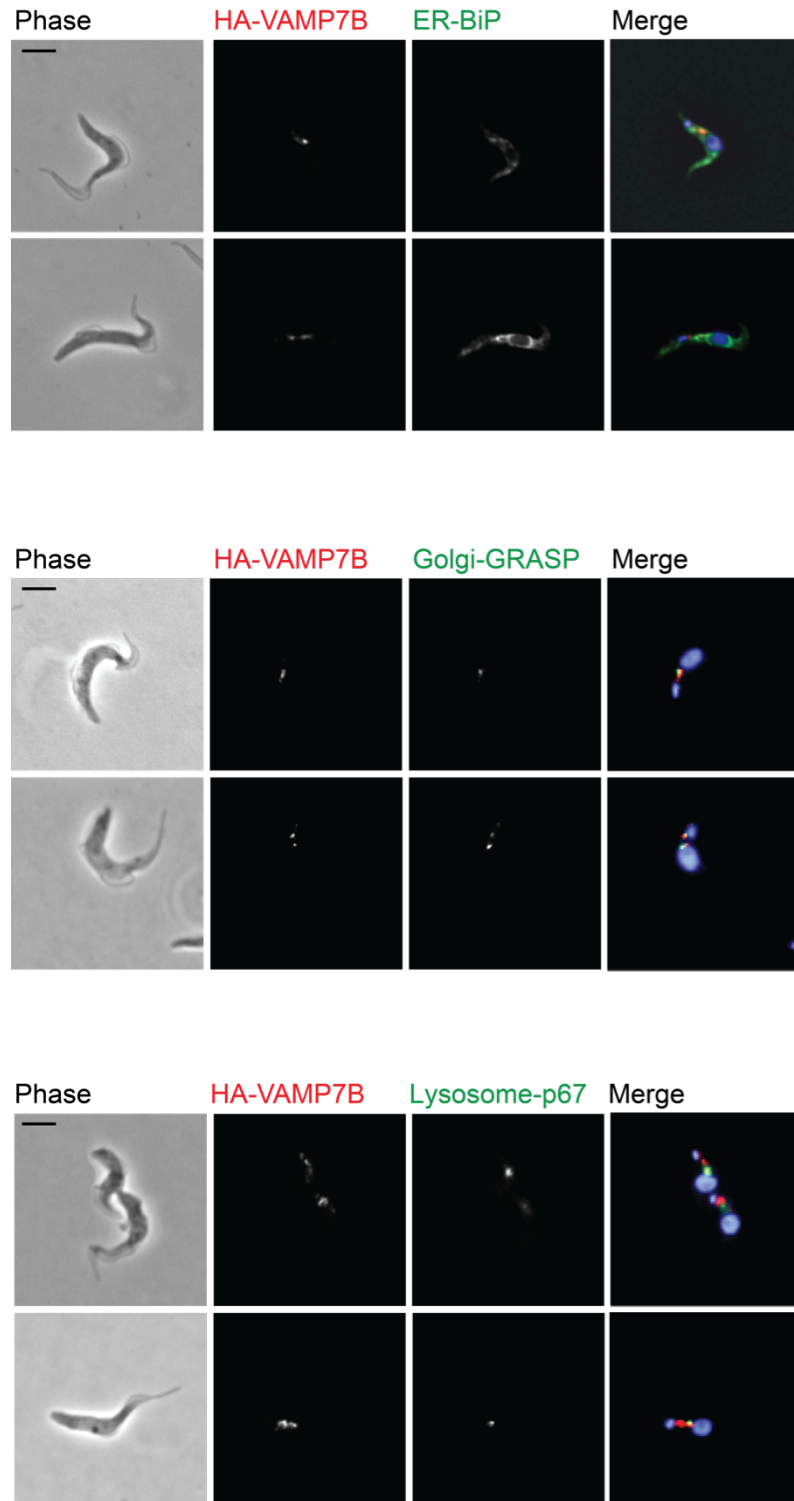
(A) HA-tagged VAMP7B (Tb427.5.3560), VAMP7C (Tb427.10.790) and Ykt6 (Tb427tmp.211.4610) and C-myc-tagged VAMP7A (Tb427.02.5120) in procyclic cells along with wild-type control. Localisation of each SNARE protein (red) is shown; nucleus and kinetoplast are stained blue with DAPI. Tagged proteins were visualised with rat anti-c-myc antibody (Santa Cruz, 9E10, 1:400) or rat anti-HA antibody (Roche, 1:1000) according to the tag. Scale bar is 2 $\mu$ m. (B) Western blot of tagged cell lines: VAMP7C and B, VAMP7A and Ykt6. Tagged proteins were visualised with rat anti-c-myc antibody (Santa Cruz, 9E10, 1:5000) or rat anti-HA antibody (Roche, 1:10000) according to the tag, untagged control cell-lines (wild-type) are shown as negative controls.



**Figure 5.3 Sub-cellular localisation of PCF R-SNARE TbVAMP7A**

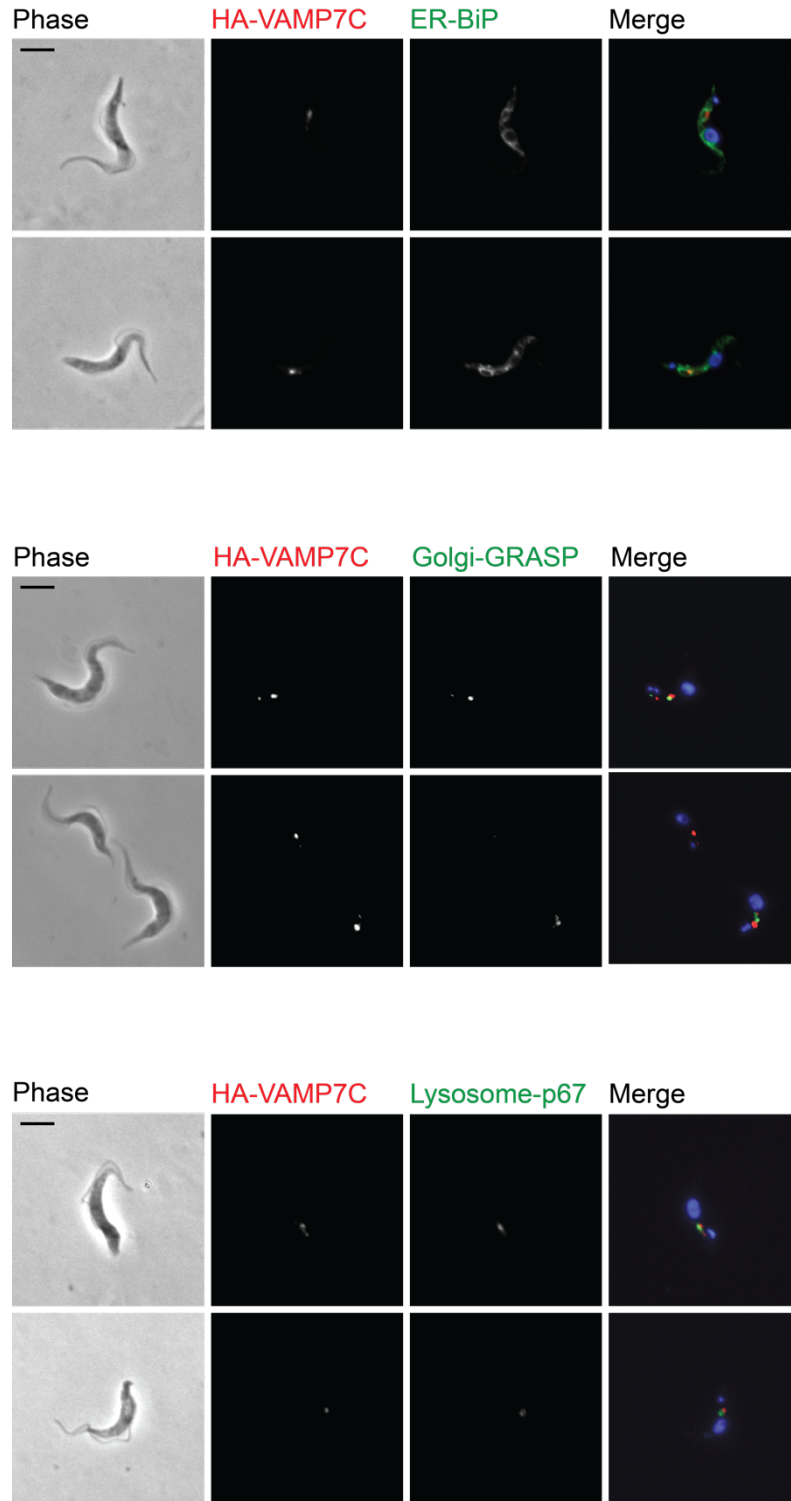
Localisation C-myc tagged TbVAMP7A (Tb427.02.5120) protein (red) is shown relative to markers (green) for endoplasmic reticulum (TbBiP) (top) and lysosome (p67) (bottom) in procyclic cells. Single stains are shown in original white for better contrast in this and all subsequent figures showing colocalisation. The nucleus and kinetoplast are stained with DAPI and appear in blue. Tagged protein was visualised with rat anti-c-myc antibody (Santa Cruz, 9E10, 1:400). Mouse anti-p67 and TbBiP (1:1000) were gift of James Bangs. Scale bar is 2 $\mu$ m.





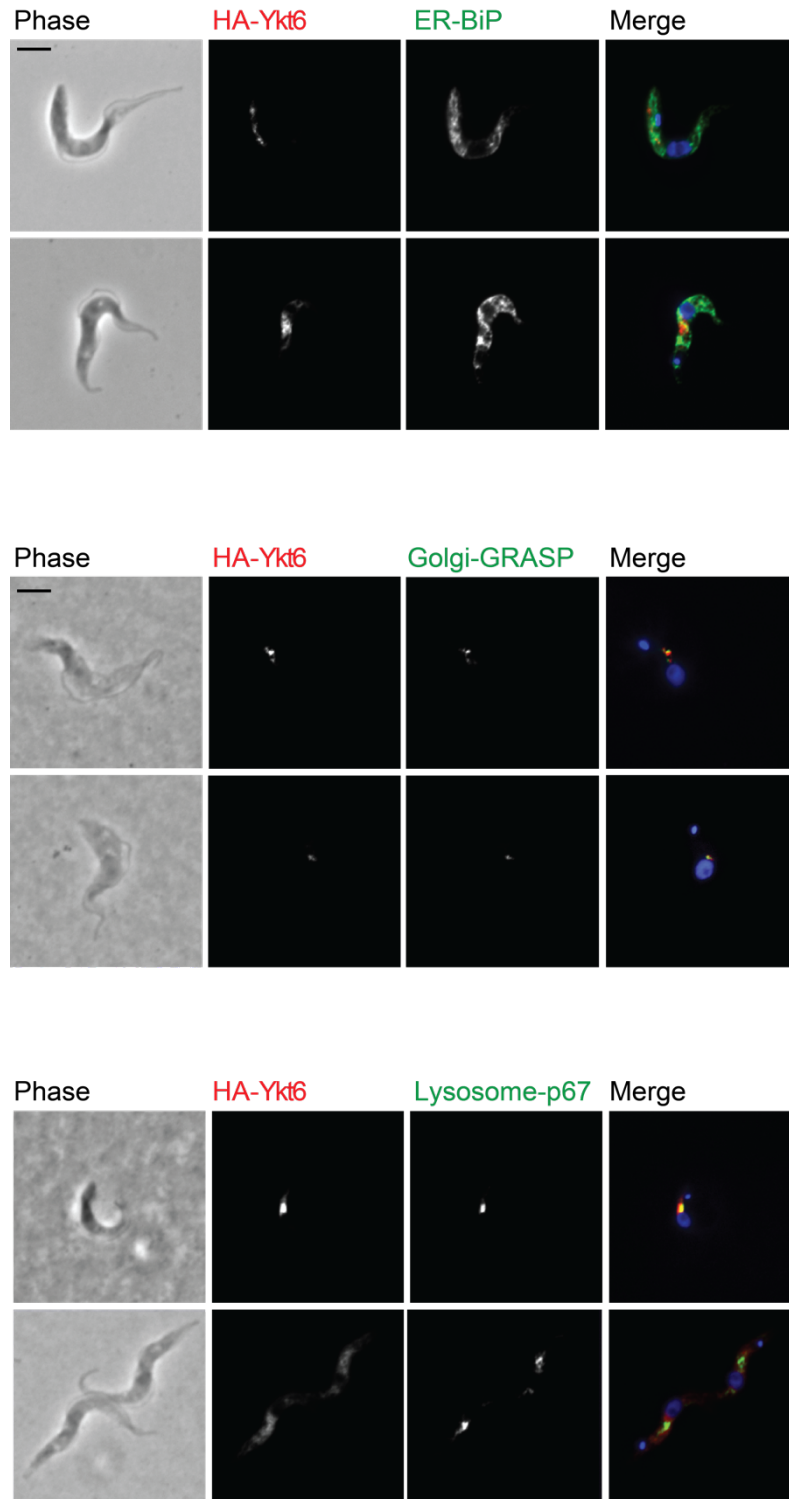
**Figure 5.4 Sub-cellular localisation of PCF R-SNARE TbVAMP7B**

Localisation of HA-tagged TbVAMP7B (Tb427.5.3560) protein (red) is shown relative to markers (green) for endoplasmic reticulum (TbBiP) (top), the Golgi complex (TbGRASP) (middle) and lysosome (p67) (bottom) in procyclic cells. The nucleus and kinetoplast are stained with DAPI and appear in blue. Tagged protein was visualised with rat anti-HA antibody (Roche, 1:1000). Mouse anti-p67 and TbBiP (1:1000) were gift of James Bangs and mouse anti-TbGRASP (1:1000) was gift from Graham Warren. Scale bar is 2 $\mu$ m.



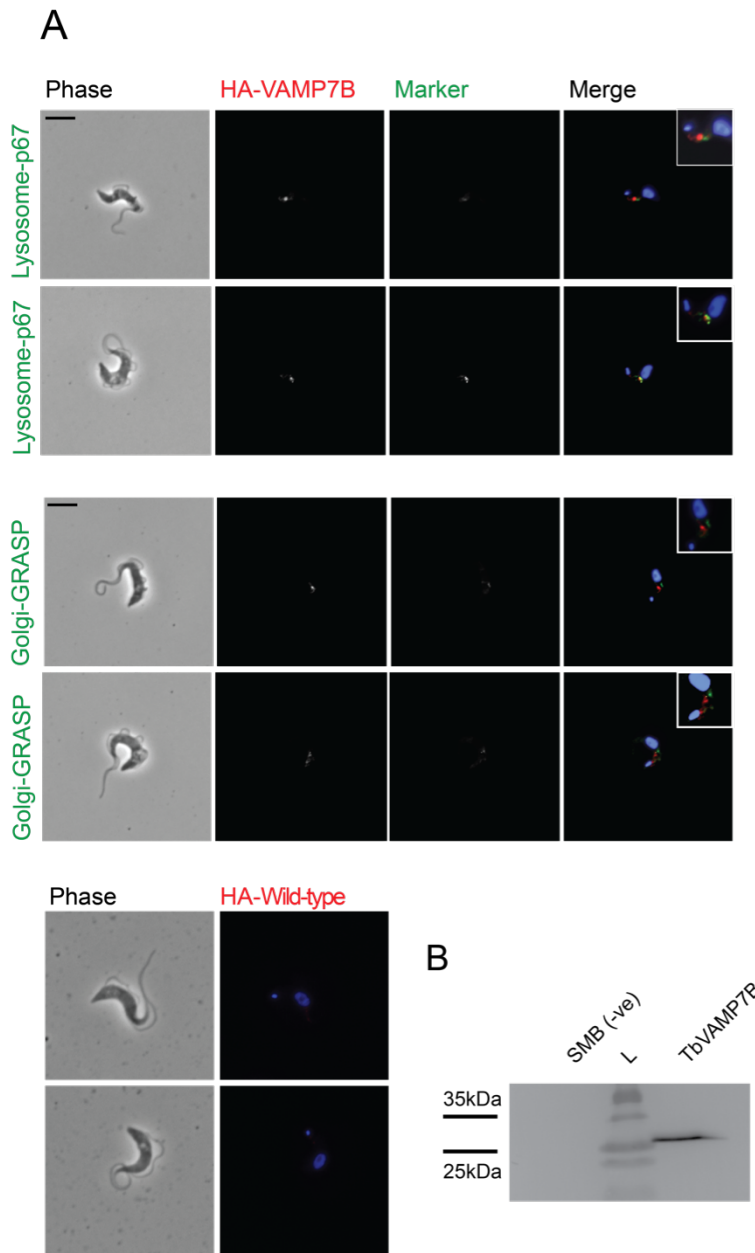
**Figure 5.5 Sub-cellular localisation of PCF R-SNARE TbVAMP7C**

Localisation of HA-tagged TbVAMP7C (Tb427.10.790) protein (red) is shown relative to markers (green) for endoplasmic reticulum (TbBiP) (top), the Golgi complex (TbGRASP) (middle) and lysosome (p67) (bottom) in procyclic cells. The nucleus and kinetoplast are stained with DAPI and appear in blue. Tagged protein was visualised with rat anti-HA antibody (Roche, 1:1000). Mouse anti-p67 and TbBiP (1:1000) were gift of James Bangs and mouse anti-TbGRASP (1:1000) was gift from Graham Warren. Scale bar is 2 $\mu$ m.



**Figure 5.6 Sub-cellular localisation of PCF R-SNARE TbYkt6**

Localisation of HA-tagged TbYkt6 (Tb427tmp.211.4610) protein (red) is shown relative to markers (green) for endoplasmic reticulum (TbBiP) (top), the Golgi complex (TbGRASP) (middle) and lysosome (p67) (bottom) in procyclic cells. The nucleus and kinetoplast are stained with DAPI and appear in blue. Tagged protein was visualised with rat anti-HA antibody (Roche, 1:1000). Mouse anti-p67 and TbBiP (1:1000) were gift of James Bangs and mouse anti-TbGRASP (1:1000) was gift from Graham Warren. Scale bar is 2 $\mu$ m.



**Figure 5.7 Validation of tagged R-SNARE TbVAMP7B in BSF**

(A) Localisation of HA-tagged TbVAMP7B (Tb427.5.3560) protein (red) is shown relative to markers (green) for the Golgi complex (TbGRASP) (top) and lysosome (p67) (bottom) in bloodstream form cells. The nucleus and kinetoplast are stained with DAPI and appear in blue. Tagged protein was visualised with rat anti-HA antibody (Roche, 1:1000). Mouse anti-p67 (1:1000) was gift of James Bangs and mouse anti-TbGRASP (1:1000) was gift from Graham Warren. A wild-type untagged cell line is also shown as negative control for anti-HA antibody. Scale bar is 2 $\mu$ m. (B) Western blot of tagged BSF TbVAMP7B cell line and negative control (SMB). Tagged protein was visualised with rat anti-HA antibody (Roche, 1:10000).

**Figure 5.7** above presents the immunofluorescence and western blot data for the VAMP7B tagged cell line raised in bloodstream form cells. The band for

VAMP7B is just above 25kDa as can be expected for a 24.4kDa protein tagged with 3xHA (+3.6 kDa) while the untagged cell-line shows an empty lane. As it is an endogenous tag, the level of expression of this protein (seen in red) in bloodstream form cell lines was markedly lower than in the procyclic cells. A magnified image of the region of interest is shown in an inset within the figures. It showed distinct puncta similar to TbVAMP7B staining in procyclic cells and no staining in the untagged cell-line. With the lysosomal marker, it showed localisation slightly distant from it as seen before in procyclic cells, but also occasionally showed closer juxtaposition and partial co-localisation. It was not possible to obtain good quality images showing co-localisation using the GRSAP antibody, New batches of antibody obtained failed to work as the previous batch and failed to generate signal in tagged or wild-type cell lines despite several attempts at troubleshooting (using a range of dilutions of antibody (1:2000, 1:1000 (recommended), 1:500, 1:100), trying a different method of cell fixation including methanol instead of paraformaldehyde, using new stocks of secondary anti-rabbit antibody and all other reagents involved). The antibody is likely to be faulty. Best images showed localisation a small distance from the Golgi as seen in procyclics, the significance of which is discussed in the next section.

