FUNCTIONS AND ORIGINS OF MITOCHONDRION-RELATED ORGANELLES IN ANAEROBIC PROTISTS

by

Courtney Weir Stairs

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
November 2014

For those who said I could.

Table of Contents

List of Tables

List of Figures

Abstract

Across the diversity of life, organisms have evolved different strategies to thrive in hypoxic environments – microbial eukaryotes (protists) are no exception; protists that experience hypoxia often possess metabolically distinct mitochondria called mitochondrion-related organelles (MROs). Here, I focus on the biochemical adaptations of poorly-studied diverse anaerobic protists with an emphasis on the evolutionary histories of pyruvate, iron-sulfur cluster and respiratory chain metabolism in MROs.

In the absence of oxygen, some organisms use the pyruvate formate lyase (PFL) system for the non-oxidative generation of acetyl-CoA. Through phylogenetic analyses, I showed that PFL is broadly, but patchily, distributed across the tree of eukaryotes. The monophyly of eukaryotes in these analyses suggest that the PFL pathway was first acquired by lateral gene transfer (LGT) into a eukaryotic lineage - from a firmicute bacterial lineage - and that it has since spread horizontally by more recent eukaryote-to-eukaryote transfer events.

Biosynthesis of Fe-S clusters via the iron-sulfur cluster (ISC) system is a near-universally conserved feature of mitochondria and MROs. In contrast, some prokaryotic anaerobes synthesize and repair oxygen-damaged Fe-S clusters using a sulfur mobilization (SUF) system. Based on a transcriptomic survey, I reconstructed the MRO proteome of the protist *Pygsuia biforma* and found no evidence for the eukaryotic ISC system but instead identified a laterally acquired archaeabacterial-type SUF. Using immunofluorescence microscopy, I showed that SUF localizes to the MRO, representing the first reported case of a mitochondrial SUF system.

In some anaerobes, the ubiquinone analog rhodoquinone (RQ) allows the respiratory complex II to function in reverse, generating succinate from fumarate. Using immunofluorescence microscopy, I demonstrated that a RQ biosynthesis protein (RQUA) localizes to *Pygsuia* MROs, suggesting the organelle participates in RQ biosynthesis. Phylogenetic analyses suggest that anaerobic eukaryotes acquired RQUA multiple times via LGT from distinct eukaryotic and bacterial donors.

These studies suggest that crucial metabolic pathways localized to the MROs of anaerobic protists - and involved in their adaption to hypoxia - have been acquired by LGT. In some cases, these laterally acquired proteins were found to interface with ancestral mitochondrial proteins (*e.g.,* acquired RQUA and the ancestrally mitochondrial complex II in *Pygsuia*) to create metabolic pathways of mosaic origins.

List of Abbreviations Used

Ace	Acetate
ACoA	Acetyl-Coa
ACS	Acetyl-Coa Synthetase
AIC	Akaike Information Criterion
AK	Adenylate Kinase
Ala	Alanine
AlaAT	Alanine Aminotransferase
AOX	Alternative Oxidase
ASCT	Acetate:Succinyl-Coa Transferase
AspAT	Aspartate Aminotransferase
ATP	Adenosine Triphosphate
BCAT	Brached Chain Amino Acid Aminotransferase
BI	Bayesian Inference
BLAST	Basic Local Alignment Search Tool
bp	Base Pairs
BV	Bootstrap Value
cDNA	Complementary DNA
CDP-DAG	Cytidine Phosphate Diacylglycerol Glycerol
CI	Complex I, NADH:Ubquinone Oxidoreductase, NADH Dehydrogenase
CIA	Cytosolic Fe-S Cluster Assembley Machinery
CII	Complex II, Succinate Dehydrogenase; Succinate:Ubiquinone Oxidoreductase
CIIaf	Succinate Dehydrogenase Assembly Factor 2,
CIII	Complex III, Ubiquinol:Cytochrome C Oxidoreductase
Cit	Citrate
CIV	Complex IV, Cytochrome C Oxidase
CL	Cardiolipin
CLS	Cardiolipin Synthase
CoA	Coenzyme A
Cpn	Mitochondrial Chaperonin
CS	Citrate Synthase-Like
CV	Complex V, ATP Synthase
Cym1	Presequence Protease
CYSJ	Sulfite Reductase Flavin Reductase Alpha Domain
DAPI	4',6-Diamidino-2-Phenylindole
DHAP	Dihydroxyacetone Phosphate
DNA	Deoxyribonucleic Acid
EGT	Endosymbiotic Gene Transfer
EST	Expressed Sequence Tag
ETC	Electron Transport Chain
ETF	Electron Transferring Flavoprotein

ETFDH	ETF Dehydrognease (ETF:Q Oxidoreductase)
$FADH_2$	Flavin Adenine Dinucleotide
Fe-S	Iron-Sulfur
FH	Fumarase, Fumarate Hydratase
FRD	Fumarate Reductase
Fum	Fumarate
G3PDH	Glycerol-3-Phosphate Dehydrogenase
GCS	Glycine Cleavage System
GFP	Green Fluorescent Protein
GK	Glycerol Kinase
Gly3p	Glycerol-3-Phosphate
GrpE-Mge1	Mitochondrial Grpe Chaperon-Mge1
GST	Glutathione S-Transferase
HADA	Trifunctional Enzyme Alpha Subunit
HADB	Trifunctional Enzyme Beta Subunit
HAR	Hydroxylamine Reductase
HMM	Hidden Markov Model
HPM	Hydrogen Producing Mitochondria
HTD2	3-Hydroxyacyl Thioester Dehydratase 2
HYDA	[Fefe]-Hydrogenase
HYDE-G	HYDA Maturase Proteins E-G
Icp55	Mitochondrial Aminopeptidase
IMM	Inner Mitochondrial Membrane
IMS	Intermembrane Space,
Ind	Indole
IND1	Fe-S Cluster Binding Protein, Ind1-Like
IPTG	Isopropyl B-D-1-Thiogalactopyranoside
ISC	Iron-Sulfur Cluster
Kar-FabG	Ketoacyl Reductase (Fabg-Like)
KBL	2-Amino-3-Ketobutyrate Coa Ligase
Lac	Lactate
LDH	Lactate Dehydrogenase
Leu	Leucine
LGT	Lateral Gene Transfer
LplA-Lip3	Lipoate-Protein Ligase
Mal	Malate
MCF	Mitochondrial Carrier Protein
ME1	Malic Enzyme NAD-Dependent
MECR	Trans-2-Enoyl-Coa Reductase
Met	Metaxin
MIA40	Mitochondrial Intermembrane Space Assemble Protein
MIP	Mitochondrial Intermediate Peptidase
ML	Maximum Likelihood
MM	Mitochondrial Matrix

MMC	Methylmalonyl-Coa
MPC	Mitochondrial Pyruvate Carrier
MPP	Mitochondrial Processing Peptidase
MRO	Mitochondrion Related Organelle
Msf1/Ups2	Slowmo Protein - Mitochondrial Protein Sorting
mtDNA	Mitochondrial DNA
mtDnaJ	Mitochondrial Dnaj Chaperone
mtHsp70	Mitochondrial Heat Shock Protein 70
MTS	Mitochondrial Targeting Signal
NA	Numerical Aperature
NADH	Nicotinamide Adenine Dinucleotide
NCBI	National Centre For Biotechnology Information
NFU1	NIFU-Like Protein
NIF	Nitrogen-Fixation
NPQDH	NAD(P)H Dehydrogenase (Quinone)
NUO	NADH:Ubiquionone Oxidoreductase (NADH Dehydrogenase)
Oaa	Oxaloacetate
OMM	Outer Mitochondrial Membrane
ORF	Open Reading Frame
Pam18	Mitochondrial Pam18-Like Protein
Pb	*Pygsuia biforma*
PCC	Propionyl-Coa Carboxylase
PDC	Pyruvate Dehydrogenase Complex
PE	Phosphatidyl Ethanolamine
PEP	Phosphoenolpyruvate
PEPM	Phosphoenolpyruvate Mutase
PFL	Pyruvate Formate Lyase
PFLA	PFL Activating Enzyme
PFO	Pyruvate:Ferredoxin Oxidoreductase
PG	Phosphatidylglycerol
PGP	Phosphatidylglycerol Phosphate
PGPS	CDP-Diacylglycerol--Glycerol-3-Phosphate 3-Phosphatidyltransferase
Pim1	Mitochondrial Serine Protease - Lon Protease, Pim1-Like
PNO	Pyruvate:NADPH Oxidoreductase
PP	Posterior Probabilities
PPA	Phosphonoacetaldehyde
PPC	Propc Carboxylase
PPD	Phosphonopyruvate Decarboxylase
PPyr	Phosphonopyruvate
Pro	Prohibitin
ProP	Propionyl-Coa
PRX	Peroredoxin
PS	Phosphatidyl Serine
PSD1	Phosphotidylserine Decarboxylase 1

PTPMT1	Protein-Tyrosine Phosphatase Mitochondrial
rDNA	Ribosomal DNA
RNA	Ribonucleic Acid
RQ	Rhodoquinone
RQUA	Rhodoquinone Biosynthesis Protein RQUA
SAM	S-Adenosyl Methionine
Sam50	Sortining And Assembly Machinery 50
SAR	Stramenopile-Alveolae-Rhizaria
SCS	Succinyl-Coa Synthetase, Succinate Thiokinase
SDHA	Succinate Dehydrogenase Flavoprotein Subunit
SDHB	Sdh Fe-S Subunit
SDHC	SDH Large Cytochrome B Subunit
SDHD	SDH Small Cytochrome B Subunit
SDR	Short Chain Dehydrogenase
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
Ser	Serine
SHMT	Serine Hydroxymethyl Transferase
SOD	Superoxide Dismutase
SSU	Small Subunit
Suc	Succinate
SUF	Sulfur Mobilization
SUFCB	Sulfur Assimilation CB-D Fusion, Cytosolic
SW802	Salt Water 802 Medium
Tam41	Translocator Assembly And Maintenance Protein 41
TBS	Tris-Buffered Saline
TCA	Tricarboxylic Acid Or Citric Acid
TDH	Threonine Dehydrogenase
Thr	Threonine
TIM	Translocator Of The Inner Mitochondrial Membrane
TnaA	Tryptophanase
Trp	Tryptophan
UBI	Ubiquionone Biosynthesis Genes
UPGMA	Unweighted Pair Group Method With Arithmetic Mean
UQ	Ubiquinone
UTR	Untranslated Region
YPAD	Yeast Extract, Peptone, Adenine And Dextrose Medium

Acknowledgements

I would like to thank Dr. Andrew Roger for being an incredible mentor, role model and friend. Your endless patience, kindness and encouragement over these past 7 years have been truly inspiring. Grad school has been the best years of my life, which I partially attribute to you giving me the freedom to explore all avenues of my research. Thank you for taking a chance on a (then quiet) undergrad that knocked on your office door -- who eventually relentlessly teased your homogenous wardrobe choices.

I would also like to thank my supervisory and thesis committees: Dr. Melanie Dobson, Dr. Barbara Karten, Dr. Claudio Slamovits and Dr. Christian Blouin. Thank you to Dr. Dobson and Joyce Chew and for teaching me the secrets of molecular biology and yeast genetics and Dr. Barbara Karten for lipid extraction expertise and discussions (even if it did not make it into the thesis).

I owe my intact sanity to my best friend and colleague Dr. Laura Eme. Your support, wisdom and contributions to our projects over these past years has been unparalleled. You were always there to help me find a coding bug, provide critical comments on my thesis or to teach me the ways of phylogenetics. I look forward to being life-long collaborators! Thank you to Michelle Leger for critical comments of my thesis, a friendship that started nearly 7 years ago, and knowledge of anaerobic metabolism. To Dr. Martin Kolisko, I owe a great deal of thanks for introducing me to python and beer as well as listening to me ramble on about crazy evolutionary scenarios. Thank you to Tommy Harding for sharing your knowledge of transcriptomics and bacteriology with me; I'm certain I could not have grown *Rhodospirillum* without you.

I would like to thank the following people: Dr. Matthew Brown for introducing me to *Pygsuia*, who knew a protist could be so interesting?; Dr. Alexandra Stechmann, Dr. Anastasious Tsaousis and Dr. Sara Diaz for mentoring me as an undergraduate; Dr. Ford Doolittle, Dr. Alastair Simpson, Dr. Michael Gray and Dr. John Archibald for critical discussions; Jacquie de Mestral for technical expertise and keeping us all in order; RogerLabCult members past and present, Susan Sharpe, Jiwon Yang, Dr. Daniel Gaston, Javier Alfaro, Dr. Eleni Gentekaki, Wanda Danilchuk, Dr. Laura Hug, and Dr. Jessica Leigh;

Dr. Conor (amazing socks) Meehan for python and KEGG discussions; Dr. David Spencer for seemingly endless knowledge of everything; Prof. Kiyoshi Kita, Dr. Daniel Inaoka and lab members at Todei for Complex II expertise and welcoming me to Tokyo for a studentship; and Dr. Vladamir Hampl, Dr. Mark van der Giezen, and Dr. Graham Dellaire for excellent collaborations.

To my fellow graduate students that I owe a great deal of gratitude to: Dennis Orton for support and understanding; Dale Corkery for critical discussions on microscopy and cell culture; Sander Roy for 4 am enzymology cram sessions and overall chemistry wisdom; Nicole Graves for a smile that can cure anything; Carolyn Robinson for letting me steal antibodies and use equipment; Tupper students for Friday discussions, Greg McCluskey, Antonietta Pietrangelo, Dave Soutar, Barry Kennedy and Marie-Laurence Tremblay. Outside of academia I owe a lot of thanks to my support team: Stan Selig, Emily Pitts, Gazheek Schmee, Amanda Schma, Mike Roy, Nick Allen, Matthew Macdowell and Sarah Bercu.

Finally thank you to my parents Heather and Dave for always believing in me, keeping my freezer stocked with food and moving my apartment countless times; to Shannon and Rick listening to me talk… a lot, wish you were here; to my siblings Patrick, Kevin, Cory, Sam and Hayley for encouraging me to stay in school and save the world. Not sure protistology will save the world, but I certainly enjoy it.

Chapter 1 Introduction

1.1 On The Origin Of Mitochondria

In the early 20th century, Russian biologist Constantin Mereschkowsky proposed that chloroplasts (or 'chromatophores') evolved from symbiosis of free-living bacteria and the eukaryotic host cell (Mereschkowsky 1905). However, Mereschokowsky vehemently denied a bacterial origin of mitochondria. In the decades to follow, 'symbiogeneticsits' began to posit that a symboitic origin of mitochondria (*i.e.,* descent from a separate organism within a host cell) might better explain their complexity than traditional Darwinian evolutionary (*i.e.,* descent by modification) processes (Portier 1918; Wallin 1927). These radical ideas were quickly dismissed by the scientific community including prominent American cell biologist Edmund Wilson (Wilson 1925). It was not until early work by Lynn Margulis many years later that this symbiotic theory began to gain serious attention by mainstream biologists. In one of her first papers, Margulis (Lynn Sagan at the time) hypothesized that mitochondria, plastids and flagellar basal bodies were once free-living prokaryotes (Sagan 1967; Schwartz 1970). Now, over 40 years later, two out of three of these predictions (*i.e.,* the origins of mitochondria and plastids) are the fundamental components of the modern 'Endosymbiotic Theory' for the origin of eukaryotic organisms. Margulis hypothesized that an anaerobic host bacterium engulfed and formed a symbiotic relationship with an aerobically respiring bacterium. The engulfed bacterium eventually lost its autonomy within the host cell cytoplasm and became the mitochondrion.

One of the strongest lines of evidence in support of the endosymbiotic theory is the presence of a vestigial genome (mtDNA) in mitochondria of extant model organisms (Nass and Nass 1963). Early research into the origins of mtDNA noted prokaryotic-like features of the genome, specifically the sequence similarity of the small subunit RNA of the mitochondrial ribosome to bacterial homologs (Bonen *et al.* 1977). As sequencing technology advanced, the gene content of mtDNA became better characterized. While content is variable across eukaryotes, mtDNA typically encodes ribosomal RNA subunits and proteins, tRNA genes and respiratory complex proteins (Anderson *et al.* 1981).

Comparative mitochondrial genomics efforts in the last few decades have since revealed that the chromosome structure, ploidy, gene number and gene content of present-day mitochondria is quite variable across the tree of eukaryotes (Gray *et al.* 1999; Gray *et al.* 2004). Gene content of mitochondrial genomes can vary from low (*e.g.,* five genes in *Plasmodium falciparum* (Wilson and Williamson 1997)) to relatively high (*e.g.,* over 100 in jakobid protists (Burger *et al.* 2013)). In the 1980s and 90s, phylogenetic analyses consistently showed that the genes encoded on mitochondrial genomes appeared to be closely related to the those of modern-day α-proteobacteria, specifically those of the obligate intracellular parasites of the Rickettsiales division such as *Rickettsia* (Yang *et al.* 1985; Olsen *et al.* 1994; Andersson *et al.* 1998; Sicheritz-Pontén *et al.* 1998). In the last few years with the availability of whole genome sequences from a vast diversity of α-proteobacteria, this question has been revisited. A recent report suggested that the closest relative to the mitochondrial genome could be the newly described SAR11 group of bacteria (including the *Pelagibacteraceae* (Thrash *et al.* 2011)), although more careful analyses suggested the latter result could represent a phylogenetic artifact and that the true affinities of mitochondrial genomes remain unclear (Rodríguez-Ezpeleta and Embley 2012).

The proteomes of modern day mitochondria of animal model systems are comprised of over 1000 proteins (Calvo and Mootha 2010; Smith *et al.* 2012) which is significantly more than the coding potential of any known mitochondrial genomes. Most mitochondrial proteins are encoded by genes of the nuclear genome and imported into mitochondria post-translationally. Some of these proteins were originally encoded by the endosymbiont genome and subsequently transferred to the host nuclear genome via endosymbiotic gene transfer (Timmis *et al.* 2004). The gene would need to acquire a signal for retargeting the protein product to the endosymbiont. In fact, studies have shown that approximately 5% of bacterial genes have predicted mitochondrial targeting sequences (Lucattini *et al.* 2004; Ueda *et al.* 2008) and random peptides enriched in leucine, arginine and serine are sufficient to target green-fluorescent protein into yeast mitochondria (Allison and Schatz 1986). Once the nucleus-encoded gene product was

successfully targeted to the mitochondrion, any mutations (*i.e.,* frame shifts, missense or nonsense mutations) on the organelle-encoded version of the gene would be selectively neutral. Eventually the nucleus-encoded gene product would be a permanent fixture of the organelle as the organelle-encoded gene degenerated and was ultimately lost.

1.2 The Archezoa Hypothesis And Amitochondriates

After Margulis' elaboration of the modern endosymbiotic theory started gaining acceptance, a number of eukaryotic microbes were discovered to be lacking typical cristate mitochondria (Cavalier-Smith 1983). This led to the proposal of the Archezoa hypothesis that suggests these protists were primitively amitochondriate and therefore diverged prior to the mitochondrial symbiosis (Cavalier-Smith 1983). These lineages included parabasalids, entamoebidae, microsporidians, and metamonads (Cavalier-Smith 1983). A few years later, phylogenetic analysis of the cytoplasmic eukaryotic small subunit ribosomal DNA (SSU rDNA) placed many of these amitochondriate taxa as basal branching lineages of the eukaryotic tree (Vossbrinck *et al.* 1987; Cavalier-Smith 1989; Sogin *et al.* 1989), consistent with a their putative 'archezoan' nature. However, in the 90s, the earliest hint that the Archezoa hypothesis may not be correct was provided in a study that showed antibodies against human or rodent HSP60 (a known mitochondrial protein) cross reacted with proteins in *Giardia lamblia* cells yielding a punctate localization pattern in immunofluorescence micrographs (Soltys and Gupta 1994) suggesting this organism could harbour vestigial mitochondrial organelles. Interestingly, when Cavalier-Smith commented on the Archezoa hypothesis in light of Soltys and Gupta's immunofluorescence findings, he recognized that identifying a *hsp60* gene in the amitochondriates (specifically *Giardia*) could help decide if these organisms were in fact primitively amitochondriate, although he cautioned the reader to "not be too hopeful" (Cavalier-Smith and Chao 1996). Towards the end of the 1990s it was the identification and phylogenetic analysis of this gene in *Giardia* and other amitochondriates that eventually led to the end of the Archezoa hypothesis (Roger *et al.* 1998). These molecular data supported earlier hypotheses regarding the potential mitochondrial ancestry of the hydrogenosomes of *Trichomonas vaginalis* (Bozner 1996;

Germot *et al.* 1996; Horner *et al.* 1996; Roger *et al.* 1996). Small double membrane bound organelles present in these diverse organisms were described and variously named hydrogenosomes, mitosomes, 'mitochondrion-like organelles', and eventually more accurately named 'mitochondrion-related organelles' (MROs) to indicate their evolutionary relationship with mitochondria.

1.3 Diversity And Function Of Mitochondria

Before discussing the metabolic diversity of MROs, I will first outline the functions of 'classical' aerobic mitochondria. Much of our current understanding of mitochondrial functions derives from studies of model organisms represented within animals, fungi or plants. As discussed above, the vast majority of mitochondrial proteins in such organisms are encoded in the nucleus and transported into the organelles post-translationally. Therefore, several essential steps in the evolutionary integration of mitochondria involved the advent of the various components of the protein import apparatus. This complex assembly of protein carriers helps direct mitochondrion targeted proteins to one of four sub-organellar components: (i) the outer mitochondrial membrane (OMM), (ii) the intermembrane space (IMS), (iii) the inner mitochondrial membrane (IMM) and (iv) the mitochondrial matrix (MM) (Geli and Glick 1990).

Important functions of mitochondria include iron-sulfur (Fe-S) cluster generation, amino and fatty acid, phospholipid, vitamin, and steroid metabolism. However, the best-known function of mitochondria is the synthesis of ATP by oxidative phosphorylation. In this pathway, pyruvate is oxidized to acetyl-CoA by pyruvate dehydrogenase (PDH) and fed into the tricarboxylic acid (TCA) cycle which in turn generates reducing equivalents (NADH and $FADH_2$) that are used by the electron transport chain (ETC) to generate a proton gradient which fuels ATP synthesis ultimately reducing O_2 to H_2O (Figure 1-1). However, the mitochondria of some organisms that occupy low oxygen environments have adapted to function without oxygen. Recently, Müller and colleagues proposed a new classification scheme for mitochondria and these anaerobic MROs (Classes I-V) based solely on the different types of energy metabolism they possess: Class I aerobic mitochondria, Class II anaerobic mitochondria, Class III hydrogen-producing

mitochondria (HPM), Class IV hydrogenosomes and Class V mitosomes. Distribution of the different MRO types found in eukaryotes is shown in Figure 1-2. General trends and specific examples of these organelles are discussed below and summarized in (Figure 1-3 to Figure 1-15).

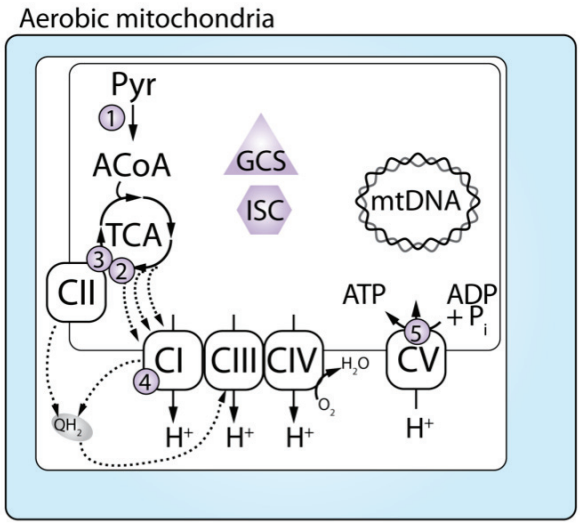

Figure 1-1: Select metabolic pathways in typical aerobic mitochondria. The cytoplasm and mitochondria are shown in blue and white respectively. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by pyruvate dehydrogenase [1] and fed into the tricarboxylic acid cycle (TCA). Within the TCA cycle, succinyl-CoA is converted to succinate by succinyl-CoA synthetase [2]. Electrons from succinate are transferred to ubiquinone (Q) generating fumarate and ubiquinol (QH_2) by Complex II (CII;][3]). NADH is oxidized to NAD+ by NADH dehydrogenase (Complex I, CI [4]) while concurrently pumping protons and passing electrons to Q generating QH_2. Electrons from ubiquinol are shuttled through complex III and complex IV ultimately reducing oxygen to water and generating a proton gradient. The proton gradient is used to fuel ATP synthesis by complex V (CV; [5]). Other pathways include the glycine cleavage system (GCS) and Fe-S cluster biogenesis system (ISC).

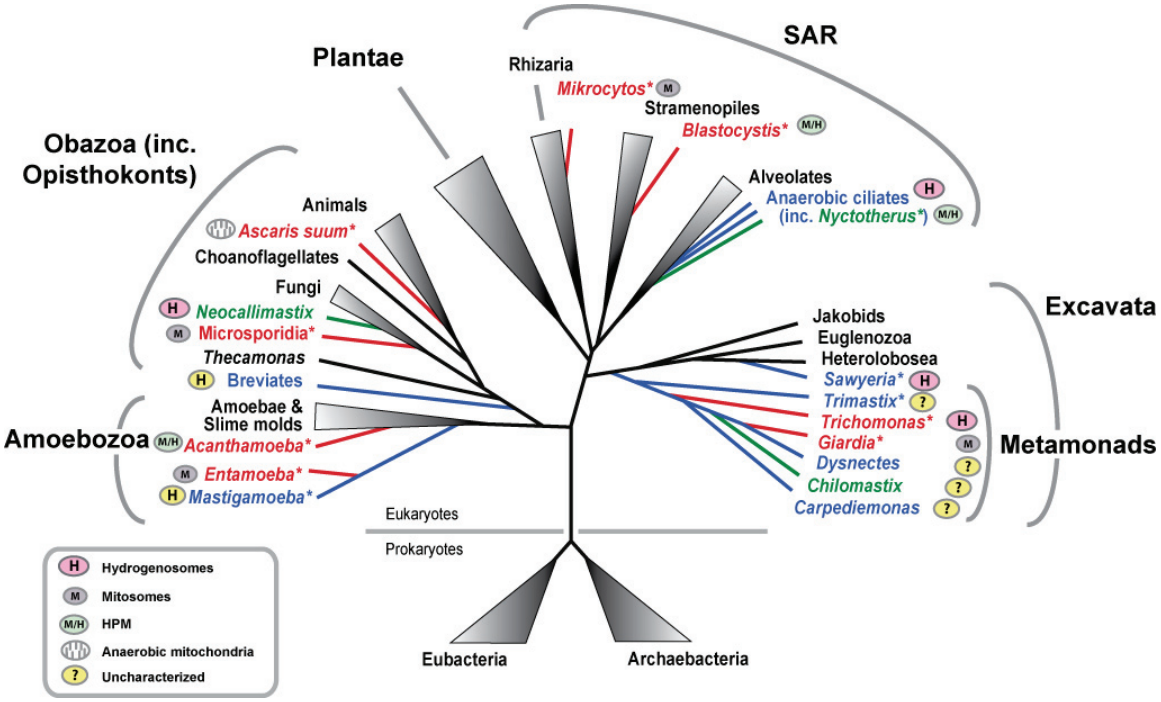

Figure 1-2: Schematic representation of mitochondrion-related organelles across the tree of eukaryotes. Next to each organism, the putative MRO is shown. Organisms that will be discussed in more detail are indicated with an asterisk. Parasitic (red), commensal (green) and free-living (blue) lineages are shown. SAR, Stramenopile-alveolate-rhizaria group; HPM hydrogen-producing mitochondria. Reproduced with permission from Dr. Alastair Simpson.

1.3.1 Anaerobic Mitochondria

Some species of fungi including *Fusarium oxysporum* are facultative anaerobes. Under low-oxygen conditions, the mitochondria of these fungi respire using nitrate as the terminal electron acceptor employing the electron transport chain components CIII and CIV together with a series of denitrification reductases (Kobayashi *et al.* 1996). Similar to aerobic respiration, these electron transfer reactions generate a proton gradient that fuels ATP synthesis via ATP synthase (Complex V, CV).

Other facultative anaerobes, such as helminths and nematodes, have life stages during which they experience hypoxic conditions (Rew 1974; Föll *et al.* 1999). Through regulated expression of various proteins, the metabolism of the anaerobic mitochondria of these organisms is reconfigured to generate ATP by substrate level phosphorylation, ATP synthase and a modified respiratory chain (Müller *et al.* 2012). As in aerobic mitochondria, electrons are transferred to CI via NADH to pump protons into the inter-

membrane space that fuel ATP synthesis via CV (Figure 1-3). However, in contrast to the metabolism of typical aerobic mitochondria, these electrons are transferred to a specialized quinone (rhodoquinone; RQ) and not ubiquinone (UQ) to generate rhodoquinol (RQH_2). A process called 'malate dismutation' helps to replenish the RQ pool using fumarate hydratase and CII that preferentially catalyze the reverse reactions to their aerobic mitochondria counterparts (*i.e.,* malate dehydration and fumarate reduction) (Tielens 1994; Tielens *et al.* 1998). In this pathway, malate is imported into the organelle and converted to fumarate via fumarate hydratase (Butler *et al.* 2012). This fumarate is reduced by CII functioning in reverse as a fumarate reductase (FRD) using electrons from RQH_2. These organisms also generate ATP via propionoyl-CoA mediated substrate-level phosphorylation (Figure 1-3). Through a series of CoA transfer reactions, succinate is converted to methylmalonyl-CoA which is used as a substrate for propionoyl-CoA carboxylase (PPC) to generate ATP and propionoyl-CoA. This cycle is restarted by transferring the CoA moiety to a new succinate molecule via acetate:succinate CoA transferase 1B (ASCT) generating propionate, the ultimate end product of metabolism.

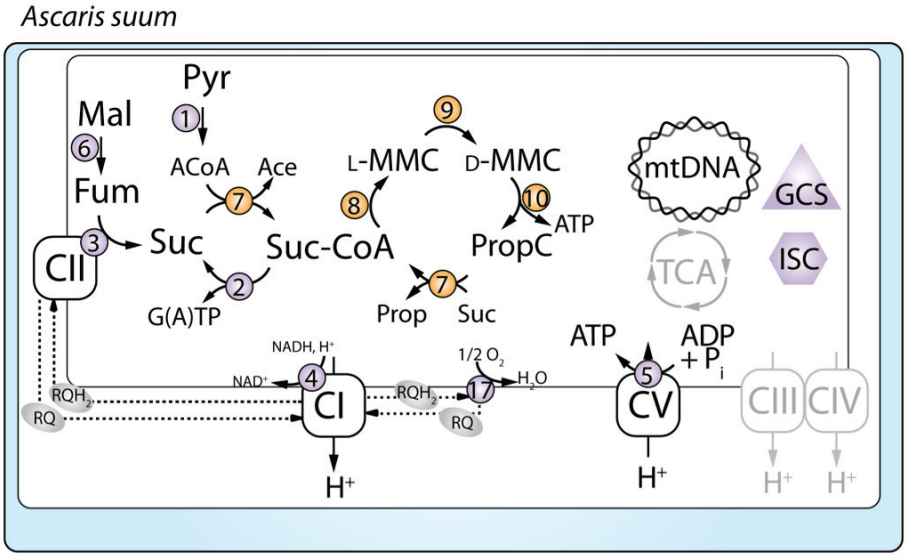

Figure 1-3: Select metabolic pathways of the facultative animal *Ascaris suum*. Numbering and labels are shown as in Figure 1-1, pathways not active at low oxygen conditions are shown in grey. Enzymes typically associated with anaerobic metabolism are shown in orange. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by pyruvate dehydrogenase [1]. A proton gradient is generated via CI generating rhodoquinol (RQH_2). Malate dismutation: Malate (Mal) is converted

to fumarate (Fum) via fumarate hydratase [6]. Electrons from rhodoquinol and transferred to fum to generate succinate (suc) vai CII. CoA from acetyl-CoA is transferred to succinate to generate acetate (Ace) and succinyl-CoA (Suc-CoA). Suc-CoA can either be converted back to suc using SCS, or enter the propionate cycle. Suc-CoA is converted to L-methylmalonyl-CoA (L-MMC) via L-MMC decarboxylase [8] which is epimerized to D-MMC via MMC epimerase [9] and finally decarboxylated to form propionyl-CoA (PropC). Oxygen can be converted to water by alternative oxidase regenerating RQ from RQH_2. The proton gradient is used to fuel ATP synthesis by complex V (CV; [5]).

1.3.2 Hydrogen-Producing Mitochondria

Hydrogen-producing mitochondria (HPMs) have been reported in the amoebozoan *Acanthamoeba castellanii*, the heterolobosean *Nagleria gruberi,* the ciliate *Nyctotherus,* and the stramenopile *Blastocystis* sp*.* (Stechmann *et al.* 2008; Fritz-Laylin *et al.* 2010; de Graaf *et al.* 2011; Leger *et al.* 2013). In general, HPM have retained an organellar genome and couple ATP generation (via substrate level phosphorylation) to hydrogen production. HPMs have retained at least CI and CII and perform malate dismutation to generate succinate much like anaerobic mitochondria. The main difference between HPM and anaerobic mitochondria is the ability of the former to make hydrogen.

Acanthamoeba

Acanthamoeba castellanii is a member of the Amoebozoa that can inhabit different environments including soil bacterial biofilms or human contact lenses (van Klink *et al.* 1992). Upon exposure to complete anoxia, *Acanthamoeba* encysts. However under low oxygen (6-16% oxygen) some species of *Acanthamoeba* grow faster than under aerated conditions (Cometa *et al.* 2011). Interestingly, the ultrastructure and general metabolism of *Acanthamoeba* mitochondria resemble classical aerobic organelles (Gawryluk *et al.* 2014). However, investigations into specific metabolic pathways revealed that this organism encodes genes responsible for hydrogen-production (Leger *et al.* 2013). Specifically, *Acanthamoeba* possesses enzymes involved in 'extended glycolysis' including [FeFe]-hydrogenase (HYDA), hydrogenase maturation proteins (HYDE, F and G) and pyruvate:ferredoxin oxidoreductase (PFO) (Figure 1-4). Unlike pyruvate metabolism in anaerobic mitochondria, in *Acanthamoeba*, pyruvate is converted to acetyl-CoA via PFO or PDH. Electrons from ferredoxin are then used by an [FeFe]-hydrogenase to

generate hydrogen gas. Most of these proteins have predicted N-terminal mitochondrial targeting sequences and some were shown experimentally to localize to the mitochondria of *Acanthamoeba* (Leger *et al.* 2013) suggesting this organelle has the hallmark features of both classical aerobic mitochondria and hydrogenosomes (discussed below in section 1.3.3). Like *Ascaris*, the *Acanthamoeba* MROs are suspected to perform malate dismutation to make succinate and ultimately succinyl-CoA via fumarate hydratase, CII, and ASCT (subtype 1A/1B) (Leger *et al.* 2013).

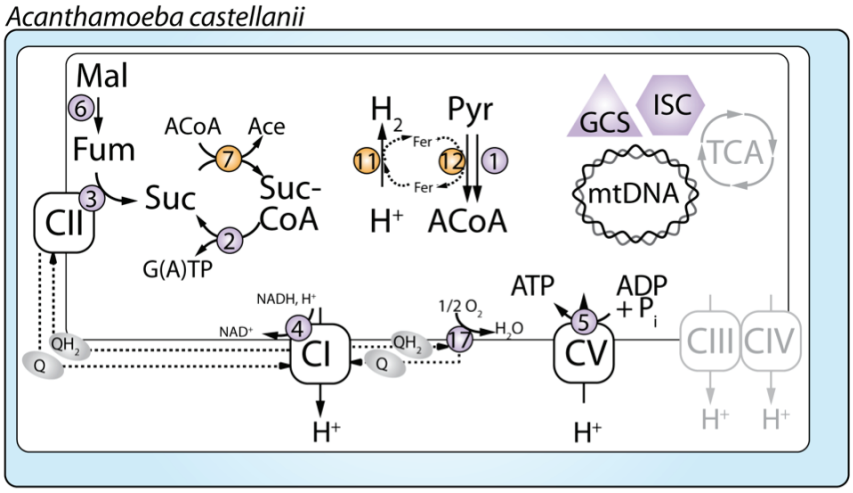

Figure 1-4: Select metabolic pathways of *Acanthaemoba castellanii* mitochondria. Numbering and labels are shown as in Figure 1-1,1-3, pathways not active at low oxygen conditions are outlined in grey. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue, electron transfer reactions are shown with dotted arrows. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by PDH [1] or pyruvate:ferredoxin oxidoreductase (PFO) [12]. Electrons from ferredoxin can be used by hydrogenase (HYDA) [11] to generate hydrogen. A proton gradient is generated via CI generating ubiquinone (QH_2). Electrons from ubiquinol and transferred to fum to generate succinate (suc) via CII. CoA from acetyl-CoA is transferred to succinate to generate acetate (Ace) and succinyl-CoA (Suc-CoA). Suc-CoA can then be converted back to succinate via SCS [2].

Naegleria gruberi

Naegleria gruberi is a free-living heterolobosean that occupies aerobic and microaerobic environments and is closely related to the deadly brain-eating amoeba *Naegleria fowleri* (Fritz-Laylin *et al.* 2010). Like *Acanthamoeba, Naegleria* has 'classical' aerobic mitochondria. However, homologs of MRO proteins such as [FeFe]-hydrogenase

were identified in the genome and were predicted to function in mitochondria, suggesting that *N. gruberi* had 'hydrogen-producing mitochondria' (Fritz-Laylin *et al.* 2010). Recent reports suggest the hydrogen production is present exclusively in the cytoplasm despite the presence of predicted N-terminal targeting sequences on hydrogenase and its maturases (Tsaousis *et al.* 2014). Nevertheless, *Naegleria gruberi* has other features in common with other organisms containing MROs including alternative energy generation machinery (via ACST) (Figure 1-5).

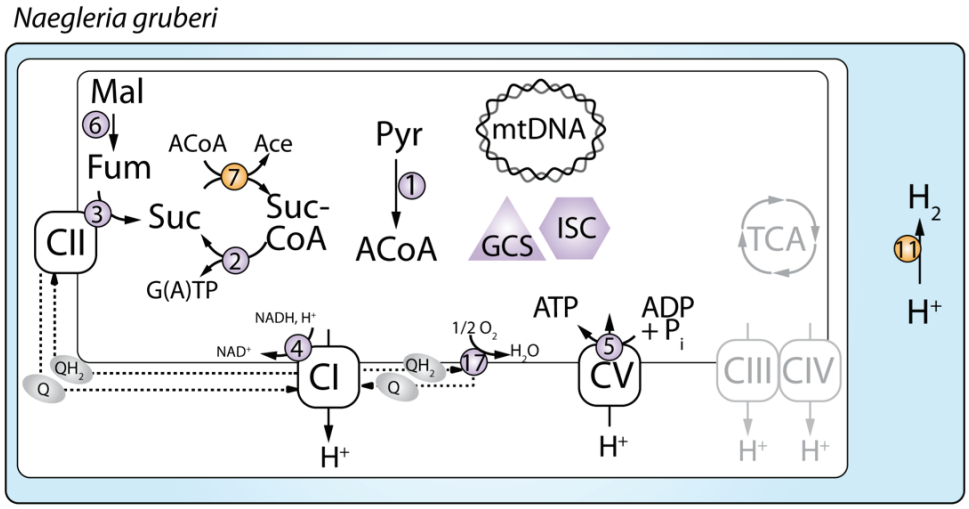

Figure 1-5: Select metabolic pathways of *Naegleria gruberi* mitochondria. Numbering and labels are shown as in Figure 1-1,1-3 to 1-4, pathways not active at low oxygen conditions are outlined in grey. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue, electron transfer reactions are shown with dotted arrows. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by PDH [1]. A proton gradient is generated via CI generating ubuquinone (QH_2). Electrons from ubiquinol and transferred to fum to generate succinate (suc) via CII. CoA from acetyl-CoA is transferred to succinate to generate acetate (Ace) and succinyl-CoA (Suc-CoA) via ASCT [7]. Suc-CoA can then be converted back to succinate via SCS [2]. Hydrogen production occurs exclusively in the cytoplasm.

***Nyctotherus* sp.**

Nyctotherus is a ciliate that inhabits the hindgut of cockroaches. Much like aerobic and anaerobic mitochondria, the HPM of *Nyctotherus* uses PDH to generate acetyl-CoA and malate dismutation to generate fumarate using RQH_2 (Boxma and Graaf 2005). In a pathway similar to anaerobic mitochondrion, the CoA moiety from acetyl-CoA is transferred to succinate via an ASCT (subtype 1A) to generate succinyl-CoA which is used by the TCA cycle enzyme succinyl-CoA synthetase to generate ATP (Boxma and Graaf

2005; de Graaf *et al.* 2011) (Figure 1-6). *Nyctotherus* also generates α-ketoglutarate (via glutamate dehydrogenase) that can be converted to succinyl-CoA (via α-ketoglutarate dehydrogenase) to be used as a substrate for SCS. The organelles of *Nyctotherus* use a specialized [FeFe]-hydrogenase that is fused to homologs of CI subunits (51 and 24 kDa subunits). This enzyme uses electrons donated from NADH to reduce protons to generate hydrogen gas. Interestingly, many rumen ciliates including *Nyctotherus* are often observed in symbiotic relationships with prokaryotes (Gijzen *et al.* 1991; van Hoek *et al.* 2000), many of which have been identified as methanogenic endosymbiotic archaea that specifically associate with the HPMs, consuming H_2 produced by the organelle (Gijzen *et al.* 1991; van Hoek *et al.* 2000).

Unfortunately, our current understanding of the *Nyctotherus* metabolism is limited due to the lack of a high coverage transcriptome or genome project. However, from the available expressed sequence tag (EST) data, we know that this organism encodes a variety of putative MRO proteins including proteins involved in amino acid metabolism, organellar protein import and folding reactive oxygen species (ROS) defense.

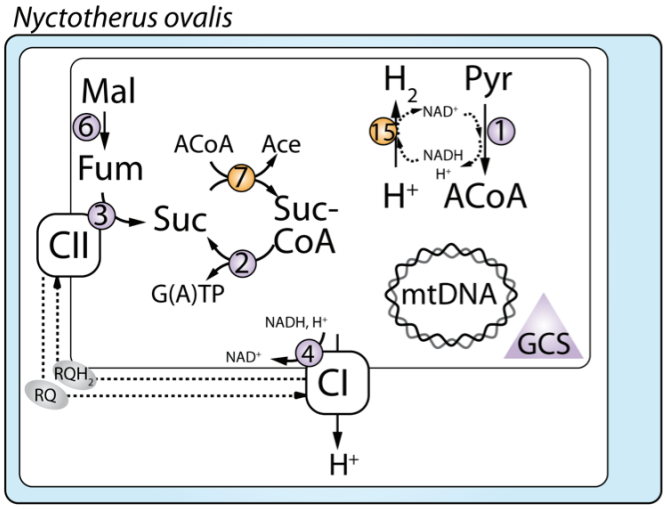

Figure 1-6: Select metabolic pathways of *Nyctotherus ovalis* mitochondria. Numbering and labels are shown as in Figure 1-1, 1-3 to 1-5, pathways not active at low oxygen conditions are shown in grey. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue, electron transfer reactions are shown with dotted arrows. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by PDH [1]. A proton gradient is generated via CI generating rhodoquinol (RQH_2). Electrons from rhodoquinol and transferred to Fum to generate Suc via CII. CoA from acetyl-CoA is transferred to succinate to generate acetate (Ace) and

succinyl-CoA (Suc-CoA) via ASCT [7]. Suc-CoA can then be converted back to succinate via SCS [2].

***Blastocystis* sp.**

Blastocystis is a unicellular stramenopile that inhabits strictly anoxic environments such as the animal gut (Zierdt 1983). From the genomic and transcriptomic data available for different subtypes of *Blastocystis* it is clear that this organism's genome encodes multiple acetyl-coA generating proteins including PDH, pyruvate:ferredoxin oxidoreductase (PFO) and pyruvate:NADP+ oxidoreductase (PNO; a specialized PFO with a C-terminal P450 reductase domain) (Lantsman *et al.* 2008; Stechmann *et al.* 2008; Denoeud *et al.* 2011) (Figure 1-7). However, to date, only PNO activity has been detected biochemically within the MROs of *Blastocystis* (Lantsman *et al.* 2008). The acetyl-CoA generated by this reaction is used by ASCT (1B and 1C subtype) to generate succinyl-CoA, which is then used by SCS to generate ATP (Figure 1-7), much like anaerobic mitochondria (Figure 1-3). While hydrogen production has not actually been detected in the MROs of *Blastocystis*, genomic and transcriptomic evidence suggests that there are multiple MRO-targeted HYDA proteins that are fused to flavodoxin domains (Stechmann *et al.* 2008). The unique domain composition of the *Blastocystis* HYDA proteins might explain the inability to detect hydrogenase activity biochemically (Lantsman *et al.* 2008). Nevertheless, the *Blastocystis* MRO is still putatively classified as a Hydrogen-producing mitochondrion (Müller *et al.* 2012). Analyses of the genome sequences of *Blastocystis* subtype 7, indicates its HPM participates in a variety of other pathways including amino acid metabolism, Fe-S cluster biogenesis, ROS defense, and fatty acid biosynthesis (Denoeud *et al.* 2011).

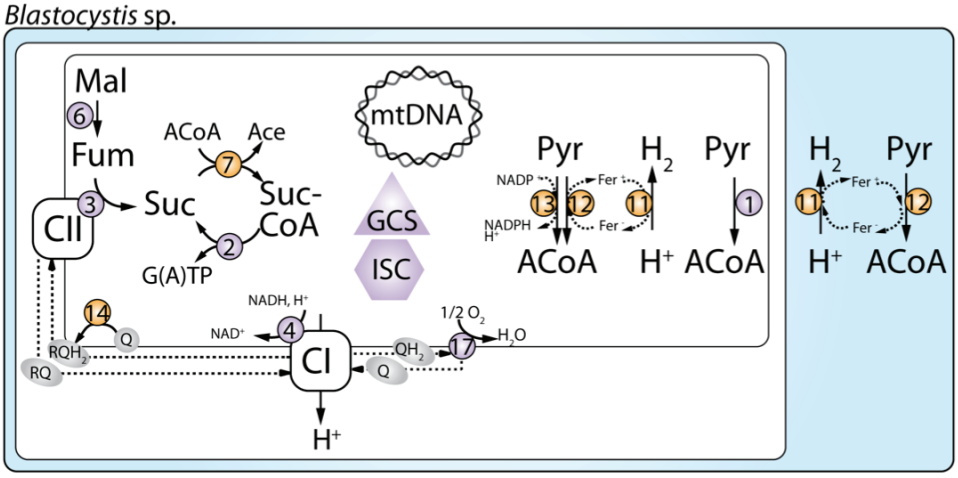

Figure 1-7: Select metabolic pathways of *Blastocystis* sp. mitochondria. Numbering and labels as in Figure 1-1, 1-3 to 1-6, pathways not active at low oxygen conditions are shown in grey. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue, electron transfer reactions are shown with dotted arrows. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by PDC [1], PFO [12] or PNO [13]. Electrons from ferreodoxin are used to produce hydrogen via hydrogenase [11]. A proton gradient is generated via CI generating rhodoquinol (RQH_2). Electrons from rhodoquinol and transferred to fum to generate suc via CII. CoA from acetyl-CoA is transferred to succinate to generate acetate (Ace) and succinyl-CoA (Suc-CoA) via ASCT [7]. Suc-CoA can then be converted back to succinate via SCS [2]. Rhodoquinone is predicted to be synthesized from ubiquinone using the rhodoquinone biosynthesis enzyme RQUA [14].

1.3.3 Hydrogenosomes

Hydrogenosomes are metabolically distinct from mitochondria, and lack most hallmark mitochondrial features such as mtDNA, oxidative phosphorylation, the ETC, the TCA cycle and other oxygen-dependent processes. These organelles couple ATP synthesis to hydrogen production without using electron transport. Hydrogenosomes have been characterized in a variety of organisms including *Trichomonas vaginalis, Spironucleus salmoncida, Mastigamoeba balamuthi,* and free-living excavates.

Trichomonas vaginalis

Trichomonas vaginalis is a parabasalid flagellate that occupies the urogenital tract of humans and is the causative agent of trichomoniasis (Petrin *et al.* 1998). In *Trichomonas* hydrogenosomes, pyruvate is metabolized via the 'extended glycolysis' pathway much like in *Acanathamoeba* where pyruvate is converted to acetyl-CoA by PFO and the

electrons are eventually used by an [FeFe]-hydrogenase to reduce protons generating hydrogen gas (Figure 1-8). The CoA from acetyl-CoA is transferred to succinate to generate succinyl-CoA via ASCT (subtype 1C). Finally, the TCA cycle enzyme succinyl-coA synthetase (SCS) generates ATP by substrate level phosphorylation (Müller and Lindmark 1978; van Grinsven *et al.* 2008; Müller *et al.* 2012) (Figure 1-8).

A proteomic survey of the *Trichomonas* proteome detected over 500 proteins contained within the organelle (Schneider *et al.* 2011). This is simpler than mammalian mitochondrial proteomes that can have upwards of 1000 proteins (Calvo and Mootha 2010) or yeast mitochondria with 750 proteins (Sickmann *et al.* 2003). In addition to the core hydrogenosomal ATP generation machinery, proteins were found in the *T. vaginalis* MRO proteome that are predicted to be involved in amino acid metabolism, Fe-S cluster biogenesis and oxygen stress. However, it is important to consider that many of the proteins represented in the *T. vaginalis* proteome are in fact in-group paralogs (*i.e.,* these are not 500 unique proteins but many are recent duplicated copies). The vast majority of the proteins identified by Schneider and colleagues were not directly related to metabolism, but instead included GTPases, hypothetical proteins and proteins potentially associated with the surface of the organelle. In any case, protein number should not be taken as a proxy for metabolic potential or complexity.

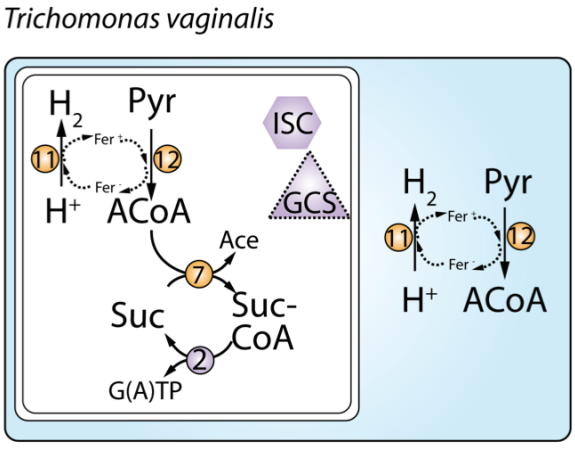

Figure 1-8: Select metabolic pathways of *Trichomonas vaginalis* hydrogenosomes. Numbering and labels are as shown in Figure 1-1, 1-3 to 1-7. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue, electron transfer reactions are shown with dotted arrows. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by PFO [12]. CoA

from acetyl-CoA is transferred to succinate to generate ace and suc-CoA via ASCT [7]. Suc-CoA can then be converted back to succinate via SCS [2].

Spironucleus salmonicida

Spironucleus is an anaerobic heterotrophic flagellate classified as a diplomonad that infects salmonid fish (Xu *et al.* 2014). Like other diplomonads, it has a small double-membrane bound organelle (Jerlström-Hultqvist *et al.* 2013). The sequencing of the nuclear genome and proteomic analysis of the hydrogenosome recently elucidated the true metabolic potential of the *Spironucleus* MRO (Jerlström-Hultqvist *et al.* 2013; Xu *et al.* 2014). These organelles participate in pyruvate oxidation via PFO, hydrogen maturation and evolution via the HYD proteins, Fe-S cluster generation and amino acid metabolism (Figure 1-9). Interestingly, in contrast to the organelles of the diplomonad *Giardia intestinalis,* the *Spironucleus* hydrogenosome appears to produce ATP. Unlike other MROs discussed above, the *Spironucleus* organelle generates ATP using acetyl-CoA synthetase and not the ASCT/SCS system. Acetyl-CoA produced by PFO is converted to acetate using acetyl-CoA synthetase (ACS) generating ATP by substrate level phosphorylation.

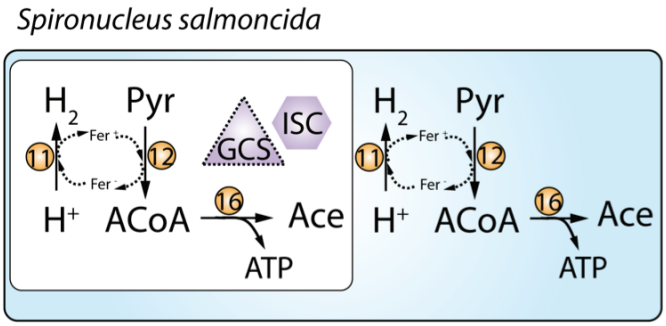

Figure 1-9: Select metabolic pathways of *Spironucleus salmonicida* hydrogenosomes. Numbering and labels are shown as in in Figure 1-1, 1-3 to 1-8. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue; electron transfer reactions are shown with dotted arrows. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by PFO [12]. CoA from acetyl-CoA is transferred to succinate to generate ace and suc-CoA via ASCT [7]. Suc-CoA can then be converted back to succinate via SCS [2]. At least one component of the glycine cleavage system was identified (dotted shape).

Free-living excavates

In recent years, many sequencing initiatives have focused on characterizing the MROs in the Excavata lineage of eukaryotes including *Psalteriomonas lanterna, Sawyeria marylandensis* and *Trimastix pyriformis*. *Psalteriomonas* and *Sawyeria* are amoeboflagellates of the heterolobosean lineages of excavates (Broers *et al.* 1990; O'Kelly *et al.* 2003). The *Psalteriomonas* EST survey reported homologs of HSP70, Hydrogenase and PFO and the 51 kDa subunit of CI (Graaf *et al.* 2009). However, due to the limited sequencing coverage, the vast majority of the *Psalteriomonas* MRO metabolic potential is unknown. The larger EST survey of *Sawyeria* revealed homologs for many of the proteins found in *Trichomonas vaginalis* hydrogenosomes including [FeFe]-Hydrogenase, PFO, chaperones, CI subunits and Fe-S cluster biosynthesis proteins. Interestingly, *Sawyeria* appears to encode a variety of pathways linked to amino acid metabolism that are not present in *Trichomonas*, such as a full glycine cleavage system, proline, serine, and ornithine metabolism and branched chain amino acid degradation.

Finally, MROs have been described in many different species of the flagellated metamonad genus *Trimastix* (Brugerolle and Patterson 1997; O'Kelly *et al.* 1999; Simpson *et al.* 2000; Stechmann *et al.* 2006; Zubáčová *et al.* 2013). A recent transcriptomic survey of *Trimastix pyriformis* revealed transcripts encoding proteins linked to hydrogen production were predicted to localize to the organelle (Figure 1-10) (Zubáčová *et al.* 2013). The *Trimastix* organelle also seems to participate in amino acid metabolism and even encodes the Kreb's cycle enzyme aconitase (not shown). To date, the typical ASCT or ACS enzymes have not been identified in the cytoplasm or organelle of *Trimastix*. Therefore the exact mechanism of ATP generation in this organism, beyond glycolysis, is unknown.

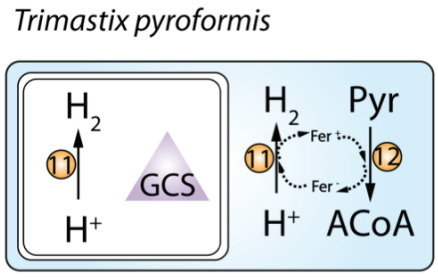

Figure 1-10: Select metabolic pathways of *Trimastix pyroformis* hydrogenosomes. Numbering and labels are shown as in in Figure 1-1, 1-3 to 1-9. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue; electron transfer reactions are shown with dotted arrows. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by PFO.

Mastigamoeba balamuthi

Mastigamoeba balamuthi is a free-living amoebozoan that inhabits low oxygen environments and is a member of the Archamoeba lineage (Chavez *et al.* 1986) that also includes the better studied human pathogen *Entamoeba histolytica*. In fact, *M. balamuthi* was one of the first non-parasitic organisms shown to have a MRO (Gill *et al.* 2007). The genome sequence of *Mastigamoeba* was recently released in Genbank as part of another project (Nyvltová *et al.* 2013), however a complete assembly and genome analysis has not been completed to date. *Mastigamoeba* MROs appear to participate in pyruvate oxidation, hydrogen evolution, sulphate activation and amino acid metabolism (Figure 1-11). Similar to *Spironucleus* hydrogenosomes, *Mastigamoeba* hydrogenosomes participate in ATP generation via ACS. Unlike the hydrogenosomes discussed above, *Mastigamoeba* encodes some proteins linked to the respiratory chain and electron transport (complex II, electron transferring flavoprotein and rhodoquinone metabolism) (Nycltová, Stairs *et al.*, unpublished). However, unlike other HPMs, this relict respiratory chain does not have canonical proton pumping components (*e.g.,* CI), therefore, it is unclear if the *M. balamuthi* MROs have a proton gradient. The *Mastigamoeba* organelle with a partial respiratory chain represents an intermediate state between HPMs and hydrogenosomes.

Mitochondria and MROs typically synthesize Fe-S clusters using the Iron Sulfur Cluster biosynthesis (ISC) system. However, in the archamoebal lineages (*i.e., Mastigamoeba* and *Entamoeba*), the ISC system has been lost and replaced with a

homologous system called the Nitrogen Fixation (NIF) system (Ali *et al.* 2004; van der Giezen *et al.* 2004; Maralikova *et al.* 2010; Nyvltová *et al.* 2013). In *Mastigamoeba*, the NIF system exists in two copies, one of which is targeted and localizes to the MRO and the other that is cytosolic, suggesting Fe-S cluster biosynthesis occurs both in the organelle and in the cytoplasm (Nyvltová *et al.* 2013).

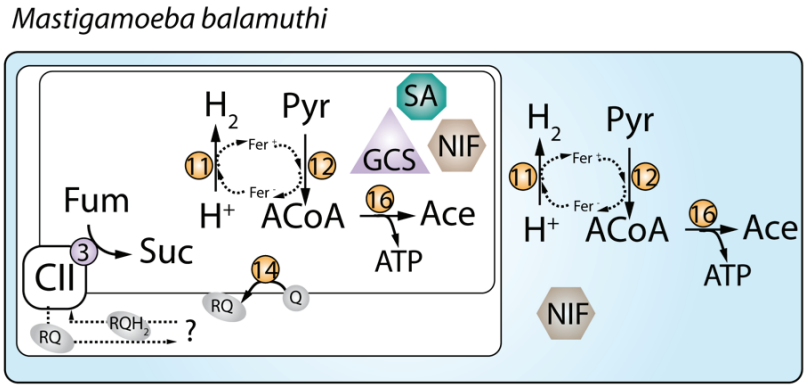

Figure 1-11: Select metabolic pathways of *Mastigamoeba balamuthi* hydrogenosomes. Numbering and labels are as shown in in Figure 1-1, 1-3 to 1-10. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue; electron transfer reactions are shown with dotted arrows. Pyruvate (Pyr) is oxidized to acetyl CoA (ACoA) by PFO [12] and coupled to hydrogen production [11]. Malate dismutation to succinate occurs as described above. Acetyl-CoA is converted to Ace via ACS to generate ATP. RQ is likely synthesized from UQ using RQUA [14]. Fe-S cluster biosynthesis, NIF; Sulfate activation pathway (SA).

1.3.4 Mitosomes

Any MRO that does not make ATP is classified as a mitosome. In contrast to the other classes of organelles, there is no universal function associated with mitosomes. Mitosomes have been identified exclusively in parasites from four of the major eukaryotic lineages: Amoebozoa (*e.g., Entamoeba,* Tovar *et al.* 1999), Stramenopiles-Alveolates-Rhizaria (*e.g., Mikrocystis* (Burki *et al.* 2013)*, Cryptosporidium* (Riordan *et al.* 2003)), Excavata (*e.g., Giardia intestinalis* (Tovar *et al.* 2003)), and Opisthokonta (*e.g., Trachipleistophora* (Williams *et al.* 2002)). As for metabolic potential, the mitosomes of most of these organisms (excluding *Entamoeba*) function in Fe-S cluster biogenesis using the ancestral mitochondrial pathway (*i.e.,* the Iron Sulphur Cluster system; ISC) (Tachezy

et al. 2001; Goldberg *et al.* 2008; Burki *et al.* 2013) and do not participate in amino acid metabolism.

Entamoeba

Entamoeba histolytica is an anaerobic parasite that infects humans and other mammals. As discussed previously, in at least two species of Archamoebae (*Entamoeba* and *Mastigamoeba*) the mitochondrial ISC system has been replaced with a nitrogen fixation system (NIF) for the biosynthesis of Fe-S clusters (Ali *et al.* 2004; van der Giezen *et al.* 2004; Maralikova *et al.* 2010; Nyvltová *et al.* 2013). While the NIF system appears to be dual localized in the hydrogenosomes and cytosol of *Mastigamoeba balamuthi*, NIF components were not identified in the recent proteome of *Entamoeba* mitosomes, suggesting this pathway is primarily cytoplasmic (Mi-ichi *et al.* 2009). Instead, these organelles house a sulfate activation pathway important for the generation of sulfolipids (Mi-ichi *et al.* 2009). Unlike hydrogenosomes, the extended glycolysis pathway (*i.e.,* PFO and HYD) and energy generation machinery (ACS) is localized exclusively in the cytoplasm (Figure 1-12).

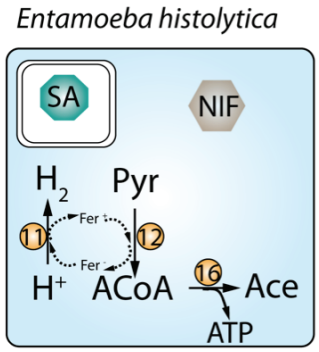

Figure 1-12: Select metabolism of *Entamoeba histolytica*. Numbering and labels are as shown in Figure 1-1,1-3 to 1-11. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue; electron transfer reactions are shown with dotted arrows. Fe-S cluster biosynthesis (NIF) is exclusively in the cytoplasm while sulfate activation (SA) occurs in the mitosome.

Mikrocytos makini

The intracellular oyster parasite *Mikrocytos* (Hine *et al.* 2001) was recently shown to be a rhizarian using multigene phylogenomic analyses (Burki *et al.* 2013). In the

transcriptomic survey of *Mikrocytos*, only four mitochondrial proteins were identified (cysteine desulphurase, ISCS; ISC scaffold protein, ISCU; ferredoxin reductase; FDXR and HSP70) (Figure 1-13). The authors were unable to identify genes that encode typical mitochondrial proteins (*e.g.,* protein import and folding) or extended glycolysis proteins suggesting the mitosomal metabolism of *Mikrocytos* is highly reduced and streamlined for Fe-S cluster biogenesis. Interestingly, the only proteins linked to energy metabolism identified in *Mikrocytos* are for a partial glycolysis pathway predicted to function in the cytosol. Like *Trimastix,* there are no obvious non-glycolytic candidates for ATP-generating enzymes in the transcriptome of *Mikrocytos* suggesting this organism (i) relies exclusively on fermentative metabolism or another ATP generation mechanism and/or (ii) imports ATP from host cytoplasms like other intracellular parasites (Schmitz-Esser *et al.* 2004; Tsaousis *et al.* 2008).

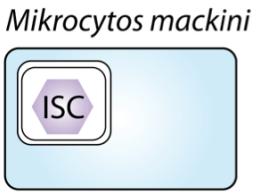

Figure 1-13: Reduced metabolism of *Mikrocytos mackini.* Only ISC-mediated Fe-S cluster biosynthesis is predicted to function in the *M. mackini* mitosomes.

Giardia intestinalis

The diplomonad *Giardia intestinalis* is a parasite of humans and thrives in the low oxygen environment of the gastrointestinal tract (Ankarklev *et al.* 2010). Similar to *Entamoeba,* the extended glycolysis and ATP generation pathways both occur in the cytoplasm of *Giardia* (Sanchez *et al.* 2000; Tovar *et al.* 2003; Emelyanov and Goldberg 2011). However, with respect to Fe-S cluster biosynthesis, *Giardia* resembles *Mikrocytos* since ancestral components of the ISC system (ISCS, ISCU, NIFU-like protein NFU1 and ISCA and monothiol glutaredoxin) have been retained and localize to its mitosomes (Tovar *et al.* 2003; Jedelský *et al.* 2011) (Figure 1-14).

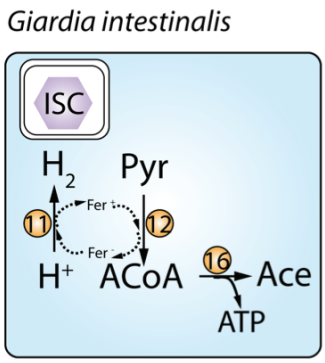

Figure 1-14: Reduced metabolism of *Giardia intestinalis*. Numbering and labels are shown as in Figure 1-1, 1-3 to 1-12. Enzymes typically associated with anaerobic metabolism are shown in orange. The cytoplasm is shaded in blue; electron transfer reactions are shown with dotted arrows. Extended glycolysis pathways are localized to the cytoplasm. ISC-mediated Fe-S cluster biosynthesis is present in the *G. intestinalis* mitosomes.

Microsporidians

Microsporidia are deep-branching fungi (Hirt *et al.* 1999) that are obligate intracellular parasites of animals (Goodgame 1996). The genomes and mitosomes of a number of microsporidians have been analyzed including pathogens of humans (*e.g., Enchephalitozoon cuniculi* (Katinka *et al.* 2001) and *Trachipleistophora hominis* (Williams *et al.* 2002)), and insects (*e.g., Nosema ceranae* (Cornman *et al.* 2009) and *Antonospora locustae* (Corradi *et al.* 2007)). In general, the mitosomes of these organisms participate in ISC-mediated Fe-S cluster biosynthesis (Katinka *et al.* 2001; Williams *et al.* 2002; Corradi *et al.* 2007; Goldberg *et al.* 2008; Cornman *et al.* 2009). Another general trend of these highly divergent organisms is their reduced genome size which has resulted in the loss of many metabolic pathways including amino acid and nucleotide metabolism (Katinka *et al.* 2001; Keeling and Fast 2002; Corradi *et al.* 2007; Keeling *et al.* 2010). In fact, some microsporidia (ex. *Enterocytozoon bieneusi*) have even lost the core sugar metabolism (*i.e.,* the majority of glycolysis, trehalose and pentose phosphate pathways are absent) (Keeling *et al.* 2010). *E. cuniculi* has significantly reduced ATP generation machinery and must instead rely on ATP import from the host cell into the parasite cytoplasm and eventually into the mitosomes using bacteria-derived ADP/ATP translocators (Tsaousis *et al.* 2008). The mitosomes of the insect pathogens (*Antonospora locustae* and *Trachipleistophora hominis)* have been hypothesized to be

involved in oxygen metabolism since they encode a mitosomal alternative oxidase (AOX) and glycerol-3-phosphate dehydrogenase and do not occupy strict anaerobic environments (Williams *et al.* 2010; Dolgikh *et al.* 2011; Heinz *et al.* 2012)(Figure 1-15). The AOX may contribute to the regeneration of cytosolic NAD^+ levels via the glycerol-3-phospate shuttle and not function as a relict respiratory chain for the purpose of generating energy (Williams *et al.* 2010; Heinz *et al.* 2012).

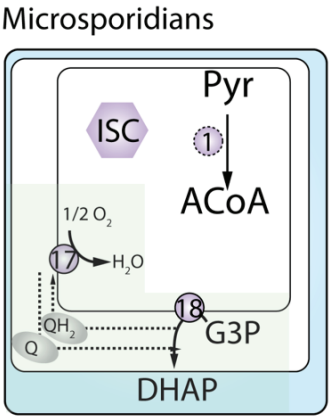

Figure 1-15: General metabolism of some microsporidia (*Enchephalitozoon cuniculi, Trachipleistophora hominis, Nosema ceranae* and *Antonospora locustae).* Numbering and labels are shown as in Figure 1-1, 1-3 to 1-14. The cytoplasm is shaded in blue; electron transfer reactions are shown with dotted arrows. Only two components of pyruvate dehydrogenase E1 are encoded [1-dotted circle]. Pathways found only in the insect parasites *A. locustae* and *N. ceranae* are shaded green: alternative oxidase [17] and glycerol-3-phosphate dehydrogenase [18].

1.3.5 Limitations Of Classification

Over the past decade, our understanding of MRO metabolism in different eukaryotes has greatly expanded (Figure 1-16). As a result, Müller and colleagues proposed the four classes of MROs (class II-V) (Müller *et al.* 2012). However, this new classification scheme was developed to accommodate organelles of primarily parasitic organisms. Yet the characterization of MROs from free-living organisms has also begun to call into question the validity of the classification boundaries proposed by Müller and colleagues. For instance, in their scheme, how does one classify an organelle that carries out malate dismutation (and therefore uses CII) but does not have CI-mediated electron transport (*Mastigamoeba balmuthi*)) for regeneration of RQ/UQ or succinyl-CoA synthetase for

ATP generation (Gill *et al.* 2007)? Furthermore, there are many other pathways and functions beyond energy generation including Fe-S cluster biogenesis, phopho(no)lipid metabolism, cofactor recycling and biosynthesis, folate pools, and fatty acid metabolism that are not captured by a classification based on energy metabolism. As we continue to gather information on MROs in diverse anaerobic lineages, particularly from free-living protists, it seems more prudent to view MRO diversity as a continuous functional spectrum, rather than binning them into a discrete set of energy metabolism-based classes.

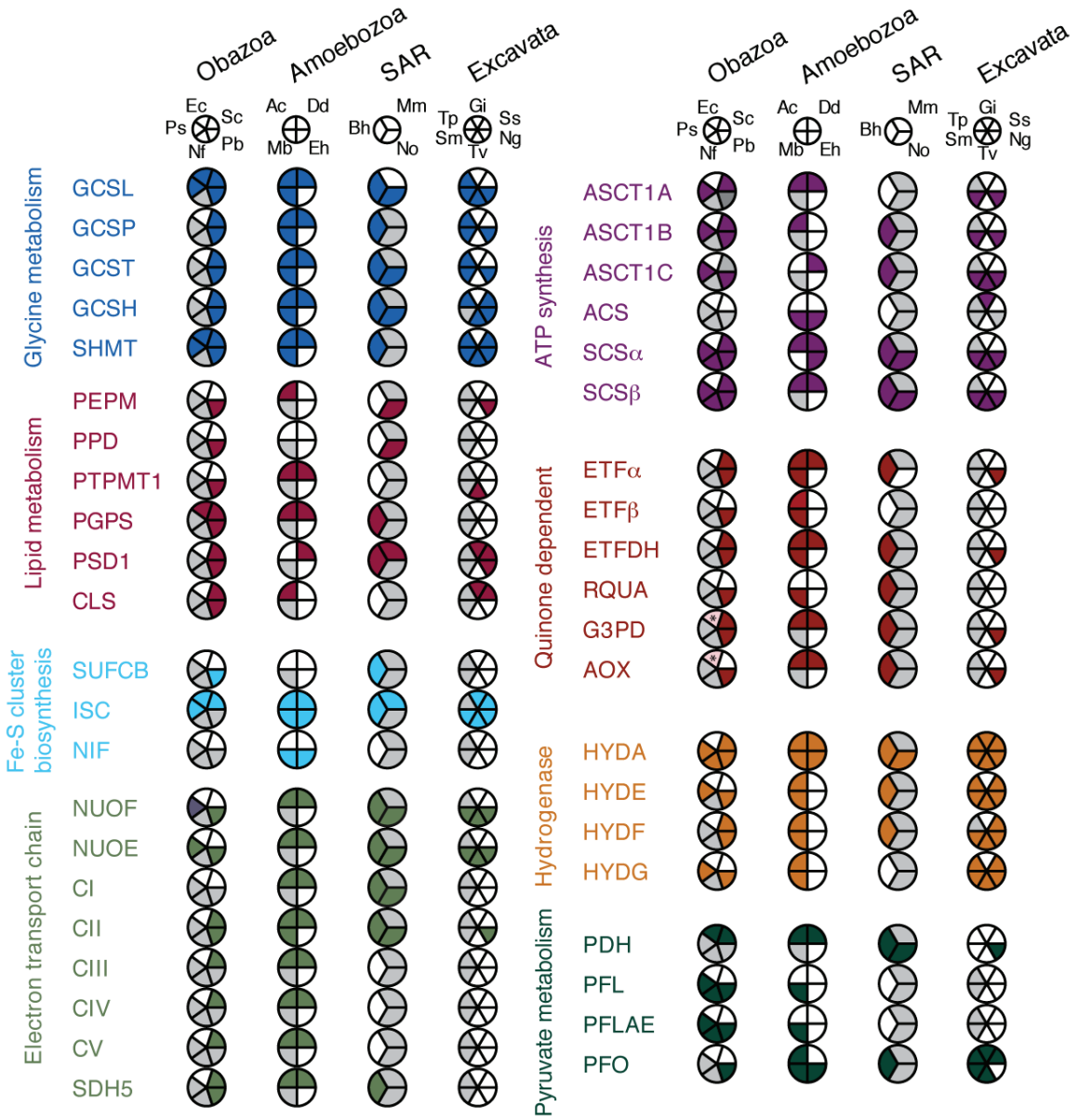

Figure 1-16: Coulson plot (Field et al. 2013) of select mitochondrial and MRO-specific genes in select eukaryotes. Mitochondrial/MRO-associated proteins were used as queries against genomic and transcriptomic sequencing pojects to identify homologs in different organisms. Each pie represents the super family of organisms surveyed: Obazoa (Ec, *Encephalitozoa cuniculi*; Sc, *Saccharomyces cerevisiae*; Pb, *Pygsuia biforma*; Nf, *Neocallimastix frontalis*; Ps, *Piromyces* sp.), Amoebozoa (Ac, *Acanthamoeba castellani*; Dd, *Dictyostelium discoideum*; Eh, *Entamoeba histolytica*; Mb, *Mastigamoeba balalmuthi*), SAR clade (Bh, *Blastocystis hominis*; Mm, *Mikcrocytos mackini;* No, *Nyctotherus ovalis*) and Excavata (Tp, *Trimastix pyroformis*; Gi, *Giardia intestinalis*; Ss, *Spironucleus salmoncida*; Ng, *Naegleria gruberi*; Tv, *Trichomonas vaginalis*; Sm, *Sawyeria marylandensis*). Coloured shading of pie pieces indicates homologs were identified by reciprocal best BLAST hits (e-value < 1 x 10^{-5}), white and grey filling indicates no homolog was identified in completed genomes or transcriptomic surveys respectively. Asterisk indicates homologs present in other microspordian genomes. Abbreviations not mentioned in text: GCS, glycine cleavage system; SHMT, serine hydroxymethyltransferase; PEPM, phosphoenolpyruvate

mutase; PPD, phonopyruvate decarboxylase; PTPMT1, phosphotidyl protein tyrosine phosphatase; PGPS, CDP-Diacylglycerol--Glycerol-3-Phosphate 3-Phosphatidyltransferase; PSD1, phosphotidylserine decarboxylase; CLS, cardiolipin synthase; SUFCB, sulfur mobilization; ISC, iron-sulfur cluster biosynthesis; NIF, nitrogen-fixation; NUO, NADH:Ubiquinone oxidoreductase E and F (24 kDa and 51 kDa respectively); ETF(DH), electron transferring flavoprotein (dehydrogenase); RQUA, rhodoquinone biosynthesis protein; and PFL(AE) pyruvate formate lyase (activating enzyme).

1.4 Origin Of Anaerobic Metabolism

Researchers interested in the early evolution of eukaryotic cells are currently divided on the evolutionary origin(s) of the aforementioned pathways of anaerobic energy generation. One possible explanation is that the ancestor of all extant eukaryotes was facultatively anaerobic and that the mitochondrion was ancestrally capable of respiring both aerobically and anaerobically, depending on the availability of oxygen. This hypothesis was first elaborated in detail by Martin and Muller in 1998 (Martin and Müller 1998). The second major hypothesis for the origins of anaerobic metabolism suggests that different genes related to anaerobic metabolism were acquired multiple times via lateral gene transfer throughout the tree of eukaryotes (Gill *et al.* 2007; Keeling and Palmer 2008; Stechmann *et al.* 2008; Hug *et al.* 2010; Stairs *et al.* 2011; Jerlström-Hultqvist *et al.* 2013). Below I hope to address the evidence in favour and against each of these hypotheses. Other hypotheses that have implications for the origins of anaerobic metabolism – *e.g.,* the syntrophy (López-García and Moreira 1999) and sulfur syntrophy (Searcy 2003) hypotheses – will not be discussed.

1.4.1 Hydrogen Hypothesis

In 1998, Martin and Müller proposed a completely novel hypothesis to explain the origin of eukaryotes and mitochondria (Martin and Müller 1998). While the hypothesis met with criticism early on (Doolittle 1998a), this was one of the first tangible hypotheses for the origin of mitochondria since Margulis's work that considered the role of anaerobiosis in early eukaryotes. This hypothesis states that the ancestor of mitochondria was a H_2-producing facultatively anaerobic α-proteobacterium (*i.e.,* that possessed the oxygen-respiring and hydrogen-producing metabolic pathways discussed

above) that formed a symbiotic relationship with a H_2-dependent methanogenic archaeon. In an anaerobic environment, the waste products of the symbiont (H_2, CO_2 and acetate) were used by the host archaeon. Over time, the host maximized surface area contact with the symbiont (without phagocytosis) to acquire the waste products. However, this new structural arrangement resulted in the symbiont having less surface area exposed to the environment to acquire the organic substrates necessary to make energy. If the genes encoding the organic transporters were transferred to the host genome and expressed on the host membrane, then the organic material from the environment could be fed to the symbiont without direct symbiont-environment contact. At that point, the host-symbiont system could exist in anaerobic and aerobic environments. This proto-eukaryote had an archaeal cytoplasm and a hydrogen-producing 'organelle' also capable of oxygen-dependent respiration, however the hydrogen hypothesis does not speculate on how or when the other eukaryotic features evolved in this proto-eukaryote lineage. Later, after the major lineages of extant eukaryotes diverged from the last eukaryotic common ancestor (LECA, a eukaryote possessing the hallmark cellular features such as a cytoskeleton, a nucleus and mitochondrion), aerobic and anaerobic metabolism was differentially lost in strict anaerobic and aerobic lineages respectively, generating the diversity of energy-metabolism and MROs that we see today.

It has been argued that multiple lines of evidence support this hypothesis. First, Müller and Martin noted that the anaerobic enzymes in eukaryotes often show a closer relationship to bacterial and not archaeal homologs. They suggested that these genes associated with anaerobic metabolism (so-called 'operational' genes) were present in the bacterial symbiont that gave rise to the mitochondria. Conversely, 'house-keeping' or 'informational' genes are more closely related to those of modern day archaea than bacteria suggesting these were present in the host cytoplasm (Zillig *et al.* 1989). A second line of evidence advanced in favour of the hydrogen hypothesis is the occurrence within contemporary Proteobacteria (although mostly within the γ- or δ-proteobacteria) of many facultative anaerobes that have genes involved in both anaerobic and aerobic

metabolism. Thus it is plausible that the proteobacterium that gave rise to mitochondria could also have been facultatively anaerobic. A third line of evidence comes from modern-day presumably H_2-dependent symbioses that are seen within the ciliate *Trimyema* in which endosymbiotic methanogenic archaea surround the H_2-producing MROs (Finlay *et al.* 1993). These, Martin and Müller argue, are modern-day 'living' analogs of how the mitochondrion first evolved.

Since its inception, the Hydrogen Hypothesis has faced criticism. The main point of contention has been the absence of molecular data supporting an α-proteobacterial origin for genes encoding anaerobic enzymes of energy metabolism within eukaryotes (Barbera *et al.* 2007; Gill *et al.* 2007; Hug *et al.* 2010; Stairs *et al.* 2011). Indeed, although all phylogenies of the pyruvate metabolizing enzymes associated with anaerobes (PFO, PNO and PFL) show eukaryotes as monophyletic, the closest bacterial taxa to eukaryotes are not α-proteobacteria but instead often δ–/ε– proteobacteria or firmicutes (Hug *et al.* 2010; Stairs *et al.* 2011; Jerlström-Hultqvist *et al.* 2013; Leger *et al.* 2013; Stairs *et al.* 2014). Furthermore, the vast majority of modern-day α-proteobacteria do not possess genes encoding enzymes for anaerobic metabolism and, those that do, appear to have recently acquired them from other bacterial lineages by LGT (Hug *et al.* 2010). This phylogenetic pattern is in sharp contrast to the typical aerobic enzymes involved in mitochondrial respiration that typically display clear phylogenetic affinities to protein sequences from modern-day α-proteobacteria (Emelyanov 2003).

In fact, in phylogenies of [FeFe]-hydrogenase, an enzyme that features prominently in the Hydrogen Hypothesis, eukaryotic sequences resolve in at least two distinct clades affiliated with separate bacterial groups, both of which are taxonomically 'mixed up' and neither of which is clearly 'α-proteobacterial' in origin (Hug *et al.* 2010). To reconcile this phylogenetic pattern with the Hydrogen Hypothesis, one would have to posit that the two eukaryotic clades are in fact paralogs that arose from an ancient duplication prior to the ancestral α-proteobacterial symbiosis. After eukaryotic lineages diversified into extant groups, the two paralogs were differentially lost in anaerobic descendant lineages or completely lost in aerobic lineages. Furthermore, proponents of

the Hydrogen Hypothesis state that the absence of obvious PFO/HYD-bearing α-proteobacteria donors is confounded by rampant lateral gene transfer amongst prokaryotes and thus the original lineage that gave rise to the mitochondrion is very different from the modern day α-proteobacteria (Müller *et al.* 2012). Conveniently, this argument neglects the observation that many typically 'aerobic' genes (*e.g.,* pyruvate dehydrogenase) shared by α-proteobacteria and mitochondria have not been affected by LGT and their mitochondrial homologs are, phylogenetically, their closest relatives (*i.e.,* many mitochondrial proteins still display a clear α-proteobacterial affinity) (Emelyanov 2003). The foregoing convoluted arguments explaining why anaerobic enzymes in eukaryotes do not show clear α-proteobacterial affinities indicate that the evidence for the Hydrogen Hypothesis is less clear than originally advanced by Martin and Müller in 1998. The main alternative explanations for anaerobic enzymes in eukaryotes involve LGT, a phenomenon that I discuss in the following sections.

1.4.2 Detecting Lateral Gene Transfer In Eukaryotes

Lateral gene transfer (LGT) is widely accepted to be one of the drivers of genome evolution in prokaryotes (Eisen 2000; Ochman *et al.* 2000; Kunin and Ouzounis 2003), but the possibility of LGT from bacteria to eukaryotes and between eukaryotes has been met with criticism (Martin 2011). LGT can be detected when the phylogeny of a particular gene (or genes) is incongruent with known organismal phylogeny (*i.e.,* a eukaryote branching closely with bacteria or distantly related eukaryotes) or - to a lesser extent - if a gene is present in the genome of an organism but not in genomes of close relatives. This pattern can also be observed in gene trees of ancient (hidden) paralogy. For example if a gene were duplicated in the common ancestor of eukaryotes and different paralogs were lost in different lineages that give rise to extant eukaryotes the resulting phylogeny could have two distinct clades of eukaryotes However in order to confidently assess LGT, the phylogeny must be not only free of phylogenetic artifact (*e.g.,* long branch attraction or compositional bias) but also well supported, which is not always achievable with single-gene trees. Topology testing can also be used to

statistically compare the likelihoods of an estimated phylogeny against the likelihoods of a set of user-provided topologies (*i.e.,* Approximate Unbiased Topology Test (Shimodaira 2002)) to detect whether the data can significantly distinguish between them.

Another obstacle in modern-day phylogenetics and phylogenomics involves the polyxenic nature of microbial culturing. Many of the free-living protists sequenced in recent years grow in mixed culture with bacteria (Graaf *et al.* 2009; Zubáčová *et al.* 2013). Therefore, the resulting genome or transcriptomic data that is generated includes both eukaryotic and bacterial sequences. Identifying putative LGT candidates that represent *bone fide* eukaryotic sequences can therefore be challenging. If one can collect sufficient genomic information that (i) reveals spliceosomal introns in the sequence or (ii) hints at a eukaryotic origin of neighbouring genes on the chromosome, then it is likely that the sequence is not a contaminant. However, in the absence of genomic data (or low intron abundance), the presence of poly-adenylation signals, biased codon usage, nucleotide composition or low sequence identity to sequenced bacterial genomes can help in determining the provenance of the sequence.

1.4.3 Laterally Acquired Anaerobic Metabolism

An alternative hypothesis for the origin of anaerobic metabolism in eukaryotes involves the process of LGT. The idea that anaerobic metabolism was acquired by eukaryotes via LGT invokes concepts from Doolittle's 'you are what you eat' model (Doolittle 1998a; Doolittle 1998b). Doolittle's model explains how genes from the endosymbiont could be transferred to the host nucleus; partial digestion of the proto-mitochondrion would lead to a constant flow of genetic material to the nucleus. Organellar genes that are incorporated into the host nuclear genome and their protein products successfully retargeted to the organelle will result in the eventual loss of the organellar copy. This ratchet-like process would continue until virtually all of the endosymbionts genes are transferred to the nucleus. Similarly, heterotrophic protists are constantly bombarded with new genetic material while feeding. Therefore, it is possible that, even though LGT is seemingly rare in eukaryotes, constant exposure to

food-derived DNA increases the opportunity for gene uptake (Doolittle 1998a; Doolittle 1998b).

Therefore, the LGT model for anaerobic metabolism suggests that the common ancestor of mitochondria was an aerobic α-proteobacterium that did not encode the genes associated with anaerobic metabolism (*i.e., pfo*, *pfl* and *hyd*) that we see in MROs today. Instead, the genes were acquired by lateral gene transfer from bacteria, potentially food bacteria, into one or more eukaryotic lineages and subsequently transferred between eukaryotes after the diversification of eukaryotes.

Support for this model is based exclusively on the phylogenetic affinities of the eukaryotic homologs of these anaerobic genes (Hug *et al.* 2010; Hampl *et al.* 2011; Stairs *et al.* 2011; Jerlström-Hultqvist *et al.* 2013; Leger *et al.* 2013). For example, in most phylogenetic analyses of hydrogenases there are two distinct clades of eukaryotes with various prokaryotic affinities, suggesting this enzyme was not present in the common ancestor of eukaryotes, or at the very least, one copy was acquired by lateral gene transfer (Hug *et al.* 2010). Furthermore, inter-eukaryotic relationships observed in the phylogeny of hydrogenases are incongruent with known organismal relationships. Another interesting observation is that many genes that are hypothesized to have been acquired by lateral gene transfer in eukaryotes are in fact fusion proteins of two or more bacterial proteins. For instance, *pfl* and *pfla* are fused into a single transcript in *Thalassiosira pseudonana* or adjacent in *Ostreococcus tauri* and *Ostreococcus lucimarinus* (Armbrust *et al.* 2004; Derelle *et al.* 2006; Lanier *et al.* 2008; Stairs *et al.* 2011). The close proximity of these genes in the eukaryotic genomes suggests that *pfl and pfla* could have originated by transfer of an operon in bacteria that was subsequently fused into a bipartite protein gene.

However, there are obvious problems with the LGT hypothesis for the origins of anaerobic metabolism in MROs. Firstly, in the initial phylogenies of PFO, eukaryotes are usually monophyletic, suggesting that eukaryotic PFO could have been vertically inherited from LECA (Horner *et al.* 1999; Hug *et al.* 2010). Moreover, if we assume that PFO was acquired by LGT into one lineage of eukaryotes, under the LGT hypothesis, the

only way to describe the relationships observed between eukaryotes would be to invoke multiple LGT events between eukaryotes. Additionally, the hydrogen-producing system of many bacteria requires at least four genes (*hydA, E, F,* and *G*) to function. Therefore, in order for a functioning copy of hydrogenase to persist in genomes, all four enzymes would have to be transferred. To reconcile this observation and the presumed difficulty of acquiring four genes independently, the LGT hypothesis must assume that the genes were transferred as a single unit (*i.e.,* an operon) and over time have been rearranged on the eukaryotic chromosome. However, some eukaryotes do not encode obvious homologs of all the maturases (Loftus *et al.* 2005; Gill *et al.* 2007; Morrison *et al.* 2007) suggesting that at least some hydrogenases can function without maturase proteins.

Finally, one mechanism to acquire genes by lateral gene transfer is from food bacteria or eukaryotes that have been consumed by phagocytosis. Not all modern day representatives of organisms with MROs or genes involved in anaerobic metabolism are capable of phagocytosis. Therefore, these organisms would have had to (i) been able to perform phagocytosis or (ii) acquire the genes independent of classical phagocytosis (examples in (Richards *et al.* 2006; Richards *et al.* 2011) and reviewed in (Keeling and Palmer 2008)).

1.5 Aims Of This Thesis

The following chapters discuss diverse anaerobic metabolisms found in mitochondrion-related organelles throughout the tree of life, with an emphasis on those pathways and functions influenced by lateral gene transfer. Each chapter represents a different approach to exploring how different lineages of eukaryotes have evolved to thrive in low oxygen environments by studying the localization and evolutionary history of anaerobic proteins.

Chapter Two analyzes the distribution and diversity of metabolism across eukaryotes with a focus on pyruvate formate lyase (PFL) and pyruvate formate lyase activating (PFLA) enzymes. Rigourous bioinformatic and phylogenetic methods were used to resolve the evolutionary relationships of eukaryotic PFL/PFLA with prokaryotes. This study determined that eukaryotic PFL and PFLA were likely acquired by LGT into one

lineage of eukaryotes from bacteria (specifically a firmicute-like bacterium) and subsequently transferred to other eukaryotes by eukaryote-eukaryote lateral gene transfer.

Chapter Three reports a large transcriptomic survey of *Pygsuia biforma* – a member of the enigmatic group of eukaryotes classified as breviates. Over 120 proteins were predicted to localize to the MRO of *Pygsuia*. The proteins predicted to function in the MRO represent a mixture of ancestral mitochondrial and hydrogenosomal features. Interestingly, some of the proteins predicted to localize to the MRO of *Pygsuia* have never been seen in mitochondria or MROs. For example, the MROs of *Pygsuia* appear to have lost the canonical ISC system for the biosynthesis of Fe-S clusters and instead use an archaeal sulfur mobilization system (SUF) likely acquired by LGT.

Finally, Chapter Four examines the cellular localization and evolutionary history of the rhodoquinone biosynthesis enzyme RQUA. RQUA is a methyltransferase protein recently identified to be essential for RQ biosynthesis in *Rhodospirillum* (Lonjers *et al.* 2012). A survey of available eukaryotic genomes revealed that RQUA is present in a variety of organisms and often possesses a predicted mitochondrial targeting sequence. Immunofluorescence microscopy was used to localize RQUA to the MROs of *Pygsuia* supporting the *in silico* predictions. Phylogenetic analysis suggests that RQUA has been acquired by eukaryotes multiple times from bacterial and eukaryotic sources.

Chapter 2 Eukaryotic Pyruvate Formate Lyase And Its Activating Enzyme Were Acquired Laterally From A Firmicute

This chapter contains work published in **Stairs, W. C.,** Hampl, V. and Roger A.J. (2011). Eukaryotic pyruvate formate lyase and its activating enzyme were acquired laterally from a Firmicute. *Mol. Biol. Evol*. 28(7):2087-99.

2.1 Abstract

Most of the major groups of eukaryotes have microbial representatives that thrive in low oxygen conditions. Those that have been studied in detail generate ATP via pathways involving anaerobically functioning enzymes of pyruvate catabolism that are typically absent in aerobic eukaryotes. These enzymes include pyruvate:ferredoxin oxidoreductase, pyruvate:$NADP^+$ oxidoreductase, and pyruvate formate lyase (PFL). PFL catalyzes the non-oxidative generation of formate and acetyl-Coenzyme A (CoA) from pyruvate and CoA and is activated by PFL activating enzyme (PFLA). Within eukaryotes, this extremely oxygen-sensitive pathway was first described in the hydrogenosomes of anaerobic chytrid fungi and has more recently been characterized in the mitochondria and chloroplasts of the chlorophyte alga *Chlamydomonas reinhardtii*. This study aims to clarify the origins of this pathway by presenting a comprehensive survey of PFL and PFLA homologues in publicly available large-scale eukaryotic genomic and cDNA sequencing data. Surprisingly, genes encoding PFL and PFLA are widely distributed in diverse facultative or obligate anaerobic eukaryotic representatives of the archaeplastidan, metazoan, amoebozoan, and haptophyte lineages. Using maximum likelihood and Bayesian phylogenetic methods, eukaryotic PFL and PFLA sequences each form monophyletic groups that are most closely related to homologs in firmicute gram-positive bacteria. Topology tests exclude both α-proteobacterial and cyanobacterial affinities for these genes suggesting that neither originated from the endosymbiotic ancestors of mitochondria or chloroplasts. Furthermore, the topologies of the eukaryote portion of the PFL and PFLA trees significantly differ from well-accepted eukaryote

relationships. Collectively, these results indicate that the PFL pathway was first acquired by lateral gene transfer into a eukaryotic lineage - most probably from a firmicute bacterial lineage - and that it has since been spread across diverse eukaryotic groups by more recent eukaryote-to-eukaryote transfer events.

2.2 Introduction

In aerobic eukaryotes, carbohydrates are catabolized glycolytically to produce pyruvate that is converted to acetyl-Coenzyme A (CoA) in mitochondria by the pyruvate dehydrogenase complex (PDC). Acetyl-CoA feeds into the tricarboxylic acid cycle generating reducing equivalents that ultimately drive ATP synthesis via the coupling of the respiratory chain and oxidative phosphorylation. Multicellular organisms that live in the absence of oxygen, such as anaerobic helminths, have evolved alternatives to aerobic energy production pathways such as the malate dismutation pathway (Tielens *et al.* 1998) and often utilize alternative terminal electron acceptors of the respiratory chain (Kobayashi *et al.* 1996). Anaerobic unicellular eukaryotes (protists) that have been studied in detail also couple pyruvate metabolism to ATP biosynthesis, often by substrate-level phosphorylation. Furthermore, they differ from aerobic eukaryotes in the enzymes they use to convert pyruvate to acetyl-CoA. Instead of using PDC, these organisms utilize enzymes such as pyruvate:ferredoxin oxidoreductase (PFO), pyruvate:$NADP^+$ oxidoreductase (PNO), and/or pyruvate formate lyase (PFL). Although PDC is known to function exclusively within mitochondria of aerobic eukaryotes, these anaerobic enzymes are sometimes localized to different subcellular compartments depending on the organism. For example, in some organisms, PFO functions within mitochondrion-related organelles (MROs) such as the hydrogenosomes of the parabasalid parasite *Trichomonas vaginalis* (Steinbüchel and Müller 1986; Hrdy and Muller 1995) where in others, such as *Entamoeba histolytica* and *Giardia intestinalis*, it is thought to primarily function within the cytosol (reviewed in (Embley and Martin 2006)). The enzyme PNO has been shown to localize in the mitochondria of *Euglena*

gracilis and *Plasmodium falciparum* and in the mitosomes of *Cryptosporidium parvum* (Rotte *et al.* 2001; Mogi and Kita 2010).

PFL (EC 2.3.1.54) is a member of the pyruvate formate lyase protein family and catalyzes the nonoxidative generation of formate and acetyl-CoA from pyruvate and CoA. PDC, PFO, and PNO (but not PFL) use redox chemistry with cofactors (NAD^+ or ferredoxins) to convert pyruvate into acetyl-CoA, whereas PFL activity involves radical chemistry. Radical generation occurs independent of catalysis and involves a second enzyme, pyruvate formate lyase activating enzyme (PFLA; EC 1.97.1.4). PFL is activated by the formation of an α-carbon centered radical in its C-terminal glycyl radical domain (pfam01228) by PFLA which utilizes the cofactors S-adenosyl methionine (SAM) and reduced flavodoxin or ferredoxin (Wagner *et al.* 1992). When pyruvate binds to PFL, the radical is transferred to two adjacent active site cysteine residues (Frey *et al.* 1994). These conserved cysteine residues along with a conserved tyrosine downstream relative to the glycine within the glycyl radical domain, can be used to distinguish pyruvate-catalyzing PFLs from other PFL protein family members (such as the benzyl succinate synthases and glycerol dehydratases) that have only one cysteine in the active site and a different radical domain architecture (Lehtiö *et al.* 2004).

The radical form of PFL upon oxygen exposure is irreversibly cleaved at the radical residue (Wagner *et al.* 1992). Presumably for this reason, the enzyme has only been identified in organisms that are either obligatory or transient anaerobes, such as the firmicutes (Thauer *et al.* 1972) and Enterobacteriaceae. PFL activity in eukaryotes has been described in cell-free extracts (Marvin-Sikkema *et al.* 1993) and hydrogenosomes of the chytrid fungus *Neocallimastix* sp. (Akhmanova *et al.* 1999) as well as its close relative *Piromyces* sp. E2 (Boxma *et al.* 2004), the chloroplasts and mitochondria of the chlorophyte algae *Chlamydomonas reinhardtii* and probably its close relative *Polytomella* sp. (Atteia *et al.* 2006; Hemschemeier *et al.* 2008). It has been suggested that PFL is part of the ATP-producing pathway in these organelles. Unlike other pyruvate-metabolizing enzymes, the action of PFL does not involve CO_2 production. Indeed, this system has been hypothesized to function as an alternative to

PFO, when the latter is inhibited by high CO_2 levels in the organism's environment (Boxma *et al.* 2004). However, PFL-mediated acetyl-CoA generation is problematic for energy production because of the SAM consumed in the initial activation of PFL; SAM requires ATP for its synthesis by SAM synthetase (Markham *et al.* 1980). Unless the energy required for SAM synthesis is regenerated elsewhere and the radical is protected, the PFL system cannot be as efficient as a PFO-based system for ATP production. Nevertheless, it is possible that PFL-generated acetyl-CoA is also important for sustaining biosynthetic pathways, such as fatty acid biosynthesis and elongation and amino acid synthesis.

The evolutionary origins of anaerobic pyruvate oxidizing enzymes within eukaryotes are contentious. It has been suggested that they may have originated with the α-proteobacterial mitochondrial symbiont that may have been a facultative anaerobe (Martin and Müller 1998). To evaluate this, as well as alternative hypotheses, Hug *et al.* (2010) surveyed the presence and absence of PFO- and PNO-related proteins, as well as [FeFe]-hydrogenases in partial and complete genomic and transcriptomic sequence data of diverse eukaryotes (Hug *et al.* 2010). Although the homologs of PFO and PNO proteins that were identified in diverse eukaryotic lineages appeared to form a monophyletic group, phylogenetic analyses did not support an α-proteobacterial origin for the eukaryotic enzymes. Similarly, a previous network analysis demonstrated that the PFL sequences from *Chlamydomonas* and the chytrid fungi form a group clustering near homologs from firmicutes (Gelius-Dietrich and Henze 2004). With recent large-scale genomic and cDNA/expressed sequence tag (EST) sequencing projects of eukaryotes, I have identified many additional sequences homologous to PFL and PFLA in a wide diversity of other eukaryotes including members of the Amoebozoa, Metazoa, Haptophyta, and Archaeplastida. In this study, I investigate the origin of eukaryotic PFL and more generally the distribution and evolution of pyruvate metabolizing enzymes within eukaryotes to trace their prokaryotic ancestry.