### 5.3.3 VAMP7 is associated with Golgi duplication

In procyclic cells both TbVAMP7B and C appeared to be associated with the Golgi even during its duplication, with specific localisation near both daughter organelles. In fact, the size of the VAMP7 staining puncta corresponded to the size of juxtaposed Golgi (as stained by marker GRASP) (**Figure 5.4, 5.5, 5.7**). This association and staining pattern was analogous to that of Centrin2 during Golgi duplication (He *et al.*, 2005). *T. brucei* has a single Golgi, and a new copy of the Golgi was assembled a fixed distance away. Centrin2, which is a basal body protein, also stains a distinct bi-lobed structure, which is associated with the Golgi. During Golgi biogenesis one lobe of Centrin2 is associated with the old Golgi and the other with the new Golgi and at least partially with the ER-exit sites at which the genesis is reported to occur (Bangs, 2011). Staining of VAMP7 shows a punctate rather than an elongated lobe-like structure seen in Centrin2. By eye, the size of the VAMP7 staining region co-related with the size of the Golgi, it was found close to – growing larger as the new Golgi also grew in size. VAMP7 is generally not associated basal bodies or centrioles. There is only one study that implicates VAMP7 in ciliogenesis in

kidney cells, where the authors speculated that the mechanism could involve delivery of proteins either to the apical membrane close to the basal body, or directly to the periciliary base (Szalinski *et al.*, 2014). Given the varied mechanisms of Golgi duplication in different taxa, it is likely to involve different protein machineries. In *T. brucei*, the expansion of the VAMP7 complement, its pattern of localisation, together with the observation that some material used to build the new Golgi is derived from the old Golgi (He *et al.*, 2004), indicate a possible role of VAMP7 given the twinned size and number of puncta of staining observed. Although, it is not known how this process of flow of materials is mediated, it could be speculated that VAMP7 positive vesicles derived from the old Golgi may be involved in delivery of materials to the new Golgi. Distinct duplicating Golgi were not observed in tagged VAMP7B bloodstream form samples to compare the double staining pattern seen in the procyclic cells.

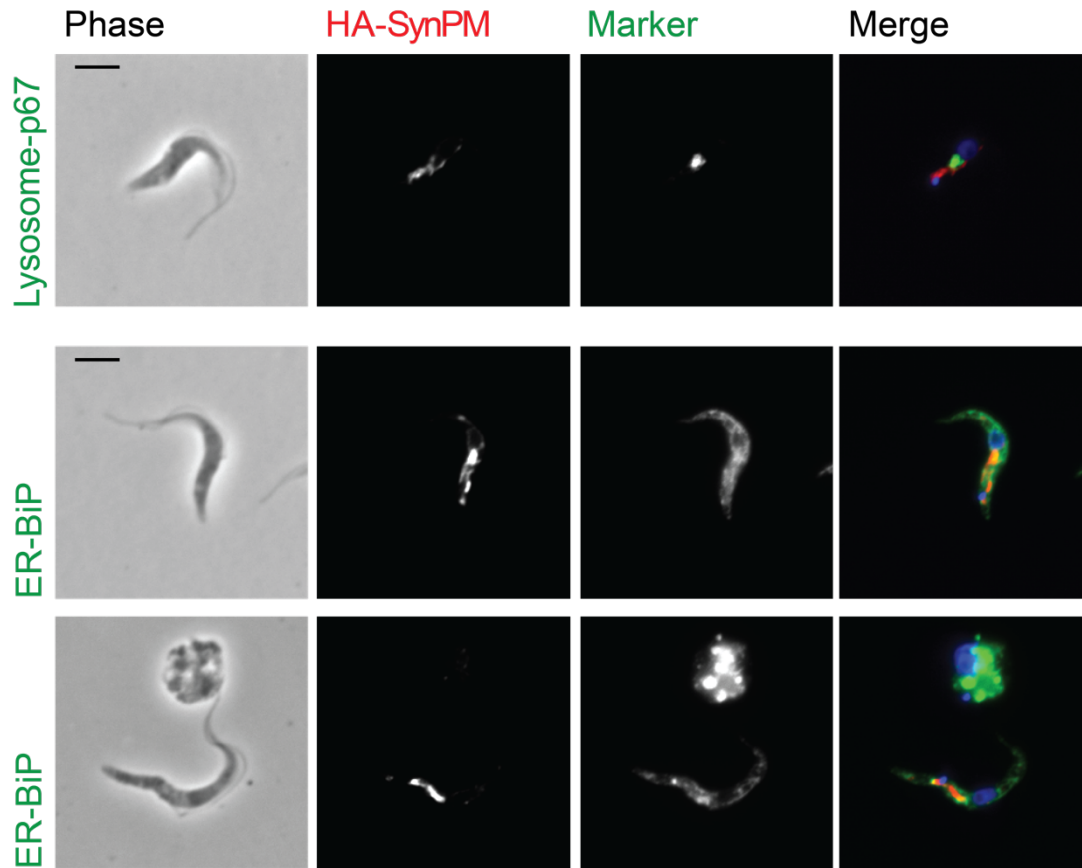
#### 5.3.4 Localisation of Qa-SNARE TbSynPM

In **Figure 5.8** below, the immunofluorescence and western blot of the SynPM-HA tag in procyclic cells is shown. TbSynPM has a low constitutive level of expression as seen by the absolute mRNA counts, and an even lower expression level in the procyclics as seen in the study by Siegel *et al.* and the relative expression analysis (**Figure 5.1** above). This is also the case in HeLa cells where the endogenous levels of expression were insufficient to determine detailed compartmental specificity (Martinez-Arca *et al.*, 2003). Attempts at *in-situ* tagging of this protein using the pMOT system also failed. Therefore, an overexpression system was used to tag SynPM with a single HA tag. TbSynPM was found broadly in the region between the nucleus and the kinetoplast. In *L. major* SynPM was similarly localised – close to but not quite overlapping with the flagellar pocket (Besteiro *et al.*, 2006). It is therefore unclear if this localisation is conserved within the kinetoplasts given that the tagged *L. major* SynPM showed a more distinct smaller area of staining than the TbSynPM construct. TbSynPM::HA also shows a partial overlapping with the ER, more so than the VAMPs as seen by eye – this was however not quantified. This partial ER co-localisation may be a feature as seen in plant Qc syntaxins that show dual localisation at the ER and plasma membrane (Suwastika *et al.*, 2008), but it could also be a fault, due to overexpression. Aberrant ER localisation of Syntaxin1 (mammalian SynPM)

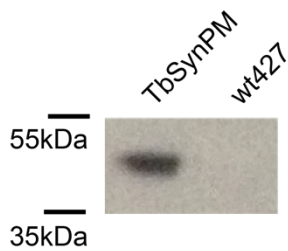
has been previously described in mammalian cell lines upon use of overexpression vector systems to tag the protein (Martinez-Arca *et al.*, 2003).

The inducible overexpression of TbVAMP7C in *T. brucei* did not cause any such problems and the construct was stably retained in culture and through freeze-thaw cycles. However, both inducible and constitutive overexpression of Ykt6, VAMP7A, B and SynPM in *T. brucei* caused a tendency to lose the tag. This is not a very unusual phenomenon in *T. brucei* when the tagged version interferes with the function of the protein and in this case, given how difficult it was to tag the proteins. Moreover an apparent increase in the level of staining of the ER marker BiP is seen in these cell lines but not in TbVAMP7C::HA cell lines. BiP is sensitive to misfolded proteins in the ER in mammalian cell lines where they have been shown to be upregulated and less mobile (available) under ER stress (Gülow *et al.*, 2002; Lai *et al.*, 2010). This indicates that the overexpression of these SNAREs may be responsible for ER stress and hence explain the difficulty in obtaining stable transformants. Despite this caveat, it appears that the overexpressed SNAREs do localise roughly in regions of the cell comparable to their yeast and human counterparts. However, exact localisation would require better markers as discussed above, as well as specific antibodies to *T. brucei* SNAREs and higher resolution microscopy so that untagged native proteins can be visualised. In absolute terms VAMP7B and C are among the most highly expressed SNAREs in procyclic *T. brucei* (data from (Siegel *et al.*, 2010)). Perhaps the unpredictability of rRNA insertion and its effect on final level of expression is a factor that may explain the different outcomes in the case of VAMP7C vs. VAMP7B. These attempts at tagging SNAREs in *T. brucei* suggest that stoichiometry of SNAREs is important for their interaction with each other and other proteins and any imbalance in the expression levels can have undesirable consequences. Alternatively, modification of the N and C termini of the proteins may affect their function, possibly rendering them dominant negative. Future studies will need more calibrated inducible expression systems.

A



B



**Figure 5.8 Validation of tagged Qa-SNARE TbSynPM in PCF**

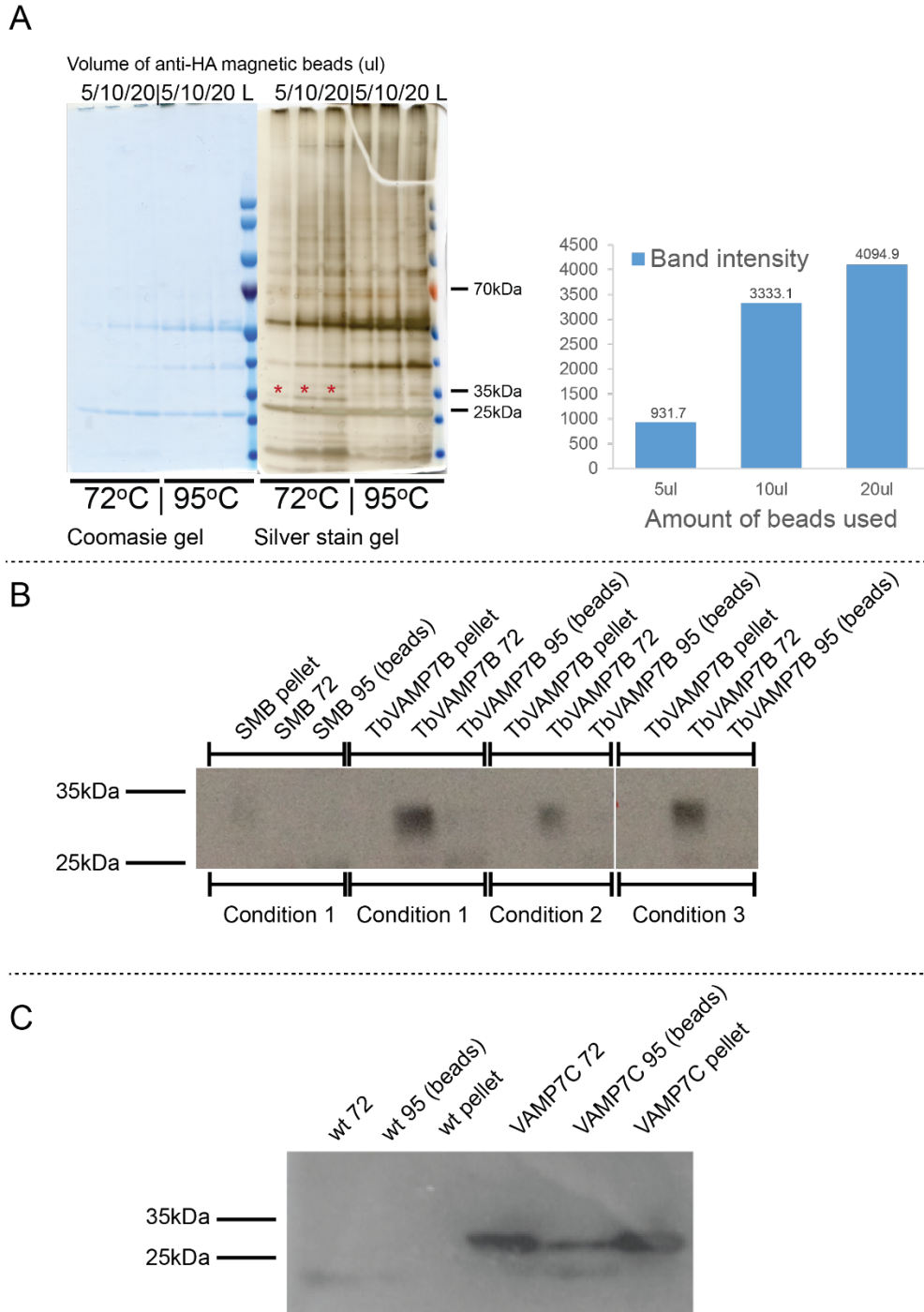
(A) Localisation of HA-tagged TbSynPM (Tb427tmp.12.0013) protein (red) is shown relative to markers (green) for the lysosome (p67) (top) and endoplasmic reticulum (BiP) (bottom) in procyclic cells. The nucleus and kinetoplast are stained with DAPI and appear in blue. Tagged protein was visualised with rat anti-HA antibody (Roche, 1:1000). Mouse anti-p67 (1:1000) was gift of James Bangs. Scale bar is 2 $\mu$ m. (B) Western blot of tagged PCF TbSynPM cell line and negative control (wt427). Tagged protein was visualised with rat anti-HA antibody (Roche, 1:10000).



## 5.4 Proteomics

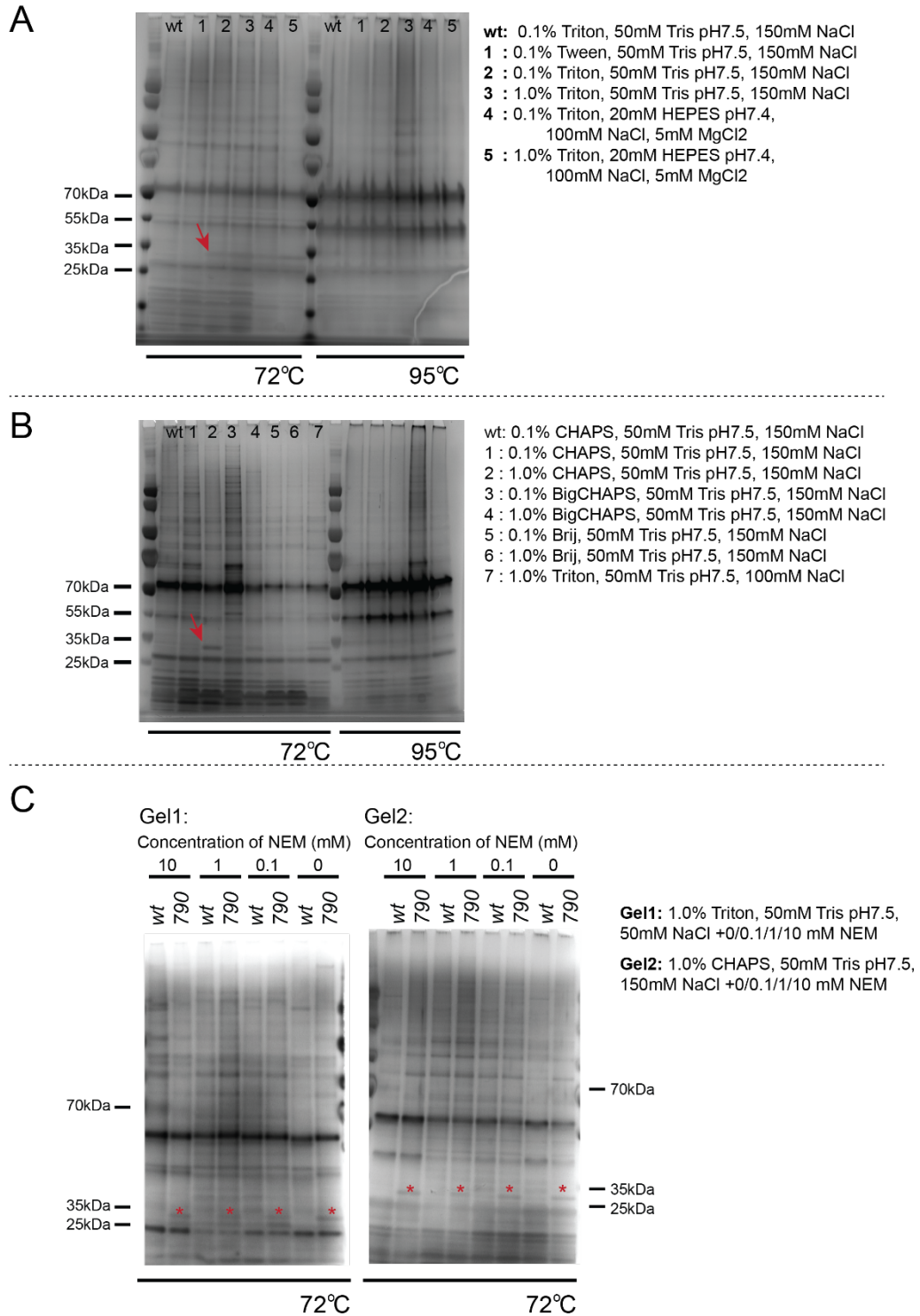
To examine the interactions of SNAREs, immunoisolation of epitope-tagged cryomilled parasites was performed, followed by mass spectrometry analysis. Several buffer conditions with different detergents, salts, crosslinking agents (DSP - dithiobis succinimidyl propionate) and NEM (N-ethyl maleimide) of varying concentrations were tested to optimize the extraction of the bait protein into solution while conserving protein-protein interactions

Pierce Anti-HA magnetic beads were used to isolate complexes associated with HA-tagged protein. First, the amount of beads needed to use in the samples was optimised. The result from this experiment is presented in panel A of **Figure 5.9** below. Lanes marked 72°C have samples from the first extraction in which at the end of the immunosolation, the washed beads are incubated at 72°C in elution buffer to extract all attached proteins from the beads. The 95°C lanes represent extraction of remnant beads in further elution buffer along with 0.1M DTT at 95°C. The latter is therefore a control to check the efficiency of the first extraction at 72°C. The extraction of the bait protein (VAMP7C::HA) was checked using 5, 10 and 20µl of the magnetic anti-HA beads. No relevant protein was detected in the Coomassie stained gel but a putative VAMP7C band (marked with a red asterisk) was found after silver staining. On quantification of band intensity using Image J software, the highest intensity was found in the sample with 20µl beads. Since the increase from 10µl to 20µl (761 units) was not as large as that seen from 5 to 10 µl (2400 units), an average of 15µl beads per sample was decided upon. The presence of the bait was confirmed by western blotting of the samples (shown in **Figure 5.9** panels B and C for TbVAMP7B and C respectively). SMB and wt are untagged cell-lines used as negative controls for BSF and PCF respectively.



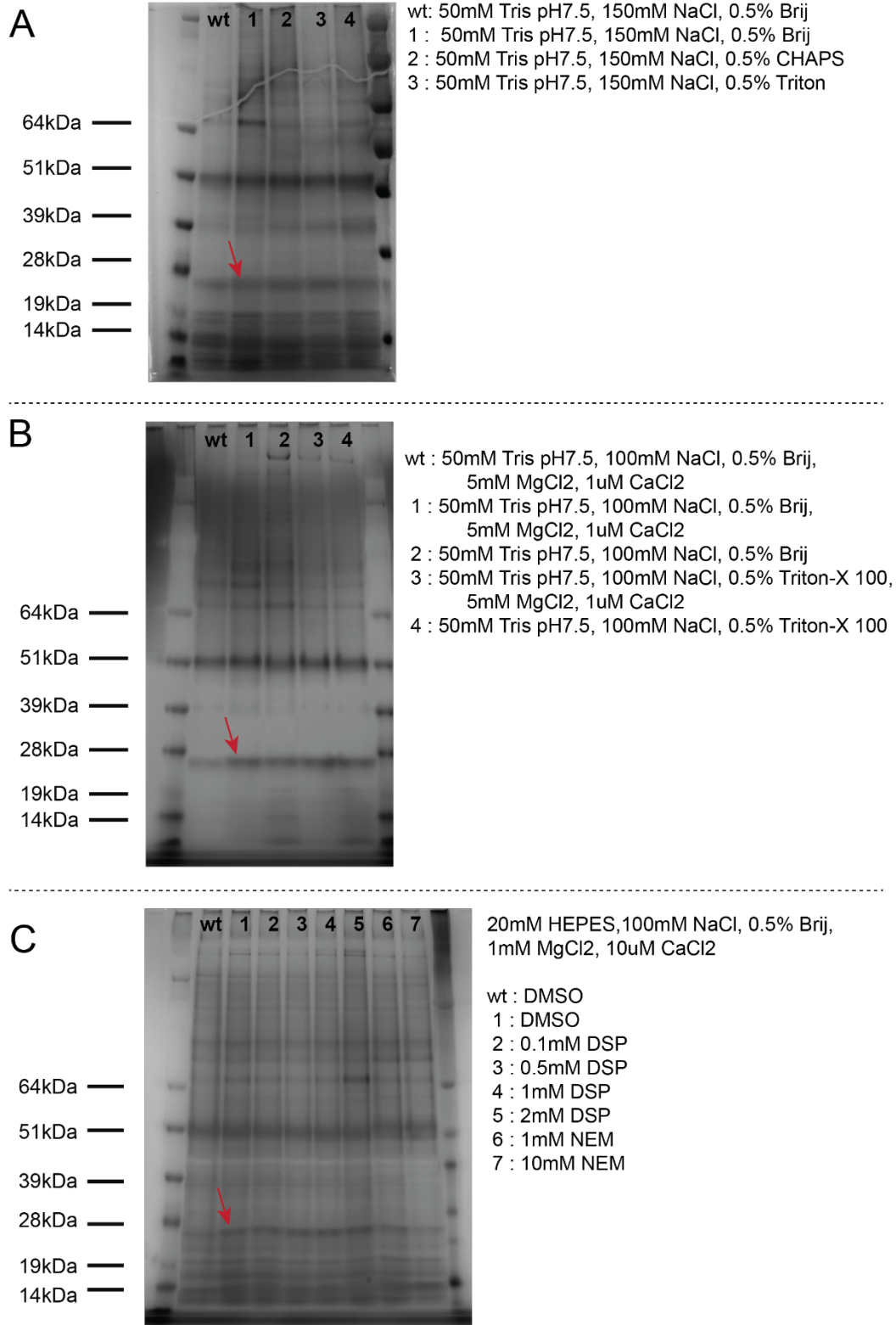
**Figure 5.9 Optimising immunoisolation of tagged SNAREs**

(A) Determination of the optimal amount of Pierce Anti-HA magnetic beads. Comparison of 5, 10, and 20 ul beads in experiment. Coomassie and silver stained gels (left) showing post-immunoprecipitation eluates extracted at 72°C (all attached proteins) and 95°C (remnants, reduced by DTT). Red asterisks indicate putative bait protein (TbVAMP7C) band, whose intensities are quantified in the bar graph (right). (B) Western blot, confirming presence of bait protein (TbVAMP7B) in the 72°C eluates in different buffer conditions (described in FigX). Samples of cytoskeletal pellet, 72°C and 95°C eluates of each condition are presented. (C) Western blot, confirming presence of bait protein (TbVAMP7C) in the 72°C eluate.



**Figure 5.10 Testing buffer conditions for extraction of TbVAMP7C::HA**

(A) and (B) Silver-stained gels showing 72 and 95°C eluates from immunoprecipitation experiments using different extraction buffers with varying buffering agents, detergents, detergent and salt concentrations. Red arrows point to bait protein (TbVAMP7C) bands (C) Silver stained gels showing 72°C eluates from immunoprecipitation experiments using extraction buffers with varying amounts of N-ethyl maleimide (NEM). Red asterisks indicate bait protein (TbVAMP7C) bands.



**Figure 5.11 Testing buffer conditions for extraction of 2HA::TbYkt6**

Silver stained gels showing 72 and 95°C eluates from immunoprecipitation experiments using different extraction buffers with varying (A) detergents (B) MgCl<sub>2</sub>, CaCl<sub>2</sub> concentrations and (C) crosslinking agent dithiobis succinimidyl propionate (DSP) and N-ethyl maleimide (NEM). Red arrows point to putative bait protein (TbYkt6) bands.

Silver-stained gels presented in **Figure 5.10A**, B and C above show the various buffer conditions that were used to extract TbVAMP7C::HA complexes. A condition that was previously used to extract SNARE complexes (Willett *et al.*, 2013) was first tried, along with varying concentrations of detergent. In panel B, the various detergents used at 0.1 and 1% concentration are shown. NEM or N-ethylmaleimide is an organic compound, which is reactive towards thiols (C-SH, R-SH groups) and can therefore modify cysteine residues in proteins and peptides. It can block vesicular trafficking by deactivating NSF (NEM sensitive factor), which is an AAA-ATPase that together with  $\alpha$ -SNAP disassembles post-fusion cis-SNARE complexes to enable the constituent SNAREs to mediate another round of fusion. Since NSF is sensitive to NEM, NEM blocks this disassembly and SNAREs are retained in their cis-quarternary complex form. The results from the addition of NEM to solubilisation buffers is shown in panel C.

It is apparent that beyond extraction of the bait protein, no other bands that could represent interacting SNAREs or other proteins were found in these analyses. Bait extracts appear indistinguishable from control untagged wild-type protein extracts except for the bait protein. This is in contrast to previous work in the lab where GFP-fused proteins such as nuclear pore complexes (NPCs) and clathrin elicited distinct putative interacting products visible by eye on the gel. The poor results in these experiments could be a result of various factors. The method was optimised for specific anti-GFP llama antibodies coupled to dynabeads, so it may take further optimisation to obtain better results for anti-HA magnetic beads used in this study. Moreover, while clathrin and NPC proteins are large abundant proteins, SNAREs are small proteins that are expressed at a relatively low level in the cell. Even though the inducible cell line produces detectable amounts of the tagged SNARE, general low-level expression of SNAREs may imply that the interacting partners are also expressed at a relatively low-level, and thus are undetectable by eye on a gel. Excess production of SNAREs, even as in the case of VAMP7C::HA that does not cause any apparent defects or overcrowding of the pathway, may well be interrupting interactions at the molecular level. Therefore two buffer conditions were chosen, one of which has been successfully used to extract SNAREs and interactors before, and another one with a different detergent. Four buffer conditions were tested in total by mass-spectrometry (See Chapter 2 for details).

**Figure 5.11** above shows three of the gels from the analysis of the 2HA::TbYkt6 cell line. Here buffer conditions with and without Calcium and Magnesium salts (panel B) as well as NEM, and the cross-linking agent DSP (panel C) are shown. Again, no band other than the putative tagged Ykt6 was recognisable in the gel. The wt gel lane also shows a band close to the 25kDa, but western blot analysis showed that this is likely not the same as the 25kDa band was only present in the tagged sample while the wt lane was clear. It is possible that it may be the light antibody chain whose molecular weight is also close to 25kDa. The silver staining and western blot analysis of gels from other cell lines also gave similar results to TbVAMP7C::HA and 2HA::TbYkt6. Positive results were obtained from the mass spectrometry analysis of TbVAMP7C::HA cell line despite not being able to discern bands other than the putative tagged bait in experimental samples vs. the wild type control in the gel. Hence, similar buffers were used to extract 2HA::TbYkt6, TbSynPM::HA, TbVAMP7B::HA in procyclic cell lines and TbVAMP7B::3HA in the bloodstream form cell line. While several putative interactors of the bait protein in the TbVAMP7C procyclic cell line and the TbVAMP7B tagged bloodstream form cell line were found, nothing was found in the mass-spectrometry analysis of any of the other cell lines (procyclic TbVAMP7B::HA, 2HA::TbYkt6, and TbSynPM::HA). The results for the former are discussed below.

#### 5.4.1 SNARE interactions of TbVAMP7C

Peak lists from the mass spectrometry experiments were submitted to ProFound and searched against an in-house *T. brucei* database using data from GeneDB. These putative interactors of TbVAMP7C were compared with the results of the same experiment (anti-HA immunoprecipitation) with untagged wild-type sample. Those proteins found to be among putative interactors of both TbVAMP7C and untagged sample were excluded from the list of putative TbVAMP7C interactors as non-specific. The list of proteins was sorted according to decreasing emPAI (exponentially modified protein abundance index) scores. The following were excluded step by step: hits with <2 significant peptide matches, common high-abundance proteins that were regularly observed as contaminants in eukaryotic proteomic datasets (Trinkle-Mulcahy *et al.*, 2008). Proteins described as hypothetical were checked for latest assignments if any on TritypDB, and associated GO terms to determine putative function, and putative domains present on the NCBI conserved

domain search (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Again, the following were excluded: regularly observed contaminants (as above), and nuclear localised proteins. This protocol was followed for each replicate experiment. The remaining proteins in the cleaned-up list were divided into a top-hits list and a second-tier list. As emPAI scores between experiments were not comparable, the lower limit for the top-hits list was set by replicability (100%) across different buffer conditions and repeat experiments, and corroboration i.e co-localisation by immunofluorescence, found within the top 50 proteins in the cleaned up dataset. Putative connection of second-tier list proteins to VAMP7 action and location specifically and membrane trafficking in general was examined in the literature, and only selected proteins of particular relevance are presented in this chapter. As an example, the top 40 proteins in peak lists from wild-type, TbVAMP7C are presented below, along with the TbVAMP7C list before and after removal of proteins common with the wild-type. (**Table 5.4**, **Table 5.5** and **Table 5.6** respectively).

Accession	emPAI	Description
CON_P04264	18.28	KRT1 Keratin, type II cytoskeletal 1
GeneDB Tb09.211.0120	15.65	nascent polypeptide associated complex subunit, putative
GeneDB Tb11.01.3675	11.74	40S ribosomal protein S17, putative
GeneDB Tb427.06.4280	9.13	glyceraldehyde 3-phosphate dehydrogenase, glycosomal
GeneDB Tb427.07.2340	6.65	40S ribosomal protein S15, putative
CON_P35527	6.48	KRT9 Keratin, type I cytoskeletal 9
GeneDB Tb427.10.5330	6.47	40S ribosomal protein S18, putative
GeneDB Tb427.01.2330	5.92	beta tubulin
GeneDB Tb927.2.4210	5.13	glycosomal phosphoenolpyruvate carboxykinase
GeneDB Tb427.08.2290	4.88	hypothetical protein, conserved
CON_P35908	4.04	KRT2 Keratin, type II cytoskeletal 2 epidermal
CON_P13645	3.88	KRT10 Keratin, type I cytoskeletal 10
GeneDB Tb427tmp.02.1085	3.8	40S ribosomal protein S4, putative
GeneDB Tb927.8.2290	3.56	hypothetical protein, conserved
GeneDB Tb427.10.2100	3.38	elongation factor 1-alpha
GeneDB Tb427.10.6630	3.35	ATP-dependent DEAD/H RNA helicase HEL64, putative
CON_P02533	3.35	KRT14 Keratin, type I cytoskeletal 14
GeneDB Tb11.03.0530	3.24	hypothetical protein SCD6.10
GeneDB Tb427.10.3930	3.15	40S ribosomal protein S3A, putative
GeneDB Tb427.01.2340	3.02	alpha tubulin
GeneDB Tb427.10.5620	3.01	fructose-bisphosphate aldolase, glycosomal
GeneDB Tb11.03.0800	2.95	hypothetical protein, conserved
GeneDB Tb11.02.1085	2.95	40S ribosomal protein S4, putative
GeneDB Tb11.01.1790	2.93	60S ribosomal protein L29, putative
GeneDB Tb427tmp.02.4150	2.92	pyruvate phosphate dikinase
GeneDB Tb427.05.4170	2.66	histone H4, putative
CON_P02538	2.55	KRT6A Keratin, type II cytoskeletal 6A
CON_P04259	2.55	KRT6B Keratin, type II cytoskeletal 6B
GeneDB Tb427.08.3530	2.54	glycerol-3-phosphate dehydrogenase [NAD <sup>+</sup> ], glycosomal
GeneDB Tb11.02.2430	2.42	60S ribosomal protein L17, putative
GeneDB Tb09.211.3550	2.42	glycerol kinase, glycosomal
GeneDB Tb11.01.3110	2.23	heat shock protein 70
GeneDB Tb427tmp.22.0012	2.09	eukaryotic peptide chain release factor subunit 1, putative
CON_P08779	2.07	KRT16 Keratin, type I cytoskeletal 16
GeneDB Tb427.05.1080	2.01	RNA-binding protein, putative
GeneDB Tb427.10.11660	1.98	hypothetical protein, conserved
GeneDB Tb427tmp.01.5570	1.85	RNA-binding protein
CON_P13647	1.79	KRT5 Keratin, type II cytoskeletal 5
GeneDB Tb427.02.5910	1.77	40S ribosomal protein S13, putative

**Table 5.4 List of putative interacting proteins - negative control**

The 40 highest ranking proteins predicted to interact with the bait protein are presented. In this case, an untagged wild-type cell-line was used and serves as negative control showing unspecific binding to antibody (magnetic anti-HA beads). The GeneDB accession, emPAI (exponentially modified protein abundance index) score as well as the GeneDB predicted protein name/description are shown.



Accession	emPAI	Description
<a href="#">GeneDB Tb427.10.790</a>	19.73	<b>vesicle-associated membrane protein, putative (Handle)</b>
CON_P04264	11.14	KRT1 Keratin, type II cytoskeletal 1
GeneDB Tb09.211.0120	9.68	nascent polypeptide associated complex subunit, putative
GeneDB Tb427.10.2340	9.12	hypothetical protein, conserved
GeneDB Tb11.01.3675	7.86	40S ribosomal protein S17, putative
GeneDB Tb427.10.5330	7.83	40S ribosomal protein S18, putative
GeneDB Tb427.01.2330	7.82	beta tubulin
GeneDB Tb427.06.4280	6.44	glyceraldehyde 3-phosphate dehydrogenase, glycosomal
GeneDB Tb427.07.2340	5.46	40S ribosomal protein S15, putative
CON_P35527	5.46	KRT9 Keratin, type I cytoskeletal 9
CON_P13645	5.3	KRT10 Keratin, type I cytoskeletal 10
GeneDB Tb427.08.3470	4.5	hypothetical protein, conserved
GeneDB Tb427.01.2340	4.12	alpha tubulin
GeneDB Tb927.2.4210	3.98	glycosomal phosphoenolpyruvate carboxykinase
GeneDB Tb427tmp.02.1085	3.8	40S ribosomal protein S4, putative
GeneDB Tb11.02.1085	3.8	40S ribosomal protein S4, putative
GeneDB Tb11.01.0355	3.74	ribosomal protein S26, putative
GeneDB Tb09.211.3830	3.24	hypothetical protein, conserved
GeneDB Tb427tmp.01.3110	3.23	heat shock protein 70
CON_P35908	3.19	KRT2 Keratin, type II cytoskeletal 2 epidermal
GeneDB Tb09.211.3550	3.02	glycerol kinase, glycosomal
GeneDB Tb427tmp.02.4150	2.92	pyruvate phosphate dikinase
GeneDB Tb427.10.2100	2.87	elongation factor 1-alpha
GeneDB Tb427.10.6630	2.6	ATP-dependent DEAD/H RNA helicase HEL64, putative
<a href="#">GeneDB Tb09.211.3920</a>	2.52	<b>QA-SNARE protein putative</b>
GeneDB Tb427.08.1120	2.5	hypothetical protein, conserved
<a href="#">GeneDB Tb427.03.5570</a>	2.42	<b>syntaxin, putative</b>
GeneDB Tb11.03.0800	2.39	hypothetical protein, conserved
GeneDB Tb427.08.4290	2.32	hypothetical protein, conserved
GeneDB Tb09.211.2570	2.25	T-complex protein 1, eta subunit, putative,t- complex protein 1 (eta subunit), putative
CON_P02533	2.24	KRT14 Keratin, type I cytoskeletal 14
GeneDB Tb427.10.14820	2.15	mitochondrial carrier protein, putative
GeneDB Tb427.08.2290	2.12	hypothetical protein, conserved
GeneDB Tb427.10.3930	2.06	40S ribosomal protein S3A, putative
GeneDB Tb427.08.3530	2.02	glycerol-3-phosphate dehydrogenase [NAD+], glycosomal
GeneDB Tb427.10.11660	1.98	hypothetical protein, conserved
GeneDB Tb11.22.0012	1.91	eukaryotic peptide chain release factor subunit 1, putative
<a href="#">GeneDB Tb427.10.1830</a>	1.84	<b>syntaxin, putative</b>
GeneDB Tb11.02.4050	1.84	60S ribosomal protein L28, putative

**Table 5.5 List of putative interacting proteins – TbVAMP7C::HA**

The 40 highest ranking proteins predicted to interact with the bait protein are presented. In this case, the TbVAMP7C::HA cell-line was used. The GeneDB accession, emPAI (exponentially modified protein abundance index) score as well as the GeneDB predicted protein name/description are shown. The bait protein is coloured red, other SNAREs are in blue and hypothetical proteins are in gray.

Accession	emPAI	Description
GeneDB Tb427.10.790	19.73	vesicle-associated membrane protein, putative (Handle)
GeneDB Tb427.10.2340	9.12	Qc SNARE Qc3 = Syx8
GeneDB Tb427.08.3470	4.5	Qb SNARE Qb2 =Vti-like
GeneDB Tb427tmp.01.3110	3.23	heat shock protein 70
GeneDB Tb09.211.3920	2.52	Qa SNARE Syx16B
GeneDB Tb427.08.1120	2.5	Qb SNARE Qb3 =Vti-like
GeneDB Tb427.03.5570	2.42	Qa SNARE SynE
GeneDB Tb11.22.0012	1.91	eukaryotic peptide chain release factor subunit 1, putative
GeneDB Tb427.10.1830	1.84	Qc SNARE Qc2a = Syx6-like
GeneDB Tb927.10.1100	1.6	60S ribosomal protein L9, putative
GeneDB Tb927.1.1720	1.45	hypothetical protein, conserved
GeneDB Tb927.10.15410	1.44	glycosomal malate dehydrogenase
GeneDB Tb427.10.6780	1.42	vacuolar protein sorting-associated protein 45, putative (SM protein)
CON_P48668	1.25	KRT6C Keratin, type II cytoskeletal 6C
GeneDB Tb427.07.6440	1.04	Qb SNARE Qb4b =Npsn
GeneDB Tb927.8.3150	1.01	T-complex protein 1, gamma subunit, putative
GeneDB Tb427.04.4090	0.96	hypothetical protein, conserved
GeneDB Tb427.08.6150	0.82	40S ribosomal protein S8, putative
GeneDB Tb927.10.9920	0.73	hypothetical protein, conserved
GeneDB Tb427.10.14200	0.63	Qa SNARE syntaxin 5
GeneDB Tb927.2.340	0.55	retrotransposon hot spot protein 4 (RHS4), putative
CON_P13646-1	0.53	KRT13 Isoform 1 of Keratin, type I cytoskeletal 13
GeneDB Tb427.10.9950	0.51	Qa SNARE Syn18
CON_P02666	0.46	Beta-casein - Bos taurus (Bovine).
GeneDB Tb09.160.0680	0.42	sec1 family transport protein, putative (SM protein)
GeneDB Tb427.10.7680	0.42	GTPase activating protein, putative (RabGAP)
GeneDB Tb427tmp.211.1750	0.42	mitochondrial carrier protein, putative
GeneDB Tb927.2.820	0.42	putative,retrotransposon hot spot protein 1 (RHS1), interrupted
GeneDB Tb09.211.2730	0.4	Gim5A protein, glycosomal membrane protein
GeneDB Tb427.04.1860	0.36	ribosomal protein S19, putative
GeneDB Tb427tmp.01.3320	0.36	Trichohyalin, putative
GeneDB Tb11.03.0580	0.35	RNA-binding protein, UBP2
GeneDB Tb09.211.2530	0.34	hypothetical protein, conserved
GeneDB Tb427.10.1900	0.32	DNA topoisomerase IA, putative
GeneDB Tb09.211.2160	0.31	hypothetical protein
GeneDB Tb427.08.1110	0.31	40S ribosomal protein S9, putative
GeneDB Tb927.1.710	0.31	phosphoglycerate kinase
GeneDB Tb927.4.4910	0.31	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor, putative
GeneDB Tb11.03.0390	0.3	protein phosphatase 2C, putative

**Table 5.6 List of putative interacting proteins (TbVAMP7C minus wild-type)**

The 40 highest ranking proteins predicted to interact with the bait protein are presented. In this case, the results in Table 5.5 are retained but non-specific hits from Table 5.4 are removed and the resulting list is presented here. The GeneDB accession, emPAI (exponentially modified protein abundance index) score as well as the GeneDB predicted protein name/description are shown. The bait protein is coloured red, other SNAREs are in blue, SM proteins are in green, and hypothetical proteins are in gray.