2.3 Methods

2.3.1 454 EST Project Of Mastigamoeba balamuthi

Mastigamoeba balamuthi cultures were maintained from an established American Type Culture Collection (ATCC) culture (ATCC #30984) in PYGC media (1% peptone, 1% yeast extract, 56 mM glucose, 86 mM NaCl, 10 mM K_2H/KH_2PO_4, pH 6.8) grown at 20–24 °C. Total RNA was isolated using Trizol (Tri-reagent) following the protocol supplied by the manufacturer (Invitrogen). A cDNA library was constructed by Vertis Biotechnologies AG (Freising, Germany), and 454 pyrosequencing was performed by Genome Québec. The Newbler assembly program (Roche) was used to assemble the reads. PFL and PFLA were identified using basic local alignment search tool (TBLASTN), and candidate reads and contiguous sequences were viewed and manually assembled and edited using Sequencher (Gene Codes corp.). Sequences were confirmed by polymerase chain reaction on cDNA and genomic DNA followed by Sanger sequencing (forward primer MbPFL_1F: ATGTCCGGCTCTATCCGGG and reverse primer MbPFL_909R: CTATGAGCACTGCACGGCGACG). The sequences of the *Mastigamoeba* PFL and PFLA genes are deposited in GenBank with the accession numbers HM590578 and HQ003218, respectively.

2.3.2 Database Searches

Bacterial PFL and PFLA protein sequences were retrieved from all whole-genome sequencing projects compiled in the microbial online analysis database (http://ratite.cs.dal.ca/moa/) using BLASTP and TBLASTN with PFL and PFLA from *C. reinhardtii* (gi:92084842, 57021069) as queries. Because only pyruvate-catalyzing PFL proteins were sought, those sequences lacking the diagnostic conserved active site motif (**CC**, Lehtiö and Goldman 2004), glycyl radical domain ([LIVM]R[LIVM]S**G**Y, Lehtiö and Goldman 2004), or those annotated as PFL4 were removed. Similarly, those PFLA sequences characterized as non-pyruvate PFL-activating enzymes that had an additional ferredoxin domain (Raynaud *et al.* 2003) or from taxa without a pyruvate-catalyzing PFL were removed. To maintain a reasonably sized data set for subsequent phylogenetic

analyses, extremely similar sequences from redundant species and strains were randomly removed.

Eukaryotic gene and protein sequences were retrieved from a variety of sources including EST and genome databases from: the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov), the Joint Genome Institute (JGI, http://www.jgi.doe.gov) PlantGDB (http://www.plantgdb.org), the Welcome Trust Sanger Institute (http://www.sanger.ac.uk), Pristionchus.org, and the Taxonomically Broad EST database (TBestDB, http://tbestdb.bcm.umontreal.ca). Because many of the retrieved sequences were suspected prokaryotic contaminants in genome sequencing projects (*i.e.,* homologs found in *Homo sapiens, Mus musculus, Oryza sativa* etc.), genes with >99% nucleotide identity to particular prokaryotic *pfl* or *pfla* sequences, or those not assembled in the final draft of the eukaryotic genome sequences were excluded from further consideration. Any gene sequences containing spliceosomal introns were maintained as true eukaryotic homologs.

The same procedures were performed to probe the presence/absence of PDC, PFO, PNO, and sulfite reductase (SR) protein families in diverse eukaryotic genomic and transcriptomic data. A PDC complex was designated as present if at least PDH (E1 subunit) was present in the data. Those proteins that resembled PFO but had a fused C-terminal CysI domain (Crane *et al.* 1995), or cytochrome P450 reductase domain (Zeghouf *et al.* 1998) were classified as SR or PNO, respectively.

2.3.3 Data Set Generation And Phylogenetic Analysis

Amino acid sequences were aligned using MUSCLE v3.2 (Edgar 2004) and manually edited to mask out regions of ambiguous alignment. The final alignments for PFL and PFLA consisted of 693 sites (166 taxa) and 194 sites (152 taxa), respectively. Maximum likelihood (ML) phylogenetic analysis was performed using RAxML version 7.2.6 (Stamatakis 2006) under the Le and Gascuel (LG) (Le and Gascuel 2008) amino acid substitution model plus gamma model of rates across sites taking into account the amino acid frequencies of the data set evolution (denoted PROTGAMMALGF in RAxML). The LG amino acid substitution matrix was identified as the optimum model for all data

sets assessed by PROTTEST (Abascal *et al.* 2005) under the Akaike and Bayesian Information Criteria. For each data set, bootstrap support for splits estimated from 500 bootstrap replicates were mapped onto splits on the best scoring ML tree (obtained by 10 heuristic search replicates). As nonparametric bootstrap support values have recently been shown not to be first order correct (Susko 2009), corrected bootstrap values (BV_c) were generated using the adjusted bootstrap probability (aBPn) program under the LG + Γ model, using alpha shape parameter values calculated using RAxML (see (Susko 2010)). The program requires nonzero branch lengths and sufficient alignment data for estimation. For this reason, taxa with extremely short branch lengths were removed from the trees and alignments and simulated data sets (10,000 sites using SeqGen http://tree.bio.ed.ac.uk/software/seqgen/) were used. aBPn values can be considered as a statistical test that the split was not present in the true phylogeny with a P value = 1 − BV_c. Bayesian inference (BI) was also conducted using PhyloBayes 3.2 (Lartillot *et al.* 2009) by running two chains under the catfix C20 model of evolution (Le *et al.* 2008). For each chain, a total of 300,000 generations were run, from which trees were sampled every 100 generations and discarding a manually determined burn-in of 50,000 generations for each (yielding a total of 2,500 trees). Posterior probabilities (PP) for splits were mapped on to the ML estimated topology.

2.3.4 Congruence Testing And Concatenation

Organisms that had multiple paralogous copies of PFL, and only one PFLA were removed prior to concatenation because each PFL could not be uniquely assigned an activating enzyme. Congruence between PFL and PFLA phylogenies was tested for the data set without paralogues using CONCATERPILLAR with an alpha-level cutoff of 0.05 (Leigh *et al.* 2008). For concatenated PFL–PFLA analyses, to assess potential artifacts introduced by gene and lineage-specific evolutionary dynamics, ML analysis was performed both with and without partitioning the data to allow for variable edge lengths and other parameters for each protein family (the RAxML—M option). In general, only minor differences in bootstrap values on splits were noted between partitioned versus nonpartitioned analyses therefore the bootstrap correction and BI PP

were calculated only for the nonpartitioned data set. To evaluate the impact of missing data on topological support values, a second analysis was performed as described above where only complete sequences and partial sequence with greater than 50% of the sites were included.

Table 2-1 AU Topology Tests for the Single Gene and Concatenated PFL and PFLA Data Sets.

Phylogeny	Monophyly Test[a]	AU Test **P** value	AU Test **P** value of Partitioned Analysis
PFL	Eukaryotes[b] and γ-proteobacteria	0.0004	
	Eukaryotes[b] and α-proteobacteria	0.0024	
	Eukaryotes[b] and α-clade	0.0130	
	Eukaryotes[b] and cyanobacteria	0.1580	
PFLA	Eukaryotes[b] and γ-proteobacteria	0.8010	
	Eukaryotes[b] and α-proteobacteria	0.0010	
	Eukaryotes[b] and α-clade	0.4660	
	Eukaryotes[b] and cyanobacteria	0.0740	
Concatenated tree[c] (PFL–PFLA)	Eukaryotes[b] and α-proteobacteria	2×10^{-32}	5×10^{-39}
	Eukaryotes[b] and α-clade	0.0130	0.0040
	Eukaryotes[b] and cyanobacteria	1×10^{-5}	0.0090
Concatenated tree (PFL–PFLA; exclusively eukaryotes[b])	Typical eukaryotic phylogeny	0.0020	0.0030
	Chlorophytes and prasinophytes	0.0100	0.0080
	Chlorophytes, prasinophytes, and *Mastigamoeba*	0.0380	0.0330
	Fungi and prasinophytes	0.5230	0.4660
	Opisthokonts	0.6610	0.6650

[a]ML tree search with the indicated constraint.

[b]Only includes eukaryotes that from a monophyletic group depicted in Figure 2-2, Figure 2-3, (and not *Brugia malayi* or *Litopenaeus vannamei*).

[c]For concatenation, the following data were removed: *Brugia malayi, Litopnaeus vannamei*, γ-proteobacteria, and paralogous sequences.

2.3.5 Topology Testing

To test alternative candidate phylogenetic positions of eukaryotes within the bacterial subtree, I used the approximately unbiased (AU) topology test implemented in CONSEL (Shimodaira and Hasegawa 2001). In brief, ML trees were estimated subject to topological constraints corresponding to alternative hypotheses (*e.g.,* α-proteobacteria as sister to the eukaryotes) using RAxML (—g option). The AU tests were performed with CONSEL (Table 2-1) based on these trees as well as a sample of 100 "ML-bootstrap" trees generated during bootstrapping using RAxML (—f g option). The 100 additional

trees were to provide a sample of "good" trees, which is necessary in order for the P values for the AU test to be accurate (Shi *et al.* 2005).

2.3.6 Gene Order And Operon Prediction

Putative operons in bacterial genomes were identified using OperonDB (Pertea *et al.* 2009). If PFL and PFLA were located at distantly separated loci on a given prokaryotic chromosome, they were not considered to be in an operon. Complete genomes of the eukaryotes *Thalassiosira pseudonana, Ostreococcus* species, and *C. reinhardtii* (jgi) were examined for the location and order of the PFL and PFLA genes on their assembled scaffolds.

2.3.7 Tests For Long-Branch Attraction, Amino Acid Composition Bias

To evaluate whether long-branch attraction contributed to the support for the grouping of the eukaryotic sequences, for each of the analyses, I: (i) removed 50% of the longest eukaryotic branches and the resulting data set was reanalyzed as described above and (ii) removed the fastest evolving sites, which are often responsible for long-branch attraction and reanalyzed the truncated alignments as described above. The fast evolving sites were identified and removed using AIR-identifier and AIR-remover (Kumar *et al.* 2009). To evaluate the possibility of phylogenetic artifacts associated with amino acid composition bias within the eukaryotic sequences, I performed a chi-squared test for deviations in amino acid compositions for each taxon implemented in TREE-PUZZLE version 5.2 (Schmidt *et al.* 2002). Taxa with significantly different amino acid composition from the overall data set frequencies were removed, and the analysis was repeated as above to check whether estimated topologies and support values changed. Additionally, I performed cluster analyses of the amino acid composition vectors of the eukaryotic sequences and the concatenated analysis. Euclidean distance matrices between composition vectors were constructed from 100 bootstrapped alignments and used to generate 100 UPGMA trees. The majority rule consensus of these trees and bootstrap support for bipartitions were visualized using CONSENSE (Felsenstein 2005).

2.3.8 In Silico Subcellular Localization Prediction

TargetP (available at http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson *et al.* 2000) and Mitoprot II (available at http://ihg2.helmholtz-muenchen.de/ihg/Mitoprot.html (Claros and Vincens 1996) were used to predict subcellular localization of proteins based on their amino acid sequences.

2.4 Results And Discussion

2.4.1 Identification Of Previously Unidentified Pfl And Pfla In Eukaryotes

Database mining of all publicly available genome and EST projects for eukaryotes revealed a large number of previously unreported PFL and PFLA homologues (Figure 2-1). A total of 16 new PFL eukaryotic homologues were identified in archaeplastids (*Scenedesmus obliquus*, *Chlorella* sp., *Acetabularia acetabulum*, *Haematococcus pluvialis, Volvox carteri, Os. tauri, Os. lucimarinus, Micromonas pusilla*, *Micromonas* sp., *Porphyra haitanensis,* and *Cyanophora paradoxa*), an opisthokont (*Amoebidium parasiticum*), an amoebozoan (*Mastigamoeba. balamuthi*), a stramenopile (*T. pseudonana* (2)), and a haptophyte (*Prymnesium parvum*). A total of seven new PFLA eukaryotic homologues were identified in archaeplastids (*Micromonas pusilla, Micromonas* sp. *Os. tauri,* and *Os. lucimarinus*), an amoebozoan (*M.balamuthi*), and a stramenopile (*T. pseudonana* (2)). The majority of these sequences were identified in both genomic and EST data. Furthermore, in most cases, spliceosomal introns were identified in the genomic sequences confirming a eukaryotic (rather than bacterial) provenance.

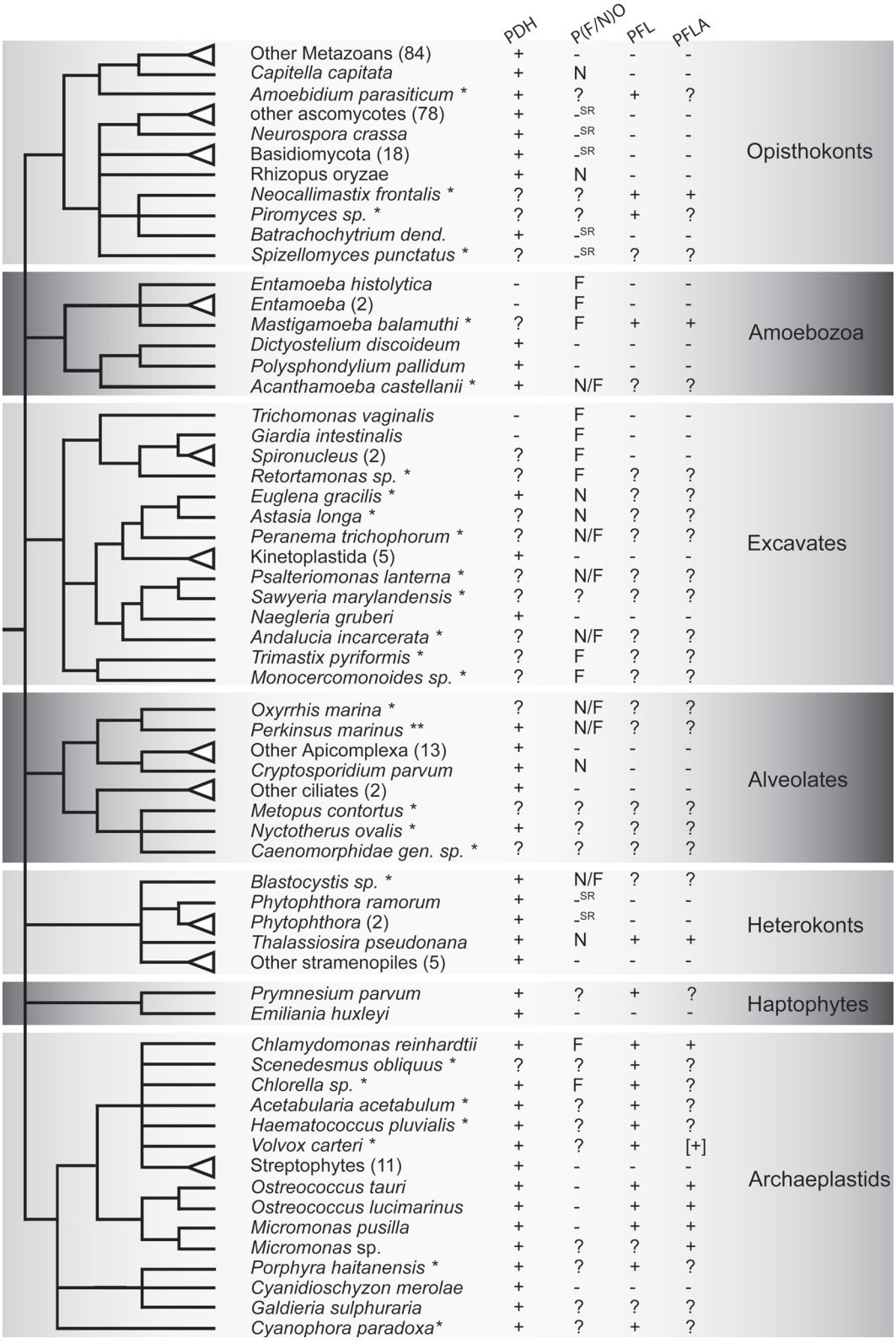

Figure 2-1: The distribution of pyruvate utilizing acetyl-CoA-generating enzymes across eukaryotic diversity. Complete genomes and EST projects (*) were surveyed for pyruvate dehydrogenase (PDC), pyruvate ferredoxin oxidoreductase (PFO), pyruvate formate lyase (PFL)

and pyruvate formate lyase activating enzyme (PFLA) for more than 200 eukaryotes from multiple databases. The presence (+) or absence (−) from complete genome projects are illustrated while "?" indicates the absence from EST data. Homologues of PFO are illustrated as F for PFO, N for those containing the cysJ domain typically associated with pyruvate:NADP+ oxidoreductases (PNO) or –SR for those sequences with cysI domain typically associates with sulfite reductases in fungi. The bracketed numbers represent the number of other taxa sampled but not displayed. Databases were mined up to October 2009.

PFL, but not PFLA, sequences were also found in the metazoan *Brugia malayi* (a filarial nematode) and *Litopenaeus vannamei* (a whiteleg shrimp) genomic databases. The *Brugia* genomic sequence corresponding to PFL was identified in an incomplete genome project, did not contain spliceosomal introns and was not identifiable in a transcriptome project (http://www.nematodes.org). The *Lit. vannamei* sequence was extremely short (~500 bp), only identified in EST data and 74% identical at the nucleotide level to *Listeria innocua pfl*. Although it remains uncertain, for these reasons, I suspect both sequences are bacterial contaminants and were not included in subsequent analyses.

2.4.2 Eukaryotic Monophyly Of Pfl And Pfla

The global phylogenetic analyses of all representative bacterial and eukaryotic pyruvate-catalyzing PFL and PFLA amino acid sequences are shown Figure 2-2 and Figure 2-3, respectively. In the PFL analysis, all eukaryote sequences form a clade with moderate support (BV = 66, BV_C = 88, PP = 0.66; Figure 2-2), showing a weak affinity (<50% BV) for a firmicutes plus *Bacteroides* grouping. Similar results were recovered for the PFLA phylogeny where eukaryotes were monophyletic with significant support (BV = 90, BV_c = 97, PP = 1.0; Figure 2-3). To test whether eukaryotic monophyly in these phylogenies were artifacts of long-branch attraction, the analyses were repeated with 50% of the longest eukaryotic branches removed. This secondary analysis did not change the overall tree topology, however it did differentially affect the support for eukaryote monophyly in the PFL and PFLA analyses (BV_{PFL} = 79, PP_{PFL} = 1.0; BV_{PFLA} = 83, PP_{PFLA} = 1.0, data not shown). For the PFL and PFLA analyses, there is very weak support for intergroup relationships especially along the "backbone" of the phylogeny (*i.e.,* the innermost branches on the tree). Although I initially suspected this could be due to the

position of the strongly supported *γ*-proteobacterial clade relative to other sequences (this is the deepest significant split in the tree), removal of the *γ*-proteobacteria did not alter the resolution amongst the other taxa.

The most common bacteria-related sequences in eukaryotic genomes are the genes that were acquired during the establishment of endosymbiotically-derived mitochondria and chloroplasts (Pisani *et al.* 2007). To test whether the position of the monophyletic eukaryotic group in the PFL and PFLA trees was consistent with acquisition during either of these endosymbioses, I tested whether ML topologies obtained by constraining the eukaryotic sequences to group with α-proteobacteria (mitochondria) or cyanobacteria (chloroplasts) respectively were significantly rejected by the data (Table 2-1). Eukaryotes + α-proteobacteria monophyly was rejected for both PFL (AU *P* value [*P*] = 0.002) and PFLA (*P* = 0.001). As I noted that the α-proteobacteria often grouped with the β-proteobacteria and actinobacteria in optimal trees I tested whether the eukaryotic affinity for this group as a whole (α-clade); this hypothesis was rejected by PFL (*P* = 0.013) but not by PFLA (*P* = 0.466) data sets (Table 2-1). The chloroplast origin hypothesis that constrained the cyanobacteria to group with the eukaryotes was not rejected for either PFL (*P* = 0.158) nor for PFLA (*P* = 0.074).

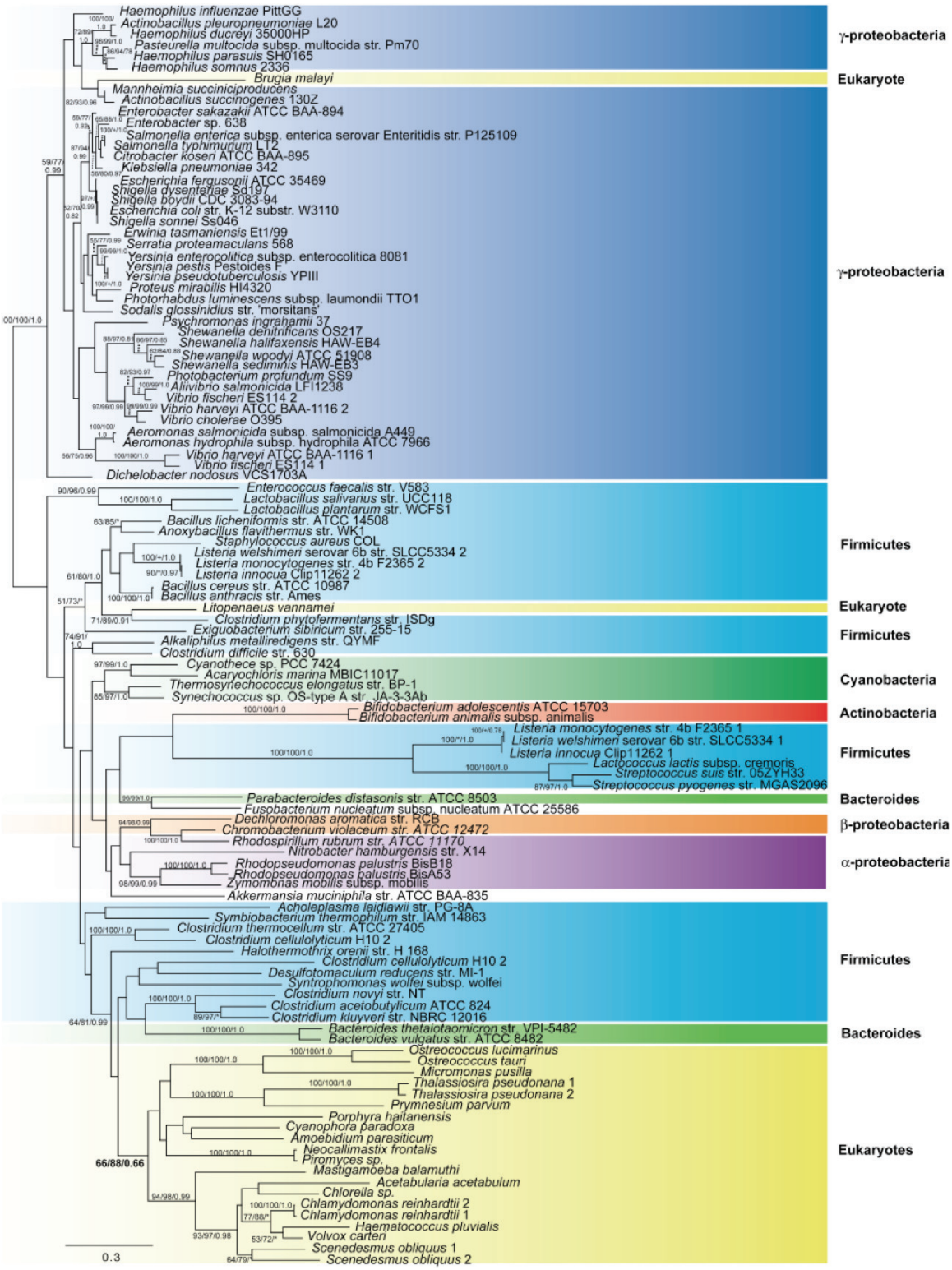

Figure 2-2: Global phylogeny of PFL in eukaryotes and bacteria. Support values (>50 or 0.5) were mapped onto the ML tree in the order of bootstrap support (BV)/corrected bootstrap support (BV_c) and Bayesian posterior probability. A total of 693 sites from 166 taxa were used for this analysis. "+" represents short branches that needed to be removed for aBPn analysis to work,

“*” represent those branches that were not found in the majority rule consensus summary of the Bayesian analysis.

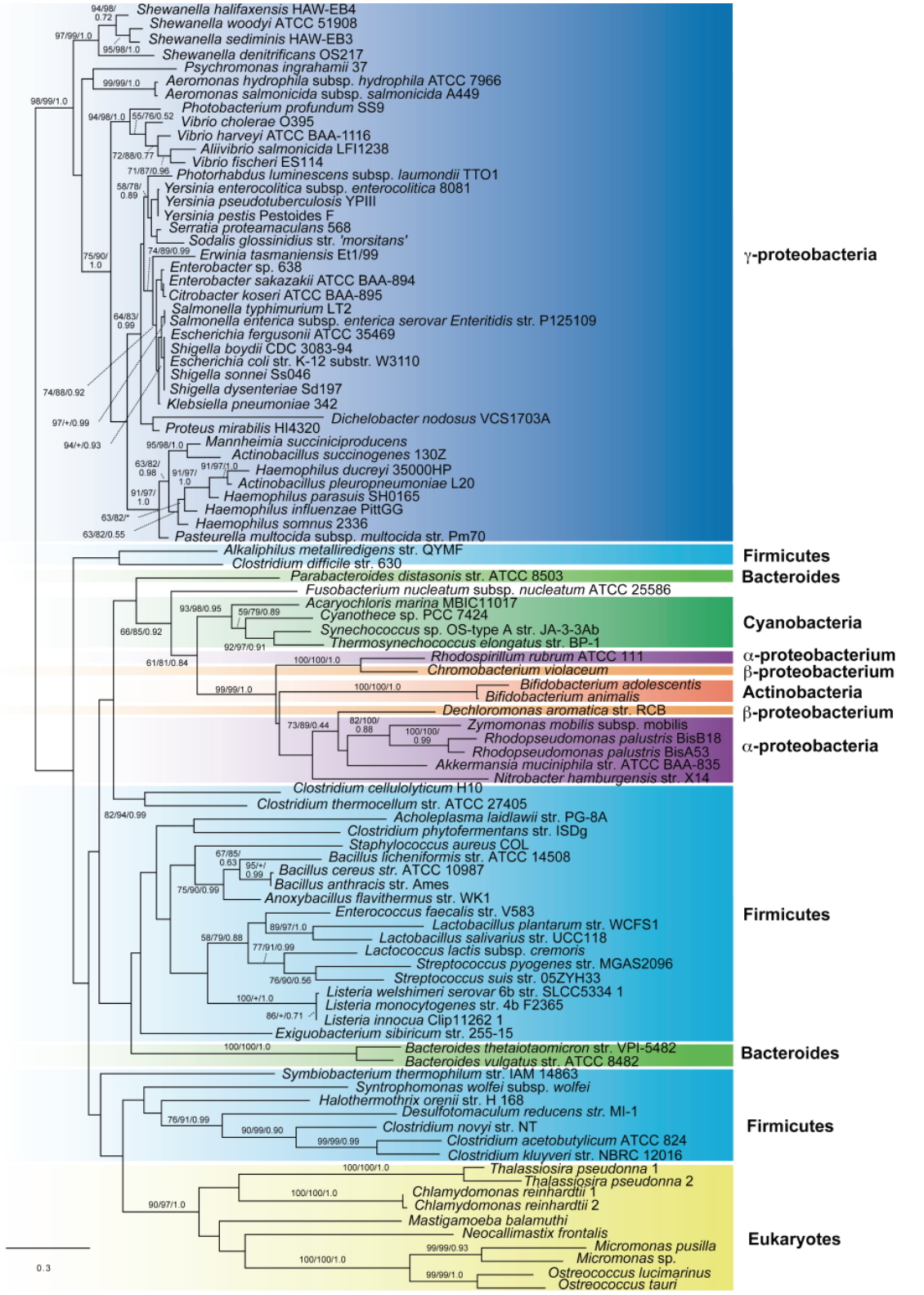

Figure 2-3: Global phylogeny of PFL activating enzyme in eukaryotes and bacteria. The topology shown was obtained by ML analysis. Support values are shown as in Figure 1. A total of 194 sites from 152 taxa were used for this analysis.

Although a cyanobacterial/chloroplast origin cannot be rejected in these analyses, it seems rather unlikely because a number of the PFL- and PFLA-containing eukaryotes, such as *Mastigamoeba*, the chytrid fungi, and *Amoebidium* are members of the so-called "unikont" eukaryotes that are likely to have diverged from eukaryotes prior to the origin of chloroplasts within eukaryotes (Lane and Archibald 2008). In any case, the limited resolution of the PFL and PFLA trees precludes us from making definitive conclusions. However, as the two proteins are tightly functionally linked, it is reasonable to suppose that they could share the same phylogenetic history. If so, then the support for a firmicute-*Bacteroides* relationship over α-proteobacterial or cyanobacterial affinity could be better assessed by concatenated analyses of these proteins.

2.4.3 Concatenation Of Pfla And Pfl Increases Monophyly Support And Suggests A Firmicute Ancestry For These Genes

I first tested the assumption that the two proteins shared the same history. In order to do this, all paralogous sequences were removed because some species had multiple PFL sequences but only one PFLA and thus each PFL could not be uniquely assigned a PFLA without duplication of the sequences. Congruence of the two protein phylogenies was tested using a likelihood ratio test implemented in CONCATERPILLAR (Leigh *et al.* 2008). Congruence was rejected (*P* value = 0.003657) with the full data sets with paralogs removed. Congruence rejection may have been the result of different topologies observed within the large γ-proteobacterial clade for the two proteins. As the latter group was strongly recovered as monophyletic (BV = 98 and PP = 1.0) and showed no affinity to the eukaryotic taxa in the phylogeny, I removed these sequences from the analysis and retested congruence; this time the null hypothesis of congruence could not be rejected (*P* = 0.056334). ML and BI were conducted on this reduced concatenated PFL–PFLA data set. For ML analyses, both partitioned (*i.e.,* allowing gene-specific branch lengths and alpha shape parameters) and nonpartitioned analyses were performed. As differences between the bootstrap values generated from these two analyses were small (*i.e.,* within Monte Carlo sampling error), I report only the unpartitioned bootstrap

values. The eukaryotic sequences remain monophyletic with borderline significant corrected bootstrap support (BV_c = 95), but with surprisingly poor posterior probability (PP = 0.5) (Figure 2-4A). Similar to the PFL analysis (Figure 2-2), the concatenated phylogeny shows a eukaryote–firmicute–*Bacteroides* grouping with strong support values (BV = 84, BV_c = 96, PP = 1.0). All alternative topologies were rejected by AU tests (Table 2-1), including cyanobacteria + eukaryote monophyly (P_{unpart} = 0.00005 and P_{part} = 0.009), which was previously not rejected.

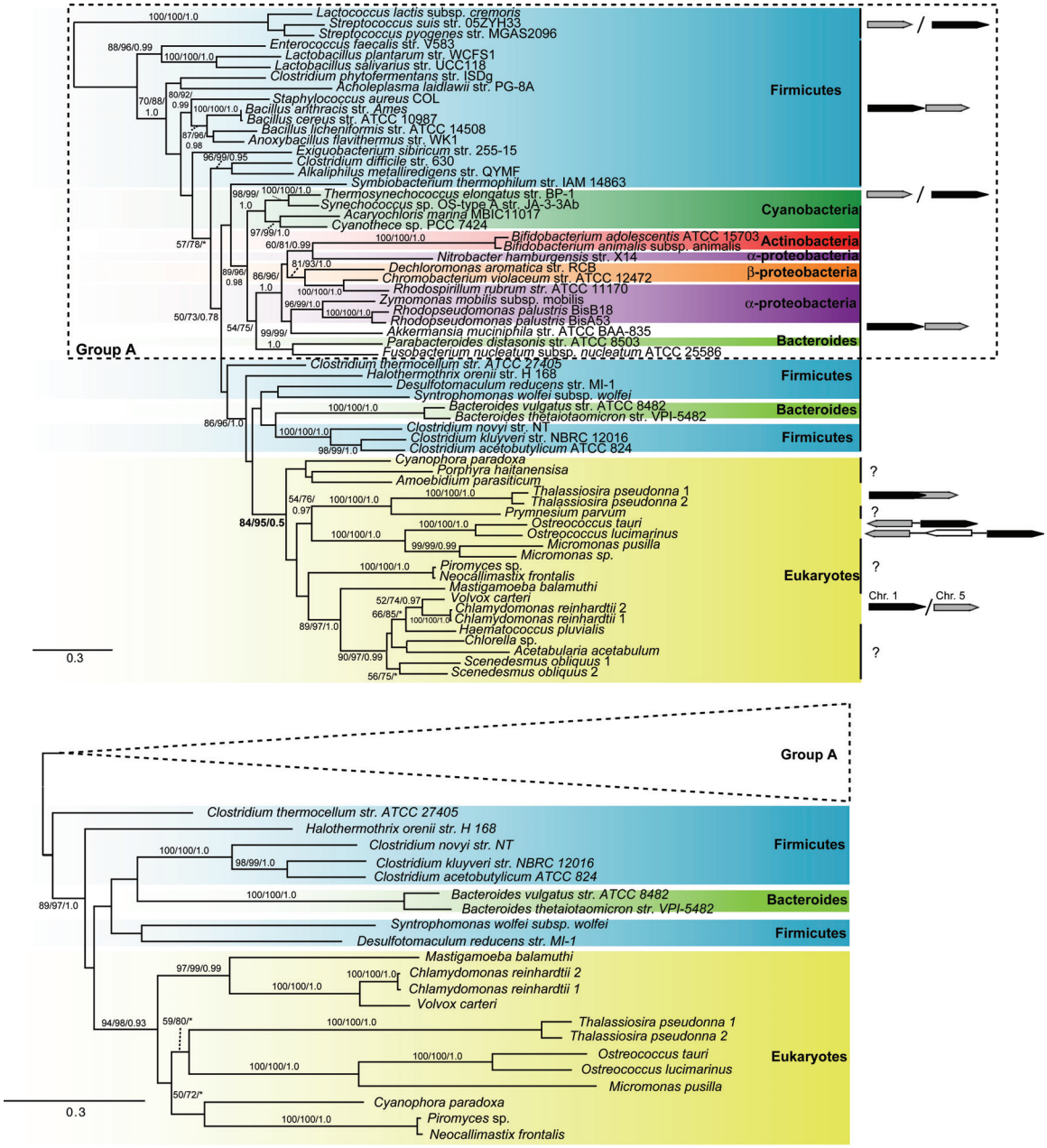

Figure 2-4 Phylogeny of concatenated pyruvate formate lyase and its activating enzyme. The topology shown was obtained by ML analysis**.** Support values for branches are shown above them in the order of bootstrap support (BV)/corrected bootstrap support (BV_c) and Bayesian posterior probability. Support values are only shown if with greater than 50% or 0.5. (*A*) A total of 887 sites from 63 taxa were used for this analysis. The conservation of operon architecture in bacteria or gene order in eukaryotes is shown in boxes to the right of the taxa where black and grey denote PFL and PFLA respectively. The dotted boxes represent those taxa that are shown as a wedge in B. (*B*) The same analysis except removing eukaryotic taxa with fewer than 50% of the sites in the total alignment. A total of 887 sites from 54 taxa were used for this analysis.

Because some of the eukaryote PFL and/or PFLA sequences in our analyses were partial with small numbers of sites, I evaluated the impact of removal of taxa with less

than 50% of the sites (Figure 2-4B). The tree topology did not change; however, support values for both eukaryotic monophyly (BV = 94, BV_c = 98, and PP = 0.93) and firmicute–*Bacteroides* affinity (BV = 89, BV_c = 97, and PP =1.0) increased (Figure 2-4B). The monophyly of eukaryotes in the both analyses suggests that eukaryotic PFL and PFLA share a single common prokaryotic origin probably resulting from a lateral gene transfer (LGT) event from either a firmicute- or Bacteroides-like organism. Interpretation of the source of transfer is complicated by the scrambled bacterial relationships observed. Because (i) bacterial "phyla" such as firmicutes, proteobacteria, and *Bacteroides* are not recovered as single groups but appear in multiple distinct regions of the tree and (ii) the presence/absence of PFL or PFLA homologs within bacterial genomes is patchily distributed across bacterial phyla, it is likely that LGT of these genes is occurring amongst these bacterial groups. Nevertheless, the most parsimonious interpretation of the placement of the eukaryotic clade within the bacteria suggests a firmicute ancestry.

2.4.4 Atypical Eukaryotic Relationships In Pfl And Pfla Phylogenies Suggest Eukaryote-To-Eukaryote LGT Events

The lack of evidence for a mitochondrial or cyanobacterial origin combined with apparent eukaryote monophyly in PFL and PFLA phylogenies can be explained most straightforwardly by three hypotheses. First, the observed pattern could be explained by several independent prokaryote-to-eukaryote LGT events from a common or closely-related prokaryotic source. Alternatively, there could have been a single origin of these enzymes by bacteria-to-eukaryote LGT along the lineage leading to the last common ancestral eukaryote followed by differential loss of the enzymes in disparate eukaryotic lineages (but retention in some facultative and obligate anaerobes). Finally, the enzymes could have been transferred from bacteria more recently to an extant eukaryote lineage and subsequently spread to other lineages by multiple eukaryote-to-eukaryote LGT events.

If the first scenario—the independent acquisition of PFL/PFLA from similar firmcute-like sources—were correct, then denser sampling of genomes from clostridial taxa could in principle show that the eukaryotic sequences are not monophyletic. Although this

scenario cannot be formally tested at this time, it seems unlikely that all relevant bacterial taxa that could disprove it are either extinct or not currently sequenced, thus this is not a favoured hypothesis. If the second "ancestral plus differential loss" hypothesis were true, then I would expect the PFL and PFLA phylogenies to resolve typical eukaryotic relationships between organisms possessing these enzymes. To test this hypothesis, ML (partitioned and nonpartitioned) and BI analyses were performed on the concatenated twenty-one eukaryotic PFL and PFLA sequences available (Figure 2-5). Again, the differences in BVs observed between partitioned and nonpartitioned analyses were within Monte Carlo error and so only the latter are shown.

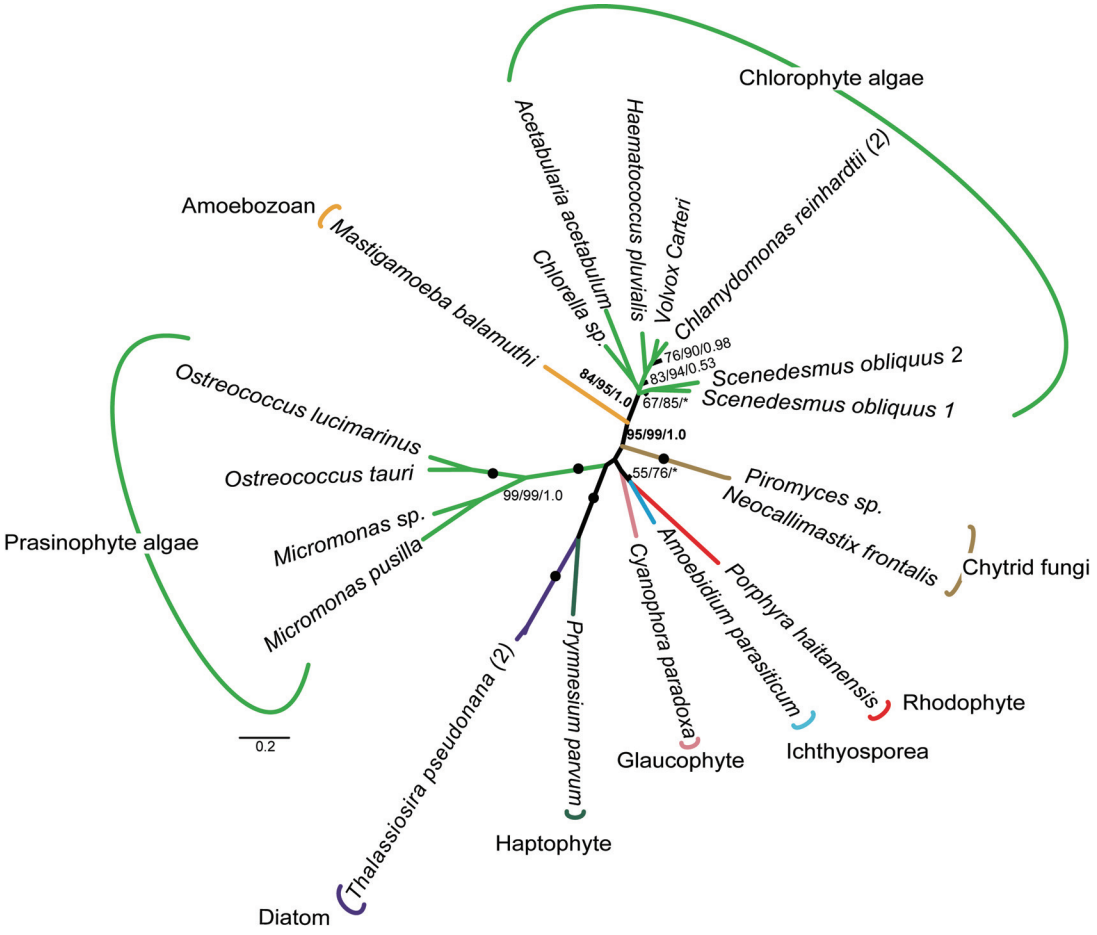

Figure 2-5: Atypical eukaryotic phylogeny of concatenated pyruvate formate lyase and its activating enzyme. Support values are as indicated in Figure 2-2. Branches with maximum support are shown with black circles. Major eukaryotic groups are labeled.

The most striking observation from these analyses is the robust recovery of both the chlorophyte and prasinophyte green algal clades as branching strongly apart (in terms of

high BV and PP values) in the tree, with all of the remaining eukaryotic lineages emerging between them. For example, the sequences from the amoebozoan *Mastigamoeba* group with the chlorophytes to the exclusion of all other eukaryotes with significant support (BV = 95, BV_c = 99, and PP = 1.0). Another odd feature of the analyses is the clustering of the ichthyosporean *Amoebidium* with the glaucophytes and rhodophytes and not the other opisthokonts (*Neocallimastix frontalis* and *Piromyces* sp.). The placement of the glaucophyte and rhodophytes in relation to the other lineages is not well resolved which has been observed previously in multiprotein eukaryotic phylogenies (Hampl *et al.* 2009; Nozaki *et al.* 2009). Finally, the haptophyte and diatoms branch together with strong support (BV = 100, BV_c = 100, and PP = 1.0).

To examine whether these unexpected eukaryotic relationships reflected a strong signal in the data versus random error, topology tests were performed in both unpartitioned and partitioned analyses (see Table 2-1 and schematic representations of topologies given in Figure 2-6). I first tested the compatibility of the data with a conservative view of known eukaryotic relationships. The following groups were enforced as topological constraints during the tree searching: monophyly of the unikonts with the amoebozoan *Mastigamoeba* to the immediate exclusion of monophyletic opisthokonts (chytrids and the icthyosporean) and monophyly of the bikonts with the haptophyte and stramenopile (diatom) to the immediate exclusion of the archaeplastid groups (prasinophytes, chlorophytes, rhodophyte, and glaucophyte). The optimal trees estimated with this "typical eukaryotic phylogeny" constraint were rejected in both nonpartitioned (P_{unpart} = 0.002) and partitioned analyses (P_{part} = 0.003). The monophyly of the green algal lineages (prasinophytes and chlorophytes) was also rejected for both analyses (P_{unpart} = 0.01 and P_{part} 0.008) as was monophyly of the green algal lineages plus *Mastigamoeba* (P_{unpart} = 0.038 and P_{part} = 0.0330) suggesting that it is not only the placement of *Mastigamoeba* that is responsible for the separation of the algae. Finally, opisthokont monophyly (*i.e., Amoebidium* + chytrid fungi) was tested but not rejected ($P_{unparat}$ = 0.6610 and P_{part} = 0.6650). Thus, only in the latter case could random error alone

be responsible for the anomalous branching pattern. Addition of more taxa may be helpful to improve resolution of this region of the phylogeny.

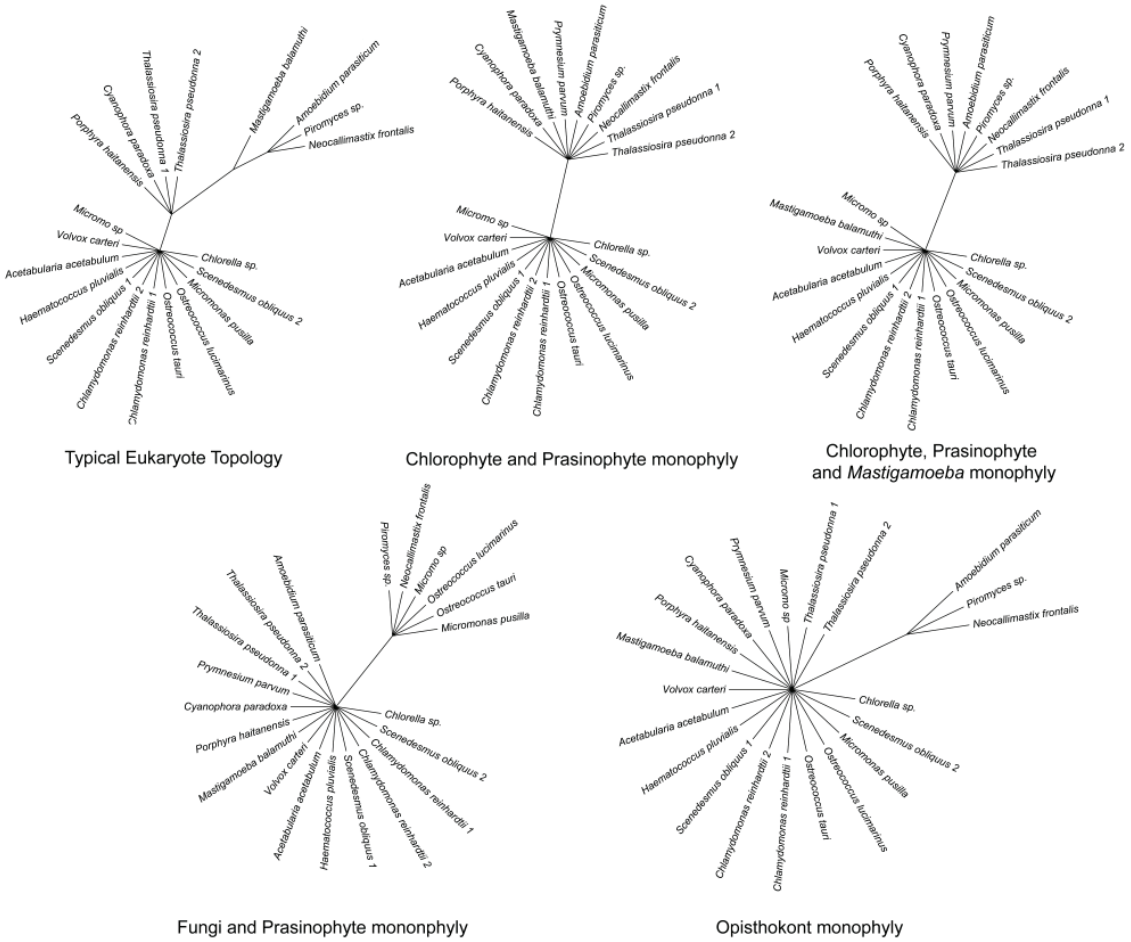

Figure 2-6: Radial diagrams of different eukaryotic relationships constrained for topology testing. The five eukaryotic topologies tested in Table 1 are demonstrated as labeled.