**Table 5.7** below reports some of the highest-ranking putative TbVAMP7C interactors identified in the mass-spectrometry analysis (a majority of which are SNARE proteins) along with lower ranking SNARE and SM proteins.

The mammalian longin VAMP7 and the yeast equivalent Snc1 and 2 have been shown to mediate transport at the cell membrane (Chaineau *et al.*, 2008; Danglot *et al.*, 2010; Shen *et al.*, 2013) by forming a complex with Syx1 (=SynPM) and Qbc SNAREs such as SNAP-25. However, no interaction of TbVAMP7C with TbSynPM was detected, and *T. brucei* and most kinetoplastids lack the Qbc SNARE. There was no indication that TbVAMP7B interacts with TbSynPM also, except for a single low confidence hit that was not replicable (See **Table 5.8** below). This leaves the possibility that it is TbVAMP7A that interacts with TbSynPM together with different Qb and Qc SNAREs in an exocytic complex. Putative candidates for these could be the plant/protist-specific SNAREs Qb Npsn and Qc Syp7. In plants, at least one paralog of each was reported to be involved in trafficking to the cell membrane together with VAMP72 and Syntaxin1 (Suwastika *et al.* 2008; Zheng *et al.* 2002; Kanazawa *et al.* 2015). TbNpsnA was found as a putative TbVAMP7C interactor across all replicates albeit with very low support, while TbSyp7 was not detected at all. It is therefore presently unclear whether TbVAMP7C mediates trafficking to the cell membrane.

Accession	ID	# of sig. seqs.	emPAI	Type	Repeat (4)	IFA?
Tb427.10.790	Handle-VAMP7C	10	5.74	SNARE	4	Y
Tb09.211.3920	Qa-Syx16B	10	1.98	SNARE	4	Y
Tb427.03.5570	Qa-Syx7	7	1.27	SNARE	4	Y
Tb427.10.1830	Qc2a-Syp7B	6	1.25	SNARE	4	Y
Tb427.08.3470	Qb2-VtiA	6	1.22	SNARE	4	Y
Tb427.08.1120	Qb3VtiB	2	0.87	SNARE	4	Y
Tb427.10.2340	Qc3-Syp5	2	0.67	SNARE	4	Y
Tb427.10.6780	VPS45	10	0.67	SM protein	4	Y
Tb427.10.9950	Qa-Syx18	4	0.32	SNARE	4	N
Tb427.10.14200	Qa-Syx5	3	0.28	SNARE	4	N
Tb09.160.0680	Sec1	8	0.09	SM protein	3	Y

**Table 5.7 Proteins identified as potential interactors for TbVAMP7C**

Data are arranged in decreasing order of emPAI score. Final column indicates interaction has been tested by co-immunofluorescence. ‘Repeat’ column depicts the number of separate proteomics experiments where the protein was identified. Only SNAREs and SM proteins identified are presented here.

VAMP7 is also known to form two well-described endosomal complexes: it interacts with Syx7, Syx8, and VtiB to mediate transport to the lysosome, and complexes with Syx6, Syx16, and VtiA to mediate intermediate endosomal trafficking (Jahn and Scheller 2006). TbVAMP7C immunoprecipitations consistently identified Vti-likeA and B, Syx8-like, Syx6-like1, Syx16 and Syx7 as top SNARE hits (see **Table 5.7** above) across all four buffer conditions that were tested. Note that only TbSyx16B paralog was found to be an interactor, but not TbSyx16A, which lacks the TM domain for membrane anchorage. To validate the interaction, these genes were genetically tagged in the TbVAMP7C::HA cell line. The immunofluorescence data showing co-localisation of these proteins, tagged with 6x-myc at the C-terminus is shown in panel A of **Figure 5.12** below. Western blot validation of these tags is also shown in the same figure in panel B at the bottom. In the western blots, multiple cell lines from each construct when present are numbered. The level of expression of various constructs varies - this is due to random insertion of the pRPAOP vector with which these gene sequences were tagged. These gene products showed between 5-approx. 15kDa gel shifting from their expected molecular weights, much like the

SynPM-HA protein did. TbSyx6-like expected to be 33kDa (25.8 + 7.2kDa(6xmyc)) was close to 39kDa; TbSynE expected at 36.4kDa (29.2 + 7.2kDa) was over 40kDa; TbSyn16B expected at 43.2 kDa (36 + 7.2 kDa) was close to 50kDa; TbVti-like A and B, expected to be at 33.5kDa (26.3 + 7.2kDa) and 21.2kDa (14 + 7.2 kDa) respectively were at 35-55kDa and just under 28kDa respectively; finally, TbSyx8-like expected at 18.4kDa (11.2 + 7.2kDa) was under 28kDa. This could be an example of a common phenomenon seen in membrane proteins due to differences in the level of SDS binding or an indication of the presence of post-translational modifications. Between replicate cell lines of the same tag, the protein size was maintained and single consistent bands were seen with no staining in the wild-type control sample, allowing confirmation of the expression of tagged constructs.

Of the six proteins, five were detectable by immunofluorescence. The Qc SNARE TbSyx8-like Tb427.10.2340::6myc was detectable in the western blot only. Absolute mRNA expression data from (Siegel *et al.*, 2010) showed that this SNARE has the lowest level of expression (among SNAREs) in the procyclic stage with only 43 counts. Perhaps attempts to overexpress such a typically low expression protein still results in undetectable amounts of protein. The Qc SNARE TbSyx6-like shows considerable overlapping with TbVAMP7C (top panel) and comparable levels of expression (as seen by eye, not quantified). In some cases, a second smaller puncta was seen, which also overlapped with a corresponding TbVAMP7C punctum. All positive TbSynE cell-lines (only one shown here) showed a much lower level of expression than TbVAMP7C, albeit overlapping with the latter in most cases. The cell line did however have a tendency to lose the tag within 3-4 passages indicating that the tag may have been unstable, perhaps due to deleterious effects on protein function. The Syx-16B showed different levels of expression in different cells, but always had at least partial overlap with TbVAMP7C. As the major TGN Qa-SNARE this protein may perhaps be involved in other SNARE complexes in other pathways as well. The two Vti-like cell-lines showed the lowest levels of overlap with TbVAMP7C – at times showed completely distinct staining. It is possible that they may be involved in other SNARE complexes (likely in homotypic fusion at the lysosome as seen in vacuoles) in addition to the complex involving VAMP7C. Therefore, all five of the proteins that could be immunolocalised showed considerable or partial co-staining

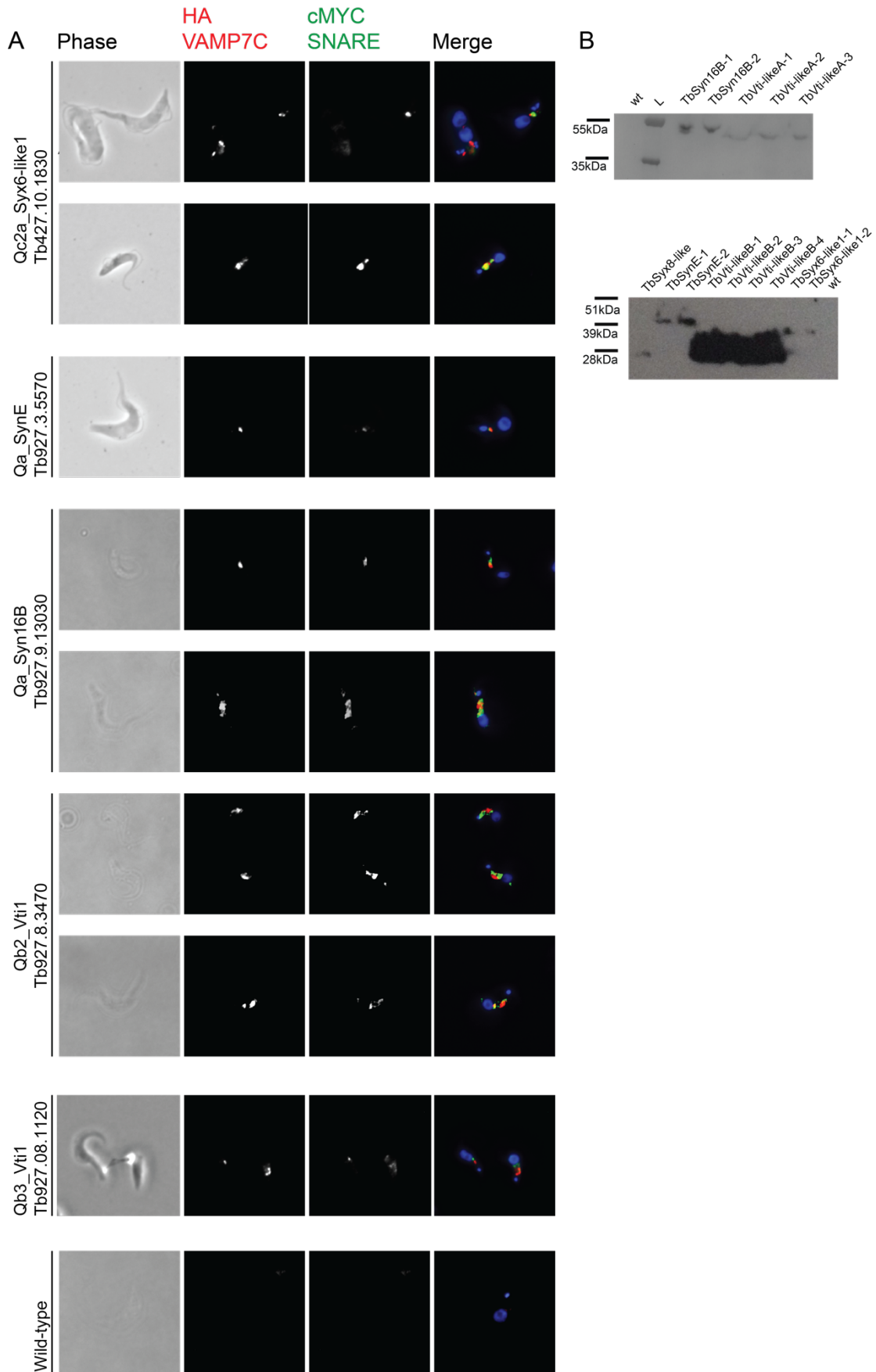
with the bait protein TbVAMP7C thus supporting the results of the mass spectrometry.

The above results support the hypothesis that the proteins interact, they cannot confirm that they do so. This is because conventional fluorescence microscopy does not provide the level of detail that can show that two proteins are physically associating. At best they can show that they are likely in the same region or compartment in the cell. Newer higher resolution microscopy methods now exist that can remedy this and confirm the mass spectrometry results. Further, because SNAREs are continually recycling between origin and target membranes via the SNARE cycle, it is likely that pools of SNAREs that interact but do not co-localise will exist within the cell. These immunofluorescence results are consistent with this observation.

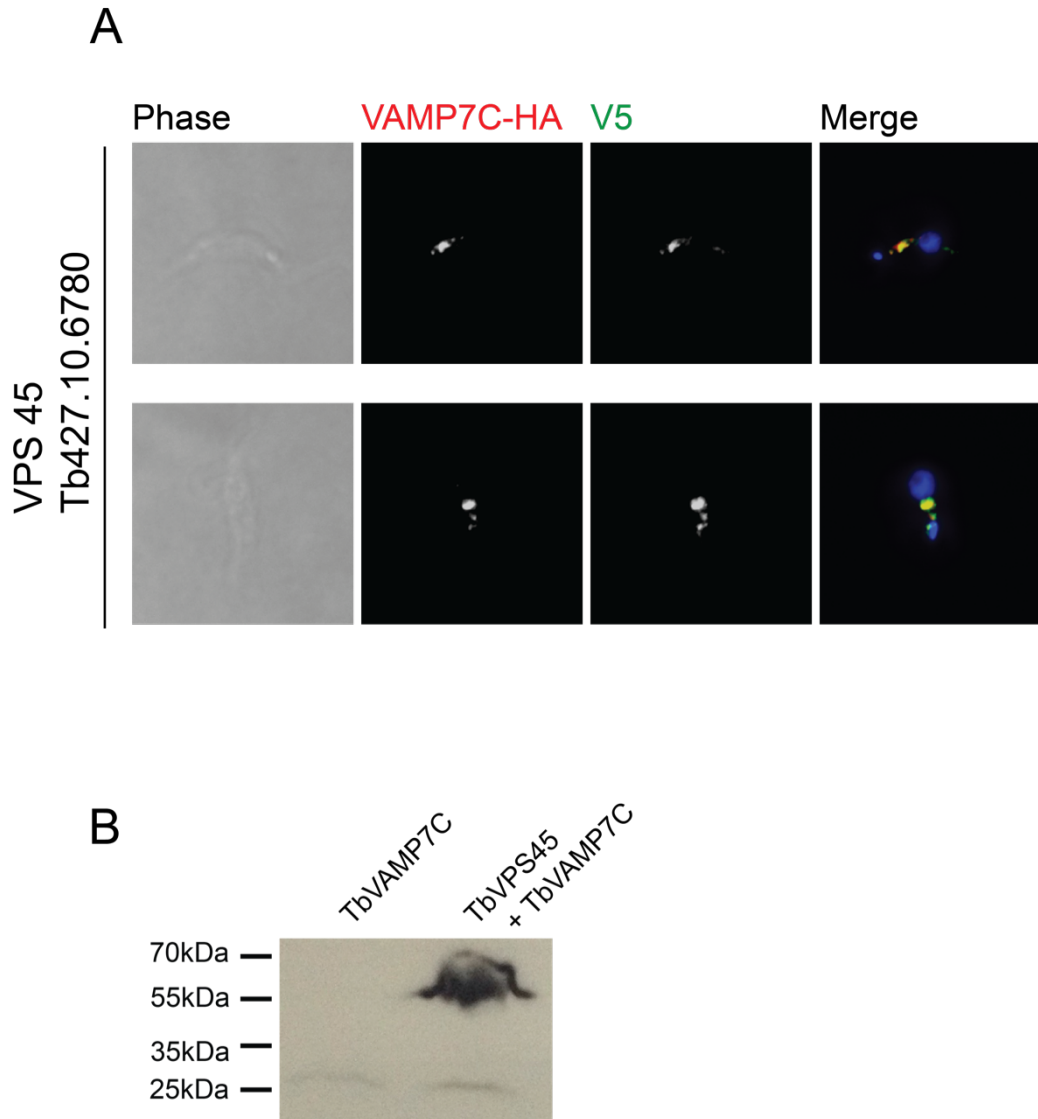
Also, the gold standard experiment that would allow us to confirm that these are indeed interactors of TbVAMP7C would be a reverse IP using one of these putative interacting SNARE proteins as the bait to check if it yields TbVAMP7C as an interactor. This was however not possible to do within the scope of this work. However, the replicability of the mass spectrometry results together with precise conservation with SNARE complexes in very divergent lineages such as yeasts and humans does indicate that these complexes in *T. brucei* are plausible.

**Figure 5.12 Validation of tagging of putative SNARE interactors of TbVAMP7C**

See page 178. (A) Localisation of candidate TbVAMP7C interactors. HA-tagged VAMP7C (red), is shown relative to putative interactors tagged with myc (green). The nucleus and kinetoplast are stained blue with DAPI. VAMP7C was stained with rat anti-HA (Roche), SNARE interactors were stained with mouse anti-c-myc antibody (Santa Cruz, 9E10, 1:400). Scale bar is 2 $\mu$ m. (B) Western blot of tagged cell lines. Tagged proteins were visualised with rat anti-c-myc antibody (Santa Cruz, 9E10, 1:5000).







**Figure 5.13 Validation of putative non-SNARE interactors of TbVAMP7C**

(A) Localisation of candidate TbVAMP7C interactor. HA-tagged VAMP7C (red), is shown relative to putative interactors tagged with V5 (green). The nucleus and kinetoplast are stained blue with DAPI. VAMP7C was stained with rat anti-HA (Roche), Putative interactor was stained with mouse anti-V5 antibody (Life Technologies, 1:500). Scale bar is 2 $\mu$ m. (B) Western blot of tagged cell lines. Tagged proteins were visualised with anti-V5 antibody (Life Technologies, 1:5000) and rat anti-HA antibody (Roche, 1:10000) according to the tag.

#### 5.4.2 Interactions of TbVAMP7B

Three mass-spectrometry analyses for the TbVAMP7B bloodstream form cell line were performed – two replicates of one condition and a second buffer condition (detailed in methods). However, no immunofluorescence validation was done, so no definite conclusions can be drawn from this data. However, with caution, a few

interesting aspects of the results are discussed here. Five of the six SNAREs that were found to be interacting with TbVAMP7C were also found in the bloodstream form TbVAMP7B immunoprecipitation (see **Table 5.8** below). Only Syx8-like protein was not identified as a binding partner. Given the low level of expression of this protein, it may be more specifically utilised. TbNpsnA, which was a low confidence hit in the TbVAMP7C immunoprecipitation, was also found in as a putative interactor of TbVAMP7B. TbSynPM was also identified; however, the interaction was not replicable in the second buffer condition. Surface proteins such as ISGs were also replicably obtained in the TbVAMP7B immunoprecipitation but not in the TbVAMP7C. Further repeats and localisation with putative binding partners will allow mapping of what appears to be a partially overlapping set of interactions.

The AP3  $\delta$  subunit, which is known to directly bind with VAMP7 longin domain in mammalian cells was not found among the interactors of TbVAMP7B (TbVAMP7C does not have a longin domain). In mammalian cells this forms the structural basis for the sorting of VAMP7 to the late endocytic compartments (Kent *et al.*, 2012). Further study of VAMP7B interactions is required to rule out this interaction. However, that VAMP7C lacking the longin domain is able to mediate this late endosomal function as shown by its interaction with other late endosomal Q-SNAREs. This indicates that other mechanisms of VAMP7 sorting may exist in *T. brucei*. Nearly identical SNARE interactions, despite differences in the N-terminal domain of the protein suggests that while N-terminal domains may regulate activity, other factors such as the SNARE domain contribute to the specificity of the interaction.

The SM protein TbVPS45 was also found as a putative binding partner for both TbVAMP7C and B. The co-localisation with TbVAMP7C and western blot validation of VPS45 tag is presented in **Figure 5.13** above. VPS45 is known to bind Syx16 directly (Dulubova *et al.*, 2002) and also the post-fusion cis-SNARE complex Syx7-Syx8-Vti-VAMP7 in yeast (Carpp *et al.*, 2006). Finding VPS45 as a binding partner for an R-SNARE indicates that the mechanism of its regulation of SNAREs as described in yeast, i.e., binding to cis-SNARE complexes, is conserved in *T. brucei*.

Accession	Identity	# sig. seqs.	emPAI	Type	Repeat (2)	in VAMP7C IP?
Tb427.05.3560	Handle R-VAMP7B	17	164.01	SNARE	2	N
Tb427.10.1830	Qc2a-Syx6-like1	6	1.01	SNARE	2	Y
Tb09.211.3920	Qa-Syx16B	8	0.96	SNARE	2	Y
Tb427.10.6780	VPS45	10	0.92	SNARE	2	Y
Tb427.08.3470	Qb2-VtiA	4	0.58	SNARE	2	Y
Tb427.10.14690	Qc1a-Syp7A	4	0.51	SNARE	2	N
Tb427.07.6440	Qb4b-NpsnA	3	0.43	SNARE	1	Y
Tb11.12.0013	Qa-Syx1*	2	0.18	SNARE	1	N
Tb427.05.350	ISG75	2	0.11	Surface protein	1	N
Tb427.05.360	ISG75	2	0.11	Surface protein	2	N

**Table 5.8 Proteins identified as potential interactions for TbVAMP7B**

Data are arranged in decreasing order of emPAI score. Final column indicates interaction has been detected for TbVAMP7C. 'Repeat' column depicts the number of separate proteomics experiments where the protein was identified. Only SNAREs, SM proteins and surface proteins are presented in this table.

Overall, these interactions indicate that, in contrast to plasma membrane, the endosomal SNARE partners of VAMP7C are well conserved in *T. brucei*. Some ER-Golgi SNAREs such as Syx5 and Syx18 and SM protein Sly1, known act in the early secretory pathway were also detected as potential interactors, albeit with low confidence. Consequently, Sly1 does not show substantial co-localisation with VAMP7C (not shown). Clathrin heavy chain was also found as a putative interactor for both VAMP7B and C. However, it was not included in the results as the protein was also found in the wild-type control results, perhaps because it is a generally very abundant protein. It was consistently found at a higher confidence level in the tagged cell line (emPAI ~0.5) than in the wild type (emPAI ~0.1). Immunoprecipitation of TbCHC (clathrin heavy chain) did reveal interaction with both VAMP7B and C (unpublished data) so the association needs to be explored further.

### 5.4.3 Others

In addition, several other proteins that have been previously been reported to interact with SNAREs have been identified. Even though these are reproducibly found in the mass-spectrometry analyses conducted on the TbVAMP7C::HA cell line, the

support for these interactions is rather low and it is difficult to distinguish between true interactions and unspecific binding. The validity of these hits was not tested by immunofluorescence but associations of these proteins with SNAREs have been reported in the literature. A selection of these are discussed below.

Accession	ID	# of sig. seqs.	emPAI	Type	Repeat (4)
Tb427.08.7982	Vacuolar H <sup>+</sup> translocating pyrophosphatase	10	0.53	H <sup>+</sup> pump	3
Tb427.05.1300	V_ATPase1 subunitA	7	0.36	H <sup>+</sup> pump	1
Tb427.08.2770	INS345Preceptor, SNARE associated domain	7	0.06	Ca <sup>2+</sup> regulation	3
Tb427.06.3500	RME-8	5	0.07	trafficking	3
Tb427tmp.02.2520	Ankyrin repeat, TPR	5	0.41	protein interaction	1
Tb427.08.4330	Rab11	3	0.46	trafficking	1
Tb427.02.6050	beta prime COP protein	3	0.1	trafficking	3
Tb427.08.1160	Vacuolar Ca <sup>2+</sup> ATPase	2	0.05	Ca <sup>2+</sup> regulation	2
Tb427.08.4610	Rab1	2	0.28	trafficking	1
Tb427.10.12960	Rab5	2	0.28	trafficking	1
Tb427.03.5060	Ankyrin repeat	2	0.26	protein interaction	1
Tb427tmp.01.4621	Calmodulin	2	0.69	Ca <sup>2+</sup> regulation	1

**Table 5.9 Proteins identified as potential interactors of TbVAMP7C**

Data are arranged in decreasing order of number of significant sequences. Final column indicates interaction was detected for TbVAMP7C. ‘Repeat’ column depicts the number of separate proteomics experiments where the protein was identified. Only SNAREs, SM proteins and surface proteins are presented in this table.

*T. cruzi* VAMP7A was recently shown to predominantly localise to acidocalcisomes, as shown by co-localisation with marker vacuolar pyrophosphatase (TcVP1) (Niyogi *et al.*, 2015), and weakly to the contractile vacuole complex (CVC) (Ulrich *et al.*, 2011). TbVP1 together with several other proteins which also have been demonstrated to localise to acidocalcisomes in *T. brucei*, such as vacuolar H<sup>+</sup> ATPase (TbVA), vacuolar Ca<sup>2+</sup> ATPase (TbPMC1), and the inositol-1,4,5-triphosphate receptor (TbIP<sub>3</sub>R) (Huang *et al.*, 2014), were identified as potential interactors of TbVAMP7C (but not TbVAMP7B) in three of four extraction conditions. Thus, it is possible that TbVAMP7C functions at this organelle. These proteins (TcVAMP7A, TcVP1, TcVA, TcPMC1, and TcIP<sub>3</sub>R) were also found to be enriched in the

contractile vacuole complex (CVC) of *T. cruzi* (Ulrich et al., 2011) under osmotic stress and are thought to be exchanged between the organelles under such conditions (Niyogi et al., 2015). Interestingly none of the Q-SNAREs, nor the TcVAMP7C was identified in these *T. cruzi* experiments even though the R-SNAREs are implicated in fusion of these membranes. It is possible that TbVAMP7A also functions redundantly at this organelle in *T. brucei* but it is yet to be tested.

Interaction with calmodulin was also found and *in-silico* analysis via <http://calcium.uhnres.utoronto.ca/ctdb/ctdb/sequence.html> showed that TbVAMP7C possessed a calmodulin-binding domain in the membrane proximal region required for this interaction. TbIP<sub>3</sub>R which is involved in Ca<sup>2+</sup> release is predicted to contain a “SNARE-associated” domain at the C-terminus, which is described in the NCBI conserved domain database as present in SNARE-associated Golgi proteins which co-localise with the Tlg2 (=Syx16) in yeast late Golgi compartments. Syx16 has been demonstrated in this study to interact with VAMP7C. TbIP<sub>3</sub>R may therefore be interacting with Syx16 when it is in complex with VAMP7C. This is consistent with the previously described model in which SNARE machinery directly interacts with Ca<sup>2+</sup> channels where unassembled SNAREs inhibit calcium influx, and SNARE-complex assembly releases the inhibition (Hay, 2007).

Other trafficking related proteins were also found. RME-8 (Tb427.06.3500) which is involved in the recycling of invariant surface glycoproteins (ISGs) to the surface (Koumandou *et al.*, 2012) was found. Several coat protein subunits, Rabs (1, 5, 11), an endosomal integral membrane protein (Tb427.08.1940) were also found. In mammalian cells, ankyrin-repeat domain of Varp protein is known to bind VAMP7 in a fusogenically inactive conformation. Two ankyrin-repeat containing proteins Tb427.03.5060 and Tb11.02.2520 were found in this study, but they appear not to be orthologs of the mammalian Varp.

## 5.5 Conclusions

SNARE mRNA levels are upregulated in the bloodstream form as compared with procyclic form *T. brucei*. Within the bloodstream form, the Golgi and post-Golgi SNAREs showed higher expression than early secretory SNAREs, a pattern not seen in procyclic forms (Siegel *et al.*, 2010). This is consistent with the rapid endocytosis of VSG and surface bound antigens seen in the bloodstream form trypanosomes but not procyclics. Several putative plasma membrane localised SNAREs were found to be upregulated, particularly the Qb and Qc SNAREs Npsn and Syp7. Unfortunately, these were not among the SNAREs localised in the experiments. Qa and R SNAREs with potential plasma membrane localisation based on phylogeny, namely Qa SynPM and R SNAREs VAMP7A, B, C, which were also upregulated in bloodstream forms, were localised. The results do not conclusively support action of these SNAREs at the cell membrane, but they do not rule it out. Further experiments are needed to determine the significance if any, of the upregulation of these SNAREs in bloodstream forms.

All the VAMP7s showed a largely endosomal location, consistent with the role of this protein in the post-Golgi trafficking pathways particularly in the intermediate endosomal compartments and in trafficking towards late endosomes and the lysosome. The three paralogs showed only a slight variation, which could be discerned by using more specific markers such as the various Rab proteins. Fresh antibody reagents to these markers would greatly facilitate further study of SNAREs and other trafficking factors in *T. brucei*. Nevertheless, the proteomics data indicated that the SNAREs that *T. brucei* VAMP7C interacted with are likely to be virtually identical to the complexes seen in yeast and humans (Jahn and Scheller, 2006). While TbVAMP7B interactions have not been verified by immunofluorescence, its putative SNARE interactors were almost identical (the exception being that it is not predicted to interact with TbSyx8-like). From this, it can be tentatively concluded that VAMP7B and C have at least partially redundant functions.

VAMP7 is a candidate for mediation of exocytosis from what we know in yeast and humans. However, one of its predicted partners (SNAP-25) is missing in kinetoplastids and the other (TbSynPM) was not identified as a binding partner for

TbVAMP7C. Outside of the opisthokonts, SNAREs have not been extensively studied experimentally. However, the studies showed SNAP-25 and longin-lacking VAMPs with altered roles in *P. teraurelia* (Schilde et al. 2010; Schilde et al. 2008), and Qbc-alternatives such as Npsn and Syp71 were also implicated in mediating transport to cell membrane in *A. thaliana* (Zheng et al. 2002; Suwastika et al. 2008). The study from this data, together with the above observations suggests that the diversity of permutations and combinations of the eukaryotic plasma membrane SNARE complex are likely to be considerable and currently unappreciated.





# 6 CONCLUSIONS

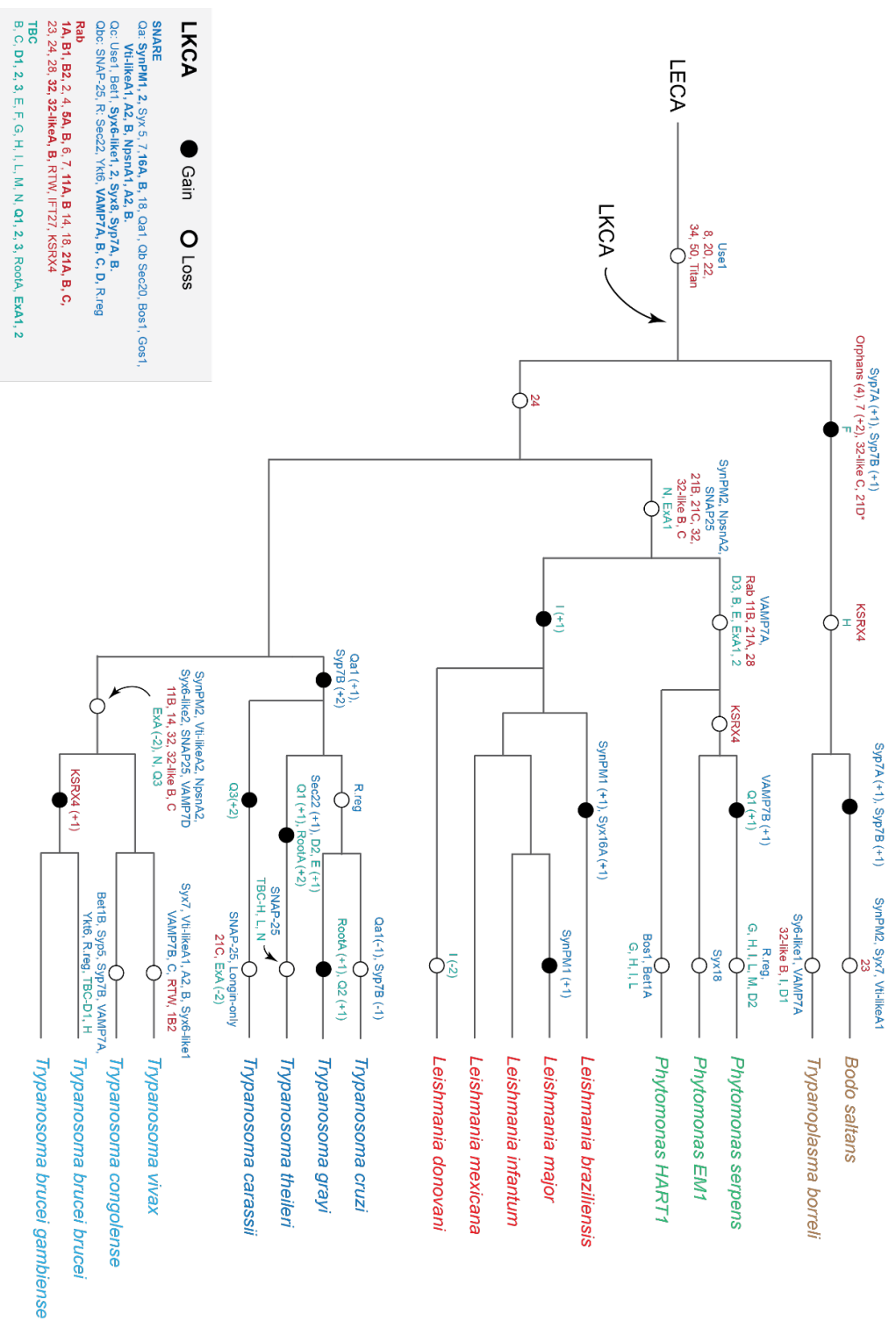
Kinetoplastids include many parasitic species, and their biology encompasses a broad range of niches and life cycles. The surface components are critical to the success of these protozoa, and this is largely reflected within the trafficking system. In this study, the evolutionary history of key proteins involved in membrane trafficking in this important lineage was reconstructed. The repertoire of Rabs, RabGAPs and SNAREs at the base of Kinetoplastida was quite similar to that of the LECA, but there was also evidence for several lineage-specific members of these protein families that likely originated in the common ancestor of kinetoplastids. Several of these orphan proteins were overlooked or misplaced in previous studies and their ancient origin was revealed. The number of paralogous pairs of proteins was high among kinetoplastids, which may indicate a genome duplication event at the base of the Kinetoplastida, or even earlier.

Rather than a precipitous decrease to trafficking complexity as the kinetoplastids transitioned from free-living phagotrophic organisms to parasitic heterotrophs, there was gradual loss of trafficking genes, likely corresponding to simplification of pathways and sorting routes. Hence, a radical remodelling did not accompany the transition to parasitism or adaptation to any other specific lifestyle. Moreover, the extensive bodonid repertoire is mainly due to retention of the ancestral gene complement, from the kinetoplastid or euglenozoan common ancestor. The most prominent expansions were seen in Syp7 and Rab32. Multiple paralogues of Vti-like, Npsn, Syx6-like and VAMP7 SNAREs, Rabs 5, 11, and 21, and TBCs D and Q appeared to have already been present in the common ancestor of the kinetoplastids (See **Figure 6.1** below), but these were asymmetrically retained across the different

lineages. In addition, multiple instances of losses and gains also appeared to take place at the taxon or species level as presented in **Figure 6.1**.

Thus, lineage-specific expansions (even if at the base of the kinetoplastids) were seen in all three trafficking families. However, the core, putative LECA derived Rab and TBC proteins remained well conserved and readily identified by phylogeny. This conservation was also true of a majority of the SNAREs, especially Qa and R SNAREs. However, several such SNAREs, mainly of the Qb and Qc groups appeared to be much diverged, with little or no phylogenetic support despite possession of relevant domains, and as seen in *T. brucei*, having conserved functions and interactions.

Unsurprisingly, the largest repertoire of Rabs, RabGAPs and SNAREs was in the free-living heterotroph *B. saltans*. There was a continued decrease in Rab and SNARE repertoire through the *Leishmania/Phytomonas* group with separate losses in the trypanosomes and to a lesser extent, in the parasitic bodonid *T. borreli*. The minimal known configuration was found in the African trypanosome clade. Indeed, the putative common ancestor of the African trypanosomes appeared to show the single instance of coordinated loss of several subunits from a SNARE complex, possibly indicating the loss of specific organelles or transport pathways. Specifically, loss of post-Golgi SNAREs was accompanied by the loss of Rabs predicted to act on similar pathways; these include the phagocytic Rab14 and the putative recycling endosomal Rab11B, which were lost in this lineage but not in *Leishmania spp.*



**Figure 6.1 Reconstruction of birth and death of SNARE, Rab, and TBC subfamilies in kinetoplastids**

The predicted LKCA repertoires are shown in the box at left. Those subtypes that have undergone an expansion are in bold italics. SNAREs are in blue, Rabs in pink and TBCs in brown. Putative points of origin (full circles) and loss (empty circles) are overlaid on a schematic kinetoplastid taxonomy. Numbers in brackets indicate the minimum number of copies of genes that are predicted lost (-n) or gained (+n). The evidence for losses and gains are based on data presented in Chapters 3 and 4.

*Leishmania* and trypanosomes however did share the loss of other endosomal trafficking proteins including Rab32, 21B and 21C, SNAREs NpsnA2 and SNAP25, and even TBC members ExA and N (See **Figure 6.1**), so that sculpting of these latter gene families was more similar. It is therefore possible that TBC-ExA and N may act as GAPs for Rab32 and 21. Again, these were gradual changes and suggested that the extracellular African trypanosomes had shed complexity that was retained by the intracellular parasites, and which may reflect the need for more profound morphological transitions. For example, the amastigote intracellular form may require endocytic systems to rapidly turnover surface membranes. This was in contrast to some of the cruzi group trypanosomes, which, like the African trypanosomes, are predicted to remain extracellular in the vertebrate host, but likely employ different, yet unidentified strategies for virulence and survival.

The cruzi group exhibited the greatest degree of inter-lineage variation, perhaps reflecting the diversity, varied lifestyles and disparate hosts and vectors that this group can use. Moreover, the bodonids and the cruzi group had several clade and species-specific gains and losses indicating a highly dynamic shaping of their trafficking gene repertoire. Phytomonads, *T. vivax* and *T. congolense* appeared to have been shaped by even further losses, although the possibility of incompletely assembled genomes cannot be absolutely ruled out. In comparison, the parasitic bodonid lost a set of trafficking proteins that did not greatly overlap with the trypanosomatid losses: Syx6-like1, VAMP7A, Rab32-likeB, as well as TBC-I and D1 (See **Figure 6.1**).

The selective retention of SNAREs, Rabs and TBCs in certain kinetoplastid lineages suggested a co-evolutionary aspect of the evolution of these trafficking families, but several other observations predict otherwise. First, the considerable expansion of the bodonid Rab repertoire did not seem to affect the SNAREs, which do not obviously exhibit such an expansion. Conversely, in lineages that had specifically lost a large number of SNAREs, such as *T. vivax* and *T. congolense*, there were no concurrent losses of Rabs or TBCs. Similarly, even though there were phytomonad-specific losses of several TBCs and two Rabs, no such trend was seen amongst the SNAREs. SNAREs showed a number domain level changes, such as loss and gain of SNARE, TM and longin domains, indicating that their mode of neofunctionalisation is distinct from Rabs in which such changes have not been reported thus far.

Recent studies in *T. cruzi* have localised TcVAMP7A (but not TcVAMP7C) to acidocalcisomes and the bladder of the contractile vacuole complex (CVC), and implicate TcVAMP7A in the biogenesis of the former by fusion of the organellar membranes (Niyogi *et al.*, 2015). Interaction analysis of TbVAMP7C indicated overlapping interacting partners with TcVAMP7A. Two further R-SNAREs were localised to the spongosome of the CVC including TcVAMP7D (specific to the cruzi group trypanosomes and *T. borreli*) and the peculiar longin-only R-SNARE protein. As the contractile vacuole is not present in *T. brucei*, it would be interesting to investigate the roles of TcVAMP7C and the *T. brucei* equivalents TbVAMP7A and longin-only protein.

Comparative genomics has revealed several other interesting proteins where further investigation could be informative. One is the continued presence of TM-lacking Qa-SNAREs (Syx16A and SynPM2) over long evolutionary distances. While TbSyx16A was expressed at the mRNA level, it is not yet known whether a functional protein results from it; according to our interaction studies, it does not seem to form complexes with the canonical partners of TbSyx16B. Others include the putative divergent Qbc protein in *T. cruzi* and the unassigned Qa1 protein, which may be involved in parasite-specific pathways.

While many of the Rabs have been extensively studied in *T. brucei*, this study reveals several Rab-like proteins that have not so far been investigated and are likely to be lineage-specific. These include KSRX4, which was not assigned to any known eukaryotic sub-type, and which appears to have undergone a rare lineage-specific duplication in the African trypanosomes (duplicates only in *T. b. brucei* and *T. b. gambiense*). There is also a singular Rab1-like, long-branching *T. brucei*-specific orphan, which may turn out to have roles outside of trafficking as seen in TbKSX1 and TbUzRX3. Other kinetoplastid specific Rabs (Rab11B, 21B and C) may be of importance in *T. cruzi* and *Leishmania* (where present). Also of interest is the comparison of duplicated Rab5 proteins, to see if the direction of neo/sub-functionalisation is distinct in organisms that do not employ rapid endocytosis like the BSF of *T. brucei*. Would they be akin to PCF, which show co-localisation, or would their functions be differentiated for example?