The foregoing analyses suggest that many of the peculiar eukaryotic relationships observed in the tree derive from strong signals in the data. To test whether the observed relationships were phylogenetic artifacts, I analyzed the amino acid compositions of each of the eukaryotic taxa relative to the full data set with a chi-squared test implemented in TREE-PUZZLE (Schmidt *et al.* 2002). I found significant deviation in the amino acid compositions of the sequences from *Micromonas pusilla* (P_{PFL} = 0.0001 P_{PFLA} = 0.00384) and *Neocallimastix* (P_{PFLA} = 0.0036). Reanalyses of the data sets after removing these two taxa from the concatenated analysis did not significantly

change the overall tree topology or support values for strongly supported branches. Furthermore, bootstrapped clustering analyses of the amino acid composition vectors of sequences in the eukaryotic and concatenated data set showed no strongly supported internal structure. The deepest clusters recovered in these dendrograms did not resemble deep groupings in ML or Bayesian phylogenies suggesting that amino acid composition alone is not biasing the estimated phylogenies. Finally, removal of fast evolving sites using the AIR package (Kumar *et al.* 2009) did not change the estimated topology, although, not surprisingly, support for eukaryotic monophyly and firmicute affinity generally decreased after removal of the ~50% fastest-evolving of the positions from the alignment (Figure 2-7)

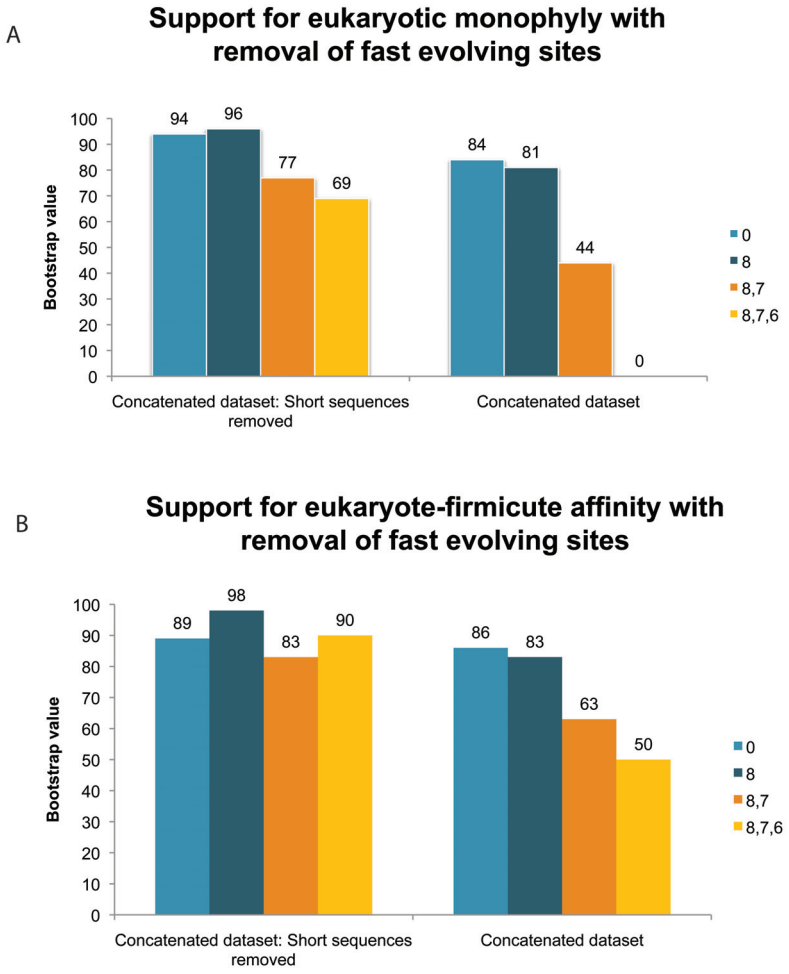

Figure 2-7: Removal of fast-evolving sites. Varying levels of fastest evolving sites were removed from the *concatenated* (Fig 2A) and *concatenated without short sequences* (Fig 2B) datasets using the AIR-identifier and -remover tools where sites classified as 8 are the fastest evolving.

The values shaded light blue (0) represent the dataset with no sites removed. 100 bootstrap replicates were generated and mapped on to the best-scoring ML tree. Eukaryotic monophlyly (**A**) and firmicute affinity (**B**) bootstrap values are reported.

Therefore, in the absence of any recognizable source of systematic error in these data sets (*e.g.,* long-branch attraction or amino acid composition bias), the atypical eukaryotic relationships observed likely reflect a true historical signal. This, combined with the extreme patchiness of PFL and PFLA genes observed across eukaryotic diversity (Figure 2-1), disfavors the "ancestral plus differential loss" scenario. A variation of this scenario that invokes large numbers of early gene duplications within eukaryotes followed by differential loss is conceivable and would explain the odd phylogenetic patterns observed. As none of the eukaryotes so far examined have retained multiple putative "ancient paralogs" expected under this scenario and as the number of parallel loss events required to explain the extant presence/absence pattern becomes even greater as does the number of paralogs that had to have been in the common eukaryotic ancestor, it seems to be a rather unlikely explanation. This leaves the hypothesis of a firmicute-to-eukaryote LGT event into an extant eukaryote lineage followed by several subsequent eukaryote-to-eukaryote gene transfers as a remaining reasonable scenario. The number of documented LGTs from prokaryotes to eukaryotes has increased steadily as more genomic data becomes available (Andersson 2009). Similarly, the frequency of reports of eukaryote-to-eukaryote LGT events have also increased; this process may occur via mechanisms, such as secondary plastid symbioses (Kamikawa *et al.* 2009), simple phagotrophy (Andersson 2009), or via parasitic genetic elements (Gilbert *et al.* 2010). Therefore, a "multiple LGT" scenario for PFL/PFLA evolution within eukaryotes seems to be at least a plausible explanation for the phylogenetic patterns observed.

Assuming, for the moment, that the foregoing LGT scenario for PFL and PFLA is correct, it is possible to make informed conjectures about the sequence of events that took place (Figure 2-8). For example, these genes could have been acquired by an ancestral archaeplastidan organism and directly inherited by *Cyano. paradoxa*, *Por. haitanensis* and the two green algal lineages. Assuming that these chlorophytes and prasinophytes form a clade to the exclusion of streptophytes (Baurain *et al.* 2010) then

loss of PFL and PFLA would have to be invoked in the latter lineage and in the red alga *Cyanidioschyzon merolae* (Figure 2-5). The remainder of eukaryotic organisms in Figure 2-5 and Figure 2-8 could have acquired the genes then by a minimum of four distinct eukaryote-to-eukaryote LGT events from the various archaeplastidan lineages. However, many other alternative scenarios that invoke a similar number of LGT events from different founding eukaryotic lineages are also possible. Ultimately, much more genomic data from many more representatives of the various microbial eukaryotic taxa involved will be needed to 1) confirm or refute the general LGT scenario I have outlined and 2) clarify the timing of the various genetic events involved in the evolution of PFL/PFLA within eukaryotes.

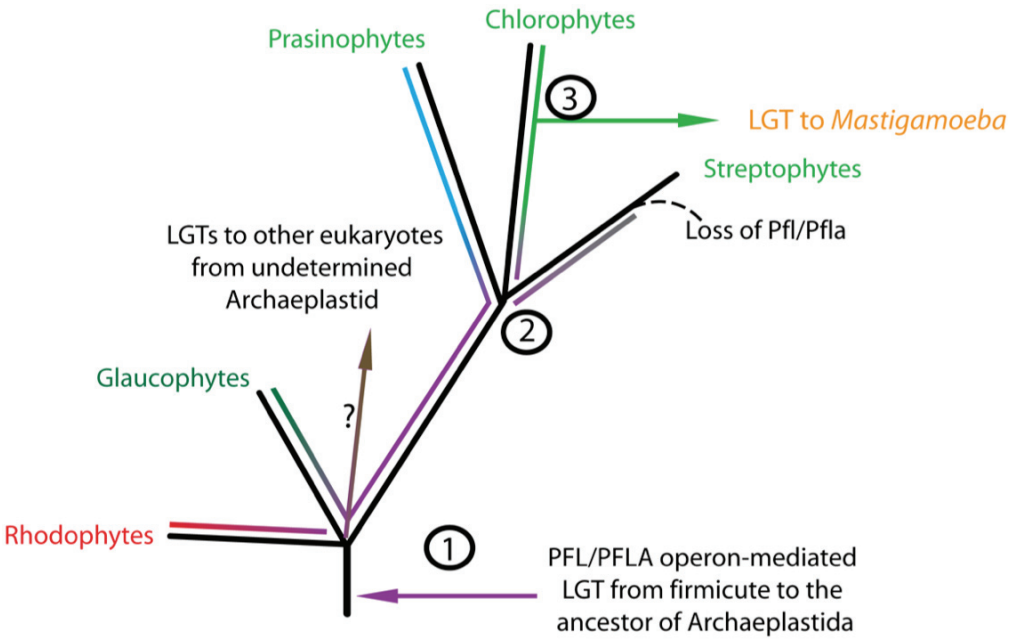

Figure 2-8: Hypothesis for the origin of eukaryotic PFL/PFLAE and its transfer amongst eukaryotes. (1) Operon-mediated transfer of PFL/PFLAE from a firmicute-like bacteria in the common ancestor of Archaeplastida and subsequent evolution of the eukaryotic PFL over speciation events (purple line). Possible LGTs from an unknown archeaplastida to other eukaryotes (2) Divergence of the main lineages of green algae and streptophytes. (3) Further evolution of the green lineages gave rise to distinct types of PFL/PFLA in the prasinophytes (blue), chlorophytes (green) and early-branching streptophytes (grey). I propose that Mastigamoeba received PFL/PFLAE from a chlorophyte alga while chytrid fungi and other eukaryotes likely acquired the enzymes from a currently unidentified extant or ancestral Archaeplastids.

2.4.5 Conservation Of Gene Order: Insights Into The Transfer Mechanism

One potential difficulty with an LGT origin scenario for PFL and PFLA genes is that for a functioning enzyme system to be "moved" between organisms, at a minimum, both genes must be transferred successfully. However, for the initial bacteria-to-eukaryote transfer, this problem is avoided if the two genes were genetically linked in an operon that could be transferred in a single genetic event (Andersson and Roger 2003). In γ-proteobacteria, such as *Escherichia coli* and most firmicutes, *pfl* and *pfla* are indeed encoded close together in a putative operon as predicted by OperonDb (Pertea *et al.* 2009). A survey of the gene organization and order is summarized in Figure 2-4 for those taxa analyzed in the concatenated gene analysis. Separation of *pfl* and *pfla* loci in bacteria was only observed in the Streptococcaceae lineage (*Lactococcus* and *Streptococcus*) and *Thermosynechococcus elongatus* (Figure 2-4); neither of these groups is closely related to the eukaryotic clade. The initial LGT into eukaryotes could therefore have plausibly occurred from a firmicute-like organism encoding *pfl* and *pfla* in an operon.

With complete genome sequences available for *Thalassiosira*, *O. lucimarinus*, *O. tauri,* and *C. reinhardtii*, the gene order of *pfl* and *pfla* could also be determined. *Thalassiosira* appears to have maintained the firmicute-like gene order with *pfl* preceding *pfla*. Curiously, however, a close examination of data from the genome project (http://genome.jgi-psf.org/Thaps3_bd/Thaps3_bd.home.html) indicates that the genes are fused into a single open reading frame (ORF). I was able to identify multiple cDNA sequences that spanned the two coding regions, and found that none of the 3' ends of the upstream partial *pfl* cDNAs had poly(A) tails as would be expected if they were separately transcribed mRNAs. Fused versions of laterally transferred genes encoded in bacteria as two genes on an operon have been described previously for the small and large subunits of glutamate synthase (Andersson and Roger 2002). However, to my knowledge, this is the first example of a PFL–PFLA fusion protein. In *O. lucimarinus* and *O. tauri*, the PFL and PFLA are not fused, however, they are neighboring genes on the chromosome that are encoded on opposite strands (Figure 2-4). *Pfl* and

pfla are separated by only 117 bp of noncoding DNA in *O. tauri* and by an ORF encoding a predicted protein (variant-specific surface protein, gi:145341972) in *O. lucimarinus*. The *pfl* and *pfla* genes in the prasinophytes could have ancestrally had the same gene order as the firmicutes and *Thalassiosira*, but then suffered a rearrangement that transposed one gene relative to the other leading to their bidirectional orientation. After speciation, either *O. lucimarinus* acquired an ORF in between the two genes or *O. tauri* lost the ORF after speciation (a homologous ORF was not identified in the *Os. tauri* genome). In *M. pusilla*, the two genes appear on different scaffolds suggesting they are not genetically linked. Similarly, the *pfl* and *pfla* genes in *Chlamydomonas* are located on completely different chromosomes.

The fusion or genetic linkage of these genes in a number of these eukaryotic lineages is consistent with their hypothetical origin by an ancestral operon-mediated gene transfer event from bacteria. Furthermore, it makes the hypothesis of subsequent eukaryote-to-eukaryote LGT more tenable because a functioning PFL system could be successfully moved between organisms in a single event if the two genes were transferred on a single piece of DNA.

2.4.6 Gene Structures And Organization Of Pfl Genes In Diverse Eukaryotes

No obvious sequence and little positional conservation of introns was observed across lineages after scanning genomic data from *O. tauri, O. lucimarinus, M. pusilla, Micromonas* sp., *T. pseudonana, C. reinhardtii, Piromyces* sp., *Chlorella* sp., and *V. carteri*. The *Ostreococcus* species did not contain any introns in *pfl* or *pfla* genes. Within the *Micromonas* species, there was intron position conservation in *pfla* however the sizes of the introns differed significantly (72 bp in *Mi. pusilla* and 293 bp in *Micromonas* sp.). This might represent an ancestral intron of the *Micromonas* lineage that was acquired after speciation from the other prasinophyte lineage (*Ostreococcus*) or, alternatively, it could be older but was lost in the *Ostreococcus* lineage.

2.4.7 In Silico Predictions Of Subcellular Localization

Previous reports have localized PFL activity to the hydrogenosomes of *Neocallimastix* and *Piromyces* and to the chloroplasts and mitochondria of *Chlamydomonas* (Boxma *et al.* 2004; Gelius-Dietrich and Henze 2004; Atteia *et al.* 2006; Hemschemeier *et al.* 2008). To elucidate the organellar localization of the newly identified eukaryotic PFL and PFLA sequences, I utilized the software tools TargetP and Mitoprot II. In all cases, except one, these programs did not strongly predict an organellar localization of PFL or PFLA for the newly described sequences. The exception was a "weakly" predicted mitochondrial targeting peptide in the PFLA of *O. tauri* ($P_{TargetP}$= 0.319 and $P_{Mitoprot}$ = 0.773). It should be noted that all of the previously characterized organellar homologs were in fact predicted to possess mitochondrial targeting peptides by one or both of these programs. Therefore, at this stage, there seems no evidence for an organellar localization of the newly described PFL and PFLA homologues; they may all function within the cytoplasm. Obviously, further molecular and biochemical experiments are needed to confirm or refute a cytosolic localization as *in silico* localization prediction has limited accuracy on non-model organisms and many organellar proteins possess cryptic targeting signals (Hurt and Schatz 1987).

2.4.8 The Distribution Of Pyruvate Catalyzing Enzymes In Eukaryotes

The discovery of enzymatic alternatives to PDC, such as PFO, PNO, and PFL in some eukaryotes that all generate acetyl-CoA raises the question of their *raison d'être*. Some organisms such as *Chlamydomonas* and *Thalassiosira* appear to contain PFO, PFL, and PDC (Figure 2-1) which may indicate differential compartmentalization of these acetyl-CoA generating enzymes within these organisms. In *Chlamydomonas*, PFL functions in both the chloroplast and the mitochondrion under anaerobic conditions suggesting that this acetyl-CoA generating enzyme might functionally replace PDC under anaerobic conditions or under conditions where high CO_2 or redox levels are inhibiting PDC- or PFO-mediated catalysis. This hypothesis is supported by the fact that PFL functions in the absence of redox components typically associated with acetyl-CoA generation, such

as $NAD(P)^+$ or ferredoxin. Indeed, activation of PFL by PFLA likely requires reduced components—more specifically ferredoxin—in species of *Clostridium* (Thauer *et al.* 1972). Therefore, the main need for PFL may arise under anaerobic conditions where reduced cofactors/proteins are at high levels, but acetyl-CoA is still needed for anabolic pathways, such as lipid and amino acid biosynthesis or even ATP generation. As mentioned earlier, initial PFL activation is dependent on SAM whose synthesis requires ATP, therefore, the radical must be protected in order to not lose this energy investment.

2.5 Conclusion

My survey of eukaryotic genomic data revealed a wide diversity of microbial eukaryotic lineages that possess PFL homologues and their activating enzymes. Many of these organisms (*e.g.,* the diatom *T. pseudonana*, the seaweed *P. haitanensis*, and the glaucophyte *C. paradoxa*) are traditionally thought of as "aerobic" eukaryotes, but it is likely that they also transiently experience low oxygen conditions and this oxygen-sensitive system may then be important for continued production of acetyl-CoA. I have shown that all PFL homologues and their activating enzymes encoded in eukaryote genomes form a monophyletic group that were likely acquired once from a bacterial source. The eukaryotic enzymes appear to be most closely related to homologues found in selected firmicute Gram-positive bacteria and show no evidence of a mitochondrial or chloroplast endosymbiotic ancestry. The most parsimonious explanation for eukaryote monophyly in this case is a single LGT event into a eukaryotic lineage, probably from a firmicute donor. The patchy distribution of PFL and PFLA across eukaryotic diversity and strongly supported atypical eukaryotic relationships recovered in the phylogenies of these enzymes argues against an origin of the enzymes in the common ancestral eukaryotic lineage followed by differential lineage-specific loss. Rather, I propose that after the initial transfer of the two genes residing in a bacterial operon into a eukaryotic lineage, subsequent eukaryote-to-eukaryote transfers of the two genes occurred giving rise to the present-day PFL and PFLA distribution in eukaryotes (Figure 2-8). This

scenario is made more plausible by my finding that the two genes are sometimes closely genetically linked (or in one case fused) in eukaryotic genomes so that single transfer events may have been sufficient to spread a functional PFL system between eukaryotes. Although I favor the latter scenario, it should be regarded as tentative as a denser and broader sampling of eukaryotic microbial genomes will be required to confirm, refine, or refute it.

The probable spread of PFL and PFLA via LGT within the eukaryotic realm questions the assumption, often made by comparative genomic analyses, that orthologues detected in a few distantly related eukaryotes can be automatically assumed to have been present in the common ancestor of all eukaryotes. Additionally, this reconstruction of the evolutionary histories of PFL and PFLA does not straightforwardly fit with the scenario for the origin of eukaryotic enzymes of anaerobic metabolism advanced in the Hydrogen Hypothesis (Martin and Müller 1998). A simple interpretation of the Hydrogen Hypothesis would predict that these enzymes were present in the eukaryotic common ancestor and should derive from the mitochondrial endosymbiont. Neither of these claims are supported by these aforementioned analyses of the PFL and PFLA data.

Finally, although the rationale for the presence of multiple pyruvate catalyzing enzymes in some eukaryotes remains uncertain, it likely involves the differential regulation of these enzymes under different oxygen tension and redox conditions. Experimental testing of gene expression, enzyme activities, and subcellular localization of these proteins under various conditions in the organisms that possess them will be required to fully understand their biochemical roles and evolutionary histories.

Chapter 3 A Mitochondrial Sulfur Mobilization (SUF) System In The Anaerobe *Pygsuia biforma*

This chapter contains work published in **Stairs, C.W**., Eme, L., Brown, M. W., Mutsaers, C., Susko, E., Dellaire, G., Soanes, D. M., van der Giezen, M. and Roger, A.J. (2014) A mitochondrial sulfur mobilization (SUF) system in the anaerobe *Pygsuia*. *Curr. Biol.* 24(11): 1176-1186.

3.1 Abstract

Many microbial eukaryotes have evolved anaerobic alternatives to mitochondria known as mitochondrion-related organelles (MROs). Yet, only a few of these have been experimentally investigated. Here I report an RNAseq-based reconstruction of the MRO proteome of *Pygsuia biforma*, an anaerobic representative of an unexplored deep-branching eukaryotic lineage. *Pygsuia*'s MRO has a completely novel suite of functions, defying existing 'function-based' organelle classifications. Most notable is the replacement of the mitochondrial iron-sulfur cluster machinery by an archaeal sulfur mobilization (SUF) system acquired *via* lateral gene transfer (LGT). Using immunolocalization in *Pygsuia* and heterologous expression in yeast, I show that the SUF system does indeed localize to the MRO. The *Pygsuia* MRO is predicted to possesses a unique assemblage of features, including: cardiolipin, phosphonolipid, amino acid, and fatty acid metabolism; a partial Kreb's cycle; a reduced respiratory chain; and a laterally acquired rhodoquinone (RQ) biosynthesis enzyme. The latter observation suggests that RQ is an electron carrier of a fumarate reductase-type complex II in this MRO. The unique functional profile of this MRO underscores the tremendous plasticity of mitochondrial function within eukaryotes and showcases the role of LGT in forging metabolic mosaics of ancestral and newly acquired organellar pathways.

3.2 Introduction

Mitochondria of modern-day eukaryotes evolved from an α-proteobacterial endosymbiont that was integrated as an organelle within a host cell prior to the last eukaryotic common ancestor (Gray *et al.* 2001). In aerobic eukaryotes, mitochondria carry out a number of important functions including pyruvate decarboxylation, oxygen-dependent ATP production, amino acid metabolism and iron-sulfur (Fe-S) cluster biosynthesis. Over the past twenty years, investigations into the mitochondria or homologous organelles of anaerobic organisms (mitochondrion-related organelles, MROs) have revealed a wide variety of different metabolic phenotypes.

Classical 'aerobic' mitochondria generate ATP by oxidative phosphorylation using ATP synthase coupled to the electron transport chain ultimately reducing O_2 to H_2O. However, anaerobically-functioning mitochondria have also been described in a number of eukaryotes (*e.g., Ascaris*) that, under hypoxic conditions, produce ATP but employ a terminal electron acceptor other than O_2 (*e.g.,* fumarate (Kita *et al.* 1988)). Radically different MROs, known as hydrogenosomes found in parasites such as *Trichomonas,* lack organellar genomes and produce ATP by an anaerobic pathway that is typically not found in classical mitochondria. In these organelles, pyruvate is oxidized to acetyl-CoA and CO_2 by a pyruvate:ferredoxin oxidoreductase (PFO) and the reduced ferredoxin is re-oxidized by an iron-only [FeFe]-hydrogenase that reduces protons to H_2 gas (Steinbüchel and Müller 1986). Acetyl-CoA is then converted to acetate by an acetate:succinate CoA transferase (ASCT) and the resulting succinyl-CoA is utilized by succinyl-CoA synthetase (SCS) to generate ATP by substrate-level phosphorylation (van Grinsven *et al.* 2008). Other anaerobic protists contain MROs called mitosomes that do not produce ATP, and function in Fe-S cluster formation via a mitochondrial-type iron-sulfur cluster (ISC) system (*e.g., Giardia* (Tovar *et al.* 2003)). In mitosome-containing protists such as *Giardia* and *Entamoeba*, ATP-production occurs by substrate-level phosphorylation in their cytoplasm (Reeves *et al.* 1977).

Recent investigations of hitherto neglected parasitic, commensal and free-living organisms have greatly expanded the spectrum of known functions of MROs (Gill *et al.*

2007; Stechmann *et al.* 2008; Barberà *et al.* 2010; Burki *et al.* 2013; Zubáčová *et al.* 2013). For example, several distantly-related protists have organelles recently described as 'hydrogen-producing mitochondria' (HPM). HPM not only have mitochondrial genomes and many canonical mitochondrial pathways (including components of the electron-transport chain, ETC), but also possess enzymes of the anaerobic 'hydrogenosomal' ATP generation pathway. Other MROs lacking mitochondrial DNA have also been described, each with a distinct combination of mitochondrial and hydrogenosomal properties. For instance, the MROs of the free-living excavate *Trimastix pyriformis,* possess several mitochondrial pathways involved in amino acid metabolism as well as enzymes for hydrogen but lack full ETC complexes (Stechmann *et al.* 2006; Zubáčová *et al.* 2013). In contrast, *Mastigamoeba balamuthi,* a free-living amoeba, has MROs with complex II (but no other ETC complexes) in addition to serine and glycine metabolic pathways as well as a [FeFe]-hydrogenase and a PFO (Gill *et al.* 2007) .

Virtually all mitochondria and MROs of all studied extant eukaryotes generate Fe-S clusters for mitochondrial Fe-S proteins using the iron-sulfur cluster system (ISC) (Stehling and Lill 2013). Fe-S clusters can also be synthesized in other cellular compartments such as the cytosol or in plastids. The cytosolic iron-sulfur cluster assembly machinery (CIA) matures cytoplasmic and nuclear Fe-S proteins and, in yeast, has been show to rely on the ISC system to supply it with an unknown sulfurous factor, so-called 'factor X' (Stehling and Lill 2013). Plastids use an endosymbiont-derived sulfur mobilization (SUF) pathway (Takahashi and Tokumoto 2002).

However, several eukaryotes have recently been shown to deviate from the foregoing patterns. *Entamoeba* and *Mastigamoeba* completely lack the ISC system and instead employ a nitrogen-fixation (NIF) related Fe-S biogenesis system that they have acquired by lateral gene transfer (LGT) from ε-proteobacteria (van der Giezen *et al.* 2004; Nyvltová *et al.* 2013). In *Mastigamoeba*, there are duplicates of the NIF genes that encode distinct paralogues that function in the cytosol and the MROs of this organism (Nyvltová *et al.* 2013). The only other known exception to the general eukaryotic pattern is in the anaerobic stramenopile *Blastocystis* sp. in which an archaeal-

like SUFCB fusion protein was shown to function in the cytosol and is induced under oxidative stress (Tsaousis *et al.* 2012).

Here, I reconstruct the proteome of the MROs of the breviate, *Pygsuia biforma*, a free-living anaerobic amoeboid flagellate from hypoxic marine sediments (Brown *et al.* 2013). The breviates have recently been shown to be an early-emerging group, branching at the base of the eukaryote supergroup Obazoa, comprised of Opisthokonts (animals and fungi) and apusomonads (Brown *et al.* 2013)**.** Our predictions reveal an extraordinarily distinct MRO in this organism. In addition to possessing several systems and pathways never before detected in MROs (*e.g.,* rhodoquinone, cardiolipin and phoshonolipid biosynthesis), it has hydrogenosomal-like anaerobic energy metabolism and a partial electron transport chain consisting of complex II, alternative oxidase (AOX) and electron transferring flavoprotein (ETF). Most unexpectedly, a mitochondrial-type ISC system appears to be completely absent. Instead *Pygsuia* expresses duplicated methanomicrobiales/*Blastocystis*-like SUFCB proteins that it has acquired by LGT. One of the SUFCB proteins localizes to its MRO suggesting that it may functionally replace the ISC system. This is only the second known lineage where the mitochondrial ISC system has apparently been lost and the first case where the SUFCB system seems to have taken over its role in Fe-S cluster biogenesis within MROs.

3.3 Methods

3.3.1 Pygsuia *RNAseq Transcriptome Analysis And Dataset Filtering*

The assembled *Pygsuia biforma* transcriptome previously reported in (Brown et al. 2013) was used for all analyses. The average coverage of all clusters was 143.542 (Brown *et al.* 2013). Despite *Klebsiella pneumoniaea* being supplemented in the culture media, the predominant bacterial sequences identified in the transcriptome was *Arcobacter.* Sequences that were greater than 96 % or 98 % similar to bacterial or obvious eukaryotic contaminants (*e.g.,* animals) at the nucleotide and amino acid level (over at least 150 nucleotides or 50 amino acids) respectively were removed. Visual inspection of the remaining putative 'eukaryotic' sequences (*i.e.,* likely to be *Pygsuia* genes) revealed an

unusual and redundant thymidine-rich pattern in the 5' untranslated region (5' UTR). The presence of these repeated thymidine stretches allowed Dr. Edward Susko and Dr. Laura Eme to implement conditional probability models trained on both 5'-UTR trinucleotide preferences and on codon usage of the coding region in order to distinguish *bona fide Pygsuia* genes from putative contaminants. Models were trained on reference datasets of (i) 159 *Pygsuia* genes (Brown *et al.* 2013) and contaminants identified using BLASTN from (ii) 150 *Arcobacter* genes; (iii) 68 other bacterial contaminants identified from BLASTN (discussed above); and (iv) 150 genes from eukaryotic contaminants identified from BLASTN. Specific parameters of this statistical are reported elsewhere (Stairs *et al.*, 2014).

For calculation of depth-of-read coverage in my *P. biforma* RNAseq assembly, the filtered raw Illumina RNAseq reads, used for the initial assembly, were mapped onto the assembled clusters using the AlignReads tool in the Trinity package. The AlignReads tool executed Bowtie as the aligning algorithm. The mapping assembly output was sorted using SAMtools (Li *et al.* 2009). From this sorted output, coverage depth of each cluster was assessed using the Qualimap analysis tool (García-Alcalde *et al.* 2012).

3.3.2 MRO Protein Prediction

Proteins were deemed MRO if one or more of the following criteria were satisfied: (i) presence of an N-terminal extension relative to bacterial homologs which was predicted to be mitochondria-targeting by Mitoprot (Claros and Vincens 1996) or TargetP (Emanuelsson *et al.* 2000) (P>0.5) and (ii) the most significant BLAST hits were against known mitochondrial or MRO proteins. Putative *Pygsuia* MRO proteins were used to mine homologs from our in-house *Breviata anathema* data as well as the recently released *Subulatomonas tetrapora* (Grant *et al.* 2012) transcriptome.

Using the algorithm designed by Dr. Edward Susko and Dr. Laura Eme discussed above, I assessed the probability of each predicted MRO protein being a true *Pygsuia* sequence and not a contaminant of bacterial (*Arcobacter*) or eukaryotic (deeply sequenced animal, plant or fungi) origin. All but seven (*tim16/pam27*, Citrate synthase-like, *hyda-cysj4*, *msf1/ups2, icp55, pcca2 and prx5*) favoured the *Pysguia* model versus

being from another source. These seven exceptions were nevertheless verified to be from *Pygsuia* by applying several other criteria. For *tim16*, although the favoured model was *Arcobacter*, orthologues of this gene have not been identified in any bacterial genomes, including *Arcobacter* genomes available on the *nr* database (NCBI). For the Citrate-synthase-like (*cs*), *msf/ups2*, *icp55*, and *pcca*2, the probabilistic model indicated that the frequencies were most consistent with the other eukaryotic contaminants in the RNAseq data. However, all of the eukaryotic contaminants I identified come from deeply sequenced genomes (*e.g., H. sapiens*) and by BLASTN (Altschul *et al.* 1990; Lander *et al.* 2001; Venter *et al.* 2001) none of these appear to possess a gene identical (or nearly identical) on the sequence level to these *cs, msf1/ups2, icp55,* and *pcca2* genes in my data. Finally, the *hyda-cysj4* and *prx5* genes were flagged as of possible bacterial (non-*Arcobacter*) origin. However, the phylogenetic analysis of HYDA-CYSJ4 revealed that it was a closely-related paralogue of four other *Pygsuia* HYDA, while a BLASTN against the *nr* database using the *prx5* gene sequence as a query did not return a hit.

3.3.3 Breviata anathema *Expressed Sequence Tag Project and* Subulatomonas tetraspora *transcriptome Analysis*

A Sanger EST project of *Breviata anathema* was performed by Dr. Mark van der Giezen (University of Exeter) consisting of a total of 6,937 sequences assembled into 1,520 clusters. Genes from *Pygsuia* were also used to search an EST database for corresponding *Breviata* orthologues. I also searched the recent 454 pyrosequencing transcriptome of another closely-related breviate, *Subulatomonas tetraspora*, using the *Pygsuia* homologues as queries. Due to the low sequence coverage and protein complement identified in the *Breviata anathema* and *Subulatomonas tetrasporta* data, the *Pygsuia* MRO protein predictions are the primary focus of this study.

3.3.4 Conserved Functional Domain Search

The identification of functional domains was carried out using HMMER 3.0 (Eddy 1998) (http://hmmer.org/) and the HMM profiles of the Pfam database (Finn *et al.* 2006) version 26.0. HMM profiles with e-values < $10e^{-5}$ were considered as significant.

3.3.5 Phylogenetic Dataset Construction

I used the *Pygsuia* amino acid sequences as queries to identify their homologues using BLASTP (Altschul 1997) and TBLASTN against the *nr* database. BLAST outputs were examined by eye to identify homologues of each protein to avoid applying an arbitrary cutoff on e-value. Eukaryotic homologues from previously published phylogenetic analyses (such as for PFO) were included when available. For each protein, homologous sequences were gathered in a dataset and aligned with MAFFT 6.903 (Katoh and Toh 2008) or MUSCLE (Edgar 2004). All the resulting alignments were manually inspected and edited if necessary using the ED program from the MUST package (Philippe 1993). Regions where homology between sites was doubtful were removed from the alignments before phylogenetic analyses using BMGE (default parameters) (Criscuolo and Gribaldo 2010).

3.3.6 Phylogenetic Analysis

Preliminary phylogenies were reconstructed using FastTree (Price et al. 2010). In order to maintain a reasonable-sized data set for subsequent phylogenetic analyses, an in-house script was used to remove extremely similar sequences from prokaryotic species of the same genus. More specifically, if these were forming a monophyletic clade in the preliminary trees, only the shortest branching sequence was kept. In addition, as I was specifically interested in the evolutionary history of the Pygsuia sequences (i.e., not in tracing back exhaustively the evolutionary events of the protein family), distant homologues were discarded based on preliminary phylogenetic trees. Following dataset trimming, alignments were recomputed and masked as outline above. The phylogenetic results presented here differ subtly from those presented in Stairs et al. (2014),

specifically the evolutionary model selection and the RAxML version. Here, final ML phylogenetic analyses were generated using RAxML version 8.0.19 (Stamatakis 2014) under the highest likelihood scoring model (LG, LG4X or LG4M) using Akaike Information Criterion (AIC) (Le et al. 2012) with gamma correction of evolutionary rates across sites heterogeneity. For each dataset, bootstrap support for splits estimated from 500 bootstrap replicates were mapped onto splits on the best scoring ML tree (obtained by 100 heuristic search replicates). Bayesian inference (BI) was also conducted using PhyloBayes 3.2 (Lartillot et al. 2009) by running four chains under the catfix C20 model of evolution and poisson + Gamma 4. For each chain, a total of 300,000 generations were run, from which trees were sampled every 100 generations and discarding a manually determined burn-in of 50,000 generations for each (yielding a total of 2,500 trees). For all analyses except [FeFe]-Hydrogenase where only 3 chains converges, all four PhyloBayes chains converged. Posterior probabilities (PP) for splits were mapped on to the ML estimated topology.

3.3.7 Culturing And Microscopy

Cultures of *Pygsuia biforma* were maintained in American Type Culture Collection medium 802 (SW802) prepared in natural seawater (Dalhousie University Aquatron) at 22 °C as described in (Brown *et al.* 2013). *Pygsuia biforma* cells were seeded into 15 mL culture tubes supplemented with a loop of *Klebsiella pneumonia* and limited (< 0.5 ml) headspace. For molecular biological manipulations, cells were collected by centrifugation (400 g x 5 min, swinging bucket) and washed with natural seawater. Transmission electron microscopy was performed as previoiusly described (Brown *et al.* 2013).

For immunofluorescence microscopy, cells were grown on Geltrex (Life Technologies) coated slides overnight in an anaerobic chamber and incubated with 500 nM of Mitotracker Orange (Molecular Probes) for 30 min prior to removal from the chamber. All subsequent manipulations were performed at room temperature under atmospheric conditions unless otherwise stated. Immediately upon removal from the anaerobic chamber, cells were fixed (4% paraformaldehyde [Aurion] in SW802) for 10

min, washed with filtered SW802 3 x 5 min, 1 x 25 mM NH_4Cl, permeabilized in 0.5% Triton X-100-SW802 (Invitrogen) for 2 minutes and immediately washed with phosphate buffered saline (PBS) 3 x 5 min. Cells were blocked in blocking solution (goat serum blocking solution [Aurion] supplemented with 0.05% saponin) for 1 h. Primary antibodies (Genscript [mSUFCB, 1:200] and gift from A.G.M. Tielens [ASCT,1:100]) were prepared in blocking solution and incubated with coverslips or chamber slides overnight at 4 °C. Slides were washed 3 x 5 min in PBS and incubated with secondary antibodies (diluted in blocking solution; 1:5000 goat α-rabbit 488, Invitrogen). Finally, cells were washed 3 x 5 min PBS, mounted in VectaSheild-DAPI and sealed. Cells were imaged on a custom-built Zeiss Cell Observer Microscope (Intelligent Imaging Innovations (3i), Boulder, Co), equipped with diode-based lasers (405, 488, 561, and 633 nm) and a spinning-disk confocal scanning unit X1 (CSU-X1)(Yokagawa, Japan) using a 63× (1.4 NA) immersion oil objective lens. Images were processed using linear adjustments (*e.g.,* brightness/contrast) and deconvolved using empirically-measured point spread functions for each wavelength and a constrained iterative algorithm in Slidebook 5.5 (Intelligent Imaging Innovations (3i), Boulder, CO), followed by three dimensional rendering using Imaris 7.1 (Bitplane/Andor,Belfast, UK).

3.3.8 Molecular Biology And Cloning

Total RNA was isolated from *Pygsuia* cells using Trizol (Life Technologies). Messenger RNA was isolated from total RNA using the Ambion Poly(A) Purist MAG Kit and reverse transcribed using oligo(dT) primers and Superscript III RT (Life Technologies) following the protocols provided by the manufacturers. Polymerase chain reactions were performed using Gotaq Green master mix (Promega) and 0.2 μM of each primer using standard thermocycling conditions (94 °C 2 min; 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1-2.5 mins, final elongation temperature of 72 °C for 10 minutes). See Table 3-1 for primer sequences. Amplicons were purified and cloned in the sequencing vector pCR4 (Life Technologies) according to the manufacturer TA-cloning protocols. Plasmids were isolated from transformed *E. coli* using the NucleoSpin plasmid purification kits (Machery Nagel) and screened for correct sequence (Genewiz).

The plasmids containing the *Pygsuia* ORFs (pCR4-*Pb-asct*, pCR4-*Pb-msufcb(x/h)*, pCR4-*Pb-msufcb-MTS*, pCR4-*Pb-csufcb*, and pCR4-*Pb-nfu1)* and destination vectors pTS395 (Ma *et al.* 1987), pGEX-4T-1 and pET16b (Novagen) were digested with the appropriate restriction enzyme (Thermo) indicated in Table 3-1 according to the protocol of the manufacturer. Resulting fragments were cloned by standard protocols (Maniatis *et al.* 1982) to generate pTS395-*Pb-asct*, pTS395-*Pb-msufcb*, pTS395-*Pb-msufcb-MTS*, pTS395-*Pb-csufcb*, pTS395-*Pb-nfu1,* pGEX-*msufcb,* and pET16b-*asct*. Destination plasmid-encoded fusion protein tags are indicated in Table 3-1.

Table 3-1 Primers used for the amplification of *Pygsuia biforma* genes encoding SUFCB, NFU1 and ASCT.

Primer Description	Direction	Sequence	Restriction Enzyme Site	Destination Plasmid
P.biforma msufcb 1F_HindIII	Forward	GGGAAGCTTATGCTGAGGGCCTTAAC	HindIII	pTS395
P.biforma msufcb 2094R_HindIII	Reverse	GCGCAAGCTTGCATTCCCCCAGCGGC	HindIII	(C-term GFP)
P.biforma msufcb 1F_HindIII	Forward	GGGAAGCTTATGCTGAGGGCCTTAAC	HindIII	pTS395
P.biforma msufcb MTS 111R_HindIII	Reverse	CGCAAGCTTGGAGACGGCGTGTCTGTGAAA	HindIII	(C-term GFP)
P. biforma csufcb 1F_HindIII	Forward	CCCAAGCTTATGGTGTACAAAGACC	HindIII	pTS395
P. biforma csufcb 2041R_HindIII	Reverse	GCGCAAGCTTGGCAGAACCCAGTGGC	HindIII	(C-term GFP)
P. biforma nfu1 1F_HindIII	Forward	GGGAAGCTTATGTCTTTGATTACTAACAT	HindIII	pTS395
P. biforma nfu1 396R_HindIII	Reverse	GCGCAAGCTTGCTCTTCGTCTTCAATT	HindIII	(C-term GFP)
pTS395_GFPR	Reverse	CTTCTCCTTTACTCAT		
P. biforma asct 1F_NdeI	Forward	GGGCCCATATGATGGCACTCTGTGCTCGTAC	NdeI	pET16b
P. biforma asct 891R_NdeI	Reverse	GCGCCATATGTTACTCCCAAGAGTCAAGCT	NdeI	(His Tag)
P. biforma msufcb 1F_XhoI	Forward	GGGCTCGAGATGCTGAGGGCCTTAAC	XhoI	pGEX-4T-1
P. biforma msufcb 2094R_XhoI	Reverse	CGCCTCGAGCATTCCCCCAGCGGCTGC	XhoI	(GST Tag)

3.3.9 Yeast Transformation And Selection

Standard protocols were used for the growth and selection of yeast (Amberg *et al.* 2005). Yeast were cultured in YPAD (1% yeast extract, BioShop; 2% bacto-peptone, Difco; 0.003% adenine, Sigma; and 2% glucose, Sigma) or synthetic defined media lacking uracil (SD-ura: 0.67% yeast nitrogen base without amino acids, Difco; 2% glucose or 2% galactose; and all auxotrophic requirements except uracil) at 28°C. All studies were performed using *Saccharomyces cerevisiae* strain W303-1A (*MATa ade2-1 ura3-1 leu2-3, 112 his3-11,15 trp1-1*). Yeast were transformed with pTS395-based plasmids (encoding *URA3* gene) by standard methods (Gietz and Schiestl 2007) and transformants selected on SD-ura solid media. GFP fusion protein expression from the pTS395-derived plasmids under the control of a galactose inducible promoter was induced by shifting

transformed yeast into SD-ura media with galactose. Mitotracker orange (500 nM) and DAPI (2.5 μg/mL) were used to stain mitochondria and nuclei, respectively, for 30 min prior to microscopic investigation. Live cells were visualized using fluorescence microcopy (Olympus BX43 with X-cite series 120 Q light source).

3.3.10 Antibody Production

The protein sequence of *Pb*-mSUFCB was provided to Genscript for antigen design. *Pb*-mSUFCB-specific (CEEKQKKDTVFSTG) peptide was selected for immunization. Antibodies were affinity purified by the manufacturer and antigenicity was assessed using ELISA (Signal/blank ratio > 2.1 at 1:64, 000 mSUFCB specificity).

3.3.11 Protein Expression And Extraction

Plasmids (pGEX-*Pb-msufcb,* and pET16b-*Pb-asct*) were transformed in *E. coli* (BL21) for protein expression (Maniatis *et al.* 1982). Protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (Sigma) to the culture medium of exponentially growing cells and incubated for an additional 4-6 h at 37 °C. Proteins were isolated from *E. coli* cells induced to express the *Pygsuia* fusion protein just the GST or a hexa-histidine tag and control cells (where protein expression was not induced). Briefly, after protein expression, *E. coli* cells were collected by centrifugation (4000 x g, 2 min, 4 °C) and lysed by French press (7000 psi). Unbroken cells and debris were removed by centrifugation (4000 x g, 2 min, 4 °C). The resulting supernatant was saved for subsequent analysis and recombinant proteins were isolated using glutathione-magnetic beads (Thermo/Pierce; GST-tag) or Talon resin (Thermo/Pierce; His tag) according to the protocols of the manufacturer.

Pygsuia cells were isolated by centrifugation (400 x g, 5 min) and washed twice with natural seawater. Cells were homogenized in PBS supplemented with protease inhibitor cocktail (Roche) using a 21G syringe. Insoluble material was removed by centrifugation at 5000 x g for 5 min.

3.3.12 Immunoblotting

Crude cell lysates and purified proteins were denatured in SDS-PAGE sample loading buffer (Sigma), boiled for 5 min and resolved by SDS-PAGE (10 % and 8 % for ASCT and SUFCB respectively). Proteins were transferred to PVDF membranes (Turbo Blot membranes, Biorad) that were then incubated in blocking buffer (5% skim milk powder, TBS (Tris-buffered saline), 0.5% Tween 20) for 1 h. Primary antibody was diluted in blocking buffer and incubated with membranes for 1 h (1:2000 a-*Pb*-mSUFCB) or overnight (1:1000 a-*Tv*-ASCT1C). Following washes in TBS-tween, membranes were incubated with horseradish peroxidase-conjugated goat anti-rabbit secondary antibodies (1:50000, Sigma), washed and incubated with enhanced chemiluminesence substrate (GE Healthcare) and visualized via autoradiography or charge coupled device chemiluminescence detector (Protein Simple).

3.4 Results

3.4.1 Metabolic Pathway Prediction In Pygsuia biforma

From the filtered transcriptomic dataset, a total of 122 proteins were putatively predicted to be localized to the MRO matrix (MM), inner membrane (IM), inter membrane space (IMS) or outer membrane (OM) of Pygsuia on the basis of homology to known mitochondrial proteins inferred from Mitominer and/or Mitoprot and TargetP prediction scores (>0.5) for mitochondrial targeting signals. The vast majority (71/76) of MM proteins have N-terminal MTSs, whereas only half of the IM/IMS proteins have one (22/43). None of the OM proteins have an MTS (0/3), as expected. A metabolic reconstruction of the various MRO pathways in Pygsuia is shown in Fig. 1 and details for each ORF and protein sequence are summarized in Table 3-2 to Table 3-9.

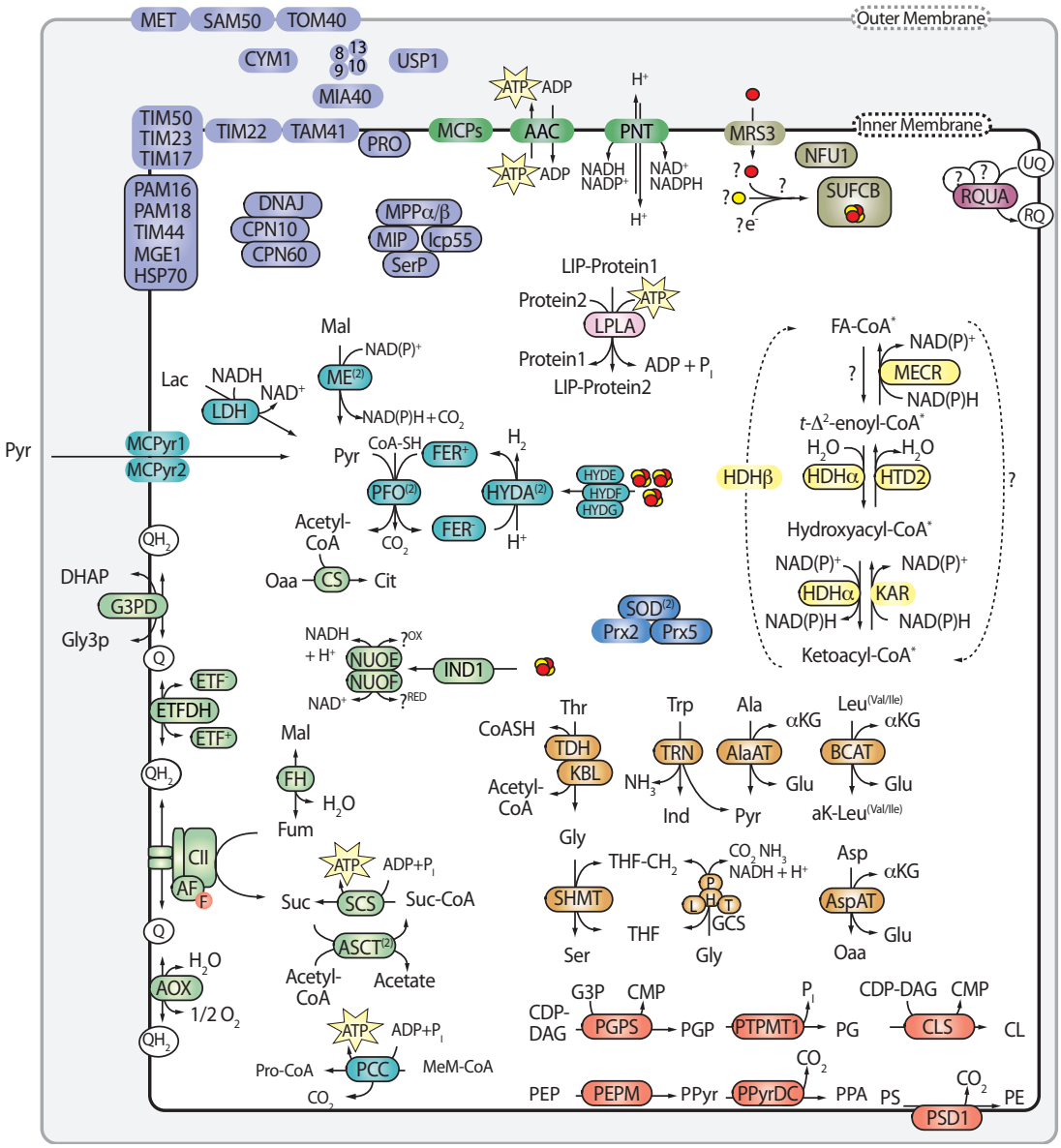

Figure 3-1: Predicted MRO metabolism based on selected *Pygsuia biforma* genes determined by BLAST-mediated homology probing. Sequences with predicted N-terminal MTS are outlined in black. Unknown proteins are shown with (?) respectively. Red and yellow circles represent Fe and S respectively. [Protein Import and Folding – (purple): Met, Metaxin; SAM50, Sorting assembly machinery 50 kDa; TOM, translocator of the OM; CYM1, presequence protease; 8, 9, 10, 13, Tiny Translocator of the IM (TIM); UPS1, IMS chaperone/slowmo protein; TAM41, translocator assembly and maintenance protein 41; PRO, Prohibitin; PAM, pre-sequence translocase-associated motor; MGE1, mitochondrial GRPE protein; CPN, chaperonin; HSP, heat shock protein; MPP, mitochondrial processing peptidase; MIP, mitochondrial intermediate peptidase; SerPr: serine protease], [Carriers (dark green): MCFs, mitochondrial carrier family proteins; AAC, ATP:ADP carrier protein, PNT, pyridine nucleotide transhydrogenase], [Fe-S cluster biosynthesis (brown): SUFCB, sulfur mobilization C/B fusion; NFU1: NIFU-like protein; MRS3: Fe^{2+} carrier], [Rhodoquinone (dark pink): UQ: ubiquinone; RQ: rhodoquinone; RQUA: RQ methyltransferase], [Lipoate metabolism (light pink) : LIP: lipoate; LPLA: lipoate-protein ligase],

[Pyruvate metabolism (blue): MPC1/2, mitochondrial pyruvate carrier (brain protein 44); Lac, lactate; LDH, lactate dehydrogenase (DH); Mal, malate; ME, malic enzyme(s); Pyr, pyruvate; FER, ferredoxin; PFO, Pyr:fer oxidoreductase (OR); HYDA, [Fe-Fe]-Hydrogenase; HYDE-G, HYDA maturases], [TCA cycle and electron transport (light green): CS, citrate synthase; OAA, oxaloacetate; NUOE/F, NADH:UQ OR; IND1, Fe-S protein required for NADH DH; FUM, fumarate; FH, fumarate hydratase; SUC, succinate; CII, succinate DH/complex II; AF, CII assembly factor; F, flavin; SCS, succinyl-CoA synthetase; ASCT; acetate:succinyl-CoA transferase(s); DHAP; dihydroxyacetone phosphate; Gly3p, glycerol-3-phosphate; G3PD, Gly3p DH; ETF: electron transferring flavoprotein; AOX: alternative oxidase; Q/QH_2, quinone/quinol], [Fatty acid metabolism (yellow): FA, fatty acid; MECR, Mitochondrial trans-2-enoyl-CoA reductase; HTD2, 3-hydroxyacyl thioester dehydratase 2; KAR, ketoacyl reductase; HDHa/b, Trifunctional enzyme, hydroxyacyl DH, *acyl-carrier protein is used in place of CoA during biosynthesis], [Amino acid metabolism (light brown): Thr, threonine; TDH, Thr DH; KBL, keto-butyrate lyase; Trp, tryptophan; TRN, tryptophanase; Ind, indole; Gly, glycine; Ala, alanine; ALAT, ala amino transferase (AT); Leu, leucine; Val, valine; Ile, isoleucine; a-KG, a-ketoglutarate; Glu, glutamate; SHMT, serine hydroxymethyl transferase; GCS, glycine cleavage system (P,H,L,T); SER, serine; THF, tetrahydrofolate; Asp, aspartate], [Oxidative stress (dark blue): SOD, superoxide dismutase; Prx2: peroredoxin; Prx5: peroxidase], [Lipid metabolism (orange): CDP-DAG, cytidine diphosphate diacylglycerol; PGPS, CDP-DAG-glycerol-3-phosphate 3-phosphatidyltransferase; PGP, phosphatidylglycerol phosphate; PG, phosphatidylglycerol; PTPMT1, protein-tyrosine phosphatase mitochondrial 1; CL, cardiolipin; CLS, CL synthase; CMP, cytidine monophosphate; PEP, phosphoenolpyruvate; PEPM, PEP mutase; PPyr, phosphonopyruvate; PPyrDC, PPyr decarboxylase; PSD1, phosphotidylserine decarboxylase 1]

Table 3-2 Proteins involved in amino acid metabolism identified in *Pygsuia biforma*. ORF completeness (C, complete; IC, incomplete) and coverage information is shown. Predicted cellular localization was calculated using Mitoprot and target (OM, outer mitochondrial membrane; IMS, inter mitochondrial membrane space; IM inner mitochondrial membrane; and MM, mitochondrial matrix).

Protein Name	Short Name	Complete	ORF Coverage	Predicted Localization	Mitoprot Prediction	TargetP prediction
Alanine Aminotransferase	AlaAT	C	285.74	MM	0.7871	0.749
Aspartate Aminotransferase	AspAT	C	166.84	MM	0.9165	0.558
Brached Chain Amino Acid Aminotransferase	BCAT	C	406.11	MM	0.5406	0.785
Lipoamide protein	GcsH	C	2,775.97	MM	0.9729	0.86
Dihydrolipoamide dehydrogenase	GcsL	C	236.52	MM	0.9925	0.823
Glycine decarboxylase 1	GcsP1	C	149.69	MM	0.9896	0.91
Glycine decarboxylase 2	GcsP2	C	353.92	MM	0.999	0.941
Aminomethyl transferase	GcsT	C	1,275.96	MM	0.977	0.927
2-amino-3-ketobutyrate CoA ligase	KBL	C	1,815.70	MM	0.9487	0.91
Serine Hydroxymethyl Transferase	SHMT	C	141.86	MM	0.9383	0.87
Threonine Dehydrogenase	TDH	C	491.41	MM	0.9447	0.917
Tryptophanase	TnaA	C	331.1	MM	0.686	0.605

Table 3-3 Mitochondrial carrier proteins identified in *Pygsuia biforma*. ORF and protein information as labeled in Table 3-2

Protein Name	Short Name	Complete	ORF Coverage	Predicted Localization	Mitoprot Prediction	TargetP prediction
MCF ADP-ATP translocase	AAC	C	59.68	IM	0.9575	0.936
LetM1-like	LetM1-like	C	138.94	IM	0.9892	0.827
Mitochondrial Carrier Family Protein (MCF) 1	MCF1	3' end IC	29.1	IM	0.7433	0.061
MCF2 triple repeat	MCF2	C	43.53	IM	0.1972	0.522
MCF3 folate carrier, partial	MCF3	C	18.39	IM	0.1508	0.554
MCF4	MCF4	C	16.18	IM	0.4892	0.162
MCF5	MCF5	3' end IC	50.05	IM	0.6541	0.373
MCF6 oxoglutarate-malate transporter	MCF6	C	69.82	IM	0.115	0.086
MCF7, partial	MCF7	5' end IC	173.57	IM	-	-
MCF8 Shm1-like	MCF8	C	3,302.82	IM	0.035	0.163
MCF9 Phosphate	MCF9	C	1,263.37	IM	0.0878	0.053
MCF10 oxoglutarate-malate transporter	MCF10	5' end IC	19.1	IM	-	-
MCF11 oxoglutarate-malate transporter	MCF11	C	69.82	IM	0.2992	0.531
MCF12 tricarboxylic acid transport protein	MCF12	C	278.34	IM	0.0771	0.073
MCF13 substrate carrier	MCF13	C	45.14	IM	0.0262	0.037
MCF14 substrate carrier	MCF14	C	151.39	IM	0.0716	0.034
MCF15 Thiamine pyrophosphate carrier	MCF15	C	93.19	IM	0.6172	0.26
Putative pyruvate carrier MCP1	MPC1	C	83.76	MM	0.7982	0.164
Putative pyruvate carrier MCP2	MPC2	C	126.25	IM	0.4562	0.034
Mitochondrial Intermembrane space Assemble protein	Mia40	C	295.08	IMS	0.3258	0.129
Mrs3_Iron transporter like partial	Mrs3	3' end IC	22.64	IM	0.0213	0.063
NAD(P) transhydrogenase	PNT	C	74.76	IM	0.9994	0.91

Table 3-4 Proteins involved in fatty acid, lipid and lipoate metabolism identified in *Pygsuia biforma*. ORF and protein information as labeled in Table 3-2

Protein Name	Short Name	Complete	ORF Coverage	Predicted Localization	Mitoprot Prediction	TargetP prediction
Trifunctional enzyme alpha subunit	HADA	C	40.43	MM	0.4216	0.435
Trifunctional enzyme beta subunit	HADB	C	3,022.23	MM	0.281	0.552
Hydroxylamine reductase	HAR	C	213.7	MM	0.9889	0.951
3-hydroxyacyl thioester dehydratase 2	HTD2	C	158.12	MM	0.8448	0.923
Ketoacyl reductase	Kar	C	45.16	MM	0.0676	0.018
Ketoacyl reductase (FabG-like)	Kar-FabG	C	128.59	MM	0.3941	0.069
Trans-2-enoyl-CoA reductase	MECR	C	176.87	MM	0.9576	0.793
Cardiolipin Synthase	CLS	C	32.38	MM	0.9912	0.872
Phosphoenolpyruvate mutase	PEPM	C	8,644.50	MM	0.9937	0.73
CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	PGPS	C	31.19	MM	0.6864	0.828
Phosphotidylserine decarboxylase 1	Psd1	C	68.02	IMS	0.8706	0.881
protein-tyrosine phosphatase mitochondrial	PTPMT1	C	83.82	MM	0.9123	0.163
Lipoate-protein ligase	LplA-Lip3	C	246.7	MM	0.8355	0.655

Table 3-5 Proteins involved in pyruvate metabolism identified in *Pygsuia biforma*. ORF and protein information as labeled in Table 3-2

Protein Name	Short Name	Complete	ORF Coverage	Predicted Localization	Mitoprot Prediction	TargetP prediction
Adrenodoxin-type Ferredoxin	Adrenodoxin	C	436.21	MM	0.9464	0.753
Ferredoxin	Fer	C	177.64	MM	0.9544	0.765
Ferredoxin 2	Fer2	C	100.69	MM	0.953	0.778
Hydrogenase Maturase E	HydE	C	123.17	MM	0.9895	0.825
Hydrogenase Maturase F	HydF	C	38.42	MM	0.9901	0.965
Hydrogenase Maturase G	HydG	C	904.63	MM	0.9801	0.891
Lactate dehydrogenase	LDH	C	28.88	MM	0.9755	0.866
Malic Enzyme NAD-dependent	ME1	C	527.69	MM	0.9944	0.916
Malic Enzyme NADP-dependent	ME2	C	250.28	MM	0.9803	0.782
Hydrogenase 1	mHydA	C	1,389.79	MM	0.9144	0.778
Pyruvate Ferredoxin Oxidoreductase 1	mPFO1	C	316.76	MM	0.7724	0.863
Pyruvate Ferredoxin Oxidoreductase 2	mPFO2	C	282.73	MM	0.9857	0.665
Phosphopyruvate decarboxylase	PPD	C	76.45	MM	0.9568	0.933
Pyruvate formate lyase	PFL	C	11.95	Cytosolic	0.2244	0.069
PFL activating enzyme	PFLA	C	544.2	Cytosolic	0.0394	0.052
Pyruvate:NADP oxidoreductase	PNO	C	2,870.92	Cytosolic	0.4316	0.054
Sulfide dehydrogenase, Hydrogenase	SD-HydA	C	113.37	Cytosolic	0.0168	0.046
[FeFe] Hydrogenase 2, CysJ	cHydA-CysJ1	C	119.71	Cytosolic	0.4316	0.054
[FeFe] Hydrogenase 3, CysJ	cHydA-CysJ2	3' end IC	2,177.89	Cytosolic	0.1114	0.064
Pyruvate Ferredoxin Oxidoreductase, cytosolic	cPFO	C	184.6	Cytosolic	0.4466	0.367

Table 3-6 Proteins involved in Fe-S cluster biosynthesis and oxidative stress identified in *Pygsuia biforma*. ORF and protein information as labeled in Table 3-2

Protein Name	Short Name	Complete	ORF Coverage	Predicted Localization	Mitoprot Prediction	TargetP prediction
Fe-S cluster binding protein, Ind1-like	Ind1	C	243.4	MM	0.9623	0.948
Sulfur Assimilation CB-D fusion	mSUFCB	C	140.88	MM	0.993	0.951
Sulfur Assimilation CB-D fusion, cytosolic	cySUFCB	C	509.78	Cytosolic	0.0218	0.07
NifU-like protein	Nfu1	C	1,596.36	MM	0.9993	0.952
Peroredoxin, Mitochondrial-like	Prx2	C	6,623.76	MM	0.2333	0.111
Glutathione amide-dependent peroxidase	Prx5	C	2,638.06	MM	0.9955	0.865
Superoxide dismutase 1 (Fe-Mn)	SOD1	C	1,460.03	MM	0.871	0.489
Superoxide dismutase 2 (Fe-Mn)	SOD2	C	420.1	MM	0.6243	0.784

Table 3-7 Proteins involved in uncategorized or incomplete metabolic processes identified in *Pygsuia biforma*. ORF and protein information as labeled in Table 3-2

Protein Name	Short Name	Complete	ORF Coverage	Predicted Localization	Mitoprot Prediction	TargetP prediction
Acyl-CoA synthetase-like protein	AcSyn	C	45.77	MM	0.994	0.835
Adenylate kinase	AK	C	27.47	MM	0.9937	0.927
Folylpolyglutamate synthase	FolC	C	8.54	MM	0.9358	0.814
Glycerol kinase	GK	C	40.89	MM	0.9971	0.715
Mitochondrial aminopeptidase	Icp55	C	905.83	MM	0.994	0.778
NAD(P)H dehydrogenase (quinone)	NPQDH	C	86.51	MM	0.9797	0.826
Short chain dehydrogenase	SDR	C	654	MM	0.9335	0.733
Propinoyl-CoA carboxylase alpha subunit	PCCa1	3' end IC	9.37	MM	0.7924	0.542
Propinoyl-CoA carboxylase beta subunit	PCCb	Complete	167.25	MM	0.4932	0.59
Propinoyl-CoA carboxylase alpha subunit	PCCa2	5' end IC	18.4	Unpredictable	-	-

Table 3-8 Proteins involved in organellar protein import and folding identified in *Pygsuia biforma*. ORF and protein information as labeled in Table 3-2.

Protein Name	Short Name	Complete	ORF Coverage	Predicted Localization	Mitoprot Prediction	TargetP prediction
Mitochondrial Chaperonin 10	Cpn10	C	651.22	MM	0.8906	0.455
Mitochondrial Chaperonin 60	Cpn60	C	142.62	MM	0.854	0.761
Presequence Protease	Cym1	C	17.06	IMS	0.7599	0.578
Mitochondrial GrpE Chaperon-Mge1	GrpE-Mge1	C	45.56	MM	0.9977	0.929
Metaxin	Met	C	662.31	OM	0.0609	0.061
Mitochondrial Intermediate Peptidase	MIP	C	27.8	MM	0.8053	0.853
Mitochondrial processing peptidase alpha subunit	MPPalpha	C	32.97	MM	0.9929	0.872
Mitochondrial processing peptidase beta subunit	MPPbeta	C	54.95	MM	0.9157	0.817
Slowmo protein - Mitochondrial protein sorting	Msf1/Ups2	C	320.77	IMS	0.3486	0.173
Mitochondrial DnaJ Chaperone	mtDnaJ	C	180.49	MM	0.9951	0.878
Mitochondrial Heat Shock Protein 70	mtHsp70	C	225.85	MM	0.9864	0.848
Mitochondrial Pam18-like protein	Pam18	C	259.97	IM	0.8841	
Mitochondrial Serine Protease - Lon protease, Pim1-like	Pim1	C	47.85	MM	0.968	0.808
Prohibitin	Pro	C	184.88	IM	0.708	0.105
Sortining and Assembly Machinery 50	Sam50	C	106.88	OM	0.118	0.04
Translocator assembly and maintenance protein 41	Tam41	C	56.82	IM	0.2363	0.084
Translocator of the inner mitochondrial membrane 10	Tim10	C	36.04	IMS	0.0948	0.174
Translocator of the inner mitochondrial membrane 13	Tim13	C	180.33	IMS	0.0553	0.072
Translocator of the inner mitochondrial membrane 16	Tim16/ Pam16	C	127.12	IM	0.7396	0.076
Translocator of the inner mitochondrial membrane 17	Tim17	C	52.97	IM	0.0405	0.047
Translocator of the inner mitochondrial membrane 22	Tim22	C	41.39	IM	0.2897	0.322
Translocator of the inner mitochondrial membrane 44	Tim44	C	59.15	IM	0.9991	0.954
Translocator of the inner mitochondrial membrane 50	Tim50	3' end IC	8.73	IM	0.9745	0.89
Translocator of the inner mitochondrial membrane 8	Tim8	C	248.55	IMS	0.0182	0.067
Translocator of the inner mitochondrial membrane 9	Tim9	C	219.19	IMS	0.0261	0.106
Translocator of the outer mitochondrial membrane 40	Tom40	C	131.2	OM	0.0984	0.112

Table 3-9 Proteins involved in electron transport, the Kreb's cycle and energy metabolism identified in *Pygsuia biforma*. ORF and protein information as labeled in Table 3-2

Protein Name	Short Name	Complete	ORF Coverage	Predicted Localization	Mitoprot Prediction	TargetP prediction
Alternative oxidase	AOX	C	181.43	IM	0.9774	np
Acetate:Succinate CoA transferase Type 1B	ASCT 1B	C	706.49	MM	0.7717	0.502
Acetate:Succinate CoA transferase Type 1C	ASCT 1C	C	643.23	MM	0.9837	0.709
Succinate Dehydrogenase Assembly Factor 2,	CIIaf	C	212.73	MM	0.846	0.821
Citrate Synthase-like	CS	C	17.42	MM	0.7565	0.403
Electron transfer flavoprotein alpha subunit	ETFa	C	48.65	IM	0.9486	0.194
Electron transfer flavoprotein beta subunit	ETFb	C	34.13	IM	0.8149	0.336
ETF:ubiquinone oxidoreductase	ETFDH	C	1,478	IM	0.6233	0.647
Fumarase	Fum	C	81.44	MM	0.96	0.571
Glycerol-3-phosphate dehydrogenase, membrane bound	G3PDH	C	226.4	IM	0.8846	0.164
NADH dehydrogenase 24 kDa Subunit, NuoE	NuoE	C	740.81	MM	0.9925	0.908
NADH dehydrogenase 51 kDa Subunit, NuoF	NuoF	C	446.61	MM	0.9489	0.899
Succinyl-CoA synthetase, alpha subunit	SCSa	C	735.35	MM	0.8893	0.825
Succinyl-CoA synthetase, beta subunit	SCSb	C	836.31	MM	0.5658	0.743
Succinate Dehydrogenase Flavoprotein Subunit	SdhA	C	20.83	MM	0.9078	0.275
SDH Fe-S Subunit	SdhB	C	335.88	MM	0.9877	0.947
SDH Large Cytochrome b Subunit	SdhC	C	97.24	IM	0.8974	0.864
SDH Small Cytochrome b Subunit	SdhD	C	1,693.76	IM	0.9861	0.918
Rhodoquinone biosynthesis protein RquA	RquA	C	849.93	MM	0.9919	0.958

The following sections discuss the major pathways/processes identified in *Pygsuia biforma* including those involved in pyruvate and ATP generation, protein import and processing, Fe-S cluster biogenesis, amino acid and lipid metabolism as well as small molecule transport. For key proteins, the subcellular localization using heterologous expression in yeast and immunofluorescence microscopy was undertaken.

3.4.2 Pyruvate And Energy Metabolism

Glycolysis-derived pyruvate is typically imported into mitochondria via the recently identified pyruvate carrier MPC1/MPC2 (brain protein 44) (Bricker et al. 2012; Herzig et al. 2012). However, pyruvate can also be generated from malate via malic enzyme (ME) (Ochoa et al. 1948; van der Giezen et al. 1997). Two putatively organellar NAD- and NADP-dependent MEs were identified in *Pygsuia*, ME1 and ME2, respectively. In mitochondria and some HPM, this pyruvate is typically oxidized via the pyruvate dehydrogenase complex (PDC) (de Graaf *et al.* 2011). However, in other MROs, acetyl-CoA is generated by a pyruvate:ferredoxin oxidoreductase (PFO), a single subunit

enzyme proposed to have been acquired by lateral gene transfer (LGT) (Hug *et al.* 2010). In *P. biforma*, I detected four transcripts encoding putative PFO/PNOs: two PFOs with predicted MTS (*Pb*-mPFO1, *Pb*-mPFO2), one without MTS (*Pb*-cPFO) and a PNO without MTS (*Pb*-cPNO). Maximum-likelihood (ML) and Bayesian phylogenetic analyses (BI) indicate all breviate sequences emerge within a monophyletic eukaryotic grouping (BV = 60). Although there is little resolution in the placement of the breviate sequences within the eukaryotic clade, at least two PFO copies appear to have been established in breviates before the divergences of *P. biforma* and *B. anathema* (*i.e., Ba*-PFOa and *Ba*-PFOb group with *Pb*-mPFO1 and *Ba*-PFOc groups with *Pb*-cPFO1, Figure 3-2). Curiously, *Pygsuia* also encodes homologs of the eukaryotic pyruvate:formate lyase (PFL) and activating enzyme (PFLA), another anaerobic enzyme catalyzing the conversion of pyruvate to acetyl-CoA sometimes found in MROs (Akhmanova *et al.* 1999; Atteia *et al.* 2006; Stairs *et al.* 2011). The lack of a targeting peptide on PFL or PFLA is suggestive of a cytosolic localization in *Pygsuia*.

In hydrogenosomes, the reduced ferredoxin generated by PFO is typically reoxidized by an [FeFe]-hydrogenase (HYDA) generating molecular hydrogen (Bui and Johnson 1996). I identified genes encoding full-length and partial canonical HYDA (*Pb*-mHYDA, Pb-HYDA4) in *P. biforma* and one (incomplete) copy in *B. anathema* (*Ba*-HYDA). In addition, I found three other putative HYDA-like proteins in *P. biforma* possessing several distinct domain architectures, including two with C-terminal flavodoxin (CYSJ) domains (*Pb*-cHYDA-CYSJ1 and 2) and one with an N-terminal sulfide dehydrogenase (*Pb*-cSD-HYDA). Of these five genes, only *Pb*-mHYDA has a predicted MTS, suggesting the other HYDA-like proteins are non-organellar. Phylogenetic analyses, although poorly supported in general, indicate the canonical and non-canonical HYDAs of the breviates branch amongst other eukaryotic sequences (Figure 3-3).

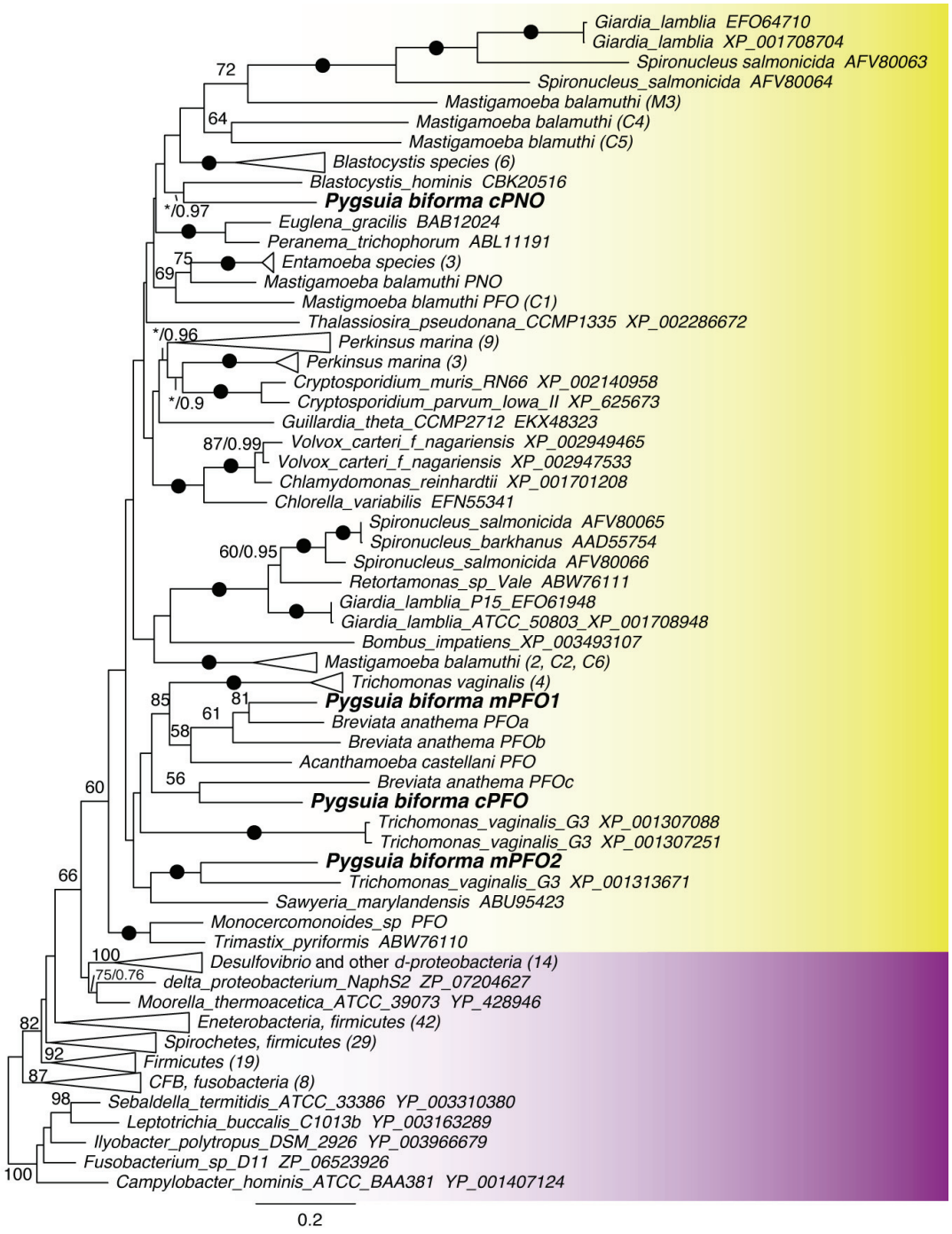

Figure 3-2: Phylogeny of pyruvate ferredoxin oxidoreductase (PFO). Maximum-likelihood (ML) tree of eukaryotic (yellow) and bacteria (purple) PFO homologs (188 sequences, 819 sites). Bootstrap support (BP) and positerior probability (PP) values on each branch were calculated using RAxML and PhyloBayes. Only values greater than 50% (BP) or 0.5 (PP) are shown, "*" denote ML values less than 50 on branches where PP values were greater 0.5. Black circles represent bipartitions with maximum support (BV = 100; PP = 1.0).

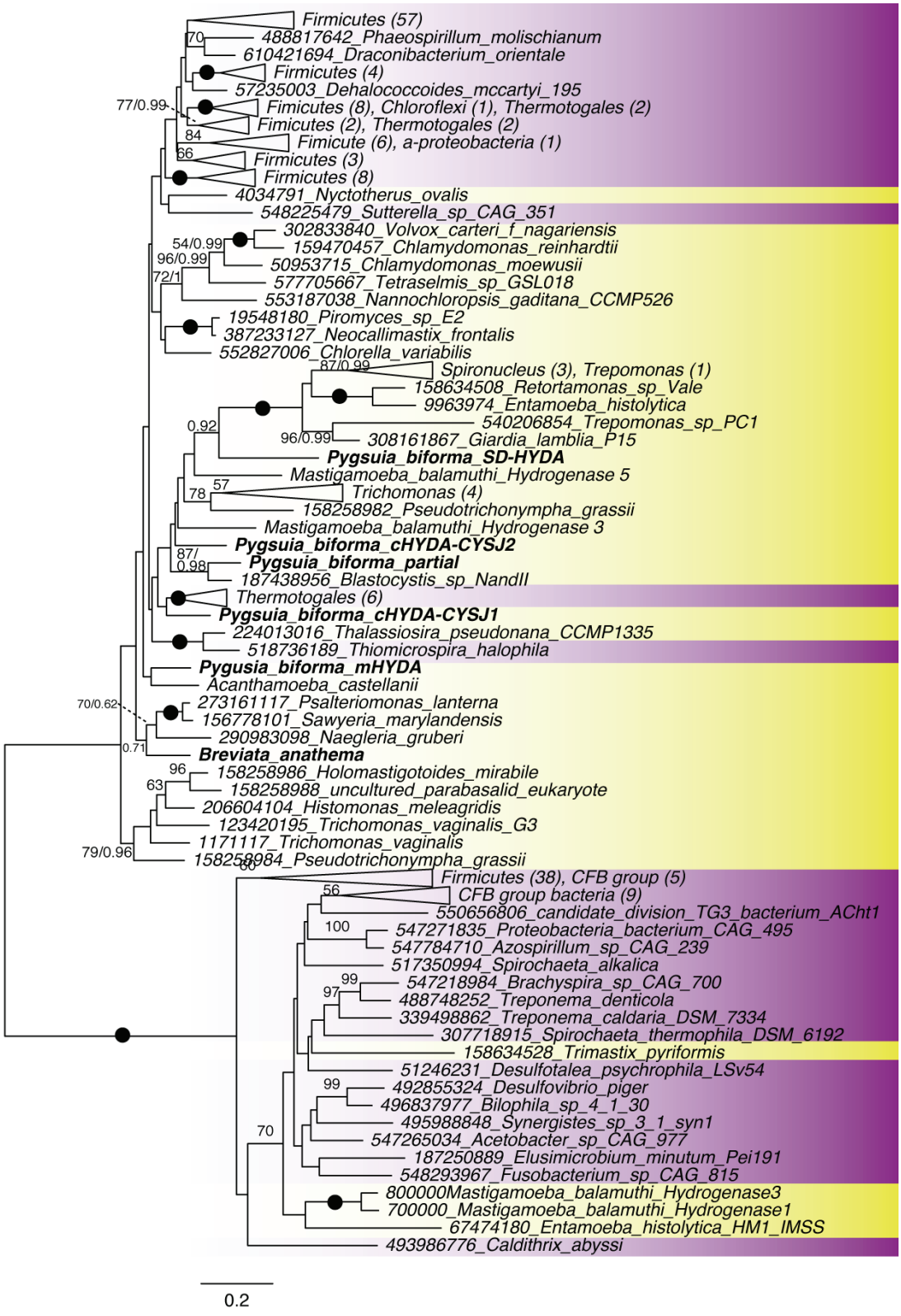

Figure 3-3: Maximum-likelihood (ML) tree of HYDA (204 sequences, 282 sites). Support values and taxa labeling as in Figure 3-2

All three HYDA maturases (HYDE, F, G) responsible for proper assembly of the H-cluster of HYDA (Mulder et al. 2010), were identified in single copy with predicted MTS. For HYDF and HYDG, the monophyly of eukaryotic homologs was recovered in ML and Bayesian phylogenetic analyses with maximum support (HYDF, Figure 3-4; HYDG, BV = 99, PP=1, Figure 3-5), however in the HYDE analysis, *Trichomonas* and *Spironucleus* sequences did not branch with *Pygsuia* and other of eukaryotes (Figure 3-6).

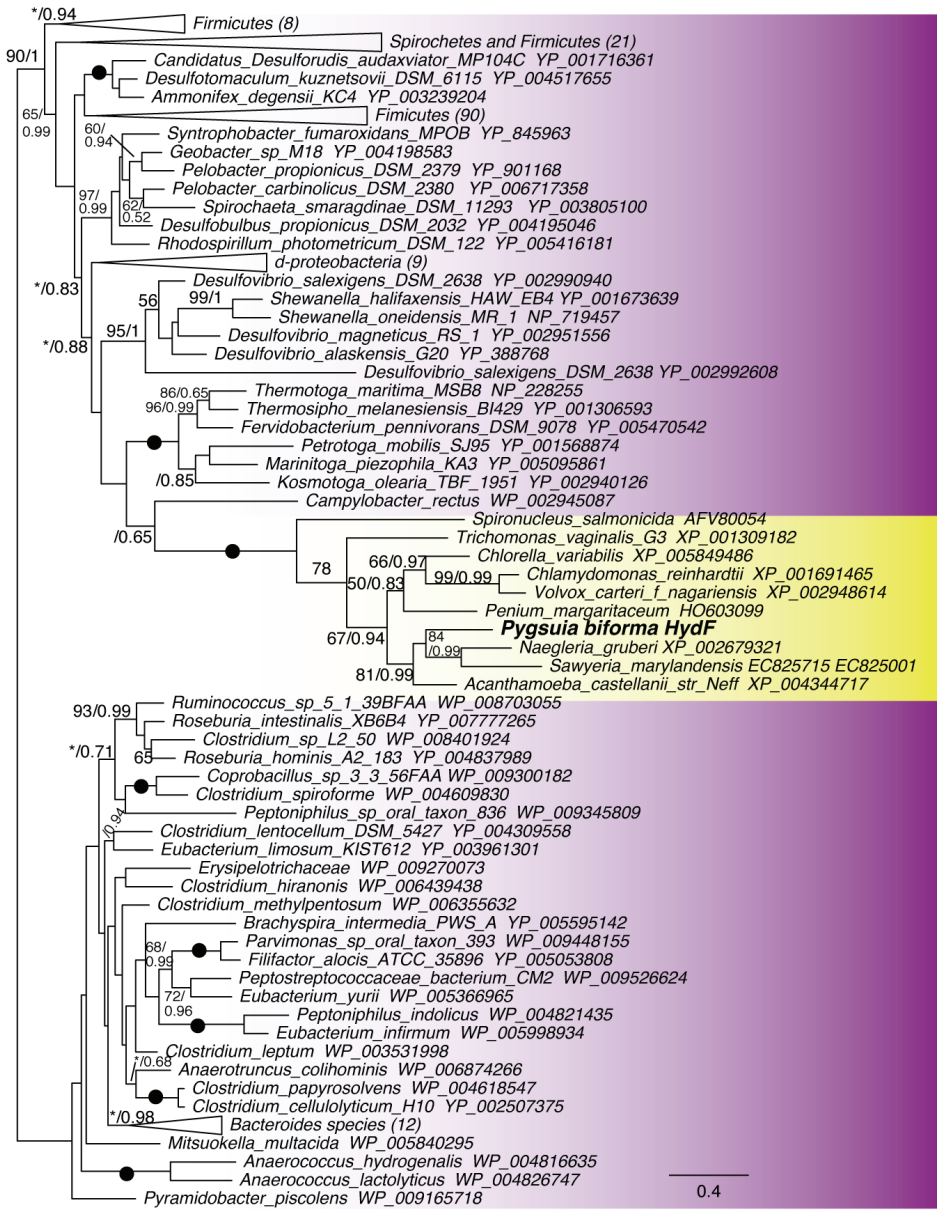

Figure 3-4: Phylogeny of hydrogenase maturase protein F (196 sequences, 232 sites). Labeling as described in Figure 3-2.

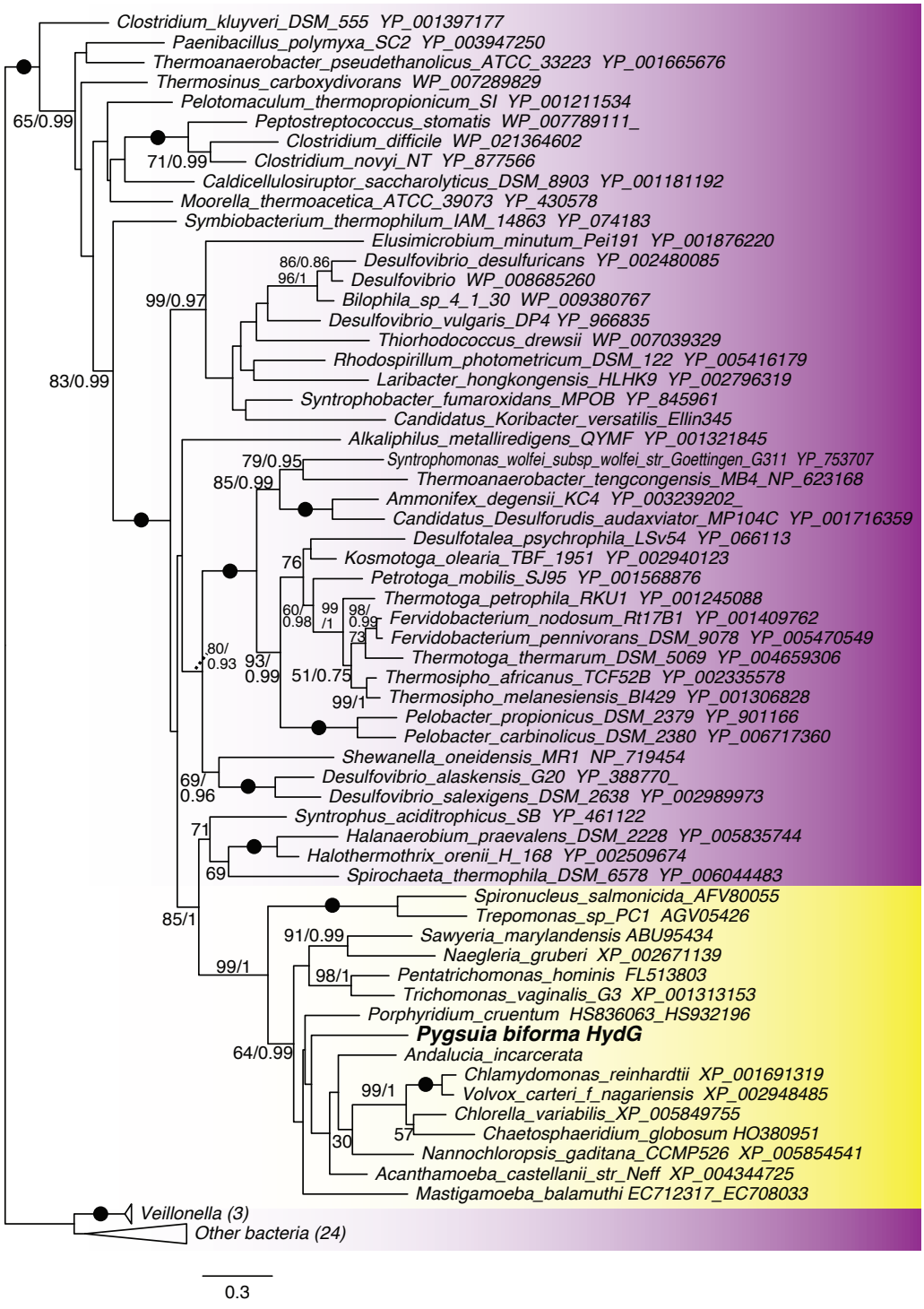

Figure 3-5: Phylogeny of hydrogenase maturase protein G (87 sequences, 327 sites). Labeling as described in Figure 3-2.

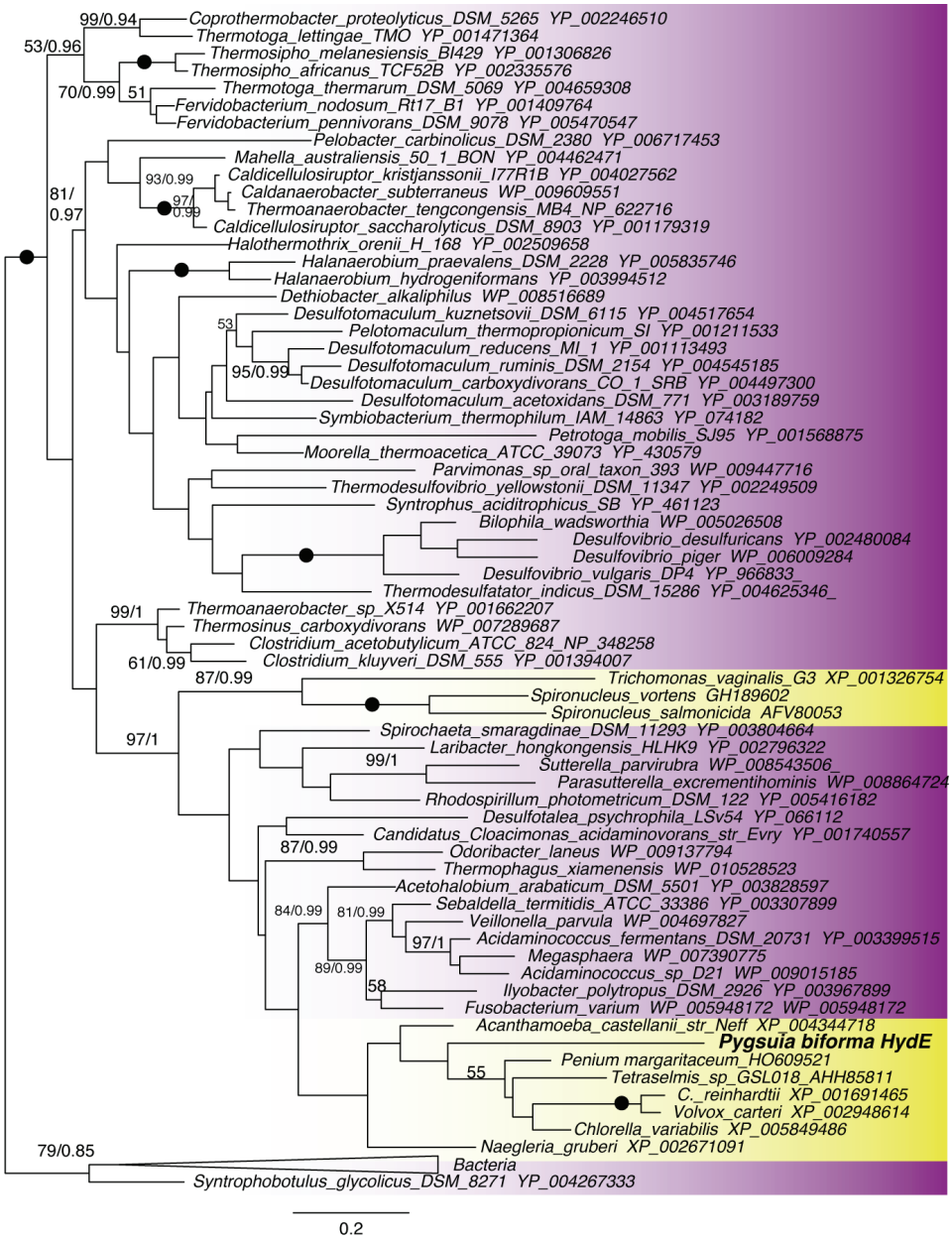

Figure 3-6: Phylogeny of hydrogenase maturase protein E (90 sequences, 187 sites). Labeling as described in Figure 3-2.

A modified tricarboxylic acid (TCA) cycle was identified in P. biforma including citrate synthase, succinyl-CoA synthetase (SCSa/b), succinate dehydrogenase/Complex II (CII; SDHA-D, and SDH assembly factor 2), fumarate hydratase (FH) and propionyl-CoA carboxylase (PCCa/b) and summarized in Table 3-9. However, aconitase, isocitrate dehydrogenase, α-ketoglutarate dehydrogenase and an organellar malate

dehydrogenase were not identified. The absence of these enzymes suggests that malate might ultimately be converted to succinate. In this scenario, CII would be functioning in reverse as a fumarate reductase (FRD) with the FRD-derived succinate is used as a CoA acceptor (from acetyl- or propionyl-CoA) by acetate:succinyl-CoA transferase (ASCT), an enzyme often found in anaerobic mitochondria, HPMs and hydrogenosomes. *Pygsuia* encodes two putative ASCTs, one corresponding to the sub-type 1B and 1C families, both with predicted MTS. The succinyl-CoA presumably generated by these enzymes could be used by the TCA cycle enzyme SCS to generate A/GTP by substrate-level phosphorylation as is the case in *Trichomonas* (van Grinsven *et al.* 2008).

Unlike the above-mentioned TCA cycle enzymes that are of mitochondrial provenance, the phylogenetic affinities of ASCT are less clear (van Grinsven *et al.* 2008). For this reason, only phylogenetic analyses of the *P. biforma* ASCT-1B and -1C homologs were conducted. The ASCT-1C ML tree shows a poorly supported eukaryotic clade within which *Pb*-ASCT-1C branches weakly as a sister-group to two trichomonad homologs (Figure 3-7); one of these – the enzyme from *T. vaginalis* – has been experimentally characterized (van Grinsven *et al.* 2008). In the ASCT-1B phylogeny, *P. biforma* emerges from within a grouping of eukaryotic sequences, but the precise branching order of the tree is not well supported (Figure 3-8).

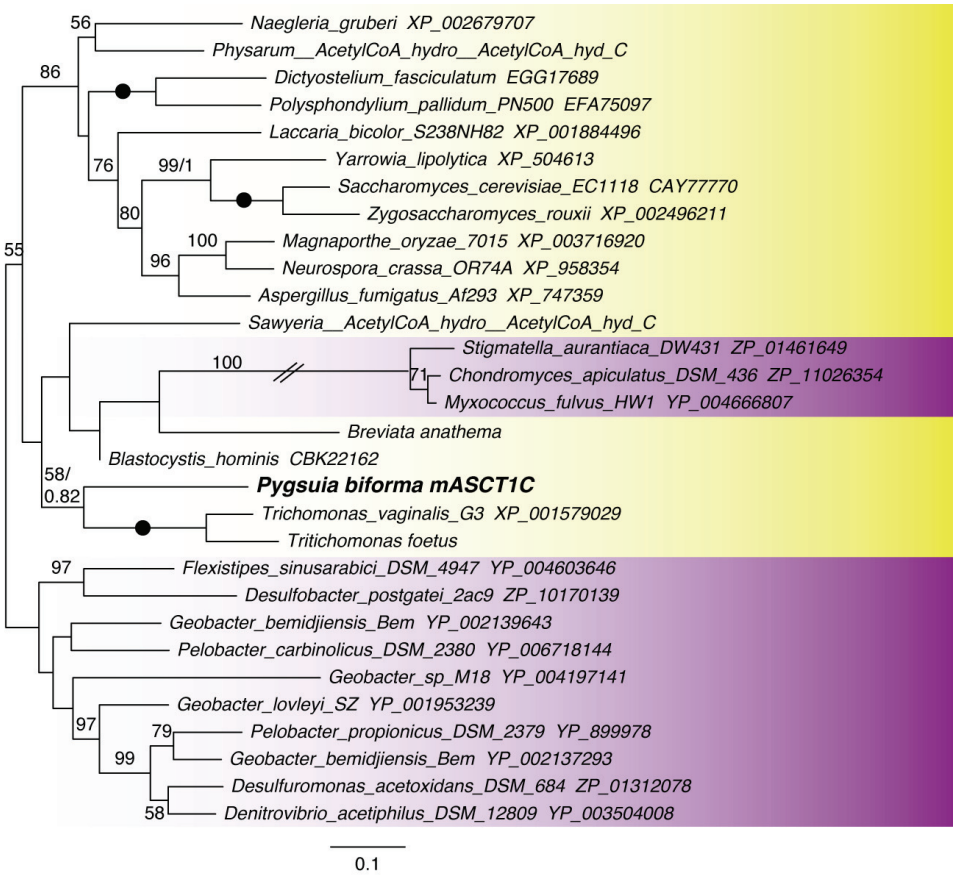

Figure 3-7: Maximum-likelihood (ML) tree of ASCT1C (25 sequences, 478 sites). Support values and taxa labeling as in Figure 3-2. Hash marks indicate where branch was shortened by 50 % for display purposes.

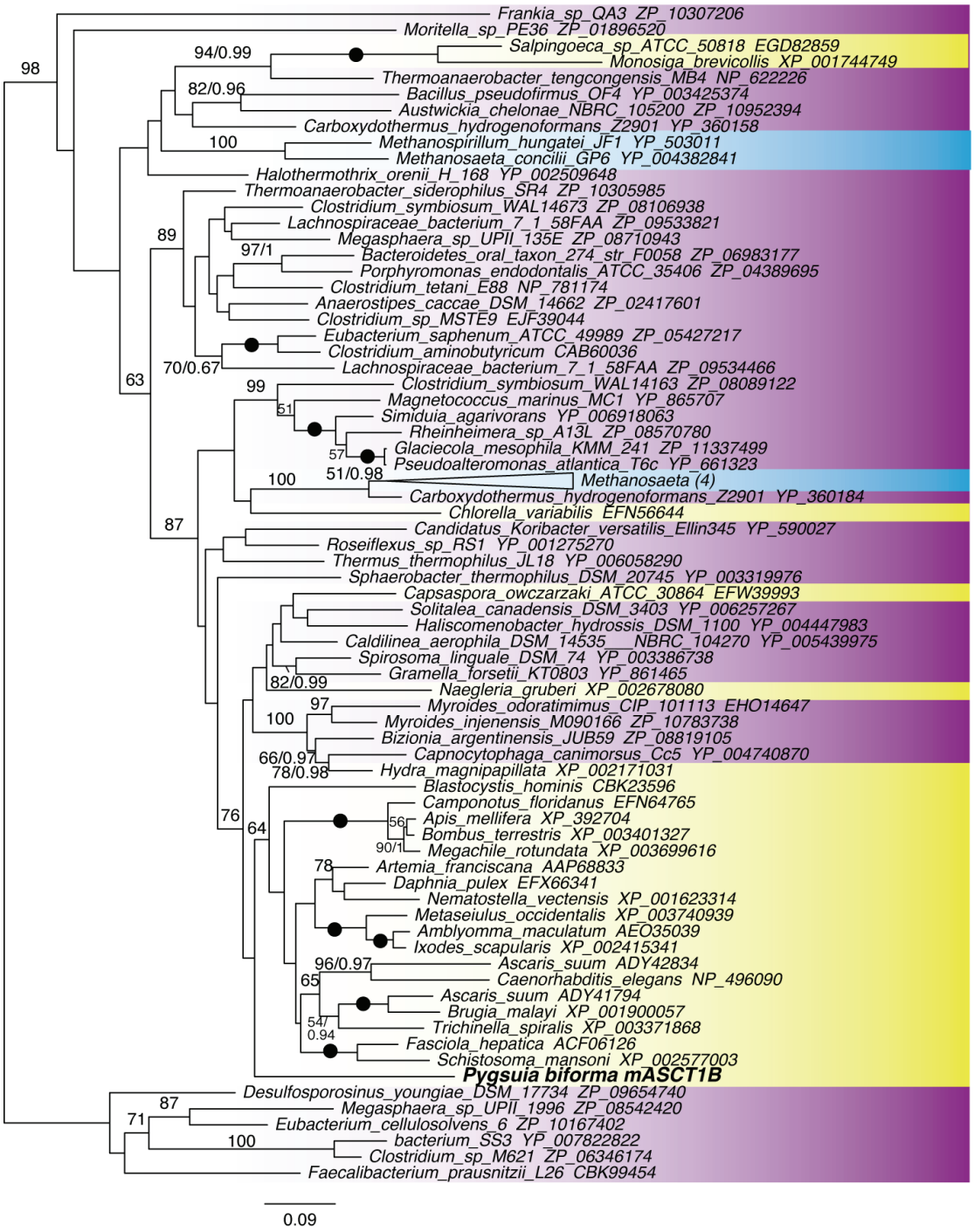

Figure 3-8: Maximum-likelihood (ML) tree of ASCT1B (73 sequences, 319 sites). Support values and taxa labeling as in Figure 3-2.