The trafficking steps associated with delivery or removal of material from the surface exhibit the greatest levels of divergence, which is possibly a reflection of the variety of surface architectures across the lineage. This interpretation was supported by

both the comparative genomics and the low level of conservation of exocytic TbVAMP7C SNARE complexes, despite a likely high degree of retention of endocytic complex composition. Cryo-immunoprecipitation of TbVAMP7C revealed interactions with other endosomal SNAREs that was almost identical to those seen in yeast and humans while for exocytosis, one of its predicted partners (SNAP-25) was missing and the other (SynPM) was not identified as a binding partner for TbVAMP7C. It would be interesting to investigate if TbNpsn and TbSyp7 proteins have a role in exocytosis as described for their plant orthologues (Zheng *et al.*, 2002; Suwastika *et al.*, 2008).

Comparative genomics also suggest that the exocytic SNARE apparatus is likely to differ considerably between different eukaryotic lineages, given the largely lineage-specific evolution of its constituents. While Syntaxin1 (=SynPM) could be identified as an ancient SNARE from which all extant SynPMs appear to originate, the Qb, Qc, Qbc, and R-SNARE components appeared to have undergone specialisations and expansions at multiple different points in eukaryotic evolution, and originating from multiple existing SNAREs. As late exocytosis is studied in diverse lineages, an update of the current definition of ‘secretory’ SNAREs will become possible. This study also provides the first evidence that the locations and compositions of many, but not all SNARE complexes are conserved across deep evolutionary time, indicating that SNAREs are potentially good markers of compartmental complexity along with Rab proteins.

This study reveals the composition and complexity of the membrane trafficking system of parasitic species as compared with a closely related and hence comparable free-living species, and finds an absence of distinctive transition; the canonical ‘LECA-origin’ repertoire of all three protein families are to a large extent, conserved across kinetoplastids. However, the exocytic and endosomal pathways exhibit the greatest levels of divergence at the sequence, phylogenetic, and functional levels, indicating that the interaction of the parasite with the host environment is likely under active selection.

# 7 REFERENCES

- Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105.
- Ackers, J. P., Dhir, V., and Field, M. C. (2005). A bioinformatic analysis of the RAB genes of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 141, 89–97.
- Adl, S. M. *et al.* (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* 52, 399–451.
- Adl, S. M. *et al.* (2012). The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* 59, 429–493.
- Akiyoshi, B., and Gull, K. (2014). Discovery of unconventional kinetochores in kinetoplastids. *Cell* 156, 1247–1258.
- Albert, S., Will, E., and Gallwitz, D. (1999). Identification of the catalytic domains and their functionally critical arginine residues of two yeast GTPase-activating proteins specific for Ypt/Rab transport GTPases. *EMBO J.* 18, 5216–5225.
- Alexander, D. L., Schwartz, K. J., Balber, A. E., and Bangs, J. D. (2002). Developmentally regulated trafficking of the lysosomal membrane protein p67 in *Trypanosoma brucei*. *J. Cell Sci.* 115, 3253–3263.
- Ali, M., Leung, K. F., and Field, M. C. (2014). The ancient small GTPase Rab21 functions in intermediate endocytic steps in trypanosomes. *Eukaryot. Cell* 13, 304–319.
- Allen, C. L., Goulding, D., and Field, M. C. (2003). Clathrin-mediated endocytosis is essential in *Trypanosoma brucei*. *EMBO J.* 22, 4991–5002.
- Almeida, I. C., Acosta-serrano, A., and Hutchinson, C. (2007). Comparison and

Evolution of the Surface Architecture of Trypanosomatid Parasites.

Alvarcz-Dominguez, C., and Stahl, P. D. (1998). Interferon- $\gamma$  selectively induces Rab5a synthesis and processing in mononuclear cells. *J. Biol. Chem.* *273*, 33901–33904.

Alvarez-Dominguez, C., and Stahl, P. D. (1999). Increased expression of Rab5a correlates directly with accelerated maturation of *Listeria monocytogenes* phagosomes. *J. Biol. Chem.* *274*, 11459–11462.

De Antoni, A., Schmitzová, J., Trepte, H. H., Gallwitz, D., and Albert, Š. (2002). Significance of GTP hydrolysis in Ypt1p-regulated endoplasmic reticulum to Golgi transport revealed by the analysis of two novel Ypt1-GAPs. *J. Biol. Chem.* *277*, 41023–41031.

Antonin, W., Fasshauer, D., Becker, S., Jahn, R., and Schneider, T. R. (2002). Crystal structure of the endosomal SNARE complex reveals common structural principles of all SNAREs. *Nat. Struct. Biol.* *9*, 107–111.

Araripe, J. R., Ramos, F. P., Cunha e Silva, N. L., Urményi, T. P., Silva, R., Leite Fontes, C. F., da Silveira, J. F., and Rondinelli, E. (2005). Characterization of a RAB5 homologue in *Trypanosoma cruzi*. *Biochem. Biophys. Res. Commun.* *329*, 638–645.

De Arras, L., Yang, I. V, Lackford, B., Riches, D. W. H., Prekeris, R., Freedman, J. H., Schwartz, D. A., and Alper, S. (2012). Spatiotemporal inhibition of innate immunity signaling by the Tbc1d23 RAB-GAP. *J. Immunol.* *188*, 2905–2913.

Attias, M., Souza, W. de, and Vommaro, R. C. (1996). Computer Aided Three-dimensional Reconstruction of the Free living Protozoan *Bodo* sp. (Kinetoplastida: Bodonidae). *Cell Struct. Funct.* *21*, 297–306.

Ballensiefen, W., Ossipov, D., and Schmitt, H. (1998). Recycling of the yeast v-SNARE Sec22p involves COPI-proteins and the ER transmembrane proteins Ufe1p and Sec20p. *J. Cell Sci.* *111*, 1507–1520.

Bangs, J. D. (2011). Replication of the ERES:Golgi junction in bloodstream-form African trypanosomes. *Mol. Microbiol.* *82*, 1433–1443.

Bangs, J. D., Brouch, E. M., Ransom, D. M., and Roggy, J. L. (1996). A soluble secretory reporter system in *Trypanosoma brucei*. Studies on endoplasmic reticulum targeting. *J. Biol. Chem.* *271*, 18387–18393.



- Bangs, J. D., Uyetake, L., Brickman, M. J., Balber, A. E., and Boothroyd, J. C. (1993). Molecular cloning and cellular localization of a BiP homologue in *Trypanosoma brucei*. Divergent ER retention signals in a lower eukaryote. *J. Cell Sci.* *105* (Pt 4), 1101–1113.
- Barbieri, M. A., Roberts, R. L., Gumusboga, A., Highfield, H., Alvarez-Dominguez, C., Wells, A., and Stahl, P. D. (2000). Epidermal growth factor and membrane trafficking: EGF receptor activation of endocytosis requires Rab5a. *J. Cell Biol.* *151*, 539–550.
- Barr, F., and Lambright, D. G. (2010). Rab GEFs and GAPs. *Curr. Opin. Cell Biol.* *22*, 461–470.
- Barry, J. D. (1979). Capping of variable antigen on *Trypanosoma brucei*, and its immunological and biological significance. *J. Cell Sci.* *37*, 287–302.
- Barry, J. D., Ginger, M. L., Burton, P., and McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? *Int. J. Parasitol.* *33*, 29–45.
- Barry, J. D., and McCulloch, R. (2001). Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv. Parasitol.* *49*, 1–70.
- Bednarek, S. Y., Ravazzola, M., Hosobuchi, M., Amherdt, M., Perrelet, A., Schekman, R., and Orci, L. (1995). COPI- and COPII-coated vesicles bud directly from the endoplasmic reticulum in yeast. *Cell* *83*, 1183–1196.
- Besteiro, S., Coombs, G. H., and Mottram, J. C. (2006). The SNARE protein family of *Leishmania major*. *BMC Genomics* *7*, 250.
- Bethani, I., Lang, T., Geumann, U., Sieber, J. J., Jahn, R., and Rizzoli, S. O. (2007). The specificity of SNARE pairing in biological membranes is mediated by both proof-reading and spatial segregation. *EMBO J.* *26*, 3981–3992.
- Birault, V., Solari, R., Hanrahan, J., and Thomas, D. Y. (2013). Correctors of the basic trafficking defect of the mutant F508del-CFTR that causes cystic fibrosis. *Curr. Opin. Chem. Biol.* *17*, 353–360.
- Blom, D., De Haan, A., Van Den Berg, M., Sloof, P., Jirků, M., Lukeš, J., and Benne, R. (1998). RNA editing in the free-living bodonid *Bodo saltans*. *Nucleic Acids Res.* *26*, 1205–1213.
- Blom, D., de Haan, A., van den Burg, J., van den Berg, M., Sloof, P., Jirku, M., Lukes, J., and Benne, R. (2000). Mitochondrial minicircles in the free-living bodonid *Bodo saltans* contain two gRNA gene cassettes and are not found in large networks. *RNA* *6*,

121–135.

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

Bonifacino, J. (2004). The Mechanisms of Vesicle Budding and Fusion. *Cell* *116*, 153–166.

Bonifacino, J. S. (2014). Vesicular transport earns a Nobel. *Trends Cell Biol.* *24*, 3–5.

Bonifacino, J. S., and Lippincott-Schwartz, J. (2003). Coat proteins: shaping membrane transport. *Nat. Rev. Mol. Cell Biol.* *4*, 409–414.

Brenndörfer, M., and Boshart, M. (2010). Selection of reference genes for mRNA quantification in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* *172*, 52–55.

Brighouse, A., Dacks, J. B., and Field, M. C. (2010). Rab protein evolution and the history of the eukaryotic endomembrane system. *Cell. Mol. Life Sci.* *67*, 3449–3465.

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

Brittingham, A., Chen, G., Mcgwire, B. S., Chang, K. P., and Mosser, D. M. (1999). Interaction of *Leishmania* gp63 with cellular receptors for fibronectin. *Infect. Immun.* *67*, 4477–4484.

Brittingham, A., Morrison, C. J., McMaster, W. R., McGwire, B. S., Chang, K. P., and Mosser, D. M. (1995). Role of the *Leishmania* surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. *J. Immunol.* *155*, 3102–3111.

Brun, R., Jenni, L., Tanner, M., Schönenberger, M., and Schell, K. F. (1979). Cultivation of vertebrate infective forms derived from metacyclic forms of pleomorphic *Trypanosoma brucei* stocks. Short communication. *Acta Trop.* *36*, 387–390.

Burki, F. *et al.* (2009). Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol. Evol.* *1*, 231–238.

Burri, L., and Lithgow, T. (2004). A Complete Set of SNAREs in Yeast. *Traffic* *5*, 45–52.

Camargo, E. P. (1999). *Phytomonas* and other trypanosomatid parasites of plants and

fruit. *Adv. Parasitol.* *42*, 29–112.

Carpp, L. N., Ciuffo, L. F., Shanks, S. G., Boyd, A., and Bryant, N. J. (2006). The Sec1p/Munc18 protein Vps45p binds its cognate SNARE proteins via two distinct modes. *J. Cell Biol.* *173*, 927–936.

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

Cavalier-Smith, T. (2013). Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa. *Eur. J. Protistol.* *49*, 115–178.

Cecílio, P., Pérez-Cabezas, B., Santarém, N., Maciel, J., Rodrigues, V., and Cordeiro da Silva, A. (2014). Deception and manipulation: the arms of leishmania, a successful parasite. *Front. Immunol.* *5*, 480.

Chaineau, M., Danglot, L., Proux-Gillardeaux, V., and Galli, T. (2008). Role of HRB in clathrin-dependent endocytosis. *J. Biol. Chem.* *283*, 34365–34373.

Chen, Y., Yang, L., Ding, Y., Zhang, S., He, T., Mao, F., Zhang, C., Zhang, H., Huo, C., and Liu, P. (2013). Tracing evolutionary footprints to identify novel gene functional linkages. *PLoS One* *8*, e66817.

Clayton, C. E. (2002). Life without transcriptional control? From fly to man and back again. *EMBO J.* *21*, 1881–1888.

Clayton, C., Häusler, T., and Blattner, J. (1995). Protein trafficking in kinetoplastid protozoa. *Microbiol. Rev.* *59*, 325–344.

Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R., and Owen, D. J. (2002). Molecular Architecture and Functional Model of the Endocytic AP2 Complex. *Cell* *109*, 523–535.

Courret, N., FREHEL, C., PRINA, E., LANG, T., and ANTOINE, J.-C. (2001). Kinetics of the intracellular differentiation of *Leishmania amazonensis* and internalization of host MHC molecules by the intermediate parasite stages. *Parasitology* *122*, 263–279.

Cramer, L. (2008). Organelle transport: dynamic actin tracks for myosin motors. *Curr. Biol.* *18*, R1066-8.

Dacks, J. B., Davis, L. A. M., Sjögren, A. M., Andersson, J. O., Roger, A. J., and

- Doolittle, W. F. (2003). Evidence for Golgi bodies in proposed 'Golgi-lacking' lineages. *Proc. Biol. Sci.* *270 Suppl*, S168-71.
- Dacks, J. B., and Doolittle, W. F. (2002). Novel syntaxin gene sequences from *Giardia*, *Trypanosoma* and algae: implications for the ancient evolution of the eukaryotic endomembrane system. *J. Cell Sci.* *115*, 1635–1642.
- Dacks, J. B., and Field, M. C. (2004). Eukaryotic Cell Evolution from a Comparative Genomic Perspective : The Endomembrane System. In: *In Organelles, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics*, 309–334.
- Dacks, J. B., and Field, M. C. (2007). Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J. Cell Sci.* *120*, 2977–2985.
- Dacks, J. B., Peden, A. a, and Field, M. C. (2009). Evolution of specificity in the eukaryotic endomembrane system. *Int. J. Biochem. Cell Biol.* *41*, 330–340.
- Dacks, J. B., Poon, P. P., and Field, M. C. (2008). Phylogeny of endocytic components yields insight into the process of nonendosymbiotic organelle evolution. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 588–593.
- Danglot, L., Chaineau, M., Dahan, M., Gendron, M.-C., Boggetto, N., Perez, F., and Galli, T. (2010). Role of TI-VAMP and CD82 in EGFR cell-surface dynamics and signaling. *J. Cell Sci.* *123*, 723–735.
- Davey, J. R., Humphrey, S. J., Junutula, J. R., Mishra, A. K., Lambright, D. G., James, D. E., and Stöckli, J. (2012). TBC1D13 is a RAB35 Specific GAP that Plays an Important Role in GLUT4 Trafficking in Adipocytes. *Traffic* *13*, 1429–1441.
- DeGrasse, J. A., DuBois, K. N., Devos, D., Siegel, T. N., Sali, A., Field, M. C., Rout, M. P., and Chait, B. T. (2009). Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. *Mol. Cell. Proteomics* *8*, 2119–2130.
- Delprato, A., and Lambright, D. G. (2007). Structural basis for Rab GTPase activation by VPS9 domain exchange factors. *Nat. Struct. Mol. Biol.* *14*, 406–412.
- Denny, P. W., Morgan, G. W., Field, M. C., and Smith, D. F. (2005). *Leishmania major*: clathrin and adaptin complexes of an intra-cellular parasite. *Exp. Parasitol.* *109*, 33–37.
- Dhir, V., Goulding, D., and Field, M. C. (2004). TbRAB1 and TbRAB2 mediate trafficking through the early secretory pathway of *Trypanosoma brucei*. *Mol. Biochem.*

- Parasitol. *137*, 253–265.
- Diekmann, Y., and Pereira-Leal, J. B. (2013). Evolution of intracellular compartmentalization. *Biochem. J.* *449*, 319–331.
- Diekmann, Y., Seixas, E., Gouw, M., Tavares-Cadete, F., Seabra, M. C., and Pereira-Leal, J. B. (2011). Thousands of Rab GTPases for the cell biologist. *PLoS Comput. Biol.* *7*.
- Docampo, R. (2015). The origin and evolution of the acidocalcisome and its interactions with other organelles. *Mol. Biochem. Parasitol.*
- Dong, J.-H., Wen, J.-F., and Tian, H.-F. (2007). Homologs of eukaryotic Ras superfamily proteins in prokaryotes and their novel phylogenetic correlation with their eukaryotic analogs. *Gene* *396*, 116–124.
- Du, L. L. L.-L., Collins, R. N., and Novick, P. J. (1998). Identification of a sec4p GTPase-activating protein (GAP) as a novel member of a rab GAP family. *J. Biol. Chem.* *273*, 3253–3256.
- Dulubova, I., Yamaguchi, T., Gao, Y., Min, S.-W., Huryeva, I., Südhof, T. C., and Rizo, J. (2002). How Tlg2p/syntaxin 16 ‘snares’ Vps45. *EMBO J.* *21*, 3620–3631.
- de Duve, C. (2007). The origin of eukaryotes: a reappraisal. *Nat. Rev. Genet.* *8*, 395–403.
- Eathiraj, S., Pan, X., Ritacco, C., and Lambright, D. G. (2005). Structural basis of family-wide Rab GTPase recognition by rabenosyn-5. *Nature* *436*, 415–419.
- El-Sayed, N. M. *et al.* (2005a). Comparative genomics of trypanosomatid parasitic protozoa. *Science* *309*, 404–409.
- El-Sayed, N. M. *et al.* (2005b). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* *309*, 409–415.
- Elias, M., Brighthouse, A., Gabernet-Castello, C., Field, M. C., and Dacks, J. B. (2012). Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. *J. Cell Sci.* *125*, 2500–2508.
- Engstler, M., Bangs, J. D., and Field, M. C. (2007). Intracellular Transport Systems in Trypanosomes: Function, Evolution, and Virulence. In: *Trypanosomes – after the Genome*.
- Fasshauer, D. (1998). Conserved structural features of the synaptic fusion complex:

- SNARE proteins reclassified as Q- and R-SNAREs. *Proc. Natl. Acad. Sci.* *95*, 15781–15786.
- Field, H., Farjah, M., Pal, A., Gull, K., and Field, M. C. (1998). Complexity of Trypanosomatid Endocytosis Pathways Revealed by Rab4 and Rab5 Isoforms in *Trypanosoma brucei*. *J. Biol. Chem.* *273*, 32102–32110.
- Field, H. I., Coulson, R. M. R., and Field, M. C. (2013). An automated graphics tool for comparative genomics: the Coulson plot generator. *BMC Bioinformatics* *14*, 141.
- Field, H., Sherwin, T., Smith, A. C., Gull, K., and Field, M. C. (2000). Cell-cycle and developmental regulation of TbRAB31 localisation, a GTP- locked Rab protein from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* *106*, 21–35.
- Field, M. C., and Carrington, M. (2004). Intracellular membrane transport systems in *Trypanosoma brucei*. *Traffic Copenhagen Denmark* *5*, 905–913.
- Field, M. C., and Carrington, M. (2009). The trypanosome flagellar pocket. *Nat. Rev. Microbiol.* *7*, 775–786.
- Field, M. C., Gabernet-Castello, C., and Dacks, J. B. (2007a). Reconstructing the evolution of the endocytic system: insights from genomics and molecular cell biology. *Adv. Exp. Med. Biol.* *607*, 84–96.
- Field, M. C., Natesan, S. K. a, Gabernet-Castello, C., and Koumandou, V. L. (2007b). Intracellular trafficking in the trypanosomatids. *Traffic* *8*, 629–639.
- Field, M. C., Sali, A., and Rout, M. P. (2011). Evolution: On a bender--BARs, ESCRTs, COPs, and finally getting your coat. *J. Cell Biol.* *193*, 963–972.
- Flávia Nardy, A., Freire-de-Lima, C. G., and Morrot, A. (2015). Immune Evasion Strategies of *Trypanosoma cruzi*. *J. Immunol. Res.* *2015*, 178947.
- De Franceschi, N., Wild, K., Schlacht, A., Dacks, J. B., Sinning, I., and Filippini, F. (2014). Longin and GAF domains: structural evolution and adaptation to the subcellular trafficking machinery. *Traffic* *15*, 104–121.
- Frasa, M. A. M., Koessmeier, K. T., Ahmadian, M. R., and Braga, V. M. M. (2012). Illuminating the functional and structural repertoire of human TBC/RABGAPs. *Nat. Rev. Mol. Cell Biol.* *13*, 67–73.
- Fritz-Laylin, L. K. *et al.* (2010). The Genome of *Naegleria gruberi* Illuminates Early Eukaryotic Versatility. *Cell* *140*, 631–642.