The presence of genes encoding putatively MRO ubiquinone-utilizing (UQ) enzymes such as alternative oxidase (AOX), SDH/FRD, electron transferring flavoprotein dehydrogenase (ETFDH), NAD(P)H dehydrogenase (NQO1) and glycerol-3-phosphate

dehydrogenase (G3PDH) prompted a search for a quinone biosynthesis pathway. Only geranyl-geranyl transferase (ISPA) and coenzyme Q methylase-like protein (COQ5) were identified, each without a MTS. While AOX and G3PD are known to use UQ as their electron acceptor (Ansell *et al.* 1997; Vanlerberghe and McIntosh 1997), CII has been shown to use rhodoquinone (RQ) when functioning as a FRD (Iwata *et al.* 2008). The exact pathway of RQ biosynthesis is unknown, however recent reports have demonstrated that in *Rhodospirillum rubrum,* RQ is synthesized from UQ *via* a number of reactions, one of which involves a putative methyltransferase (RQUA) (Brajcich *et al.* 2010; Lonjers *et al.* 2012). Unexpectedly, I identified a homolog of RQUA in *Pygsuia* that possesses a MTS suggesting that RQ is synthesized in its MRO. A survey of the *nr* database revealed that a number of other eukaryotic lineages have RQUA homologs, including obligate (*Blastocystis*) and facultative (*Euglena*) anaerobes, the latter of which is known to synthesize RQ (Hoffmeister et al. 2004; Castro-Guerrero et al. 2005). Phylogenetic analysis of RQUA indicates a patchy and limited distribution in a small number of α- and β-proteobacteria and eukaryotes. The extremely limited distribution of this enzyme within eukaryotes coupled with the atypical phylogenetic groupings observed (Figure 3-9) strongly suggests the enzymes have been acquired multiple times by eukaryotes via LGT from distinct bacterial or eukaryotic donors.

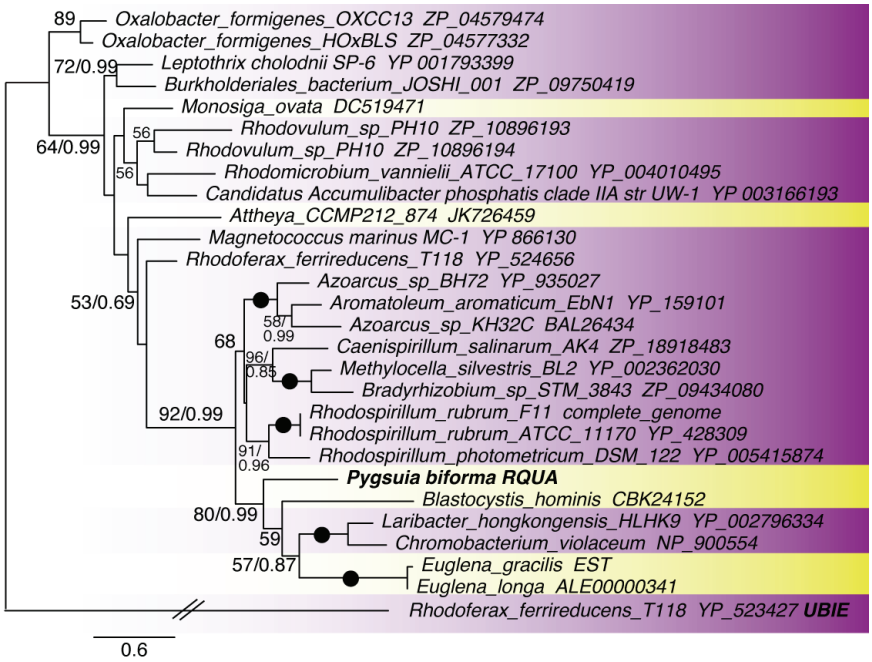

Figure 3-9: Maximum-likelihood (ML) tree of RQUA (28 sequences, 193 sites) rooted with UbiE/COQ5 methyltransferase from *Rhodoferax ferrireducens.* Support values and taxa labeling as in Figure 3-2. Hash marks indicate where branch was shortened by 50 % for display purposes.

The only other respiratory complex identified was the two soluble subunits of NADH:ubquinone oxidoreductase (NUOE and NUOF) from complex I, along with the putative assembly factor (IND1/MRP1-like). These two subunits of NUO are often found in hydrogenosomes of protists and are presumed to function in Q-independent electron transfer reactions (Hrdy et al. 2004).

3.4.3 Pygsuia *MROs Contain Canonical Protein Import/Processing Machinery*

Mitochondrial-derived genes residing on the nuclear genome are recognized and imported to the MM or IMS via the protein import machinery. *Pygsuia* encodes two components of the translocator of the outer mitochondrial matrix complex (TOM40 and SAM50), all four tiny translocators of the inner mitochondrial membrane (TIM8, 9, 10, 13), IMS import and assembly protein 40 (MIA40), IMS sorting protein (UPS2), TIM22, the majority of the TIM23 and presequence translocase-associated motor complex (PAM) complex (TIM50, TIM23, TIM17, HSP70, MGE1, TIM44, PAM16 and PAM18), an

assembly/maintenance factor for the translocator machinery (TAM41), chaperonins (CPN10, CPN60) and membrane integrity protein prohibitin (PHB/PRO). MTSs were identified for all matrix-associated import proteins and internal MIA40 targeting sequences [[MILENKOVIK]] (Milenkovic et al. 2009)were identified for IMS proteins TIM9 and TIM10. A variety of proteases and processing peptidases were also identified including presequence peptidase (CYM1), serine protease (SERPr) and mitochondrial processing peptidase (MPPα and MPPβ) responsible for cleaving the N-terminal pre-sequence of mitochondrial proteins following import into the organelle. I used sequence logo analysis to graphically represent the level of conservation of the predicted MTS and observed the typical features, including arginine or lysine in the penultimate position (Claros and Vincens 1996; Emanuelsson *et al.* 2000) (Figure 3-10). There was also a preference for leucine or phenylalanine immediately following the methionine (Figure 3-10).

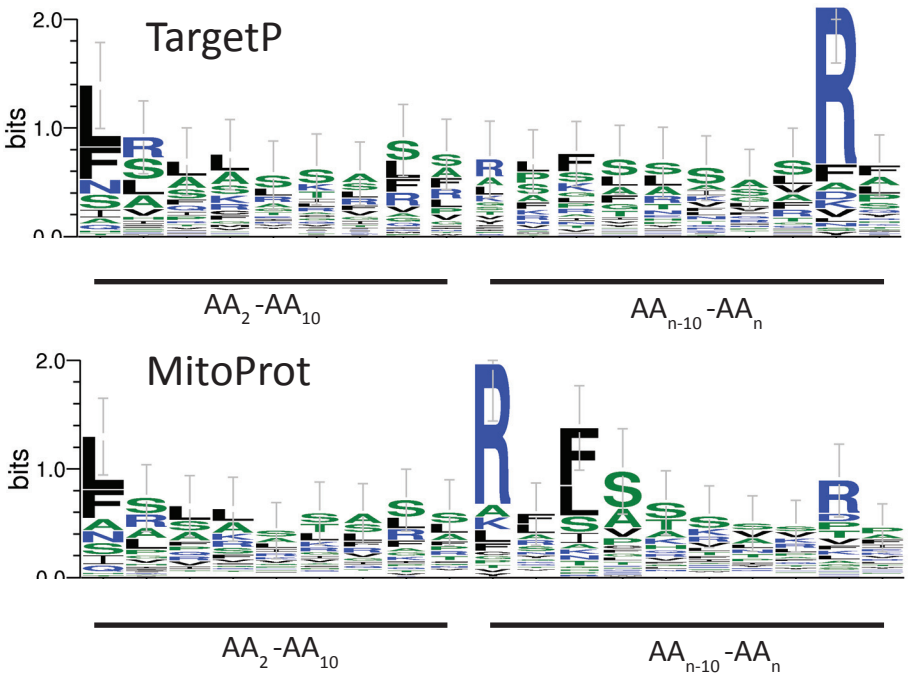

Figure 3-10: Sequence logo analysis of MTS from *Pygsuia biforma* matrix proteins. Composition of the first 9 residues after the methionine and last 10 residues of each MTS from predicted MRO matrix proteins revealed preference for leucine or phenylalanine in the second position. The starting methionine residue was removed for simplicity. Image generated using webserver available at http://weblogo.threeplusone.com/

3.4.4 Fe-S Cluster Biogenesis

All eukaryotes studied to date, with the exception of *E. histolytica* (van der Giezen *et al.* 2004) and *M. balamuthi* (Gill *et al.* 2007; Nyvltová *et al.* 2013) utilize the mitochondrial iron sulfur cluster (ISC) system for mitochondrial/MRO Fe-S cluster biogenesis. Considering this, the most intriguing result from the *P. biforma* RNAseq data was the apparent absence of the vast majority components of the ISC machinery (*i.e.,* genes encoding ISCA, ISCU, Frataxin, ISPG/H, ISCR, YAH1 or ARH1 proteins) and their associated proteins involved in factor X transport and iron homeostasis (*i.e.,* genes encoding ERV1, ATM1/ABC7 or ABCB6/MtABC3 proteins) (Lill 2009). As factor X is thought to be indispensable for the function of CIA-mediated cytoplasmic Fe-S cluster assembly, the lack of the CIA-components predicted to interact with factor X (*i.e.,* the TAH18/DRE2 complex) correlates with the absence of its transporters. However, *Pygusia* does encode the remaining components of the CIA system (genes encoding CIA1, NPB35, CFD1, NAR1, CIA2 and MET18 proteins).

Despite the absence of all other ISC components *Pygsuia* encodes a protein containing an ISCS-like domain fused C-terminal to a 4-thiouridine biosynthesis protein (THII; PbISCS-THII). A partial sequence encoding ISCS was identified in *B. anathema* without the THII domain. Interestingly, *Leishmania* species also encode the *Pygusia*-type THII-ISCS fusion protein. Phylogenetic analyses showed that nearly all eukaryotic mitochondrion-targeted ISCS sequences, including those of *Breviata anathema*, form a clade with α-proteobacteria (BV = 70; PP = 1.0) (Figure 3-11). In phylogenetic analysis the *P. biforma* and some *Leishmania* sequences are clearly separated from the rest of eukaryotes by at least one highly supported bipartition (BV = 95; PP = 1.0, Figure 3-11). This observation combined with the absence of a MTS suggests that the *P. biforma* and *Leishmania* sequences are unlikely to be of mitochondrial origin and may instead have been acquired though LGT from a prokaryotic source, although the donor lineage is currently unclear. Furthermore, in contrast to the *Breviata* sequence, the *Pygsuia* sequence lacks the mitochondrial and bacterial ISCS-specific residues (Figure 3-12). The lack of a predicted MTS on this protein, and its distinct evolutionary origin from

mitochondrial ISCS homologs of other eukaryotes, suggests PbISCS-THII protein is unlikely to be involved in Fe-S cluster biogenesis in the MROs of *Pygsuia.*

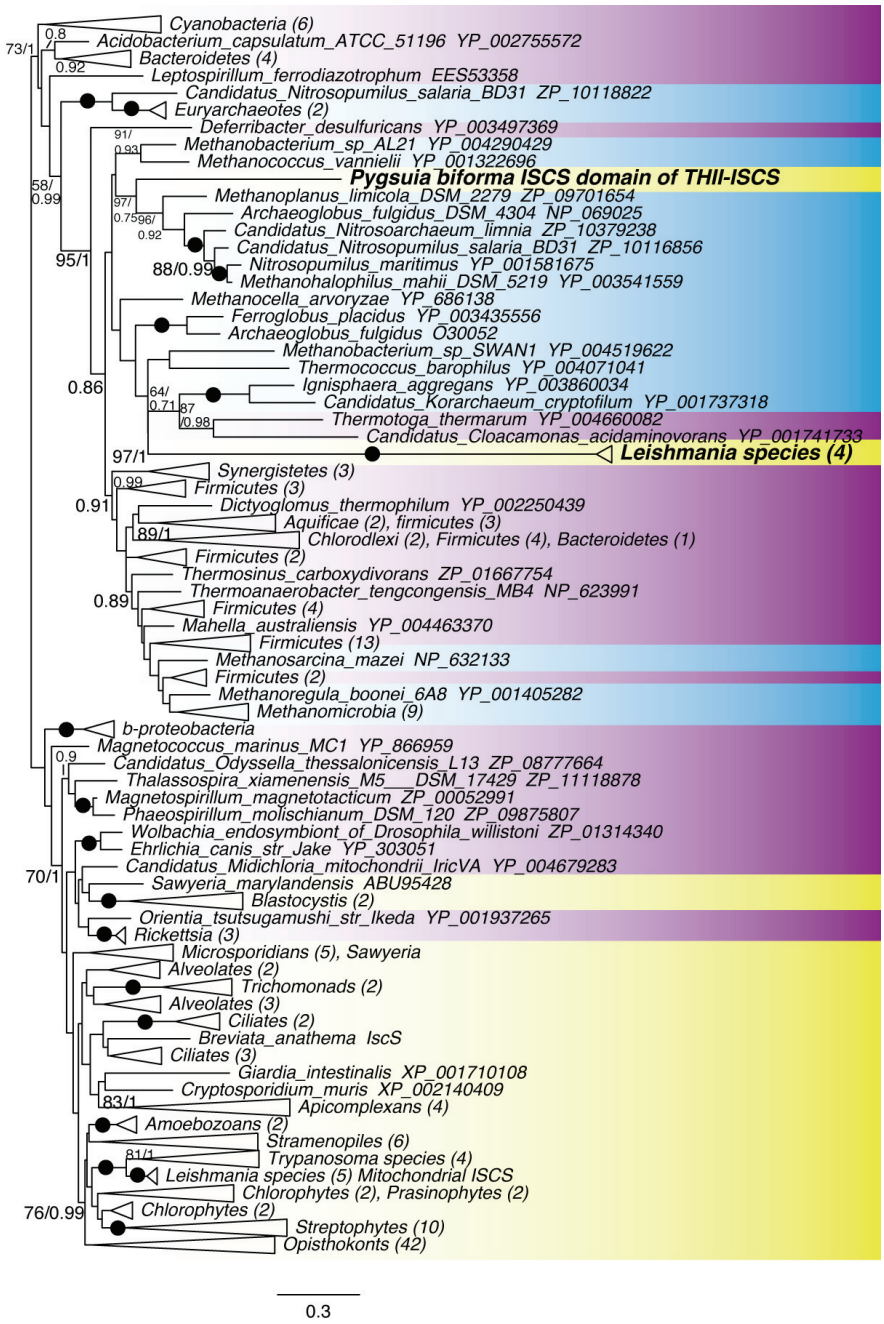

Figure 3-11: Maximum-likelihood (ML) tree of ISCS (210 sequences, 327 sites). Support values and taxa labeling as in Figure 3-2.

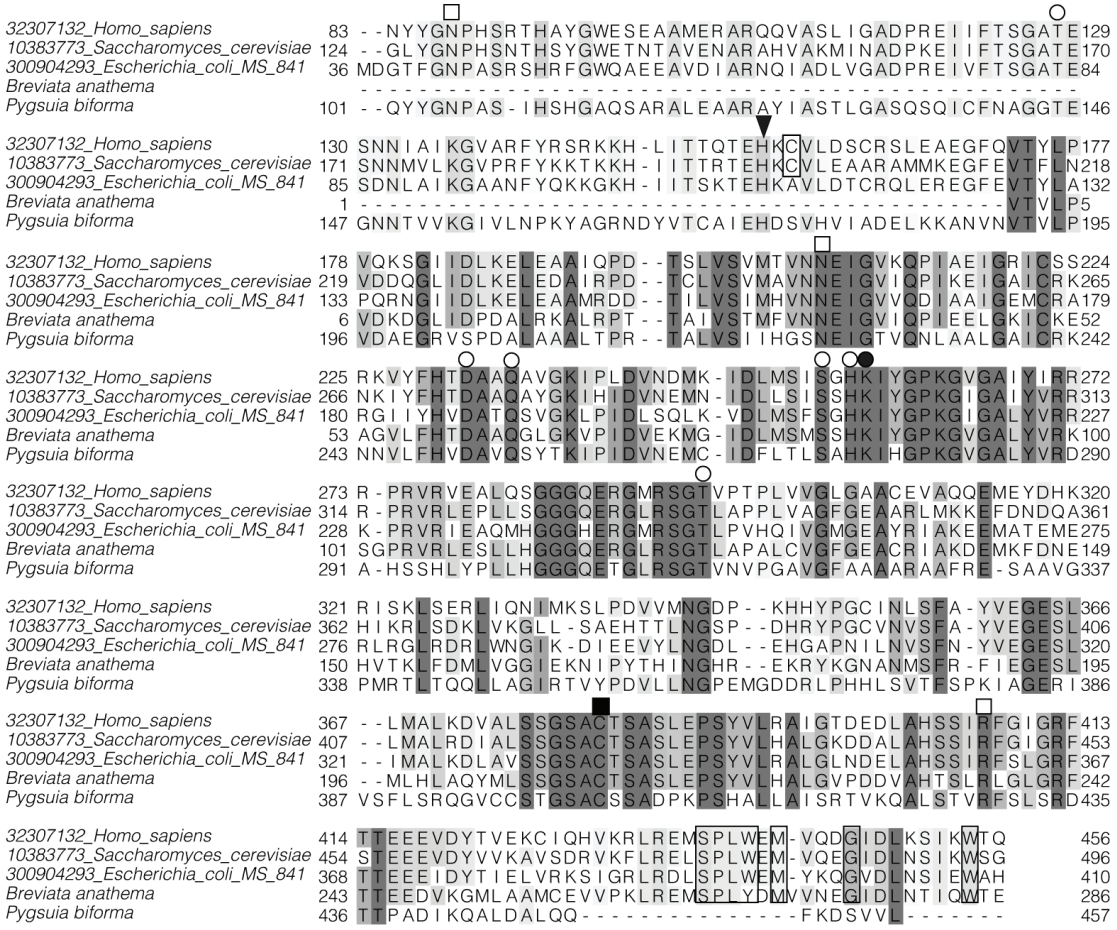

Figure 3-12: Alignment of mitochondrial, bacterial and Breviate IscS sequences. Closed squares indicate cysteine residues essential for disulfide bridge formation with scaffold proteins, open squares indicate residues essential for substrate (cysteine) binding, closed and open circle represents lysine residue and other residues essential for pyridoxal-5'-phosphate binding respectively, arrow indicates conserved histidine residue necessary for substrate deprotonation. Conserved eukaryotic/prokaryotic cysteine and C-terminal region are boxed. Residues are shaded according to relative conservation.

Apart from the ISCS-like sequence, the only other putative dedicated ISC components identifiable in *P. biforma* are NFU1 and IND1. Both proteins possess canonical MTS and in phylogenetic analysis NFU1 groups with other eukaryotic homologs (Figure 3-13). In contrast, in *Breviata* I identified the Fe-S scaffold of the ISC system, (ISCU) in addition to IND1.

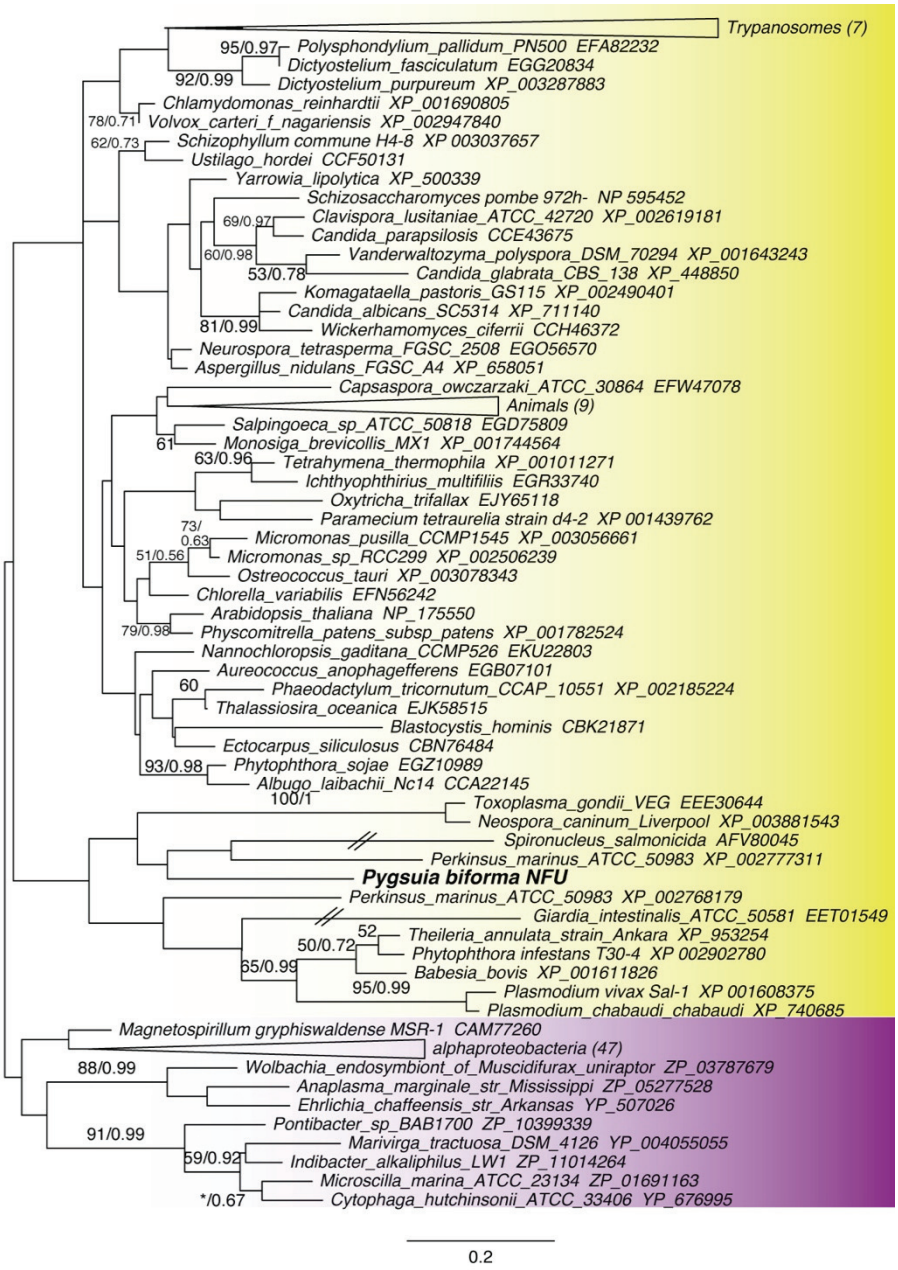

Figure 3-13: Maximum-likelihood (ML) tree of NFU (123 sequences, 75 sites). Support values and taxa labeling as in Figure 3-2. Hash marks indicate where branch was shortened by 50 % for display purposes.

Although I could not identify core ISC system orthologues in *Pygsuia*, I did find two putative homologs of the SUF system in the form of a SUFCB fusion protein. Unexpectedly, one of the two SUFCB homologs (*Pb*-mSUFCB) possesses a MTS suggesting that it functions within the MRO. The other homolog, *Pb*-cSUFCB, lacks the putative MTS. As this fusion of SUFC and SUFB is only observed in *Blastocystis* sp. and

Pygsuia, I performed separate phylogenetic analyses of SUFB (Figure 3-14) and SUFC (Figure 3-15) regions and their respective prokaryotic homologs. The two analyses resolved broadly congruent phylogenies reflecting similar evolutionary histories of the two proteins. To improve the signal, I analyzed a concatenation of the two datasets (Figure 3-16). Unexpectedly, the two *Pygsuia* copies form a clade that branches with the *Blastocystis* SUFCB with maximal support. This group of fused SUFCB proteins emerges as sister to homologs from Methanomicrobiales (a division of Euryarchaea), with maximal support.

The close relationship between *Pygsuia* SUFCBs and *Blastocystis* homologs suggests that they descend from a unique fusion event (although there is no obvious similarity between these sequences in the fusion region). Nevertheless, their common ancestry is supported by the existence of an insertion in the SUFB region shared between the *Pygsuia* and *Blastocystis* sequences to the exclusion of all closely related sequences (Figure 3-17). The *Pygsuia* and *Blastocystis* SUFB domains and all of their close homologues lack the $FADH_2$-binding motif that exists in the *E. coli* homologue. The *Pygsuia* SUFC sequences possess the functionally important residues for metal binding and ATPase activity (including the Walker A/P-loop, Walker B and D-loop motifs (Roche *et al.* 2013)). There is also a CX_nCX_2C motif toward the C-terminal part of SUFC shared only by *Blastocystis* and close prokaryotic homologues (data not shown).

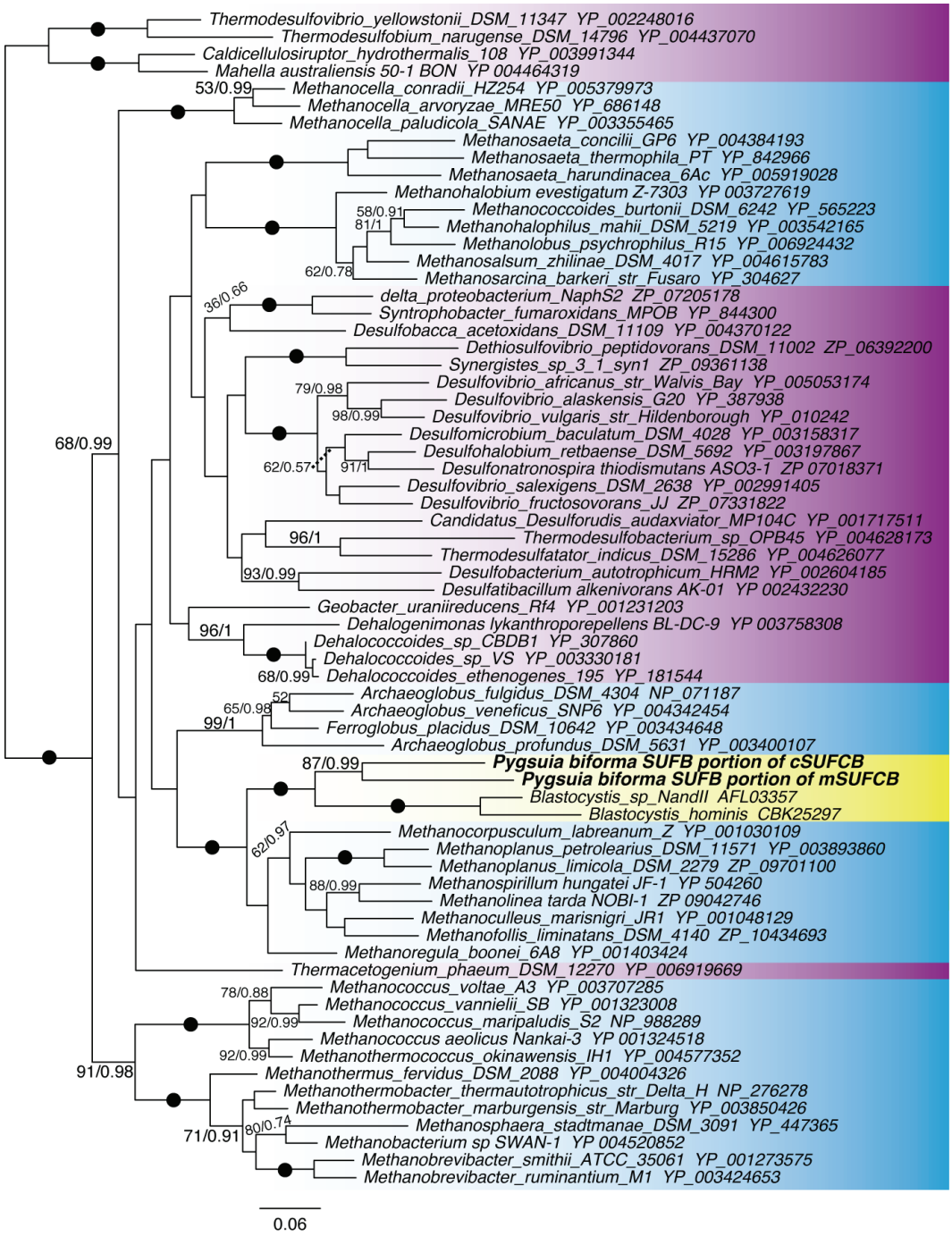

Figure 3-14: Maximum-likelihood (ML) tree of SUFB (68 sequences, 301 sites). Support values and taxa labeling as in Figure 3-2.

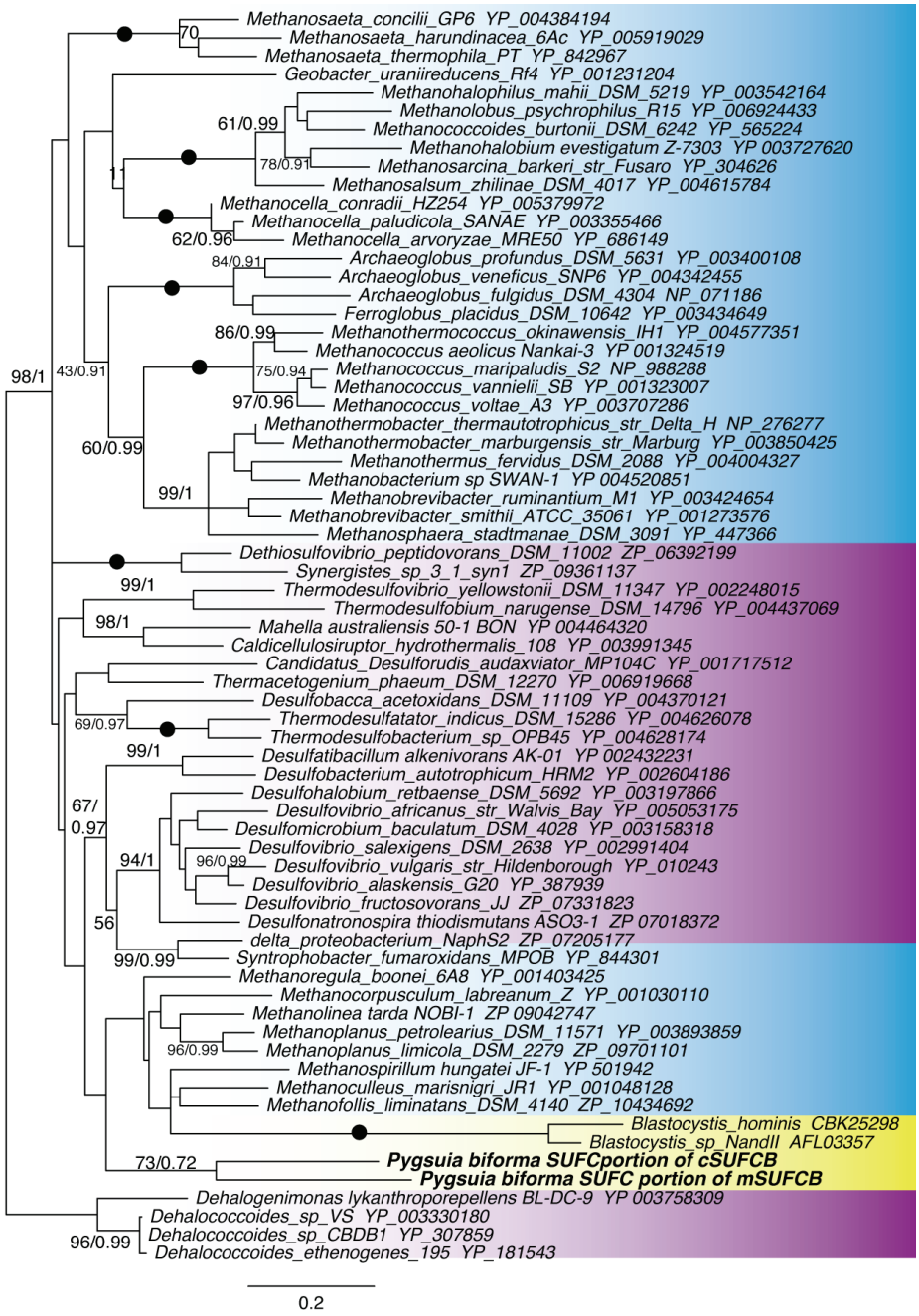

Figure 3-15: Maximum-likelihood (ML) tree of SUFC (68 sequences, 195 sites). Support values and taxa labeling as in Figure 3-2.

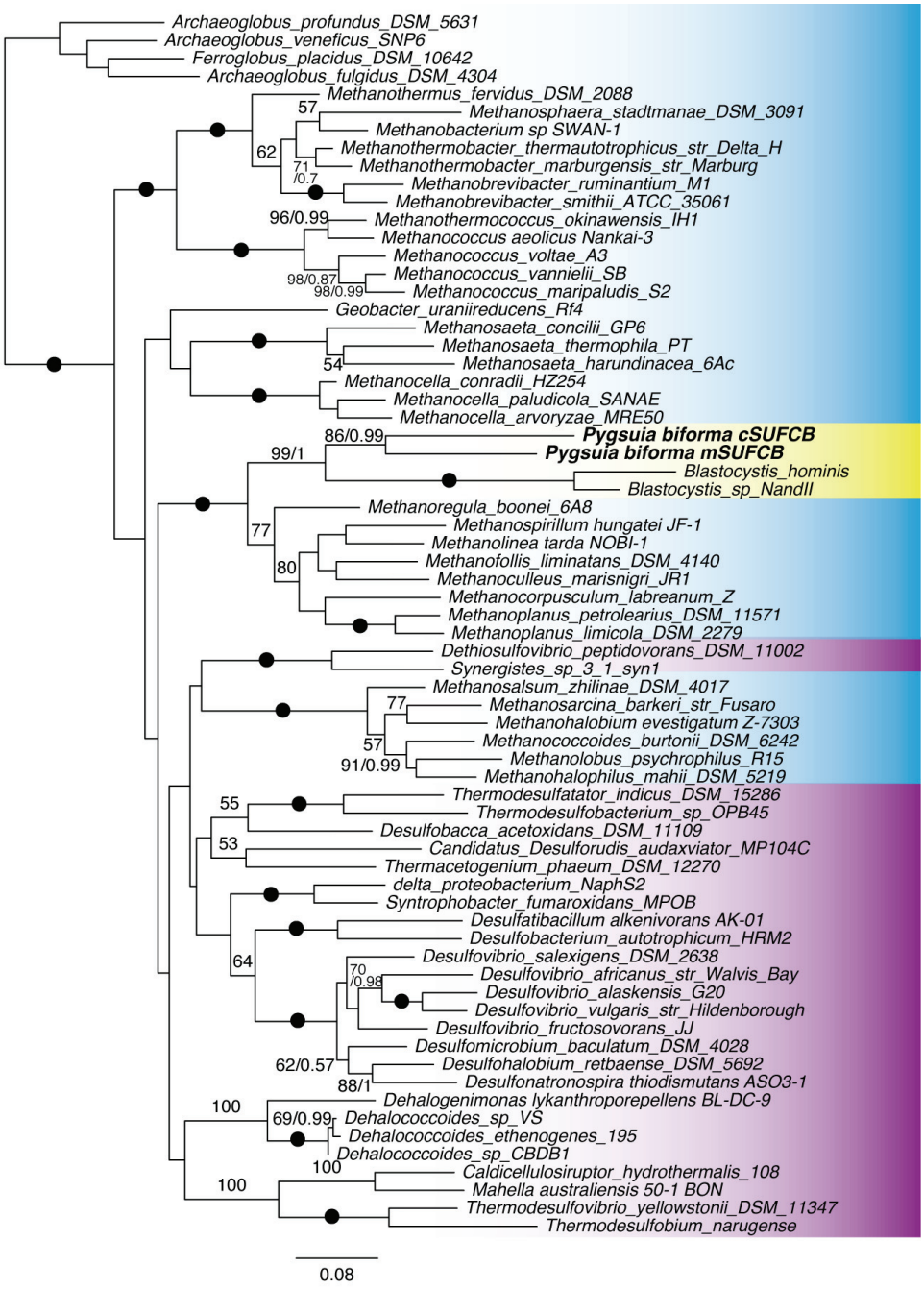

Figure 3-16: Maximum-likelihood (ML) tree of concatenated SUFC and SUFB (68 sequences, 496 sites). Support values and taxa labeling as in Figure 3-2.

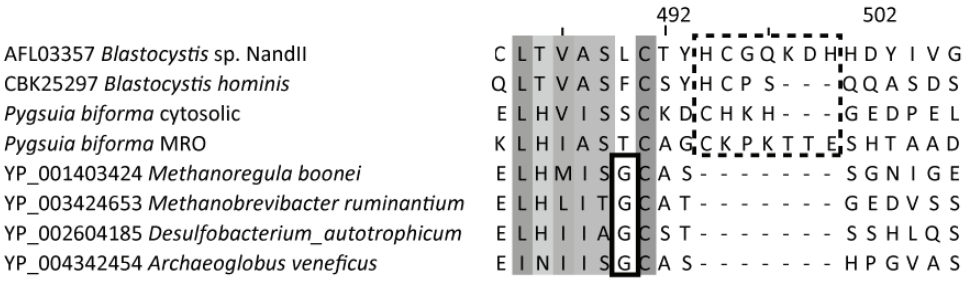

Figure 3-17: Alignment of select region of SUFB indicating shared insertion in *Pygsuia* and *Blastocystis*. Residues conserved in all taxa except *Pygsuia* and *Blastocystis* are boxed with solid black lines. Residues exclusive to these eukaryotes are boxed with dotted lines**.**

3.4.5 Solute Transport, Amino Acid And Lipid Metabolism

Pygsuia encodes a variety of solute transporters including 16 mitochondrial carrier family proteins (MCFs) including an ADP/ATP carrier. Although none of the classical proton pumping respiratory complexes were identified, *Pygsuia* encodes two proteins that could be important for generating a proton gradient: LETM1 (a putative Ca^{+}_{in}/H^{+}_{out} antiporter (Jiang *et al.* 2013)) and pyridine nucleotide transhydrogenase (PNT; responsible for the interconversion of NADH/NADP to NAD/NADPH with the concomitant export of protons (Arkblad *et al.* 2002)).

Enzymes responsible for the synthesis and catabolism of amino acids such as glycine, serine, threonine, tryptophan, alanine, leucine, isoleucine and valine were identified (Figure 3-1). Elements of biosynthesis and β-oxidation of fatty acids were identified, however two components do not have predicted MTS (HDHβ and KAR) (no outline Figure 3-1). Furthermore, there was no evidence for a fatty acyl-CoA dehydrogenase (responsible for the formation of the enoyl-CoA moiety) or a carnitine shuttle (responsible for fatty acid transport from the cytoplasm to the MRO). I identified enzymes involved in folate biosynthesis (folylpolyglutamate synthase) and various other reactions (short chain dehydrogenase, glutathione amine-dependent peroxidase, acyl-CoA synthetase) (Table 3-4 and Table 3-7).

Finally, *Pygsuia* encodes a complete cardiolipin biosynthesis pathway which was recently suggested to be absent in MRO-containing organisms (Tian *et al.* 2012).

Furthermore, *Pygusia* encodes enzymes involved in the synthesis of phosphotidylethanolamine and 2-amino-3-phosphonic acid (2-AEP), a head group for phosphonolipids. The latter pathway is rare in eukaryotes having only been documented in molluscs, trypanosomes and ciliates (Liang and Rosenberg 1968; Adosraku *et al.* 1996; Sarkar *et al.* 2003; Metcalf and van der Donk 2009). In *Tetrahymena*, phosphoenol pyruvate (PEP) is converted to 3-phosphonopyruvate (PPyr) by a PEP mutase (PEPM) and is subsequently converted to phosphonoacetaldehyde (PPA) by a PPyr decarboxylase (PPYRDC); both of these enzymes are present in the mitochondrial proteome (Smith *et al.* 2007). The final step in the pathway is performed by a putative PPA transaminase, which does not appear to be organellar in *Tetrahymena*. In *Pygsuia,* I identified transcripts encoding all three enzymes, and, as in *Tetrahymena*, only PEPM and PPYRDC possess predicted MTS.

I examined the evolutionary history and predicted cellular localization of PEPM, PPYRDC and PPA transaminase across eukaryote diversity. I found homologues within Amoebozoa, the Stramenopile-Alveolate-Rhizaria (SAR) clade, Holozoa and kinetoplastids although not all species have all three enzymes. Only some of these homologues have predicted MTS. The remaining members of these eukaryote groups lack these enzymes. In phylogenetic analyses, the *Pygsuia* PEPM homologue branches within a larger eukaryotic clade that robustly emerges as sister to a predominantly firmicute group (BV = 82, PP = 0.99; Figure 3-18). A generally similar pattern is observed in the PPA transaminase phylogeny with *Pygsuia* grouping in a predominantly eukaryote group that also comprises sequences from a few miscellaneous bacteria (BV = 84, PP = 0.99; Figure 3-19) cluster with other eukaryotic sequences, whereas its PPYRDC homolog branches separate from other eukaryotes with a heterogeneous collection of prokaryotes (Figure 3-20). In contrast, the PPYRDC phylogeny shows *Pygsuia* emerging within a taxonomically heterogeneous prokaryotic assemblage separated from an otherwise monophyletic eukaryotic grouping (BV = 65, PP = 0.98; Figure 3-20). In each case, the eukaryotic homologs of these enzymes do not have α-proteobacterial affinities that would indicate a mitochondrial origin.

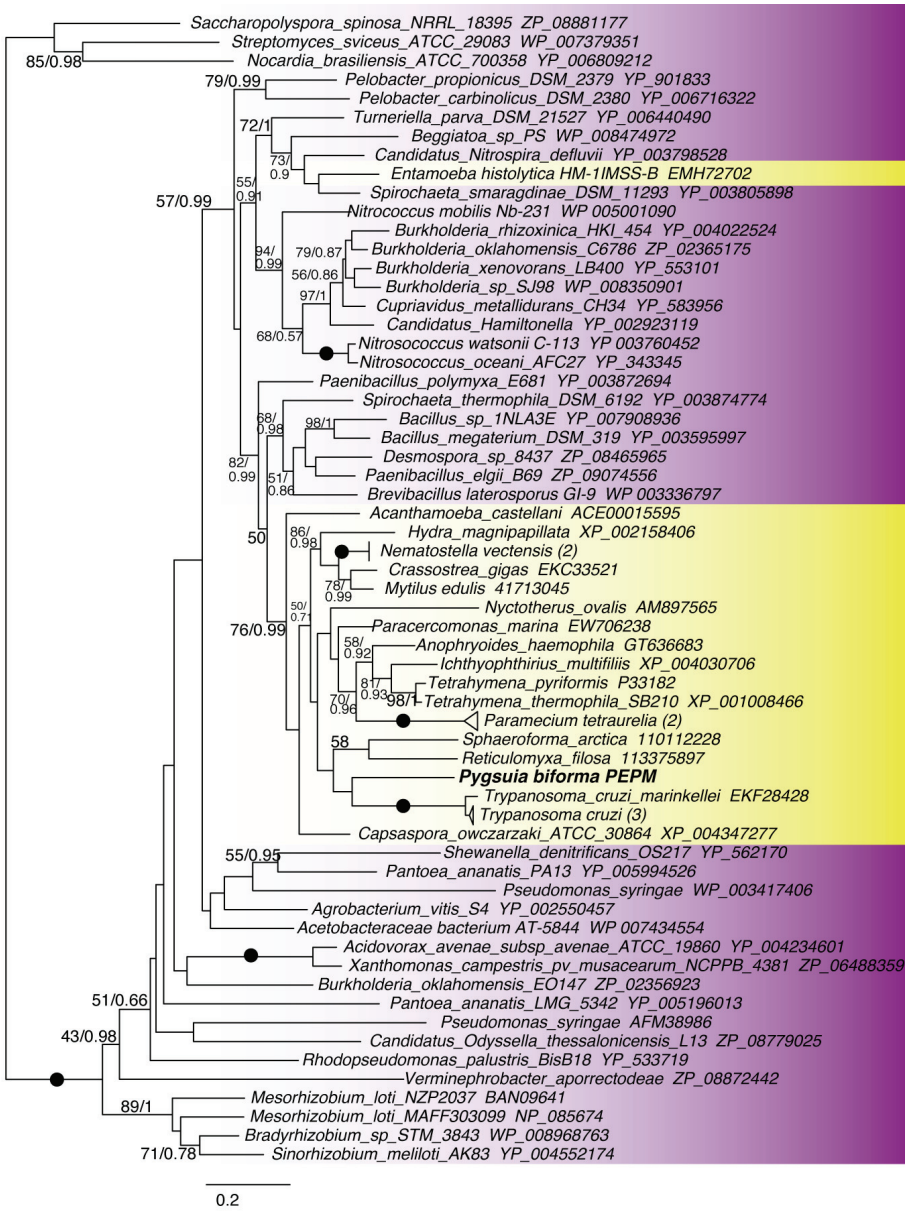

Figure 3-18: Maximum-likelihood (ML) tree of phosphoenolpyruvate mutase (PEPM; 65 sequences, 284 sites). Support values and taxa labeling as in Figure 3-2.

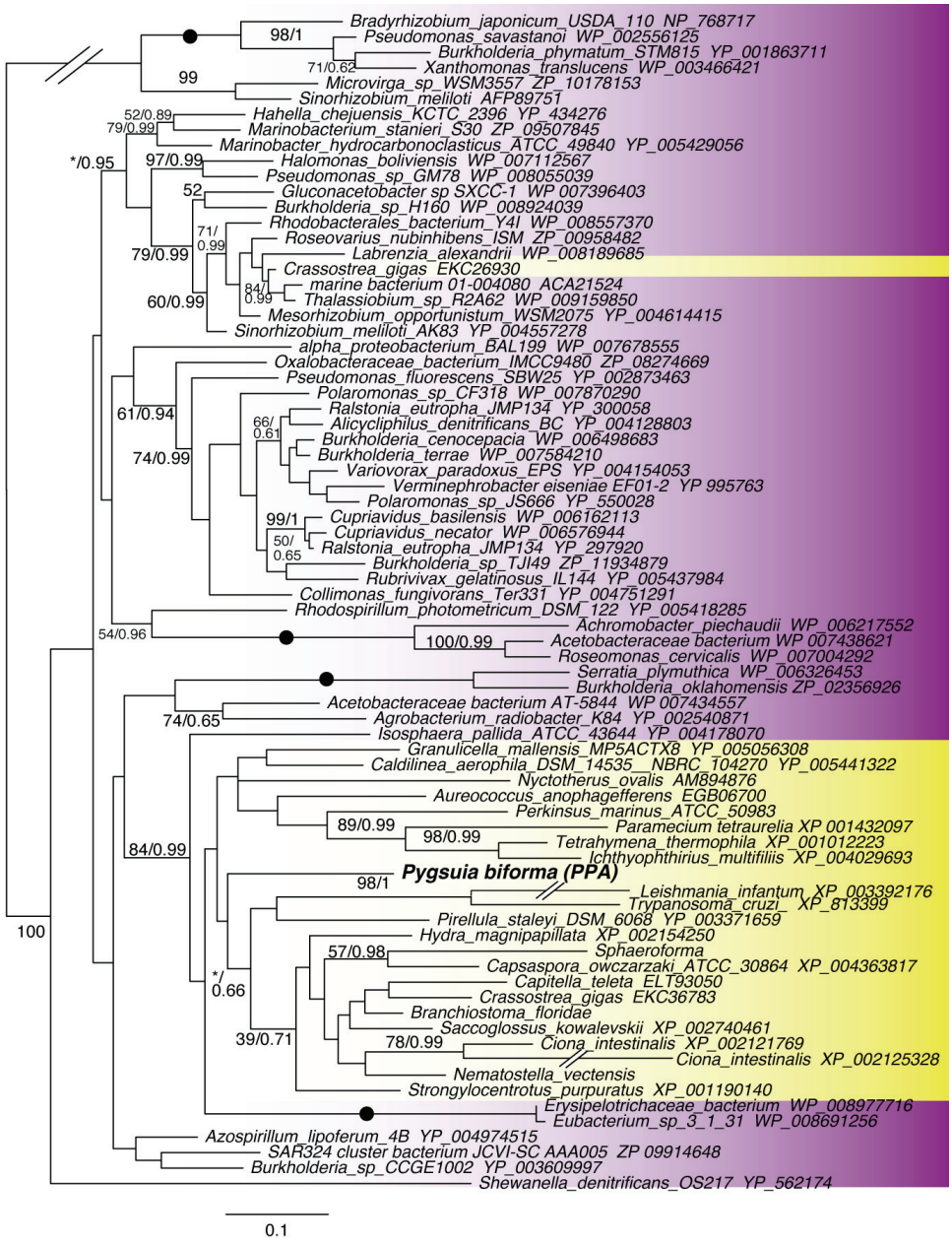

Figure 3-19: Maximum-likelihood (ML) tree of concatenated Phosphonoacetaldyhyde aminotransferase (PPA; 76 sequences, 218 sites). Long branches that were shortenend are indicated with hash marks. Support values and taxa labeling as in Figure 3-2.