- Fruth, U., Solioz, N., and Louis, J. A. (1993). Leishmania major interferes with antigen presentation by infected macrophages. *J. Immunol.* *150*, 1857–1864.
- Fukasawa, M., Varlamov, O., Eng, W. S., Söllner, T. H., and Rothman, J. E. (2004). Localization and activity of the SNARE Ykt6 determined by its regulatory domain and palmitoylation. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 4815–4820.
- Fukuda, M. (2011). TBC proteins: GAPs for mammalian small GTPase Rab? *Biosci. Rep.* *31*, 159–168.
- Gabernet-Castello, C., O'Reilly, A. J., Dacks, J. B., and Field, M. C. (2013). Evolution of Tre-2/Bub2/Cdc16 (TBC) Rab GTPase-activating proteins. *Mol. Biol. Cell* *24*, 1574–1583.
- Gadelha, C., Holden, J. M., Allison, H. C., and Field, M. C. (2011). Specializations in a successful parasite: what makes the bloodstream-form African trypanosome so deadly? *Mol. Biochem. Parasitol.* *179*, 51–58.
- Garcia-Boronat, M., Diez-Rivero, C. M., Reinherz, E. L., and Reche, P. A. (2008). PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res.* *36*, W35-41.
- Ghalib, H. W., Whittle, J. A., Kubin, M., Hashim, F. A., el-Hassan, A. M., Grabstein, K. H., Trinchieri, G., and Reed, S. G. (1995). IL-12 enhances Th1-type responses in human Leishmania donovani infections. *J. Immunol.* *154*, 4623–4629.
- Ghedini, E. *et al.* (2004). Gene synteny and evolution of genome architecture in trypanosomatids. *Mol. Biochem. Parasitol.* *134*, 183–191.
- Ginger, M. L. (2005). Trypanosomatid biology and euglenozoan evolution: new insights and shifting paradigms revealed through genome sequencing. *Protist* *156*, 377–392.
- Gonzalo, S., and Linder, M. E. (1998). SNAP-25 palmitoylation and plasma membrane targeting require a functional secretory pathway. *Mol. Biol. Cell* *9*, 585–597.
- Gordon, D. E., Mirza, M., Sahlender, D. A., Jakovleska, J., and Peden, A. A. (2009). Coiled-coil interactions are required for post-Golgi R-SNARE trafficking. *EMBO Rep.* *10*, 851–856.
- Gossing, M., Chidambaram, S., and Fischer von Mollard, G. (2013). Importance of the N-terminal domain of the Qb-SNARE Vti1p for different membrane transport steps in the yeast endosomal system. *PLoS One* *8*, e66304.

- Goulding, C. W., Parseghian, A., Sawaya, M. R., Cascio, D., Apostol, M. I., Gennaro, M. L., and Eisenberg, D. (2002). Crystal structure of a major secreted protein of *Mycobacterium tuberculosis*-MPT63 at 1.5-Å resolution. *Protein Sci.* *11*, 2887–2893.
- Grab, D. J., Webster, P., Ito, S., Fish, W. R., Verjee, Y., and Lonsdale-Eccles, J. D. (1987). Subcellular localization of a variable surface glycoprotein phosphatidylinositol-specific phospholipase-C in African trypanosomes. *J. Cell Biol.* *105*, 737–746.
- Grazia Cappiello, M., Sutterwala, F. S., Trinchieri, G., Mosser, D. M., and Ma, X. (2001). Suppression of Il-12 transcription in macrophages following Fc gamma receptor ligation. *J. Immunol.* *166*, 4498–4506.
- Grünfelder, C. G. *et al.* (2003). Endocytosis of a glycosylphosphatidylinositol-anchored protein via clathrin-coated vesicles, sorting by default in endosomes, and exocytosis via RAB11-positive carriers. *Mol. Biol. Cell* *14*, 2029–2040.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* *59*, 307–321.
- Gülow, K., Bienert, D., and Haas, I. G. (2002). BiP is feed-back regulated by control of protein translation efficiency. *J. Cell Sci.* *115*, 2443–2452.
- Gurkan, C., Koulov, A. V, and Balch, W. E. (2007). An evolutionary perspective on eukaryotic membrane trafficking. *Adv. Exp. Med. Biol.* *607*, 73–83.
- Haas, A. K., Fuchs, E., Kopajtich, R., and Barr, F. A. (2005). A GTPase-activating protein controls Rab5 function in endocytic trafficking. *Nat. Cell Biol.* *7*, 887–893.
- Haas, A. K., Yoshimura, S., Stephens, D. J., Preisinger, C., Fuchs, E., and Barr, F. A. (2007). Analysis of GTPase-activating proteins: Rab1 and Rab43 are key Rabs required to maintain a functional Golgi complex in human cells. *J. Cell Sci.* *120*, 2997–3010.
- Hall, B., Allen, C. L., Goulding, D., and Field, M. C. (2004). Both of the Rab5 subfamily small GTPases of *Trypanosoma brucei* are essential and required for endocytosis. *Mol. Biochem. Parasitol.* *138*, 67–77.
- Handman, E., Osborn, A. H., Symons, F., Van Driel, R., and Cappai, R. (1995). The *Leishmania* promastigote surface antigen 2 complex is differentially expressed during the parasite life cycle. *Mol. Biochem. Parasitol.* *74*, 189–200.
- Hardwick, K. G., and Pelham, H. R. (1992). SED5 encodes a 39-kD integral membrane



- protein required for vesicular transport between the ER and the Golgi complex. *J. Cell Biol.* *119*, 513–521.
- Hartzell, P. L. (1997). Complementation of sporulation and motility defects in a prokaryote by a eukaryotic GTPase. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 9881–9886.
- Hay, J. C. (2007). Calcium: a fundamental regulator of intracellular membrane fusion? *EMBO Rep.* *8*, 236–240.
- He, C. Y., Ho, H. H., Malsam, J., Chalouni, C., West, C. M., Ullu, E., Toomre, D., and Warren, G. (2004). Golgi duplication in *Trypanosoma brucei*. *J. Cell Biol.* *165*, 313–321.
- He, C. Y., Pypaert, M., and Warren, G. (2005). Golgi duplication in *Trypanosoma brucei* requires Centrin2. *Science* *310*, 1196–1198.
- He, D., Fiz-Palacios, O., Fu, C.-J., Fehling, J., Tsai, C.-C., and Baldauf, S. L. (2014). An alternative root for the eukaryote tree of life. *Curr. Biol.* *24*, 465–470.
- Hirota, Y., and 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.
- Hirumi, H., and Hirumi, K. (1994). Axenic culture of African trypanosome bloodstream forms. *Parasitol. Today* *10*, 80–84.
- Huang, G., Ulrich, P. N., Storey, M., Johnson, D., Tischer, J., Tovar, J. A., Moreno, S. N. J., Orlando, R., and Docampo, R. (2014). Proteomic analysis of the acidocalcisome, an organelle conserved from bacteria to human cells. *PLoS Pathog.* *10*, e1004555.
- Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* *17*, 754–755.
- Huet, D., Blisnick, T., Perrot, S., and Bastin, P. (2014). The GTPase IFT27 is involved in both anterograde and retrograde intraflagellar transport. *Elife* *3*, e02419.
- Hyttinen, J. M. T., Niittykoski, M., Salminen, A., and Kaarniranta, K. (2013). Maturation of autophagosomes and endosomes: a key role for Rab7. *Biochim. Biophys. Acta* *1833*, 503–510.
- Itoh, T., Satoh, M., Kanno, E., and Fukuda, M. (2006). Screening for target Rabs of TBC (Tre-2/Bub2/Cdc16) domain-containing proteins based on their Rab-binding activity. *Genes Cells* *11*, 1023–1037.

- Ivens, A. C. *et al.* (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science* *309*, 436–442.
- Jackson, A. P. (2015). Genome evolution in trypanosomatid parasites. *Parasitology* *142 Suppl*, S40-56.
- Jackson, A. P., Quail, M. a, and Berriman, M. (2008). Insights into the genome sequence of a free-living Kinetoplastid: *Bodo saltans* (Kinetoplastida: Euglenozoa). *BMC Genomics* *9*, 594.
- Jahn, R., Lang, T., and Südhof, T. C. (2003). Membrane Fusion. *Cell* *112*, 519–533.
- Jahn, R., and Scheller, R. H. (2006). SNAREs--engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* *7*, 631–643.
- Jang, S. B., Kim, Y.-G., Cho, Y.-S., Suh, P.-G., Kim, K.-H., and Oh, B.-H. (2002). Crystal structure of SEDL and its implications for a genetic disease spondyloepiphyseal dysplasia tarda. *J. Biol. Chem.* *277*, 49863–49869.
- Jaskowska, E., Butler, C., Preston, G., and Kelly, S. (2015). *Phytomonas*: Trypanosomatids Adapted to Plant Environments. *PLOS Pathog.* *11*, e1004484.
- Jeffries, T. R., Morgan, G. W., and Field, M. C. (2001). A developmentally regulated rab11 homologue in *Trypanosoma brucei* is involved in recycling processes. *J. Cell Sci.* *114*, 2617–2626.
- Jeffries, T. R., Morgan, G. W., and Field, M. C. (2002). TbRAB18, a developmentally regulated Golgi GTPase from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* *121*, 63–74.
- Jensen, B. C., Sivam, D., Kifer, C. T., Myler, P. J., and Parsons, M. (2009). Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*. *BMC Genomics* *10*, 482.
- Jiang, J., Qi, Y.-X., Zhang, P., Gu, W.-T., Yan, Z.-Q., Shen, B.-R., Yao, Q.-P., Kong, H., Chien, S., and Jiang, Z.-L. (2013). Involvement of Rab28 in NF- $\kappa$ B Nuclear Transport in Endothelial Cells. *PLoS One* *8*, e56076.
- Kanno, E., Ishibashi, K., Kobayashi, H., Matsui, T., Ohbayashi, N., and Fukuda, M. (2010). Comprehensive screening for novel rab-binding proteins by GST pull-down assay using 60 different mammalian rabs. *Traffic* *11*, 491–507.
- Kasmapour, B., Gronow, A., Bleck, C. K. E., Hong, W., and Gutierrez, M. G. (2012).

- Size-dependent mechanism of cargo sorting during lysosome-phagosome fusion is controlled by Rab34. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 20485–20490.
- Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* *33*, 511–518.
- Kelly, S. *et al.* (2007). Functional genomics in *Trypanosoma brucei*: a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Mol. Biochem. Parasitol.* *154*, 103–109.
- Kent, H. M., Evans, P. R., Schäfer, I. B., Gray, S. R., Sanderson, C. M., Luzio, J. P., Peden, A. A., and Owen, D. J. (2012). Structural basis of the intracellular sorting of the SNARE VAMP7 by the AP3 adaptor complex. *Dev. Cell* *22*, 979–988.
- Khurana, T., Brzostowski, J. A., and Kimmel, A. R. (2005). A Rab21/LIM-only/CH-LIM complex regulates phagocytosis via both activating and inhibitory mechanisms. *EMBO J.* *24*, 2254–2264.
- Kieft, R., Capewell, P., Turner, C. M. R., Veitch, N. J., MacLeod, A., and Hajduk, S. (2010). Mechanism of *Trypanosoma brucei gambiense* (group 1) resistance to human trypanosome lytic factor. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 16137–16141.
- Kienle, N., Klopper, T. H., and Fasshauer, D. (2009a). Differences in the SNARE evolution of fungi and metazoa. *Biochem. Soc. Trans.* *37*, 787–791.
- Kienle, N., Klopper, T. H., and Fasshauer, D. (2009b). Phylogeny of the SNARE vesicle fusion machinery yields insights into the conservation of the secretory pathway in fungi. *BMC Evol. Biol.* *9*, 19.
- Kima, P. E., Soong, L., Chicharro, C., Ruddle, N. H., and McMahon-Pratt, D. (1996). Leishmania-infected macrophages sequester endogenously synthesized parasite antigens from presentation to CD4<sup>+</sup> T cells. *Eur. J. Immunol.* *26*, 3163–3169.
- Kirkham, M., and Parton, R. G. (2005). Clathrin-independent endocytosis: New insights into caveolae and non-caveolar lipid raft carriers. *Biochim. Biophys. Acta - Mol. Cell Res.* *1745*, 273–286.
- Klopper, T. H., Kienle, C. N., and Fasshauer, D. (2007). An elaborate classification of SNARE proteins sheds light on the conservation of the eukaryotic endomembrane system. *Mol. Biol. Cell* *18*, 3463–3471.
- Klopper, T. H., Kienle, C. N., and Fasshauer, D. (2008). SNAREing the basis of

multicellularity: consequences of protein family expansion during evolution. *Mol. Biol. Evol.* *25*, 2055–2068.

Klöpffer, T. H., Kienle, N., Fasshauer, D., and Munro, S. (2012). Untangling the evolution of Rab G proteins: implications of a comprehensive genomic analysis. *BMC Biol.* *10*, 71.

Koonin, E. V., and Aravind, L. (2000). Dynein light chains of the Roadblock/LC7 group belong to an ancient protein superfamily implicated in NTPase regulation. *Curr. Biol.* *10*, R774-6.

Koumandou, V. L., Boehm, C., Horder, K. A., and Field, M. C. (2012). Evidence for Recycling of Invariant Surface Transmembrane Domain Proteins in African Trypanosomes. *Eukaryot. Cell* *12*, 330–342.

Koumandou, V. L., Dacks, J. B., Coulson, R. M. R., and 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.

Koumandou, V. L., Natesan, S. K. A., Sergeenko, T., and Field, M. C. (2008). The trypanosome transcriptome is remodelled during differentiation but displays limited responsiveness within life stages. *BMC Genomics* *9*, 298.

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

Kouranti, I., Sachse, M., Arouche, N., Goud, B., and Echard, A. (2006). Rab35 Regulates an Endocytic Recycling Pathway Essential for the Terminal Steps of Cytokinesis. *Curr. Biol.* *16*, 1719–1725.

Lafourcade, C., Galan, J.-M., Gloor, Y., Haguenaer-Tsapis, R., and Peter, M. (2004). The GTPase-Activating Enzyme Gyp1p Is Required for Recycling of Internalized Membrane Material by Inactivation of the Rab/Ypt GTPase Ypt1p. *Mol. Cell. Biol.* *24*, 3815–3826.

Lai, C. W., Aronson, D. E., and Snapp, E. L. (2010). BiP Availability Distinguishes States of Homeostasis and Stress in the Endoplasmic Reticulum of Living Cells. *Mol. Biol. Cell* *21*, 1909–1921.

Lanzetti, L., Palamidessi, A., Areces, L., Scita, G., and Di Fiore, P. P. (2004). Rab5 is a

- signalling GTPase involved in actin remodelling by receptor tyrosine kinases. *Nature* *429*, 309–314.
- Lapierre, L. A., Dorn, M. C., Zimmerman, C. F. F., Navarre, J., Burnette, J. O., and Goldenring, J. R. (2003). Rab11b resides in a vesicular compartment distinct from Rab11a in parietal cells and other epithelial cells. *Exp. Cell Res.* *290*, 322–331.
- Leung, K. F., Dacks, J. B., and Field, M. C. (2008). Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. *Traffic* *9*, 1698–1716.
- Li, F., Kümmel, D., Coleman, J., Reinisch, K. M., Rothman, J. E., and Pincet, F. (2014). A half-zipped SNARE complex represents a functional intermediate in membrane fusion. *J. Am. Chem. Soc.* *136*, 3456–3464.
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000a). Golgi and Post-Golgi Protein Sorting and Proteolytic Processing.
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000b). Post-Translational Modifications and Quality Control in the Rough ER.
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000c). Protein Glycosylation in the ER and Golgi Complex.
- Longatti, A., Lamb, C. A., Razi, M., Yoshimura, S. I., Barr, F. A., and Tooze, S. A. (2012). TBC1D14 regulates autophagosome formation via Rab11- and ULK1-positive recycling endosomes. *J. Cell Biol.* *197*, 659–675.
- Longatti, A., and Tooze, S. A. (2012). Recycling endosomes contribute to autophagosome formation. *Autophagy* *8*, 1682–1683.
- Lukes, J., Guilbride, D. L., Votýpka, J., Zíková, A., Benne, R., Englund, P. T., Voty, J., and Zíkova, A. (2002). Kinetoplast DNA Network: Evolution of an Improbable Structure. *Eukaryot. Cell* *1*, 495–502.
- Lumb, J. H., and Field, M. C. (2011). Rab23 is a flagellar protein in *Trypanosoma brucei*. *BMC Res. Notes* *4*, 190.
- Lumb, J. H., Leung, K. F., DuBois, K. N., and Field, M. C. (2011). Rab28 function in trypanosomes: interactions with retromer and ESCRT pathways. *J. Cell Sci.* *124*, 3771–3783.
- Luo, Z., and Gallwitz, D. (2003). Biochemical and Genetic Evidence for the Involvement of Yeast Ypt6-GTPase in Protein Retrieval to Different Golgi

Compartments \*. 278, 791–799.

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* 252, 1162–1164.

Macara, I. G., and Brondyk, W. H. (1995). *Small GTPases and Their Regulators Part C: Proteins Involved in Transport*, Elsevier.

Mai, A., Veltel, S., Pellinen, T., Padzik, A., Coffey, E., Marjomäki, V., and Ivaska, J. (2011). Competitive binding of Rab21 and p120RasGAP to integrins regulates receptor traffic and migration. *J. Cell Biol.* 194, 291–306.

Maier, A. G., Webb, H., Ding, M., Bremser, M., Carrington, M., and Clayton, C. (2001). The coatomer of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 115, 55–61.

Malsam, J., Kreye, S., and Söllner, T. H. (2008). Membrane traffic in the secretory pathway: Membrane fusion: SNAREs and regulation. *Cell. Mol. Life Sci.* 65, 2814–2832.

Manna, P. T., Boehm, C., Leung, K. F., Natesan, S. K., and Field, M. C. (2014). Life and times: synthesis, trafficking, and evolution of VSG. *Trends Parasitol.* 30, 251–258.

Manna, P. T., Kelly, S., and Field, M. C. (2013). Adaptin evolution in kinetoplastids and emergence of the variant surface glycoprotein coat in African trypanosomatids. *Mol. Phylogenet. Evol.* 67, 123–128.

Mantel, P.-Y., and Marti, M. (2014). The role of extracellular vesicles in *Plasmodium* and other protozoan parasites. *Cell. Microbiol.* 16, 344–354.

Marth, T., and Kelsall, B. L. (1997). Regulation of interleukin-12 by complement receptor 3 signaling. *J. Exp. Med.* 185, 1987–1995.

Martinez-Arca, S., Proux-Gillardeaux, V., Alberts, P., Louvard, D., and Galli, T. (2003). Ectopic expression of syntaxin 1 in the ER redirects TI-VAMP- and cellubrevin-containing vesicles. *J. Cell Sci.* 116, 2805–2816.

Martínez-Calvillo, S., Nguyen, D., Stuart, K., and Myler, P. J. (2004). Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryot. Cell* 3, 506–517.

Martínez-Calvillo, S., Vizuet-de-Rueda, J. C., Florencio-Martínez, L. E., Manning-Cela, R. G., and Figueroa-Angulo, E. E. (2010). Gene expression in trypanosomatid parasites. *J. Biomed. Biotechnol.* 2010, 525241.