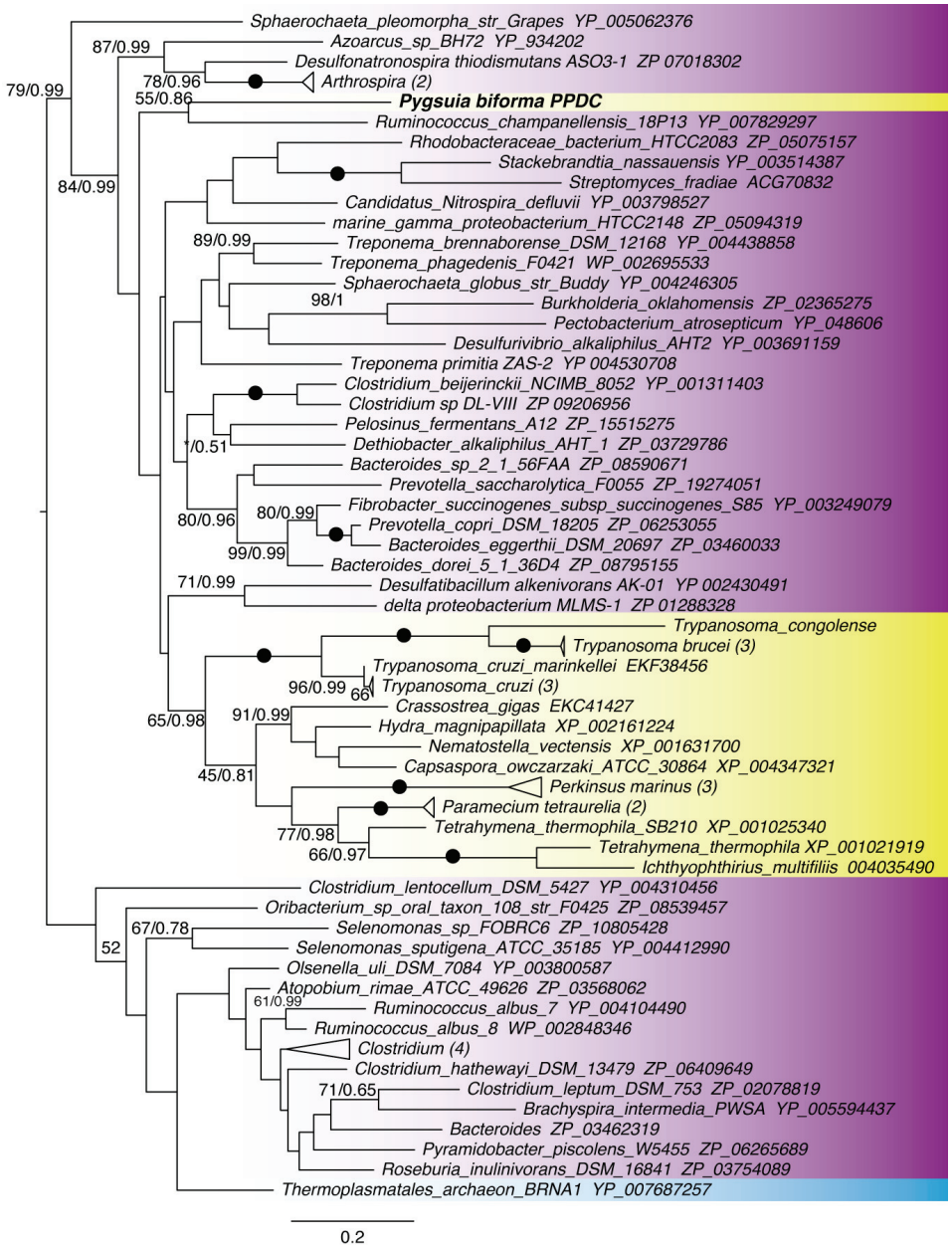

Figure 3-20: Maximum-likelihood (ML) tree of phosphonopyruvate decarboxylase (PPDC; 69 sequences, 268 sites). Support values and taxa labeling as in Figure 3-2.

3.4.6 Localization And Morphology Studies Of Pygsuia biforma *Proteins In Yeast And* in vivo

In the absence of a genetic system in *Pygsuia*, Pygsuia MRO proteins (fused to GFP) were heterologously expressed in yeast to assess the localization of putative MRO proteins. GFP fusion protein constructs of NIFU-like protein (NFU1), the putative targeting peptide from mSUFCB (mSUFCB-MTS) and the full-length mSUFCB GFP fusion

proteins were localized to the mitochondrion of yeast whereas cSUFCB-GFP was cytosolic (Figure 3-21). Note that the full-length mSUFCB-GFP appears to alter the morphology of the yeast mitochondria when expressed for more than one hour.

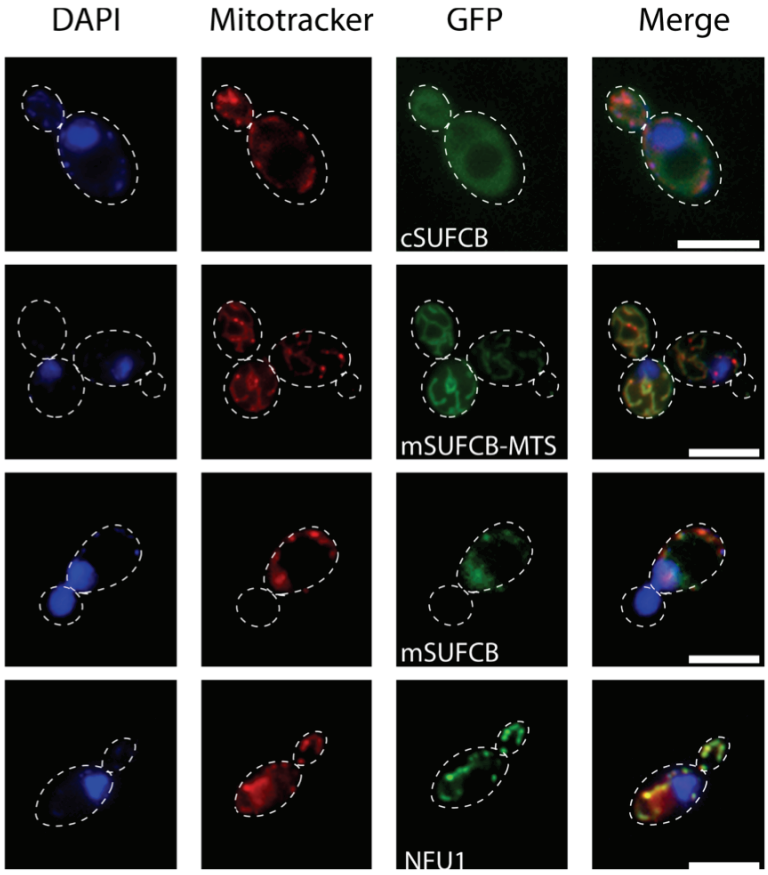

Figure 3-21: Localization of Pygsuia cSUFCB, mSUFCB, mSUFCB-MTS and NFU1 GFP fusion proteins in yeast. Indicated GFP-fusion proteins were expressed in yeast (green). Mitochondria and nucleic acid were co-stained with Mitotracker orange and DAPI respectively. Scale bar = 10 µm.

Antibodies raised against a peptide specific to *Pb*-mSUFCB recognized recombinant mSUFCB (α-mSUFCBpep; Figure 3-22A). Similarly, heterologous antibodies raised against *Trichomonas vaginalis* ASCT (α-*Tv*-ASCT, type 1C) recognized native and purified recombinant *Pb*-ASCT1C in immunoblots (Figure 3-22A). Spinning-disc immunofluorescence confocal and electron microscopy were used to explore the 3D morphology of *Pygsuia* MROs (single image slices can been seen in Figure 3-22C). MitoTracker Orange recognizes an elongated organelle located along the dorsum of the cell subtending the flagellum and was typically observed wrapping around the periphery

of the cytoplasm. Both α-*Tv*-ASCT and α-mSUFCBpep antibodies co-localized with MitoTracker (Figure 3-22C and Figure 3-23B,C).

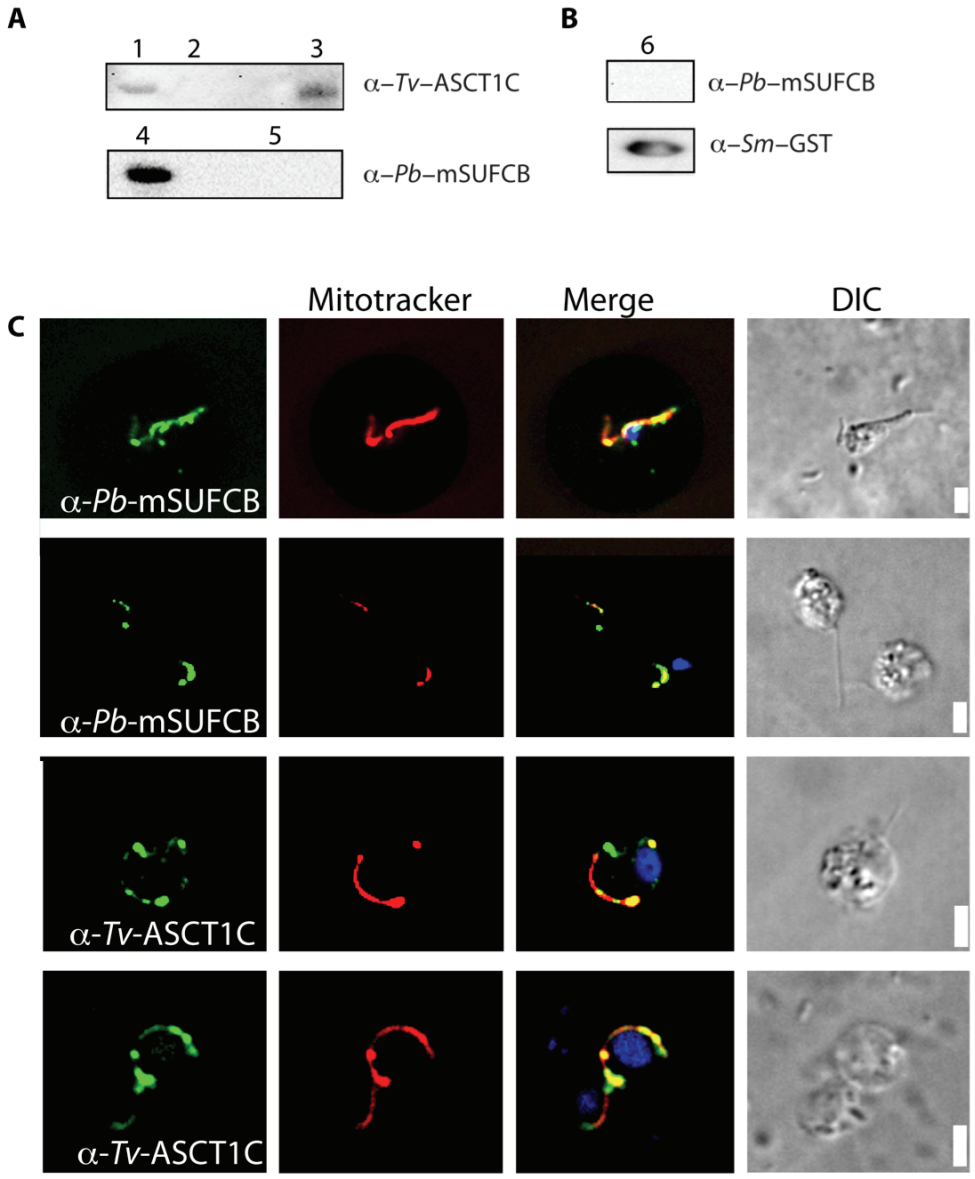

Figure 3-22: Antibodies raised against mSUFCB and ASCT recognize *Pygsuia* mSUFCB and ASCT proteins and localize to *Pygsuia* MROs. **A**: Protein extracted from *Pygsuia* (1) or *Klebsiella* (2) purified recombinant Pb-ASCT (3), purified recombinant *Pb*-mSUFCB (4) or proteins extracted from *E. coli* transformed with pGEX-GST (5) were analyzed by western blotting using indicated antibodies against *Trichomonas vaginalis* ASCT (α-*Tv*-ASCT1C) and *Pygsuia biforma* mSUFCB (α-*Pb*-mSUFCB), respectively. **B:** To ensure the antibodies were not recognizing the GST-tag of recombinant *Pb*-mSUFCB, proteins from *E. coli* expressing just the GST-tag (6) were analyzed by western blotting. Antibodies raised against GST (α-*Sm*-GST) and not antibodies raised against SUFCB (α–*Pb*-SUFCB) recognized the 26 kDa GST. **C:** α-*Pb*-mSUFCB and α-*Tv*-ASCT1C antibodies (green) were incubated with *Pygsuia* cells and images captured represent single 0.3 μm sections

slices of *Pygsuia* cells stained with mitotracker orange (red) and DAPI (blue). DIC, differential interference contrast; scale bar represents 5μm.

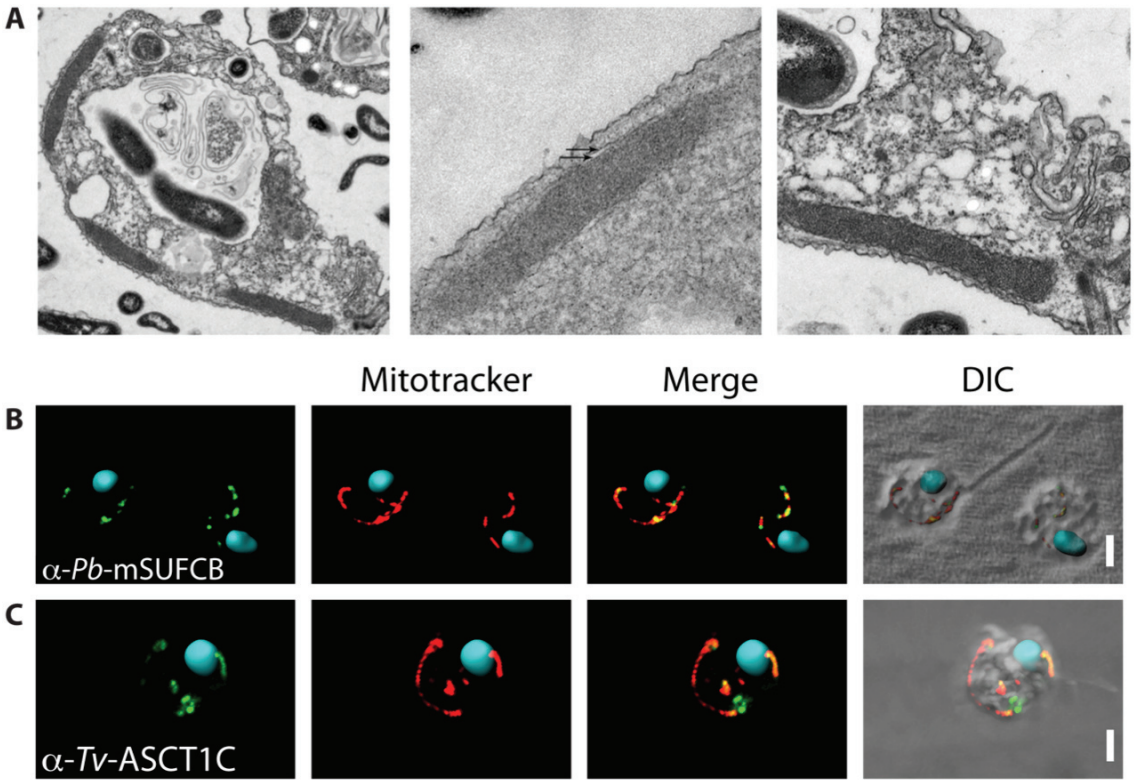

Figure 3-23: **A:** Transmission electron microscopy of *Pygsuia* cells. MRO (M) and Bacteria (B) injected in a vacuole are labeled, arrows indicate the presence of a double membrane (middle). Scale bars 1000, 200, 500 nm. Antibodies raised against mSUFCB (green; **B**) and ASCT (green; **C**) colocalize with MRO-dye Mitotracker (red) in *Pygsuia* cells Confocal slices (0.3 μm) were deconvoluted and combined to generate a 3D-image and DAPI-stained nucleus were rendered using Imaris. Scale bars 5 μm.

Transmission electron microscopy of a *Pygsuia* cell revealed an electron-dense double membrane-bound organelle without canonical cristae similar to those reported in *Breviata anathema* (Heiss *et al.* 2013) (Figure 3-23A). Like many other MROs, there is no evidence for an organellar genome since fluorescent nucleic acid dyes such as DAPI and DRAQ5 did not detect nucleic acid inside the organelle. Furthermore, *Pygsuia* does not possess genes typically described as organellar genome encoded (such as respiratory chain proteins) or organelle specific genome maintenance proteins (such as RNA polymerase or mitochondrial ribosomal proteins).

3.4.7 *Proteins Predicted In* Breviata Anathema *And Not* Pygsuia biforma

A total of 46 of the 122 *Pygsuia* MRO proteins were identified in *Breviata*. Most of these sequences are incomplete. *Breviata* encodes three ISCU and one ISCS proteins. I identified a malate dehydrogenase-like protein in *Breivata* that I did not find in *Pygsuia*. Mitoprot (PP = 0.6226) but not TargetP predicted a MTS. However, further inspection of this sequence revealed that there was not an extension relative to bacterial and cytoplasmic isoforms suggesting this protein might not function in the MRO (data not shown). I also identified a thioredoxin-like protein in *Breviata* with a predicted MTS that does not have a MTS in *Pygsuia* (data not shown, see supplementary file 1 from Stairs *et al.,* 2014).

3.5 Discussion

Microscopic analysis of Pygsuia identified a double membrane-bounded, structure reminiscent of the hydrogenosomes of Trichomonas vaginalis. However, unlike the T. vaginalis organelle which is often present in multiple copy within the cell, the Pygsuia MRO is structurally unique and appears to be restricted to one organelle per cell. Next-generation sequence technology has allowed for the characterization of the transciptome of P. biforma and inference of 122 putative MRO proteins, the majority of which have predicted MTS (Figure 3-1, Table 3-2 to Table 3-9). This unique set of functions predicted for the *Pygsuia* MRO not only bridges the gap between hydrogen-producing mitochondria (HPM) and hydrogenosomes (Müller *et al.* 2012) but also reveals completely novel biochemical properties associated with an MRO.

3.5.1 Energy Metabolism In MROs

HPM and hydrogenosomes are ATP- and hydrogen-producing organelles (via SCS, ASCT and HYDA) that also participate in pyruvate oxidation (via PNO or PDC in HPM; PFL or PFO in hydrogenosomes). In these organelles, HYDA function is dependent on H-cluster maturation by three HYDA maturases (HYDE, F, G). I identified putative organellar PFO, HYDA, ASCT, SCS, and all three HYDA maturases in *Pygsuia*. Since I predict dual

localization of HYDA in both the MRO (canonical *Pb*-mHYDA) and cytoplasm (*Pb*-cHYDA-CYSJ1-3, *Pb*-cSD-HYDA) of *Pygsuia,* it is unclear how the cytoplasmic HYDAs are matured. Either the maturases are dual targeted as described for other proteins in some model systems (Baudisch *et al.* 2014) or the putatively cytosolic hydrogenases can function without maturation. The latter possibility may be related to the presence of additional SD or CYSJ domains on the cytoplasmic hydrogenases.

Similar to the HPM of *Blastocystis, Pygsuia* organelles appear to have an incomplete TCA cycle possessing only FH, CII and SCS suggesting malate is ultimately converted to succinate. In this scenario, CII would function in the reductive reverse direction (as a fumarate reductase) and require a quinone with a lower electron potential such as RQ (Amino *et al.* 2003; Hoffmeister *et al.* 2004), and a corresponding RQ reductase. In fact, I identified a gene encoding a recently described RQ biosynthesis enzyme (RQUA, (Lonjers *et al.* 2012)) with an MTS in *Pygsuia biforma* and other eukaryotes that appears to have been laterally acquired (Figure 3-9). The exact pathway for RQ biosynthesis remains elusive, however, recent reports suggest it can be synthesized from UQ (Brajcich *et al.* 2010; Lonjers *et al.* 2012). Since I was unable to identify all the components for ubiquinone biosynthesis, *P. biforma* might rely on exogenous UQ - much like UQ-deficient yeast (Padilla-López *et al.* 2009) or humans (Lagier-Tourenne *et al.* 2008; Quinzii and Hirano 2010). Transport of exogenous UQ to mitochondria is not well understood. However, in UQ-deficient mice, exogenously supplied UQ is specifically transported to the IM of the mitochondrion (Lapointe *et al.* 2012). I hypothesize that a similar transport mechanism could exist in *Pygsuia* where bacteria-derived UQ is transported from the food vacuole or plasma membrane to the MRO and specifically incorporated into the IM where it is converted to RQ. The following chapter provides a more in-depth analysis of RQ biosynthesis in eukaryotes.

The MROs of *Pygsuia* seem to blur the boundaries between HPM and hydrogenosomes since its proteins are predicted to be involved in pyruvate oxidation and hydrogen production like hydrogenomes (PFO, HYDA, ASCT and SCS), but also HPM features such as quinol-reoxidation (CII and AOX but not CI).

3.5.2 Conservation Of Mitochondrial Protein Import

In eukaryotes, the vast majority of mitochondrial matrix proteins are encoded by the nucleus and subsequently transported into the organelle (Dolezal et al. 2006). When comparing protein import components across eukaryotic diversity, the complement of proteins identified in *Pygsuia* is similar to other non-Opisthokonts (*e.g., Dictyostelium*) with the exception of the outer membrane complex (*Pygsuia* only encodes TOM40 and SAM50). *Pygsuia* encodes many of the same components as other well studied MRO-bearing organisms (*T. vaginalis*, *E. histolytica,* and microsporidians (Liu *et al.* 2011; Heinz and Lithgow 2012)) such as SAM50, TOM40, TIM23, TIM17 and PAM complex. The widespread conservation of the aforementioned proteins in otherwise 'reduced' MROs from diverse and distantly-related organisms suggests that they represent the 'core' components of protein import. However, *Pygsuia* appears to have a more elaborate import apparatus compared to other MRO-bearing organisms since it encodes components of the IMS disulfide relay system (TIM8, 9, 10, 13, MIA40) and TIM50 (Liu *et al.* 2011).

3.5.3 Acquisition Of SUF-Like Fe-S Cluster Biosynthesis And Loss Of Isc Machineries In The **Pygsuia** *Lineage*

Until now, amongst eukaryotes, the archamoebae lineage (*i.e., Entamoeba* and *Mastigamoeba*) is known to have lost the organellar ISC system for Fe-S cluster biogenesis and possess instead a homologous nitrogen-fixation system (NIF) acquired by LGT from ε-proteobacteria (van der Giezen *et al.* 2004; Gill *et al.* 2007; Nyvltová *et al.* 2013). In some bacteria and these amoebae, the simpler NIF system is the only system present for the synthesis of Fe-S clusters (Olson *et al.* 2000). A recent report demonstrated that *Mastigamoeba balamuthi* actually encodes two copies of each component of the NIF system: two targeted to the MRO (*Mb*-NIFS-M and *Mb*-NIFU-M) and two destined for the cytoplasm (*Mb*-NIFS-C and *Mb*-NIFU-C) (Nyvltová *et al.* 2013). Here, I report another apparent loss of the ISC system in *Pygsuia biforma.* The high depth of Illumina sequencing coverage I have obtained for ISC-related genes such as

nfu1 and *ind1* (1596.36X and 243.4X respectively) suggest that the lack of reads corresponding to any *isc* homologs in our *Pygsuia* transcriptome likely represents genuine absences of the genes. A complete genome sequence for *Pygsuia* would be useful to confirm this observation.

I identified two fused SUFC/SUFB scaffold proteins (SUFCB) in *Pygsuia biforma* and showed that the isoform possessing a predicted mitochondrial targeting peptide (*Pb*-mSUFCB) is in fact localized to the MROs by immunolocalization in *Pygsuia* (Figure 3-23). This SUFCB fusion protein is also present in *Blastocystis,* however the gene exists only in single copy and the protein product was shown to localize to the cytoplasm (Tsaousis *et al.* 2012). Phylogenetic analyses of SUFCB from *Blastocystis* and *Pygsuia* suggest that the SUFC/SUFB operon was acquired by one of these eukaryotic lineages from a methanomicrobiales archaeon donor and subsequently fused into a single open reading frame in the recipient genome. The SUFCB fusion gene was then transferred to the other eukaryotic lineage through a eukaryote-to-eukaryote LGT. Since SUFCB does not appear to be encoded by *Breviata anathema* this LGT event might have happened after the divergence of *Pygsuia* from other breviates. I hypothesize that the *Pygsuia* SUFCB was duplicated and that one of the copies eventually acquired a MTS. Over time, the MRO SUFCB system may have functionally replaced the ancestral mitochondrial ISC system resulting in the loss of all ISC components including the nuclear transcription factor ISCR.

Unlike the NIF system present in *Mastigamoeba* and *Entamoeba,* the SUF system of *Pygsuia biforma* (SUFCB) is not homologous to any component of the ISC. This non-homologous replacement scenario requires the co-evolution of chaperone proteins (NFU1 and IND1) that have to now interact with the SUFCB proteins in order to transfer the Fe-S clusters to apoproteins. This hypothesis is consistent with the observation that, in bacteria that harbor ISC and SUF systems (*e.g., E. coli*), the typically ISC-associated NFU1 has been shown to transfer Fe-S clusters from $SUFBC_2D$ to apoproteins (Py *et al.* 2012). Like the SUFCB proteins of *Blastocystis* and methanomicrobiales, the residues responsible for binding flavin are absent in the *Pygsuia biforma* SUFCB sequences. The flavin cofactor of the *E. coli* $SUFBC_2D$ complex has been hypothesized to be important

for the acquisition of ferric iron from various iron donors (ferritin, ferric citrate and frataxin) (Roche *et al.* 2013). However, lack of flavin binding does not prevent removal of the Fe-S cluster from the $SUFBC_2D$ complex of *E. coli* (Wollers *et al.* 2010). Moreover, ferritin and frataxin have not been identified in *Pygsuia*. Therefore, the apparent absence of the flavin binding residues of the SUFCB proteins of *Pygsuia, Blastocystis* and methanomicrobiales might be related to the absence of such electron-dependent iron donors. The source and means by which the SUFCB proteins of these organisms acquire iron remains unknown. Similarly, the traditional components of the SUF operon (SUFA-E, SUFS) have not been identified in *Pygsuia, Blastocystis* and some archaea suggesting these organisms employ a yet unknown process of Fe-S cluster biosynthesis.

It is unclear why an ISC system would be replaced by the non-homologous SUF system in the *Pygsuia* MRO. Clearly after acquisition of the SUF system by LGT, the ancestral breviate must have possessed both Fe-S biosynthetic systems. The SUF operon in prokaryotes is typically upregulated under (and is more tolerant to) iron starvation and oxidative stress (Outten *et al.* 2004). If the ancestors of *Pygsuia* were periodically exposed to such conditions, this could have favoured the maintenance of the acquired SUF system over the ancestral ISC system.

3.6 Conclusion

Here I report a unique collection of functions associated with the mitochondrion-related organelles of the breviate flagellate *Pysuia biforma.* In addition to the typical MRO and mitochondrial processes, I identified genes involved in functions previously unknown in mitochondria or MROs. Some of these genes were likely acquired by LGT including a rhodoquinone biosynthesis enyzme and a SUFCB protein involved in Fe-S cluster biosynthesis. These are striking examples of how lateral gene transfer can remodel MRO function in adaptation to hypoxia.

As more mitochondria and MROs are characterized from a greater diversity of eukaryotic lineages, it is becoming clear that at least one kind of Fe-S cluster biosynthesis system is essential. While most eukaryotes have retained the ISC system in their MROs, the two clear exceptions are the MROs of the archamoeba *Mastigamoeba*

that has a NIF system and the MROs of *Pygsuia biforma* that have SUF system. This strongly suggests that the reactions needed for the synthesis of Fe-S clusters, regardless of their evolutionary origin, demand compartmentalization. This further highlights the fundamental role and widespread conservation of Fe-S cluster biosynthesis in mitochondria and MROs.

The novel combination of properties of the *Pygsuia* organelles cannot be easily fit into any of the classes of MRO functions recently proposed by Müller and colleagues (Müller *et al.* 2012). As more lineages of anaerobic/microaerophilic protists are studied, the diversity of MRO properties will likely increase, suggestive of a continuous spectrum of metabolic phenotypes rather than well-defined classes, revealing the plasticity of these endosymbiont-derived organelles.

Chapter 4 Evolution And Cellular Localization Of Rhodoquinone Biosynthesis In *Pygsuia biforma* And Other Anaerobic Eukaryotes

4.1 Abstract

Complex II (succinate dehydrogenase, SDH) of the respiratory chain typically catalyzes the conversion of succinate to fumarate with the concomitant reduction of ubiquinone (UQ). In some anaerobic bacteria and eukaryotes complex II (CII) functions as a fumarate reductase (FRD) to convert fumarate to succinate with the oxidation of a different quinone species (rhodoquinone, RQ). RQ is structurally similar to UQ, but has a lower electron potential favouring CII-catalyzed fumarate reduction over succinate oxidation. Recently, a putative methyltransferase homolog was discovered in *Rhodospirillum rubrum* (named RQUA) that was shown to be involved in RQ biosynthesis (Lonjers et al. 2012). To investigate the prevalence and evolutionary history of RQUA in bacteria and eukaryotes, I performed similarity searches of all publicly available genome sequences. RQUA is rare within the bacterial domain, and is encoded by only 50 distinct bacterial genomes from the α- and β-proteobacterial divisions. It is completely absent in Archaea. A number of novel RQUA homologs were identified in a variety of anaerobic or facultatively anaerobic microbial eukaryotic genomes including genomes of two subtypes of *Blastocystis* sp*., Mastigamoeba balamuthi, Pygsuia biforma, Euglena gracilis* and *Monosiga ovata*. All of the full-length RQUA sequences from these protists were found to contain a predicted mitochondrial targeting sequence, suggesting they function within mitochondria or related organelles. The patchy phylogenetic distribution and relationships of the eukaryote and bacterial RQUA sequences suggest that the gene has been laterally transferred multiple times between Domains of Life. Using immunofluorescence microscopy, homologous antibodies directed against RQUA localized to *Pygsuia biforma* mitochondrion-related organelles. Organisms that acquired the *rqua* gene are likely able to synthesize RQ. This suggests that the transfer of the

gene between protists is probably selectively advantageous by allowing the mitochondrial electron transport chain to function in hypoxia via an RQ-utilizing CII.

4.2 Introduction

In aerobic eukaryotes, complexes I (CI) and II (CII) of the mitochondrial respiratory chain oxidize reduced cofactors generated by the tricarboxylic acid cycle (NADH and $FADH_2$ respectively) and reduce the lipid-soluble electron carrier ubiquinone (UQ, Figure 4-1) to ubiquinol (UQH_2). UQH_2 is then oxidized by complex III (CIII) and electrons are then shuttled, via cytochrome c, to complex IV (CIV) that reduces O_2 to H_2O (See Chapter 1, Figure 1-1). In the process CI, CIII, and CIV generate a proton gradient across the inner mitochondrial membrane that fuels oxidative phosphorylation of ADP to ATP by the $F1\text{-}F_o$ ATPase (CV). In the presence of oxygen, the respiratory chain of the facultative anaerobes, *Ascaris suum*, resembles that of other aerobes using CI-CV and the electron carriers UQ and cytochrome c ultimately reducing O_2 to H_2O. However, under anaerobic conditions, the *Ascaris* respiratory chain shifts to use the electron carrier rhodoquinone, (RQ, Figure 4-1) and only CI, CII and CV (Figure 4-2). In these conditions, CI reduces RQ to rhodoquinol (RQH_2) and CII functions as a fumarate reductase (FRD), reducing fumarate to succinate, using electrons donated from RQH_2. RQ is structurally similar to UQ possessing an amino group instead of a methoxy group on the quinone ring (Figure 4-1) and its lower electron potential RQ (-63 mV) compared to UQ (+100 mv) favours the FRD reaction. Thus the *Ascaris* mitochondrial respiratory chain is still able to generate a proton gradient to fuel ATP synthesis even in the absence of oxygen (Figure 4-2). RQ has also been detected in *Caeorhabditis elegans* (Takamiya *et al.* 1999), *Euglena gracilis* (Hoffmeister *et al.* 2004) and *Nyctotherus ovalis* (Boxma and Graaf 2005), however the presence of RQ in the mitochondria of other anaerobic eukaryotes has not been established.

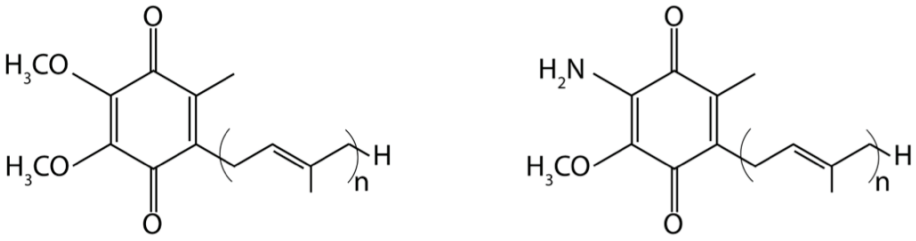

Figure 4-1: Structure and electron potential of ubiquinone and rhodoquinone. Variable number of isoprenyl chains are shown with 'n'.

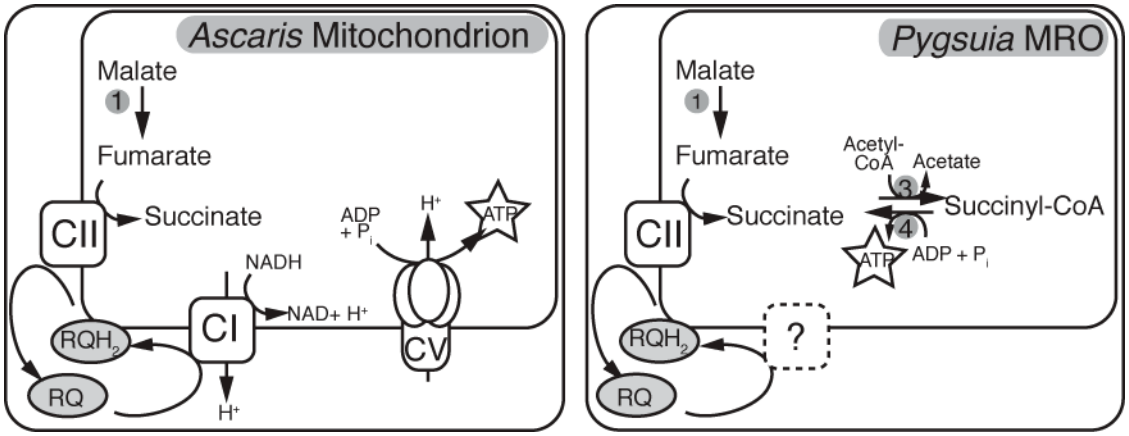

Figure 4-2: Malate dismutation and energy metabolism in mitochondria and related organelles. In *Ascaris suum*, like aerobically functioning mitochondria, NADH from the Kreb's cycle is oxidized by complex I (CI) concomitantly pumping protons into the mitochondria intermembrane space to fuel ATP synthesis by Complex V (CV). Unlike aerobic mitochondria, CI transfers electrons from NADH to the electron carrier rhodoquinone (RQ) to generate rhodoquinol (RQH_2). Fumarate hydratase (1) catalyzes the conversion of malate to fumarate. Fumarate is reduced by complex II (CII) using electrons from RQH_2 to generate succinate and RQ. *Pygsuia* does not encode CI or CV subunits but instead relies on substrate-level phosphorylation to generate ATP. CoA is transferred from acetyl-CoA to succinate by acetate:succinate CoA transferase (2) generating succinyl-CoA which in turn is converted back to succinate by the Kreb's cycle enzyme succinyl-CoA synthetase (3) ultimately generating ATP (or GTP). The enzyme responsible for generating RQH_2 in *Pygsuia* is unknown (?).

Mitochondrion-related organelles (MROs) are specialized mitochondria found in anaerobic protistan lineages that have evolved to cope with low oxygen conditions. The metabolism of MROs from distinct lineages are surprisingly similar despite their independent evolutionary origins (details outlined in Chapter 1, Section 1.3). The properties of MROs vary between organisms ranging from the simple 'mitosomes' of

Giardia intestinalis, that function solely in Fe-S cluster generation, to the more complex 'hydrogenomes' of *Trichomonas vaginalis* or 'hydrogen-producing mitochondria' of *Nyctotherus ovalis* that function in ATP generation. Of the various adaptations to anaerobiosis, the reduction or remodeling of the respiratory chain is of central importance to the ATP generating function of the resulting organelles. Some anaerobic protists have completely lost all components of the respiratory chain while others have maintained all or parts of CI and CII (Williams *et al.* 2002; Tovar *et al.* 2003; Hrdy *et al.* 2004; Boxma and Graaf 2005; Loftus *et al.* 2005; Gill *et al.* 2007; Stechmann *et al.* 2008; Barberà *et al.* 2010; Burki *et al.* 2013; Stairs *et al.* 2014). However, all of these organisms have lost the ability to perform ATP synthesis by oxidative phosphorylation. While RQ has been proposed to function as an electron carrier in the respiratory chains of at least some of these protists (de Graaf *et al.* 2011), direct biochemical evidence for genes involved in RQ biosynthesis have not been described.

Our current knowledge regarding the biosynthesis of RQ comes from studies of the α-proteobacterium *Rhodospirillum rubrum*. Previous reports have suggested that UQ is a precursor to RQ (Brajcich *et al.* 2010) in *Rhodospirillum*. More recently a putative methyltransferase named RQUA was shown to be involved in RQ biosynthesis in this organism (Lonjers *et al.* 2012). This study found that the *R. rubrum* F11 mutant was able to synthesize UQ, but was incapable of synthesizing RQ or growing anaerobically. Sequencing of the genome of the F11 isolate revealed a single nonsense mutation in the *rqua* gene. The mutant could be rescued by complementation with the wildtype *rqua* suggesting RQUA is essential for RQ biosynthesis.

In Chapter 3 and published manuscript (Stairs *et al.* 2014) I showed that RQUA is found in less than 50 different species of bacteria and eukaryotes and emerges from within a clade of ubiquinone biosynthesis methyltransferases (UBIE). The following chapter expands on this observation by examining the evolutionary history, sequence composition and subcellular localization of eukaryotic RQUA homologs.

4.3 Methods

4.3.1 Molecular Biology

Total RNA extraction, cDNA construction and PCR were performed as described in Chapter 3. The *Pygsuia rqua* gene was amplified using primers with BamHI restriction enzyme sites (*Pb-rqua*-forward 5'-CCGGATCCATGAATTCTTTAAGAATTAC-3' and *Pb-rqua*-reverse 5'-CCCGGATCCTGCAATGCGGTGTGCAACAACC-3'; restriction enzyme recognition sites are underlined), purified and cloned into the sequencing vector pCR4 (Life Technologies) by TA-cloning. Plasmids (pCR4-Pb-*rqua*) were purified from transformed *E. coli* using the Nucleospin plasmid purification kit (Machery Nagel) and screened for correct sequence (Genewiz). Destination plasmid pGEX-4T-1 (GE) and pCR4-Pb-*rqua* were digested with BamHI (Thermo). Fragments were purified using the Extract II kit (Machery Nagel) and cloned by standard protocols (Maniatis *et al.* 1982) to generate pGEX-Pb-*rqua.*

4.3.2 Heterologous Expression Of Proteins In E. coli *And Immunoblotting*

Plasmids were transformed into *E. coli* (BL21) for protein expression. Protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (Sigma) to the culture medium of exponentially growing cells and incubating for an additional 4-6 h. Proteins were isolated from *E. coli* cells induced to express the *Pygsuia* protein (GST-RQUA), from plasmid pGEX-*Pb-rqua,* GST tag from plasmid pGEX-*Pb-4T1* and control cells (where preotein expression was not induced). After protein expression, *E. coli* cells were collected by centrifugation (4000 x g, 2 min, 4 °C) and lysed by French press (7000 psi). Unbroken cells and debris were removed by centrifugation (4000 x g, 2 min, 4 °C). The resulting supernatant was saved for subsequent analysis and recombinant protein was isolated using glutathione-magnetic beads (Thermo/Pierce; GST-tag) according to the protocol of the manufacturer. Crude cell lysates and purified RQUA were denatured in sample loading buffer (Sigma), boiled for 5 min and resolved by SDS-PAGE (10%). Proteins were transferred to PVDF membranes (Turbo Blot membranes, Biorad) that were then incubated in blocking buffer (5% skim milk powder,

TBS, 0.5% Tween 20) for 1 h. Anti-*Pygsuia* RQUA antibodies were diluted in blocking buffer (1:500) and incubated with membranes overnight. Following 3 washes in TBS-tween (TBS, 0.5% Tween 20), membranes were incubated with horseradish peroxidase-conjugated goat anti-rabbit secondary antibodies (1:50000, Sigma), washed in TBS-tween and incubated with enhanced chemiluminesence substrate (GE Healthcare) and visualized using a charge-coupled-device chemiluminescence detector (Protein Simple).

4.3.3 Phylogenetic Dataset Construction And Sequence Analysis

Eukaryotic and prokaryotic homologs of RQUA, and close UBIE homologs, were retrieved via BLAST(Altschul 1997) using the *R. rubrum* RQUA sequence as a query against the expressed sequence tag (EST), whole genome shotgun contigs, transcriptome shotgun assemblies and *nr* databases available at the National Centre for Biotechnology and Information (http://www.ncbi.nlm.nih.gov/), the Marine Microbial Eukaryotic Transcriptome Project available through the Camera database (http://camera.calit2.net/) and in-house sequencing projects for *Mastigamoeba balamuthi* (described in Chapter 2), *Blastocystis hominis*, *Condylostoma magnum* (provided by Dr. Eleni Gentekaki and Dr. Denis Lynn) and *Copromyxa protea* (provided by Dr. Matthew Brown). Information regarding the database from which each eukaryotic sequence was retrieved can be found in Table 4-1. Mitochondrial targeting sequences were predicted for each sequence using Mitoprot and TargetP (Claros and Vincens 1996; Emanuelsson *et al.* 2000). Sequences were tentatively annotated as 'mitochondrial' if either Mitoprot or TargetP returned mitochondrial-targeting scores of greater than 0.5. The gene context of *rqua* in bacterial genomes was determined by manual investigation of the relevant genome sequences deposited in GenBank.

Table 4-1 Accession numbers, sequence completeness and mitochondrial targeting predictions for eukaryotic *rqua* squences identified from publicly available and in-house sequencing projects.

Phylum	Organism	Accession	database	Complete[b]	MitoProt	TargetP
Archamoeba	*Mastigamoeba balamuthi*	KM186976	KM186976	C	0.8324	0.918
Discosea	*Paramoeba atlantica*	0201483178	CAMERA	IC	n.d.	n.d.
Discosea	*Neoparamoeba aestuarina*	0201515656	CAMERA	C	0.709	0.66
Tubulinea	*Copromyxa protea*	this study	this study	C	0.8457	0.879
Heterokonta	*Blastocystis_hominis*	300176487	Genbank *nr*	C	0.9684	0.868
Heterokonta	*Blastocystis* sp.	this study	this study	C	0.9231	0.885
Heterokonta	*Attheya septenetrionalis*	0198293608	CAMERA	IC	n.d.	n.d.
Heterokonta	*Attheya* sp.	371497227	Genbank EST	IC	n.d.	n.d.
Heterokonta	*Staurosira complex*	0202500160	CAMERA	IC	n.d.	n.d.
Ciliophora	*Condylostoma magnum*	KM186977	this study	C	0.9182	0.878
Ciliophora	*Fabrea fabrea salina*	0202430720	CAMERA	C	0.7987	0.596
Mollusca	*Aplysia californica*[a]	613494758	Genbank TSA	IC	n.d.	n.d.
Chaonoflagellata	*Monosiga ovata*	163101368	Genbank EST	IC	n.d.	n.d.
Breviatea	*Pygsuia biforma*	KM186978	Genbank SRA	C	0.9919	0.958
Euglenida	*Euglena gracilis*	EC670846	Genbank EST	C	0.9697	0.886
Euglenida	*Eutreptiella gymnastica*	0113734402	CAMERA	C	0.9955	0.922
Euglenida	*Eutriptiella-like*	0174352920	CAMERA	C	0.9986	0.947
Foraminifera	*Reticulomyxa filosa*	569420904	Genbank *nr*	C	0.6108	0.82
Foraminifera	*Ammonia* sp.	0197027732	CAMERA	C	0.5539	0.554
Foraminifera	*Elphidium margaritaceum*	0202694266	CAMERA	C	0.9908	0.881

[a]Potential contaminant
[b]Completeness of sequence (C, complete; IC, incomplete)
n.d. Insufficient data

4.3.4 Phylogenetic Analysis

Sequences were aligned using MAFFT-linsi (Katoh and Toh 2008) and regions of ambiguous alignment were removed using BMGE (Criscuolo and Gribaldo 2010) with default settings. Phylogenies were estimated with RAxML version 8.0.19 (Stamatakis 2014) under the LG, LG4X and LG4M models of evolution. The best-fitting model was determined to be LG4X based on the Akaike Information Criterion (Akaike 1974). Previous reports hypothesized that RQUA evolved from a methyltransferase (UBIE) involved in ubiquinone biosynthesis (Lonjers *et al.* 2012). Indeed, initial phylogenies of RQUA and UBIE sequences revealed RQUA to emerge from within a larger clade of UBIE sequences (Chapter 3 and Stairs *et al.* 2014). Only 5 representative UBIE sequences were included in the final analyses. Final phylogenies were estimated under the LG4X model. Bootstrap support for branches was determined from a total of 500 bootstrap replicates and bootstrap values were mapped onto the best-scoring ML tree that was retrieved from 100 heuristic searches. Bayesian inference was conducted using

PhyloBayes 3.2 (Lartillot *et al.* 2009) by running four Markov chain Monte Carlo (MCMC) chains under the C20 (Poisson) model of evolution (Le *et al.* 2008) with a four-category discretized gamma distribution to account for variable rates-across-sites. MCMC chains were run sampling every 100th tree until all 4 chains converged with a maximum-difference (max-diff) of 0.03. The final consensus tree with posterior probabilities was generated from 4000 trees with a manually determined burn-in. Posterior probabilities (PP) for splits were mapped onto the ML estimated topology using the Dendropy package (Sukumaran and Holder 2010).

4.3.5 Identification Of Quinone-Utilizing Enzymes In Eukaryotes

Sequences from *Dictyostelium discodeum* and *Chlamydomonas reindhartii* respiratory chain complexes (CI, CIII, CIV), quinone biosynthesis enzymes (COQ1-10), alternative oxidase (AOX), electron transferring flavoprotein, (ETFα and β) and ETF dehydrogenase (ETFDH) were manually retrieved from the Kyoto Encyclopedia of Genes and Genomes (Kanehisa 2000). These Q-utilizing enzyme sequences were used as queries to search each eukaryotic genome or transcriptome than also encoded *rqua* using BLAST. Candidate RQUA sequences that were returned by the foregoing BLAST searches with e-values less than 0.001 were then used as queries to search the *nr* database with BLAST to ensure that *Rhodospirillum rubrum* RQUA was one of the most similar homologs in *nr.* To represent the presence or absence of each protein in each organism, a Coulson plot was generated using the Coulson Plot Generator software (Field *et al.* 2013).

4.3.6 Antibody Production

The protein sequence of *Pb*-RQUA was provided to Genscript for antigen design. The series of peptide sequences were provided by the company and the peptide CGGKAVFIDYGRPST was selected as optimal for immunization since it represented a conserved region of the protein and was predicted to be immunogenic by the company.

4.3.7 Culturing And Microscopy

Cultures of *Pygsuia biforma* were maintained in American Type Culture Collection medium 802 prepared in natural seawater as described in (Brown *et al.* 2013) (Stairs *et al.* 2014) at 22 °C. Cells were grown in 15 mL culture tubes filled with media and supplemented with *Klebsiella pneumonia*. Immunofluorescence analysis was performed as described previously in Chapter 3, Section 0.

4.4 Results

4.4.1 Distribution Of RQUA In Bacteria And Eukaryotes

To assess the presence/absence of RQUA homologs in prokaryotic and eukaryotic organisms, RQUA homologs were mined from various publicly available databases. Interestingly, the *rqua* gene is relatively rare in bacteria and eukaryotes and not found in any publicly available archaeal genomes. *RQUA* genes are only found in seven different orders of α–, β– and γ-proteobacteria (*Burkholderiales, Magnetococcales, Neisseriales, Rhiobiales, Rhodobacteriales, Rhodocyclales and Rhodospirillales)*. However, within eukaryotes, I identified at least one *rqua* homolog in five of the six super-groups of eukaryotes (Table 4-1). Some of the sequences are incomplete as they were retrieved from ongoing genomic and transcriptomic sequencing projects. However, all of the full-length sequences encode an N-terminal mitochondrial targeting sequences (MTS) as indicated in Table 4-1.

4.4.2 Phylogentic Analysis Of Bacterial And Eukaryotic RQUA Homologs

Preliminary phylogenetic analyses that included RQUA and related sequences showed that prokaryotic and eukaryotic RQUA orthologs formed a clade that emerged from within a larger group of ubiquionone/menaquinone biosynthesis C-methyltrasnferase proteins (UBIE). After reducing the dataset to include only the RQUA sequences and their closest related UBIE homologs, ML and Bayesian analyses were conducted. In the resulting phylogenies, the branch separating RQUA homologs from UBIE sequences was strongly supported (Figure 4-3, Bootstrap value (BV) = 93, and

posterior probability (PP) = 1.0). Within the RQUA clade, there seems to be two distinct monophyletic groups of homologs composed of prokaryotic and eukaryotic sequences of mixed taxonomic affinities (Group A; PP =0.61, and Group B; BV = 86, PP =0.98 (Figure 4-3)). Group A includes proteobacteria, ciliates (*Fabrea* and *Condylostoma*), foramineriferans (*Ammonia*, *Reticulomyxa* and *Elphidium*), diatoms (*Attheya* and *Staurosira*) and a chanoflagellate (*Monosiga*). While support is generally poor for internal branches in the backbone of this group, it is clear that the ciliates and foraminiferans are nested within a clade of α, β, γ proteobacteria with moderate support (BV=79, PP = 0.99, Figure 4-3,), suggesting that these sequences were likely transferred from bacteria to eukaryotes via LGT at least once. Group B includes the remaining eukaryotic taxa from Amoebozoa (*Mastigamoeba, Paramoebea* and *Copromyxa*), Obazoa (*Pygsuia* and *Aplysia*), Excavata (*Eutriptiella* and *Euglena*) and a stramenopile (*Blastocystis*). Support is poor throughout the clade, however the majority of the eukaryotic sequences (excluding *Copromyxa*) branch together with a group of β-proteobacteria. The topology yielded by Bayesian analysis is slightly different; *Mastigamoeba* branches at the base of Group B and the aforementioned eukaryotes (excluding *Copromyxa*) form a monophyletic group (PP=0.84) sister to β-proteobacteria (PP = 0.99; Figure 4-4). In general, the lack of strong support and incongruence between ML and BI analyses makes it difficult to disentangle the history of these genes in eukaryotes in detail. The close relationship between the mollusk parasite *Neoparamoeba* and sea slug *Aplysia* (BV =100, PP = 1.0) raises the possibility that the *Aplysia* specimens from which genomic sequences were obtained might have been infected with *Neoparamoeba*-like organisms and therefore this sequence is a contaminant (further investigation is necessary to confirm or refute this possibility, but is beyond the scope of this study).