- Martínez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K. D., Myler, P. J., and Street, N. (2003). Transcription of *Leishmania major* Friedlin Chromosome 1 Initiates in Both Directions within a Single Region. *Mol. Cell* *11*, 1291–1299.
- Matthews, K. R. (2005). The developmental cell biology of *Trypanosoma brucei*. *J. Cell Sci.* *118*, 283–290.
- McConville, M. J., Mullin, K. A., Ilgoutz, S. C., and Teasdale, R. D. (2002). Secretory pathway of trypanosomatid parasites. *Microbiol. Mol. Biol. Rev.* *66*, 122–154.
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* *41*, W597-600.
- Moreira, D., López-García, P., and Vickerman, K. (2004). An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int. J. Syst. Evol. Microbiol.* *54*, 1861–1875.
- Morgan, G. W., Allen, C. L., Jeffries, T. R., Hollinshead, M., and Field, M. C. (2001). Developmental and morphological regulation of clathrin-mediated endocytosis in *Trypanosoma brucei*. *J. Cell Sci.* *114*, 2605–2615.
- Natesan, S. K. A., Peacock, L., Leung, K. F., Matthews, K. R., Gibson, W., and Field, M. C. (2009). The trypanosome rab-related proteins RabX1 and RabX2 play no role in intracellular trafficking but may be involved in fly infectivity. *PLoS One* *4*.
- Navarro, M., and Gull, K. (2001). A pol I transcriptional body associated with VSG monoallelic expression in *Trypanosoma brucei*. *Nature* *414*, 759–763.
- Neuwald, A. F. (1997). A shared domain between a spindle assembly checkpoint protein and Ypt/Rab-specific GTPase-activators. *Trends Biochem. Sci.* *22*, 243–244.
- Nevin, W. D., and Dacks, J. B. (2009). Repeated secondary loss of adaptin complex genes in the Apicomplexa. *Parasitol. Int.* *58*, 86–94.
- Newman, A. P., Shim, J., and 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.
- Nikolova, L., SOMAN, K., NICHOLS, J. C., DANIEL, D. S., DICKEY, B. F., and HOFFENBERG, S. (1998). Conformationally variable Rab protein surface regions

- mapped by limited proteolysis and homology modelling. *Biochem. J.* *336*, 461–469.
- Niyogi, S., Jimenez, V., Girard-Dias, W., de Souza, W., Miranda, K., and Docampo, R. (2015). Rab32 is essential for maintaining functional acidocalcisomes, and for growth and infectivity of *Trypanosoma cruzi*. *J. Cell Sci.* *128*, 2363–2373.
- Norris, K. a, Bradt, B., Cooper, N. R., and So, M. (1991). Characterization of a *Trypanosoma cruzi* C3 binding protein with functional and genetic similarities to the human complement regulatory protein, decay-accelerating factor. *J. Immunol.* *147*, 2240–2247.
- Oberholzer, M., Morand, S., Kunz, S., and Seebeck, T. (2006). A vector series for rapid PCR-mediated C-terminal in situ tagging of *Trypanosoma brucei* genes. *Mol. Biochem. Parasitol.* *145*, 117–120.
- Ogbadoyi, E., Robinson, D., and Gull, K. (2003). A High-Order Trans-Membrane Structural Linkage Is Responcible for MItochondrial Genome Positioning and Segregation by Flagellar Basal Bodies in Trypanosomes. *Mol. Biol. Cell* *14*, 2559–2569.
- Olkkonen, V. M., and Ikonen, E. (2006). When intracellular logistics fails--genetic defects in membrane trafficking. *J. Cell Sci.* *119*, 5031–5045.
- Opdam, F. J., Kamps, G., Croes, H., van Bokhoven, H., Ginsel, L. A., and Fransen, J. A. (2000). Expression of Rab small GTPases in epithelial Caco-2 cells: Rab21 is an apically located GTP-binding protein in polarised intestinal epithelial cells. *Eur. J. Cell Biol.* *79*, 308–316.
- Ortiz-Ordóñez, J. C., and Seed, J. R. (1995). The removal of trypanolytic activity from human serum by *Trypanosoma brucei gambiense* and its subsequent recovery in trypanosome lysates. *J. Parasitol.* *81*, 555–558.
- Pal, A., HALL, B. S., JEFFRIES, T. R., and FIELD, M. C. (2003). Rab5 and Rab11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in *Trypanosoma brucei*. *Biochem. J.* *374*, 443–451.
- Pal, A., Hall, B. S., Nesbeth, D. N., Field, H. I., and Field, M. C. (2002). Differential endocytic functions of *Trypanosoma brucei* Rab5 isoforms reveal a glycosylphosphatidylinositol-specific endosomal pathway. *J. Biol. Chem.* *277*, 9529–9539.
- Palade, G. (1975). Intracellular aspects of the process of protein synthesis. *Science* *189*,



867.

Parfrey, L. W., Lahr, D. J. G., Knoll, A. H., and Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 13624–13629.

Park, H. H. (2013). Structural basis of membrane trafficking by rab family small G protein. *Int. J. Mol. Sci.* *14*, 8912–8923.

Parlati, F., McNew, J. A., Fukuda, R., Miller, R., Söllner, T. H., and Rothman, J. E. (2000). Topological restriction of SNARE-dependent membrane fusion. *Nature* *407*, 194–198.

Parlati, F., Varlamov, O., Paz, K., McNew, J. A., Hurtado, D., Söllner, T. H., and Rothman, J. E. (2002). Distinct SNARE complexes mediating membrane fusion in Golgi transport based on combinatorial specificity. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 5424–5429.

Van de Peer, Y., and De Wachter, R. (1997). Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *J. Mol. Evol.* *45*, 619–630.

Pellinen, T. *et al.* (2008). Integrin trafficking regulated by Rab21 is necessary for cytokinesis. *Dev. Cell* *15*, 371–385.

Pellinen, T., Arjonen, A., Vuoriluoto, K., Kallio, K., Fransén, J. A. M., and Ivaska, J. (2006). Small GTPase Rab21 regulates cell adhesion and controls endosomal traffic of beta1-integrins. *J. Cell Biol.* *173*, 767–780.

Peralta, E. R., Martin, B. C., and Edinger, A. L. (2010). Differential Effects of TBC1D15 and Mammalian Vps39 on Rab7 Activation State, Lysosomal Morphology, and Growth Factor Dependence. *J. Biol. Chem.* *285*, 16814–16821.

Pereira-Leal, J. B. (2008). The Ypt/Rab family and the evolution of trafficking in fungi. *Traffic* *9*, 27–38.

Pereira-Leal, J. B., and Seabra, M. C. (2000). The mammalian Rab family of small GTPases: definition of family and subfamily sequence motifs suggests a mechanism for functional specificity in the Ras superfamily. *J. Mol. Biol.* *301*, 1077–1087.

Pereira-Leal, J. B., and Seabra, M. C. (2001). Evolution of the Rab family of small GTP-binding proteins. *J. Mol. Biol.* *313*, 889–901.

- Pérez-Morga, D. *et al.* (2005). Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science* *309*, 469–472.
- Petrželková, R., and Eliáš, M. (2014). Contrasting patterns in the evolution of the Rab GTPase family in Archaeplastida. *Acta Soc. Bot. Pol.* *83*, 303–315.
- Podar, M., Wall, M. A., Makarova, K. S., and Koonin, E. V (2008). The prokaryotic V4R domain is the likely ancestor of a key component of the eukaryotic vesicle transport system. *Biol Direct* *3*, 2.
- Popovic, D., and Dikic, I. (2014). TBC1D5 and the AP2 complex regulate ATG9 trafficking and initiation of autophagy. *EMBO Rep.* *15*, 392–401.
- Porto-Carreiro, I., Attias, M., Miranda, K., De Souza, W., and Cunha-e-Silva, N. (2000). Trypanosoma cruzi epimastigote endocytic pathway: cargo enters the cytosome and passes through an early endosomal network before storage in reservosomes. *Eur. J. Cell Biol.* *79*, 858–869.
- Rajendran, L., and Annaert, W. (2012). Membrane trafficking pathways in Alzheimer’s disease. *Traffic* *13*, 759–770.
- Ramadas, R., and Thattai, M. (2013). New organelles by gene duplication in a biophysical model of eukaryote endomembrane evolution. *Biophys. J.* *104*, 2553–2563.
- Rambaut, A. (2009). FigTree v1.3.1. 2006-2009. Accessed Novemb. 29, 2012, Program package available at <http://tree.bio.ed.ac>.
- Rasband, W. (2012). ImageJ. U. S. Natl. Institutes Heal. Bethesda, Maryland, USA, [//imagej.nih.gov/ij/](http://imagej.nih.gov/ij/).
- Richardson, P. M., and Zon, L. I. (1995). Molecular cloning of a cDNA with a novel domain present in the tre-2 oncogene and the yeast cell cycle regulators BUB2 and cdc16. *Oncogene* *11*, 1139–1148.
- Rivera-Molina, F. E., and Novick, P. J. (2009). A Rab GAP cascade defines the boundary between two Rab GTPases on the secretory pathway. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 14408–14413.
- Roger, A. (1999). Reconstructing Early Events in Eukaryotic Evolution. *Am. Nat.* *154*, S146–S163.
- Rojas, A. M., Fuentes, G., Rausell, A., and Valencia, A. (2012). The Ras protein superfamily: evolutionary tree and role of conserved amino acids. *J. Cell Biol.* *196*,

189–201.

Roux, A. (2014). Reaching a consensus on the mechanism of dynamin? *F1000Prime Rep.* *6*, 86.

Saito-Nakano, Y., Mitra, B. N., Nakada-Tsukui, K., Sato, D., and Nozaki, T. (2007). Two Rab7 isotypes, EhRab7A and EhRab7B, play distinct roles in biogenesis of lysosomes and phagosomes in the enteric protozoan parasite *Entamoeba histolytica*. *Cell. Microbiol.* *9*, 1796–1808.

Sanderfoot, A. (2007). Increases in the number of SNARE genes parallels the rise of multicellularity among the green plants. *Plant Physiol.* *144*, 6–17.

Sato, M., Sato, K., Liou, W., Pant, S., Harada, A., and Grant, B. D. (2008). Regulation of endocytic recycling by *C. elegans* Rab35 and its regulator RME-4, a coated-pit protein. *EMBO J.* *27*, 1183–1196.

Schilde, C., Lutter, K., Kissmehl, R., and Plattner, H. (2008). Molecular identification of a SNAP-25-like SNARE protein in *Paramecium*. *Eukaryot. Cell* *7*, 1387–1402.

Schilde, C., Schönemann, B., Sehring, I. M., and Plattner, H. (2010). Distinct subcellular localization of a group of synaptobrevin-like SNAREs in *Paramecium tetraurelia* and effects of silencing SNARE-specific chaperone NSF. *Eukaryot. Cell* *9*, 288–305.

Schimanski, B., Nguyen, T. N., and Günzl, A. (2005). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*. *Mol. Cell. Biol.* *25*, 7303–7313.

Schlacht, A., Herman, E. K., Klute, M. J., Field, M. C., and Dacks, J. B. (2014). Missing Pieces of an Ancient Puzzle: Evolution of the Eukaryotic Membrane-Trafficking System. *Cold Spring Harb. Perspect. Biol.* *6*, a016048.

Schledzewski, K., Brinkmann, H., and Mendel, R. R. (1999). Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin of adaptor protein complexes 1, 2, and 3 and the F subcomplex of the coatomer COPI. *J. Mol. Evol.* *48*, 770–778.

Schwartz, K. J., Peck, R. F., and Bangs, J. D. (2013). Intracellular trafficking and glycobiology of TbPDI2, a stage-specific protein disulfide isomerase in *Trypanosoma brucei*. *Eukaryot. Cell* *12*, 132–141.

- Seaman, M. N. J., Harbour, M. E., Tattersall, D., Read, E., and Bright, N. (2009). Membrane recruitment of the cargo-selective retromer subcomplex is catalysed by the small GTPase Rab7 and inhibited by the Rab-GAP TBC1D5. *J. Cell Sci.* *122*, 2371–2382.
- Seixas, E., Barros, M., Seabra, M. C., and Barral, D. C. (2013). Rab and Arf proteins in genetic diseases. *Traffic* *14*, 871–885.
- Sevova, E. S., and Bangs, J. D. (2009). Streamlined architecture and glycosylphosphatidylinositol-dependent trafficking in the early secretory pathway of African trypanosomes. *Mol. Biol. Cell* *20*, 4739–4750.
- Shen, D., Yuan, H., Hutagalung, A., Verma, A., Kümmel, D., Wu, X., Reinisch, K., McNew, J. A., and Novick, P. (2013). The synaptobrevin homologue Snc2p recruits the exocyst to secretory vesicles by binding to Sec6p. *J. Cell Biol.* *202*, 509–526.
- Short, B., Preisinger, C., Korner, R., Kopajtich, R., Byron, O., and Barr, F. A. (2001). A GRASP55-rab2 effector complex linking Golgi structure to membrane traffic. *J. Cell Biol.* *155*, 877–884.
- Siddiqi, S. A., Mahan, J., Siddiqi, S., Gorelick, F. S., and Mansbach, C. M. (2006). Vesicle-associated membrane protein 7 is expressed in intestinal ER. *J. Cell Sci.* *119*, 943–950.
- Siegel, T. N., Hekstra, D. R., Wang, X., Dewell, S., and Cross, G. A. M. (2010). Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res.* *38*, 4946–4957.
- Simpson, A. G. B., Stevens, J. R., and Lukes, J. (2006). The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol.* *22*, 168–174.
- Singer-Kruger, B. (1994). Role of three rab5-like GTPases, Ypt51p, Ypt52p, and Ypt53p, in the endocytic and vacuolar protein sorting pathways of yeast. *J. Cell Biol.* *125*, 283–298.
- Singh, S. B., Tandon, R., Krishnamurthy, G., Vikram, R., Sharma, N., Basu, S. K., and Mukhopadhyay, A. (2003). Rab5-mediated endosome-endosome fusion regulates hemoglobin endocytosis in *Leishmania donovani*. *EMBO J.* *22*, 5712–5722.
- Sogin, M. L. (1991). Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* *1*, 457–463.

- Sousa, R., and Lafer, E. M. (2015). The role of molecular chaperones in clathrin mediated vesicular trafficking. *Front. Mol. Biosci.* 2, 26.
- Souza, W. DE, CARREIRO, I. P., MIRANDA, K., and SILVA, N. L. C. E. (2000). Two special organelles found in *Trypanosoma cruzi*. *An. Acad. Bras. Cienc.* 72, 421–432.
- de Souza, W., Sant'Anna, C., and Cunha-e-Silva, N. L. (2009). Electron microscopy and cytochemistry analysis of the endocytic pathway of pathogenic protozoa. *Prog. Histochem. Cytochem.* 44, 67–124.
- Stenmark, H. (2009). Rab GTPases as coordinators of vesicle traffic. *Nat. Rev. Mol. Cell Biol.* 10, 513–525.
- Stevens, J. R. (2008). Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes. *Parasite* 15, 226–232.
- Stuart, K., Brun, R., Croft, S., Fairlamb, A., Gürtler, R. E., McKerrow, J., Reed, S., and Tarleton, R. (2008). Kinetoplastids: related protozoan pathogens, different diseases. *J. Clin. Invest.* 118, 1301–1310.
- Suda, Y., Kurokawa, K., Hirata, R., and Nakano, A. (2013). Rab GAP cascade regulates dynamics of Ypt6 in the Golgi traffic. *Proc. ...* 110, 18976–18981.
- Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, a T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* 395, 347–353.
- Suwastika, I. N., Uemura, T., Shiina, T., Sato, M. H., and Takeyasu, K. (2008). SYP71, a plant-specific Qc-SNARE protein, reveals dual localization to the plasma membrane and the endoplasmic reticulum in *Arabidopsis*. *Cell Struct. Funct.* 33, 185–192.
- Szalinski, C. M., Labilloy, A., Bruns, J. R., and Weisz, O. A. (2014). VAMP7 modulates ciliary biogenesis in kidney cells. *PLoS One* 9, e86425.
- Tang, B. L., Low, D. Y., and Hong, W. (1998). Hsec22c: a homolog of yeast Sec22p and mammalian rsec22a and msec22b/ERS-24. *Biochem. Biophys. Res. Commun.* 243, 885–891.
- Tanigawa, G. (1993). Hydrolysis of bound GTP by ARF protein triggers uncoating of Golgi- derived COP-coated vesicles. *J. Cell Biol.* 123, 1365–1371.
- Thomson, R., Molina-Portela, P., Mott, H., Carrington, M., and Raper, J. (2009). Hydrodynamic gene delivery of baboon trypanosome lytic factor eliminates both animal

and human-infective African trypanosomes. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 19509–19514.

Trinkle-Mulcahy, L. *et al.* (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* *183*, 223–239.

Uemura, T., Ueda, T., Ohniwa, R. L., Nakano, A., Takeyasu, K., and Sato, M. H. (2004). Systematic analysis of SNARE molecules in *Arabidopsis*: dissection of the post-Golgi network in plant cells. *Cell Struct. Funct.* *29*, 49–65.

Ulrich, P. N. *et al.* (2011). Identification of contractile vacuole proteins in *Trypanosoma cruzi*. *PLoS One* *6*, e18013.

Vacca, M., Albania, L., Della Ragione, F., Carpi, A., Rossi, V., Strazzullo, M., De Franceschi, N., Rossetto, O., Filippini, F., and D'Esposito, M. (2011). Alternative splicing of the human gene SYBL1 modulates protein domain architecture of Longin VAMP7/TI-VAMP, showing both non-SNARE and synaptobrevin-like isoforms. *BMC Mol. Biol.* *12*, 26.

Valdez-Taubas, J., and Pelham, H. (2005). Swf1-dependent palmitoylation of the SNARE Tlg1 prevents its ubiquitination and degradation. *EMBO J.* *24*, 2524–2532.

Vanhamme, L. *et al.* (2003). Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* *422*, 83–87.

Vanhollebeke, B., Lecordier, L., Perez-Morga, D., Amiguet-Vercher, A., and Pays, E. (2007). Human serum lyses *Trypanosoma brucei* by triggering uncontrolled swelling of the parasite lysosome. *J. Eukaryot. Microbiol.* *54*, 448–451.

Varlamov, O. *et al.* (2004). i-SNAREs: inhibitory SNAREs that fine-tune the specificity of membrane fusion. *J. Cell Biol.* *164*, 79–88.

Vedovato, M., Rossi, V., Dacks, J. B., and Filippini, F. (2009). Comparative analysis of plant genomes allows the definition of the ‘Phytolongins’: a novel non-SNARE longin domain protein family. *BMC Genomics* *10*, 510.

Vidugiriene, J., Sharma, D. K., Smith, T. K., Baumann, N. a, and Menon, a K. (1999). Segregation of glycosylphosphatidylinositol biosynthetic reactions in a subcompartment of the endoplasmic reticulum. *J. Biol. Chem.* *274*, 15203–15212.

Weise, F., Stierhof, Y. D., Kühn, C., Wiese, M., and Overath, P. (2000). Distribution of GPI-anchored proteins in the protozoan parasite *Leishmania*, based on an improved

- ultrastructural description using high-pressure frozen cells. *J. Cell Sci.* *113 Pt 24*, 4587–4603.
- Whyte, J. R. C., and Munro, S. (2002). Vesicle tethering complexes in membrane traffic. *J. Cell Sci.* *115*, 2627–2637.
- Willett, R., Kudlyk, T., Pokrovskaya, I., Schönherr, R., Ungar, D., Duden, R., and Lupashin, V. (2013). COG complexes form spatial landmarks for distinct SNARE complexes. *Nat. Commun.* *4*, 1553.
- Wilson, M. E., and Pearson, R. D. (1986). Evidence that *Leishmania donovani* utilizes a mannose receptor on human mononuclear phagocytes to establish intracellular parasitism. *J. Immunol.* *136*, 4681–4688.
- Wirtz, E., Leal, S., Ochatt, C., and Cross, G. M. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* *99*, 89–101.
- Woollard, A. A. D., and Moore, I. (2008). The functions of Rab GTPases in plant membrane traffic. *Curr. Opin. Plant Biol.* *11*, 610–619.
- Van Xong, H., Vanhamme, L., Chamekh, M., Chimfwembe, C. E., Van Den Abbeele, J., Pays, A., Van Melvenne, N., Hamers, R., De Baetselier, P., and Pays, E. (1998). A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell* *95*, 839–846.
- Yao, C., Gaur Dixit, U., Barker, J. H., Teesch, L. M., Love-Homan, L., Donelson, J. E., and Wilson, M. E. (2013). Attenuation of *Leishmania infantum chagasi* Metacyclic Promastigotes by Sterol Depletion. *Infect. Immun.* *81*, 2507–2517.
- Zheng, H., Bednarek, S. Y., Sanderfoot, A. A., Alonso, J., Ecker, J. R., and Raikhel, N. V. (2002). NPSN11 is a cell plate-associated SNARE protein that interacts with the syntaxin KNOLLE. *Plant Physiol.* *129*, 530–539.
- Zick, M., and Wickner, W. (2012). Phosphorylation of the effector complex HOPS by the vacuolar kinase Yck3p confers Rab nucleotide specificity for vacuole docking and fusion. *Mol. Biol. Cell* *23*, 3429–3437.