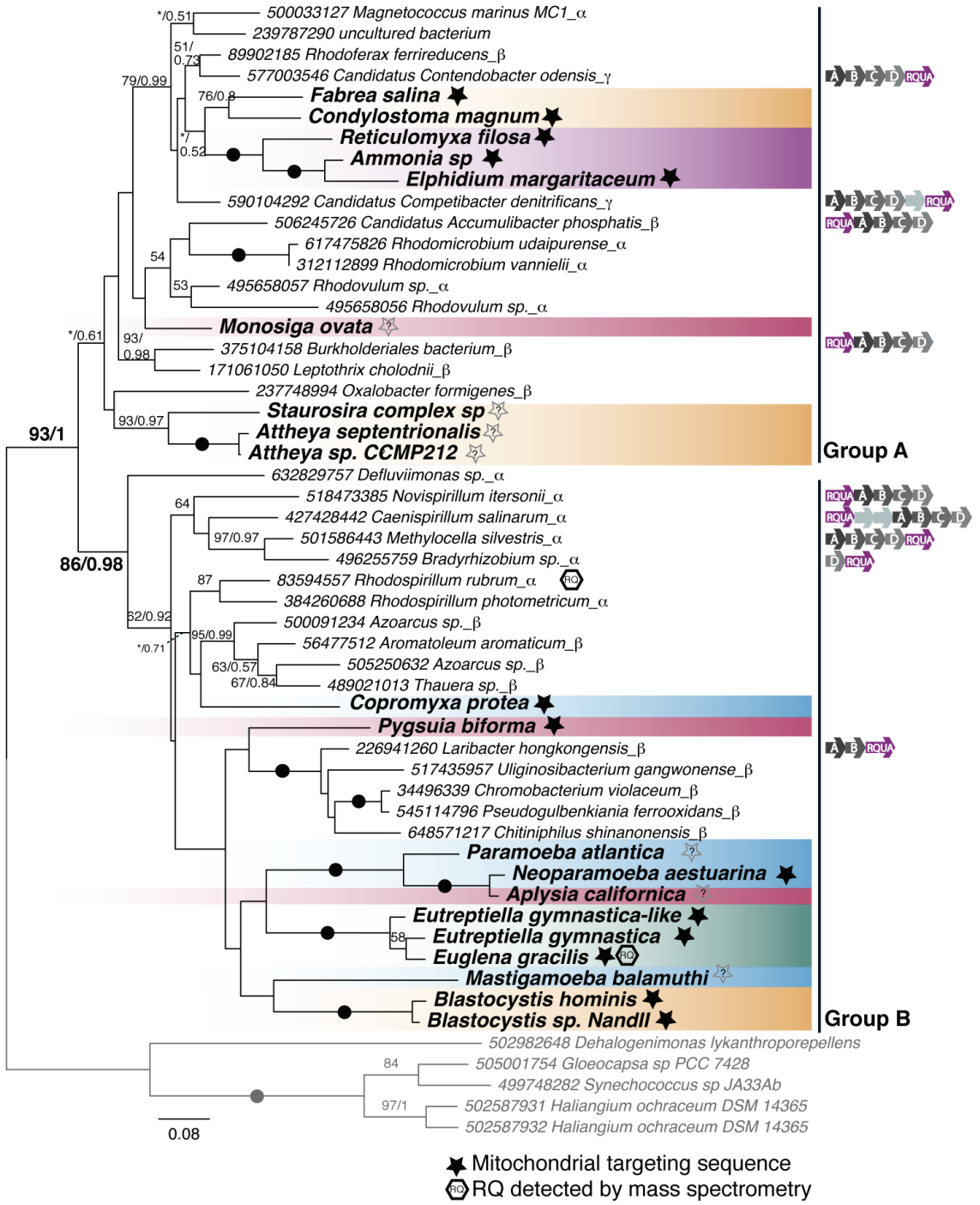

Figure 4-3: Phylogenetic analysis of all eukaryotic and bacterial RQUA homologs (54 sequences and 180 sites). Maximum likelihood (ML) bootstrap values (BV) and Bayesian posterior probabilities (PP) were mapped onto the best scoring ML tree (BV/PP). Only BV or PP values greater than 50 or 0.5 respectively are shown where '*' indicates bipartitions with BV support less than 50 but PP support greater than 0.5. Bipartitions with maximum support (BV = 100 and PP = 1.0) are shown with closed circles. Organisms are coloured based on eukaryotic 'super-group' classification (Amoebozoa, blue; Obazoa, red; Excavata, green; Heterokonts & alveolates, orange; and Rhizaria, purple). Bacterial class (α, β, γ-proteobacteria) are labeled next to each

taxa. Closed black stars indicate presence of a mitochondrial targeting sequence and hexagons represent organisms where RQ has been detected by mass spectrometry. The gene context of *rqua* relative to fumarate reductase (*frd*)/succinate dehydrogenase (*sdh*) subunits A-D are indicated with arrows next to bacteria where applicable. Non- *frd* /*sdh* genes are shown with light blue arrows and *rqua* shown in purple. Select UBIE outgroup was used to root the tree (grey sequences).

To determine if *rqua* is genetically linked to other potential RQ biosynthesis genes, the genomic context of *rqua* in the various bacterial genomes was investigated manually. The *rqua* gene does not appear to be linked to genes encoding other hypothetical proteins or candidate quinone metabolism enzymes. However, in at least nine of the bacterial genomes, *rqua* is encoded physically close (*i.e.,* within one gene or 1 kilobase pair) to the genes annotated as anaerobic fumarate reductase (*frd*)/succinate dehydrogenase (*sdh*) subunits (A-D) (Figure 4-3). Futhermore, in 12 of these bacteria (*Competibacter denitrificans, Candidatus Contendobacter odensis, Rhodomicrobium udaipurense, Rhodomicrobium udaipurense, Rhodomicrobium vannielii, Burkholderiales, Caenispirillum salinarum, Novispirillum itersonii, Methylocella silvestris, Thauera sp., Aromatoleum aromaticum, Laribacter hongkongensis, Pseudogulbenkiania ferrooxidans,* and *Chromobacterium violaceum*) there is second *frd/sdh* operon located elsewhere in the genome. The genetic linkage I observed between *rqua* and the anaerobic complex II suggests they could be transcriptionally linked in an operon in these bacteria.

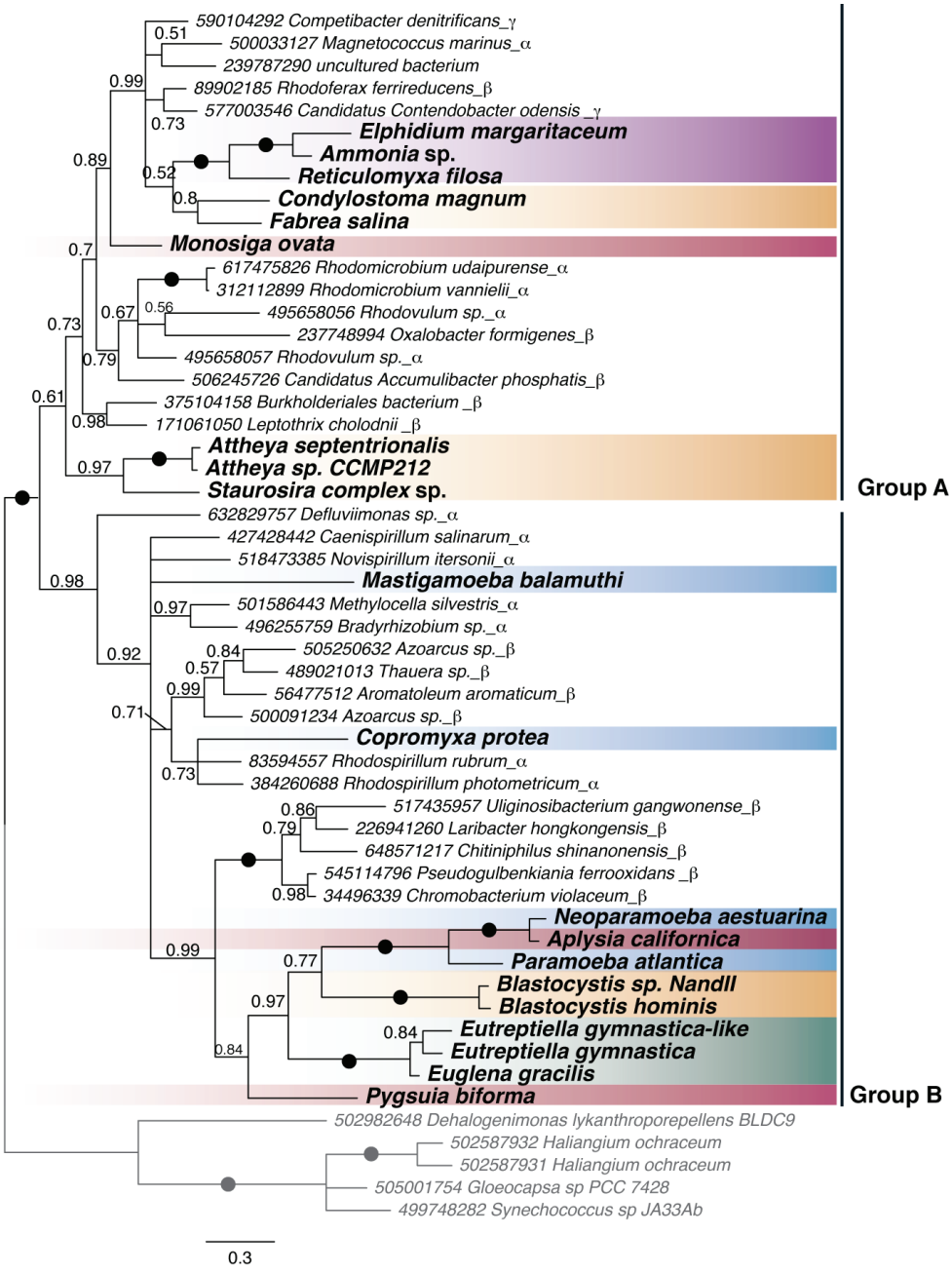

Figure 4-4: Bayesian analysis of eukaryotic and bacterial RQUA sequences. Bipartitions less than 0.5 were collapsed. Organisms are coloured and symbols are as described in Figure 4-3.

4.4.3 Primary Sequence Analysis Of Eukaryotic RQUA

The closest related annotated homologs of the RQUA sequences belong to a family of Class I S-adenosyl methionine (SAM) methyltransferases including the ubiquinone methyltransferases UBIE. A recent survey of methyltransferases identified four distinct motifs common to most class I SAM methyltransferases (motif I, motif post-

I, motif II and motif post-II) that are crucial for protein folding and SAM binding (Katz *et al.* 2003; Petrossian and Clarke 2009). Three of these motifs are conserved in RQUA sequences. However, the SAM-binding motif I (bold text indicates the site that is a glutamine or methionine in RQUA sequences compared to UBI sequences; [VILFG]-[LIVCS]-[**DENLV**]-[VALMIT]-[GLYCFA]-[CSTAYFPAG]-[GA]-[PTSGNKRMV]-[GD]) is different in RQUA sequences when compared to other Class I SAM-methyltransferases (Figure 4-5, open circle). Like the bacterial RQUA sequences discussed by Lonjers and colleagues, the eukaryotic RQUA sequences have substitutions in key SAM binding sites shown in Figure 4-5. While further experimental studies are needed, these substitutions suggest that RQUA does not bind SAM or binds and uses SAM in a different manner than known SAM-dependent methyltransferases (Lonjers *et al.* 2012).

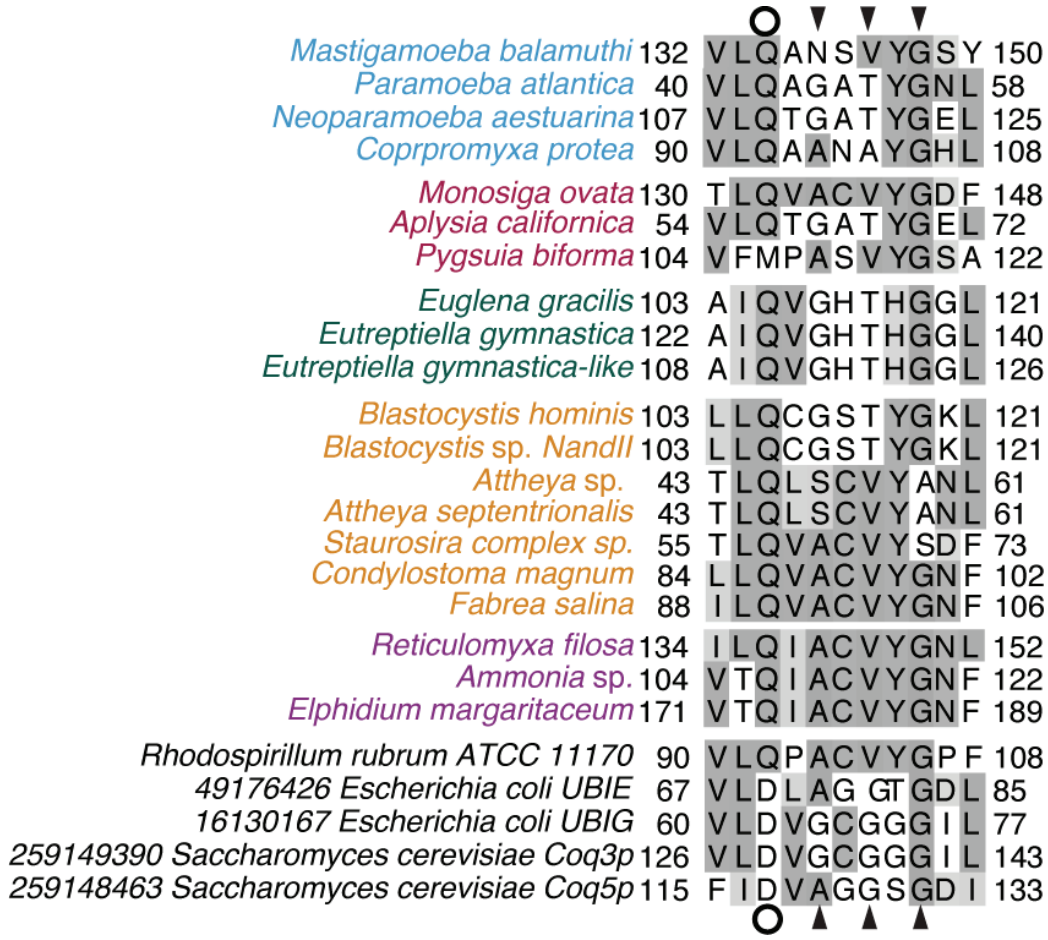

Figure 4-5: An alignment of a selected region of eukaryotic and bacterial RQUA and UBI/COQ homologs representing the SAM binding Motif I (Petrossian and Clarke 2009). RQUA-specific conserved 'signature' substitutions are denoted with open circles (interaction with SAM) and

triangles (interactions with the carbonyl group of SAM). Taxa are coloured by super-group Amoebozoa (blue), Obazoa (red), Excavata (green), heterokonts and alveolates (orange), Rhizaria (purple).

4.4.4 The Distribution Of Quinone-Utilizing Enzymes In Eukaryotes

RQ is known to function as an electron carrier with complex II and other Q-utilizing enzymes. If RQUA is in fact synthesizing RQ in the eukaryotes presented here, then these organisms must encode two types of RQ-utilizing enzymes – those that reduce RQ and others that reoxidize it. To test this hypothesis, I searched for Q-utilizing enzymes (respiratory complexes, CI, CIII, CIV; quinone biosynthesis enzymes, COQ1-7; alternative oxidase, AOX; electron transferring flavoprotein, ETFα and β; and ETF dehydrogenase, ETFDH) in the *rqua*-containing eukaryotic genomes and transcriptomes FIGURE using homologs of these enzymes from *Dictyostelium discodeum* and *Arabidopsis thaliana* as query sequences. Interestingly, most MRO-containing protists (*Pygsuia, Mastigamoeba* and *Blastocystis*) identified to date that encode complex II and at least one other quinone-utilizing enzyme (complex I, AOX or ETF) also have RQUA.

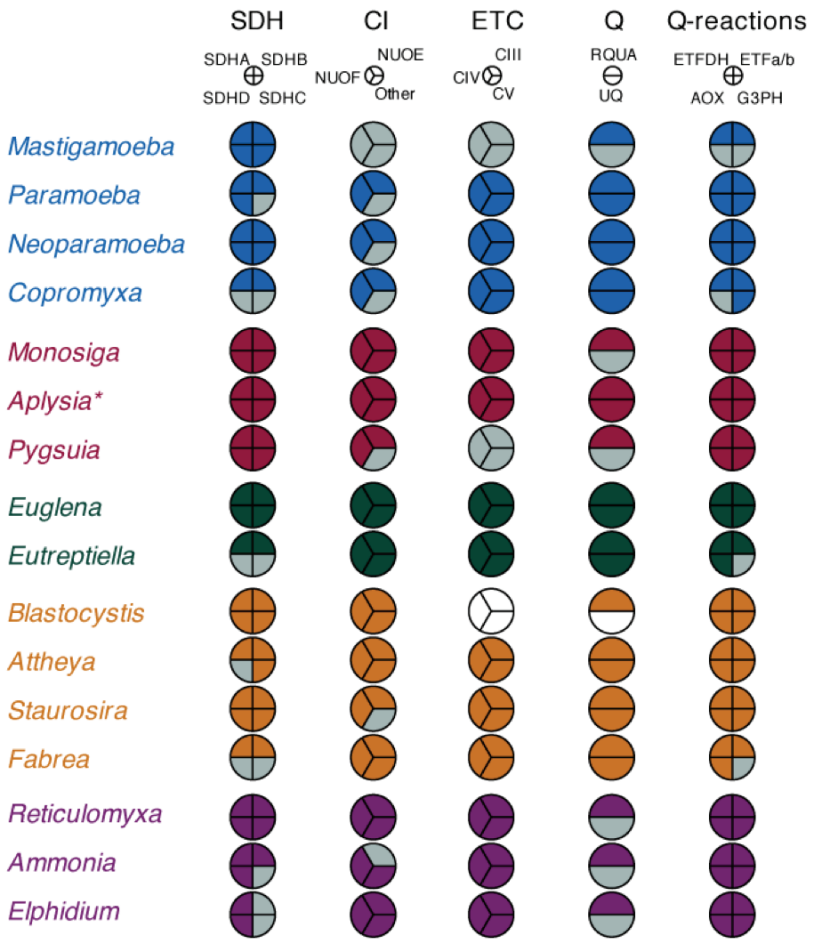

Figure 4-6: Coulson plot of Q-utilizing and Q biosynthesis enzymes in eukaryotic organisms that encode *rqua.* Organisms are coloured based on super-group classification (Amoebozoa, blue; Obazoa, red; Excavata, green; Chromalevolata, organge; and Rhizaria, purple). Presence of homologs of enzymes encoded in the genome of a given organism are indicated with coloured pie wedges. Uncoloured wedges indicate that a homolog of that enzyme was not identified in complete genome (white) or incomplete (grey) sequencing projects. A suspected *Paramoeba* contaminant in the *Aplysia* genome is indicted with an asterisk.

4.4.5 Subcellular Localization Of RQUA In Pygsuia biforma

In most eukaryotes, ubiquinone biosynthesis occurs in the mitochondrion (Robinson 1997; González-Mariscal *et al.* 2014) and since UQ is a known precursor to RQ in *R. rubrum*, it is possible that RQ biosynthesis (and therefore RQUA) would also be localized to the mitochondrion (or MRO). This is supported by my observation that all of the eukaryotic RQUA complete protein sequences possess putative N-terminal targeting peptides for mitochondrial localization as predicted by Mitoprot and TargetP (Table 4-1). To experimentally test these predictions in *Pygsuia biforma*, I used immunofluorescence

confocal microscopy to determine the localization of RQUA. Anti-*Pygsuia biforma* RQUA antibodies co-localized with mitochondrion-reactive stain MitoTracker in *Pygsuia* (Figure 4-7) suggesting this protein functions in the MROs of this organism. Although experimental confirmation of RQUA function is required, it seems probable that, RQ biosynthesis occurs in the mitochondrion (or MRO) of all of the eukaryotes possessing RQUA.

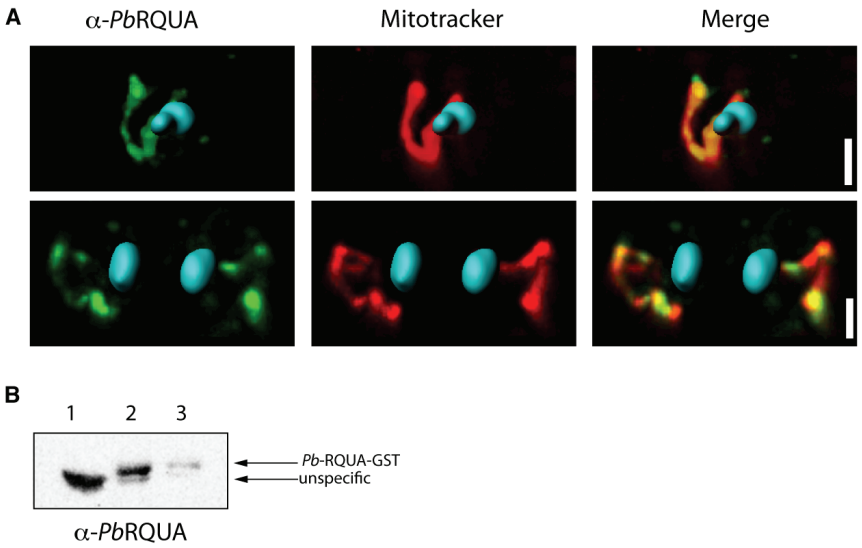

Figure 4-7: Antibodies raised against ASCT and SUFCB localize to *Pygsuia* MROs using immunofluorescence confocal microscopy. A. RQUA (green) co-localized with Mitotracker (red) in *Pygsuia biforma*. Nuclei were stained using DAPI (blue). Confocal slices (0.3 μm) were deconvoluted and combined to render a 3D image. DAPI-stained nuclei (blue) were rendered in Imaris. Scale bars (5 μm). B. A western blot showing that anti-*Pb*RQUA antibodies recognize purified recombinant *Pb*RQUA. Protein extracts were isolated from whole cell extracts of *E. coli* transformed with empty plasmid pGEX4T-1 (1) or pGEX-*Pb*RQUA (2) and cultured on conditions to induce expression of the encoded protein and GST-*Pb*RQUA purified using glutathione magnetic beads (3) were resolved by SDS-PAGE and probed by immunoblotting using anti-*Pb*RQUA antibodies. The estimated molecular weight of *Pygsuia* GST-RQUA is 56 kDa (*i.e.,* 26 kDa for GST plus 30 kDa for RQUA).

4.5 Discussion

4.5.1 Origin Of RQUA And Rhodoquinone Biosynthesis

The ability to perform fumarate reduction using rhodoquinone appears to be an adaptation in some eukaryotes that helped them thrive under low oxygen conditions.

After a thorough examination of publicly available genome and transcriptome sequencing projects, the RQ biosynthesis protein RQUA was identified in multiple lineages of eukaryotes and bacteria. Our preliminary phylogenetic analysis revealed that all RQUA homologs emerge as a monophyletic group from within a larger clade of proteobacterial Class I SAM-dependent methyltransferases (not shown). The closest homologs of RQUA are in fact members of the ubiquinone biosynthesis pathways. Interestingly, previous reports have demonstrated that RQ is synthesized from UQ and not a UQ precursor. Therefore it is possible that RQUA evolved from a family of proteins already capable of binding UQ or similar molecules and eventually gained new enzymatic activity to function in RQ biosynthesis.

Within the eukaryote domain, enzymatic interactions with RQ have only been investigated in detail in *Ascaris suum, Caenorhabditis elegans*, and *Euglena gracilis*. While RQUA does appear to be encoded by *E. gracilis*, RQUA was not identified in the genomes of *A. suum, C. elegans* or closely related helminths and nematodes. This suggests that the RQ biosynthesis pathway of these organisms is not detectably related to the RQUA-based system and they have convergently evolved to the capacity to synthesize and utilize RQ.

4.5.2 Dispersion Of RQUA By Lateral Gene Transfer (LGT)

Closer examination of the RQUA clade revealed that the relationships between bacterial and eukaryotic sequences are not congruent with expected organismal relationships for either the eukaryotes or bacteria (Figure 4-3 and Figure 4-4). If this enzyme was in fact ancestral to, and inherited vertically by, α–, β– and γ–proteobacteria, one would expect to recover monophyly of these bacteria to the exclusion of eukaryotes, or at least recovery of different groups of bacteria according to their phylogenetic affinities (α–, β– and γ–proteobacteria). However, in all phylogenies, the bacterial sequences are not monophyletic but rather distributed amongst eukaryotes (Figure 4-3 and Figure 4-4). Interestingly, some of the bacterial groups do cluster together to the exclusion of eukaryotes and other proteobacteria (*e.g.,* there is a clade of α–proteobacteria encompassing some members of the orders Rhodospirillales

and Rhizobiales BV=64; and a clade of β–proteobacteria of grouping together the orders Neisseriales and Rhodocylales, BV = 100, PP =1.0). The remaining bacterial sequences are spread throughout the tree with their branching patterns that are at odds with presumed bacterial relationships. Recovering 'scrambled' phylogenetic affinities amongst bacterial sequences is relatively common as LGT is known to be a frequently occurring phenomenon in prokaryotic genome evolution (Eisen 2000; Ochman *et al.* 2000; Kunin and Ouzounis 2003).

The 'patchiness' of *rqua* occurrence within bacterial genomes is also consistent with extensive LGT of this gene. If one were to try to explain this *rqua* distribution by vertical inheritance alone, the gene would have had to have been present in the common ancestor of α–, β–, and γ– proteobacteria, and been lost potentially hundreds of times independently in the various members of these orders that lack the enzyme. For example, RQUA is found in 4 out of 13 orders of α-proteobacteria (9/796 genomes), 3 out of 10 orders of β-proteobacteria (13/473 genomes) and 1 unclassified order of γ-proteobacteria (2/3190 genomes) present in Genbank. Therefore, if RQUA was present in the common ancestor of proteobacteria it was lost in all extant δ/ε-proteobacteria (460 genomes), ζ-protebacteria (10) and remaining α–,β–,γ–proteobacteria (42 different orders). Although this scenario is unparsimonious and very unlikely, it cannot be definitively excluded as an explanation.

Similarly, the intra-eukaryote relationships observed in the phylogeny do not reflect typical organismal relationships. However, the fact that the eukaryotic RQUA sequences likely function within mitochondria (Figure 4-7) combined with the broad distribution of the gene in genomes of diverse eukaryotes could indicate that *rqua* was in the genome the common endosymbiotic ancestor of mitochondria (Figure 4-8). If so, and RQUA traces its ancestry to the last common ancestral eukaryote through vertical inheritance alone, one would expect the eukaryote sequences to cluster together to the exclusion of bacterial sequences. However, this is not observed. Instead, the eukaryotic sequences are scattered throughout the tree intermixed with bacterial homologs. For example, the Obazoa sequences (*Monosiga, Pygsuia* and *Aplysia*) are separated by at least one

highly supported branch (BV=86 and PP =0.98) and the stramenopile (ciliates, *Blastocystis* species and diatoms) are separated by multiple highly supported branches (Figure 4-3 and Figure 4-4). Furthermore, the rhizarian (*Ammonia, Reticulomya* and *Elphidium*) and ciliate (*Fabrea* and *Condylostoma*) sequences are nested within a moderately supported clade of α- and γ-proteobacteria in ML and Bayesian analyses (Figure 4-3 and Figure 4-4). These observations suggest that *rqua* was not present in the common endosymbiotic ancestor of mitochondria but instead suggests that the various eukaryote lineages acquired *rqua* multiple times by LGT from bacterial and/or other eukaryotic sources (Figure 4-8A). Following LGT from bacteria, the eukaryotic *rqua* gene sequences must eventually have acquired mitochondrial targeting signals. This scenario is consistent with growing evidence for extensive bacterial-to-eukaryote and eukaryote-to-eukaryote LGT that has been reported in recent years (Andersson and Roger 2003; Hug *et al.* 2010; Takishita *et al.* 2012; Leger *et al.* 2013; Nyvltová *et al.* 2013). Nevertheless, it remains possible (if unlikely) that *rqua* was ancestrally present in eukaryotes and has been lost multiple times independently in the majority of eukaryotic lineages. In this view the 'mixing' of bacterial and eukaryotic taxa in the phylogeny could either be due to multiple events of eukaryote-to-bacterial LGT or phylogenetic artifacts of an unknown source combined with poor phylogenetic resolution.

The idea that the ancestral eukaryote was capable of a broad array of both anaerobic and aerobic functions is not new. The Hydrogen Hypothesis posits that the ancestral eukaryote possessed an α-proteobacterial-derived mitochondrion capable of respiring by oxidative phosphorylation, but also, under anoxic conditions, respiring anaerobically via hydrogen production and/or fumarate reduction (see Chapter 1). If this were true, the ancestral mitochondrion would likely have had RQ and an RQ-synthesizing system such as RQUA (or another enzyme that carries out this function), although it is possible that the ancestral respiratory system could function in both directions without aid of different electron carriers. For instance, the soluble subunits of bovine complex II subunits can be forced to perform fumarate reduction with an artificial electron acceptor (benzyl viologen) *in vitro*, although the electrochemistry of

this reaction is noticeably different than known fumarate reductases (Sucheta *et al.* 1992; Ackrell *et al.* 1993). In any case, it is possible that the common ancestor of eukaryotes that possessed a CII capable of performing both succinate oxidation and fumarate reduction and the electron acceptor/donor might have been UQ under aerobic conditions and RQ under anaerobic conditions. In this scenario, over time, numerous lineages of eukaryotes lost RQUA (or an analogous enzyme) as they adapted to a predominantly aerobic lifestyle and preferentially used UQ.

Arguments against scenarios such as the Hydrogen Hypothesis have been accumulating since it was first proposed (Doolittle 1998a; López-García and Moreira 1999; de Duve 2007; Poole and Penny 2007; Hug *et al.* 2010). These arguments are discussed in greater detail in Chapter 1. Focusing specifically on RQUA, the Hydrogen Hypothesis predicts that eukaryote homologs should most closely be related to those of the α-proteobacteria that gave rise to the mitochondrion. Although determining the true phylogenetic affinities of the α-proteobacterial endosymbiont has been confounded by phylogenetic artifacts (Rodríguez-Ezpeleta and Embley 2012), several affinities have been proposed including the Rickettsiales (Andersson *et al.* 1998; Ferla *et al.* 2013), *Pelagibacter* (Thrash *et al.* 2011), or a deeper-branching position within the α -proteobacteria (Rodríguez-Ezpeleta and Embley 2012). Characterized members of the former two groups of α-proteobacteria completely lack a RQUA homolog and if either is the correct sister group to mitochondria, this observation would make a mitochondrial origin for RQUA less likely.

If RQUA was present in the common ancestor of eukaryotes as predicted by the Hydrogen Hypothesis (Figure 4-8), it is surprising that it was secondarily lost in the ancestors of facultatively anaerobic animals such as sponges, helminthes and nematodes. While it is hypothesized by some that global oxygenation sparked the evolution of animals (Mills and Canfield 2014), recent evidence suggests that basal animal lineages can 'thrive' at 0.5% to 4% of present oxygen levels (Mentel *et al.* 2014; Mills *et al.* 2014). Mills and Canfield (2014) postulate that the ancestor of animals was likely a facultative anaerobe that could at least 'tolerate episodes of complete anoxia'.

Others have also implied that the anaerobic metabolism observed in modern-day animals is a direct 'hold-over' from the ancestors of animals (Mentel *et al.* 2014). Unfortunately these studies failed to derive these hypotheses in light of the Hydrogen Hypothesis and the last common ancestor of eukaryotes (LECA). For instance, if the LECA was a facultative anaerobe (with hydrogen-producing mitochondria, malate dismutation and RQ biosynthesis), and by these accounts the ancestor of animals was also a facultative anaerobe, then why have the mitochondria of anaerobic animals lost the ability to perform these ancestral functions? Why have facultatively anaerobic animals whose extant relatives have RQ (*i.e., A. suum* and *C. elegans*) lost RQUA but then subsequently acquired a different enzyme system for synthesizing RQ? It is possible that hypoxia tolerance was lost early in animal evolution and secondarily developed later, though this directly contradicts the findings and conclusions of the recent studies of sponges (Mentel *et al.* 2014; Mills and Canfield 2014). Nevertheless, the absence of anaerobic genes, such as *rqua* in facultatively anaerobic animal lineages is hard to explain if their anaerobic mitochondrial biochemical capacity is in fact a 'hold-over' from the last eukaryotic common ancestor as proposed by advocates of the Hydrogen Hypothesis (Martin and Müller 1998; Mentel and Martin 2010; Mills and Canfield 2014).

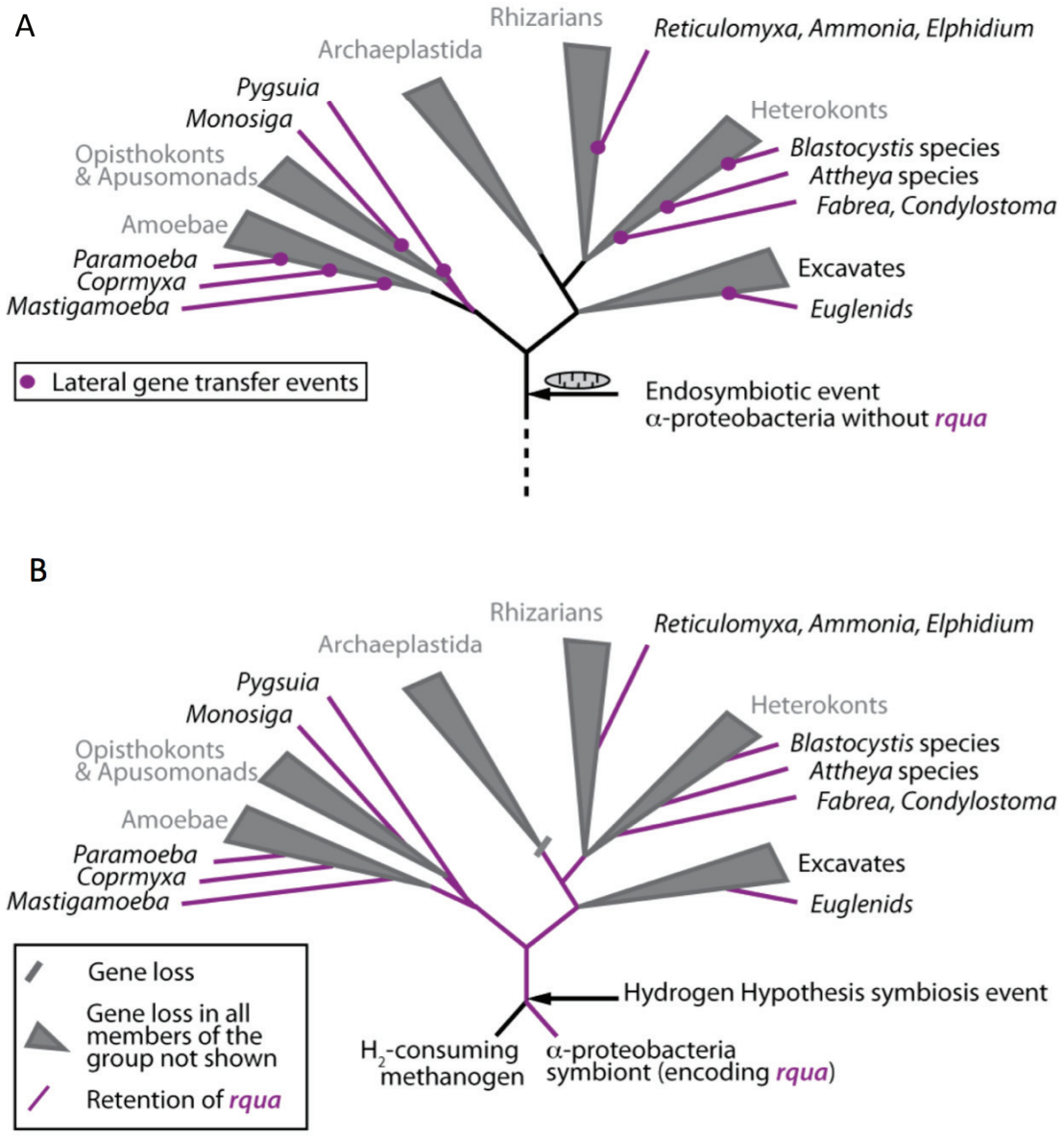

Figure 4-8: Two hypotheses for the origin of eukaryotic RQUA. (A) *rqua* was acquired by lateral gene transfer in only some lineages of eukaryotes as illustrated with purple circles. (B) *rqua* was present in the common ancestor of eukaryotes under the Hydrogen Hypothesis. Lineages where *rqua* was maintained are illustrated with purple lines. Gene loss events are denoted in grey. Grey triangles represent all other gene losses in the extant members of indicated super-groups.

4.5.3 Eukaryotic RQUA Likely Functions In The Mitochondrion

The mitochondrion is the location of the terminal steps of UQ biosynthesis (Wang and Hekimi 2012). Since UQ is known to be a precursor of RQ (Brajcich *et al.* 2010), RQ biosynthesis could also occur in the mitochondrion in eukaryotes. All of the eukaryotic homologs of RQUA with complete N-termini have predicted mitochondrial targeting

signals suggesting mitochondrial localization. These predictions were experimentally validated for *Pygsuia biforma* using immunofluorescence microscopy; RQUA co-localized with mitochondrial marker mitotracker in the MRO these cells (Figure 4-7).

Most of the enzymes that use UQ are also found in mitochondria such as CI, CII, alternative oxidase and the electron transferring flavoprotein system. Interestingly, every eukaryotic organism that encodes RQUA (including anaerobes) encodes CII and at least one other Q-utilizing enzyme. Some of these proteins have been shown to interact with RQ in *Ascaris suum* (Ma *et al.* 1993) and I therefore hypothesize that the Q-utilizing complexes of other RQUA-containing eukaryotes, might also be capable of using RQ as a substrate under certain conditions (*i.e.,* anoxia).

4.5.4 RQUA Function And The 'Transferability' Of RQ Biosynthesis Between Organisms

The conversion of the methoxy group of UQ to the amine of RQ proceeds via an unknown mechanism with an unknown number of intermediates (Brajcich *et al.* 2010; Lonjers *et al.* 2012). Furthermore, the role and enzymatic activity of RQUA in this reaction is currently unknown. Lonjers *et al.* propose three hypotheses for the role of RQUA and RQ biosynthesis. Firstly, they suggest RQUA could serve a regulatory role in the expression or function of RQ biosynthesis proteins. However, my evidence that RQUA functions in mitochondria, makes such a putative role in regulation of nuclear-encoded quinone biosynthesis genes very unlikely.

The second hypothesis is that RQUA evolved a novel function to catalyze the methoxy to amino reaction and therefore directly converts UQ to RQ. This hypothesis is is consistent with the observation that, like the closely-related ubiquinone biosynthesis enzymes (UBIE), RQUA can likely bind and interact with quinone species. Furthermore, the residues involved in SAM binding are not conserved in RQUA sequences suggesting that RQUA functions in a different manner than UBIE proteins.

Lonjers and colleagues final hypothesis is that UQ and RQ are synthesized by different multi-enzyme complexes in *R. rubrum* that share *some* components but not all. Therefore the removal of one of these components (RQUA) results in a non-functional

complex in their *rqua* knockout experiments (Lonjers *et al.* 2012). This is based on the observation that some components of the yeast ubiquinone biosynthesis complex (COQ3, 4, 5, 6, 7 and 9) are necessary for stabilization of the complex independent of their catalytic role (Baba *et al.* 2004). If this were the case for RQ biosynthesis, one would expect all organisms that encoded RQUA would also have to encode UQ biosynthesis proteins. However, many of the eukaryotes that encode RQUA do not encode any component of the ubiquinone biosynthesis pathway (Figure 4-6). These organisms either synthesize UQ by an unknown mechanism or rely on exogenous sources of UQ (Padilla-López *et al.* 2009). Therefore, it is unlikely that RQUA serves to stabilize the ubiquinone biosynthesis complex in these organisms.

In the future, phylogenomic profiling (for examples see (Rodionov and Gelfand 2005; Wu *et al.* 2006)) of genomes encoding RQUA and additional genome-wide mutation screens will be essential for validating these hypotheses. I have also established a heterologous expression system in *Rhodospirillum rubrum rqua* mutants (incapable of synthesizing RQ) to test if eukaryotic *rqua* can complement RQ biosynthesis (data not shown). Enzyme assays to determine the exact function of RQUA (*i.e.,* methyltransferase or aminotransferase) will be necessary to understand how RQUA convert UQ to RQ.

In summary, my data is most consistent with the second hypothesis of Lonjers and colleagues that RQUA directly converts UQ to RQ in one step. If true, this explains why RQUA is not genetically linked to (thus not expressed in an operon with) any other quinone biosynthetic enzyme in bacteria. Furthermore, it would make the acquisition of RQ biosynthesis by LGT relatively straightforward – only a single gene need be acquired by an organism to allow it to convert UQ to RQ.

One obstacle to overcome is the potential for futile cycling of electrons by CII in organisms that have UQ and RQ. To help regulate the directionality of CII, *Euglena* and *Caenorhabditis* maintain different relative levels of UQ and RQ under aerobic (higher UQ) and anaerobic (higher RQ) conditions (Takamiya *et al.* 1999; Hoffmeister *et al.* 2004). Interestingly, *Ascaris suum* encodes stage-specific isoforms of CII subunits that

favour fumarate reduction, specifically the succinate/fumarate binding flavoprotein subunit (SDHA) and the small cytochrome b subunit (SDHD) (Amino *et al.* 2003; Iwata *et al.* 2008). Future investigations are needed to determine which mechanism is used by the protists presented here.

4.6 Conclusions

The ability to synthesize RQ using RQUA was likely a prokaryotic invention that was transferred to eukaryotes by multiple independent events of LGT well after the establishment of mitochondria within eukaryotes. This is supported by the scrambled relationships observed between eukaryotes and bacteria. RQUA has then been transferred between eukaryotes via eukaryote-eukaryote LGT. After establishing RQ biosynthesis in mitochondria or related organelles, complex II and other Q-utilizing enzymes of these eukaryotes changed over time to use RQ for the purposes of fumarate reduction allowing the eukaryotes to perform respiration independent of oxygen. This is an example of how laterally acquired functions can interface with ancestral functions of mitochondria to create metabolisms of mosaic origins.

Rhodoquinone:fumarate oxidoreduction is one of many ways anaerobic organism tolerate low oxygen environments. However, as more organisms are studied, it is becoming increasingly difficult to test for various enzymatic activities in a high throughput manner. Therefore, the identification of proteins that synthesize cofactors important for anaerobic metabolism (such as RQ) could one day serve as an *in silico* proxy for enzyme activity (*e.g.,* fumarate reductase), thus avoiding tedious experimental approaches to determine metabolic potential.

Chapter 5 Conclusions

The previous chapters highlight both the incredible diversity yet also the extensive similarities in metabolism of mitochondrion-related organelles across the tree of eukaryotes. Although MROs have evolved independently in dozens of lineages, obvious convergences in function have occurred. Several different enzymes appear to co-occur in different anaerobic lineages across the tree of eukaryotes and behave almost as functional and evolutionary modules. While there are obvious exceptions to these observations, some common modules include: (i) PFO and hydrogenase (ii) PFO and hydrogenase with a complete or partial complex I, (iii) ASCT-mediated ATP generation accompanying SCS, or (iv) the presence of an ACS. Using these modules of laterally acquired and ancestral proteins, the MROs of anaerobic protists have convergently acquired similar mechanisms to oxidize pyruvate, generate hydrogen and synthesize ATP. However, the exact sequences of events that resulted in the retailoring of the aerobic mitochondria to the reduced organelles I observe today are unknown. Below I outline possible evolutionary pathways that could explain the spectrum of extant organelle diversity.

5.1 A Few Assumptions

The endosymbiotic theory for the origin of mitochondria suggests that the bacteria that gave rise to mitochondria were alpha-proteobacteria capable of aerobic respiration. In the following hypothesis, I will make the uncontroversial assumption that the mitochondria of the last eukaryotic common ancestor (LECA) had the hallmark features of aerobic mitochondria (*i.e.,* similar to aerobic mammalian mitochondria) capable of (i) pyruvate oxidation by pyruvate dehydrogenase, (ii) electron transport and oxidative phosphorylation with oxygen as the terminal electron acceptor, and posseses (iii) mitochondrial DNA.

5.2 Reduction Of The Electron Transport Chain In Response To Hypoxia

A first major step in the evolution of anaerobic MROs was the adaptation of mitochondrion-containing protists to be able to continue to make ATP when oxygen levels were low. Similar to metazoans like *Ascaris,* low oxygen conditions would likely lead to down-regulation and reduced efficiency of Complex III and Complex IV. If these protists eventually adapted to a predominantly anaerobic existence (*e.g.,* perhaps exploiting nutrient-rich anaerobic environments such as marine or intertidal sediments or animal guts), there would be less purifying selection on genes encoding the subunits of CIII and CIV; with neutral mutational drift these were lost from the mitochondrial and nuclear genomes. In this scenario, the first half of the TCA cycle was still functioning to produce succinate (starting at citrate) and NADH. The remaining part of the TCA cycle could be functioning in reverse in a malate dismutation pathway where malate is imported into the organelle and converted to fumarate. Complex I consumed NADH to pump protons to maintain a proton gradient for ATP synthesis via ATP synthase. However, because of the lack of CIII and CIV, there would have been a build-up of reduced ubiquinol generated by Complex I. This buildup of ubiquinol could favour the functioning of CII in the reverse direction as a fumarate reductase to regenerate ubiquinone. This is plausible as some aerobic CIIs can be forced to perform fumarate reduction with sufficient concentrations of substrate (Ackrell *et al.* 1993; Sakai *et al.* 2012). Thus, succinate – and not water – would be the end point of metabolism and likely excreted from the organelle via a succinate/fumarate antiporter (analogous to anaerobically-upregulated dicarboxylic acid antiporters in *E. coli* (Engel *et al.* 1992), although there are no discernable homologs of this kind in eukaryotic genomes). This organelle would partially resemble the mitochondria of some animals (*e.g., Ascaris*) under anaerobic conditions, except the genes for CIII and CIV subunits would have been completely lost. In some lineages of eukaryotes, the ability to synthesize and use rhodoquinone (RQ) by the acquisition of RQUA or analogous protein by LGT could greatly increase the efficiency of the fumarate reductase activity of CII.

5.3 Changes To Pyruvate And ATP Generation

Perhaps concomitantly with the evolutionary 'remodeling' of the respiratory chain, the pyruvate oxidation system was also changed in this protist adapting to anoxic conditions. If the ancestor of mitochondria only possessed one pyruvate-metabolizing enzyme (the PDC), during the transition to anaerobiosis, pyruvate:ferredoxin oxidoreductase and hydrogenase must have been acquired by LGT at some point and co-existed with PDH. Some MRO-containing organisms still encode complete or partial PDH complexes in addition to PFO and/or PNO (*e.g., Acanthamoeba castellanii* (Leger *et al.* 2013) *Blastocystis* sp., (Stechmann *et al.* 2008), *Nyctotherus ovalis* (Boxma and Graaf 2005)). However, under highly reductive conditions (elevated NADH levels due to an inefficient electron transport chain missing CIII and CIV) the ancestral mitochondrial PDC would likely be inhibited (Bremmer 1969; Kerbey *et al.* 1976). In fact, in rat liver mitochondria, the K_i of PDC for NADH is lower than the K_m of PDC for NAD+ (Bremmer 1969) suggesting that PDH is very sensitive to NADH/NAD+ ratios. This would decrease the energy output of MROs under low oxygen conditions similar to the mitochondria of patients with PDC deficiency (Brown *et al.* 1994). In turn this might limit the efficiency of pathways in mitochondria that rely on acetyl-CoA as a substrate (*e.g.,* some long-chain fatty acid biosynthesis pathways (Hiltunen *et al.* 2010)). Under these conditions, there would be a selective advantage to acquiring and maintaining an NADH-tolerant pyruvate oxidizing enzyme such as PFO instead of PDC. Once acquired, as long as the organism was living under predominantly hypoxic conditions (as PFO is an oxygen-sensitive enzyme (Chabrière *et al.* 1999)), PDC would be unnecessary. The multiple subunits of PDH and its allosteric regulatory PDC kinase (Patel and Korotchkina 2006) would be energetically 'expensive' to express and the loss of these genes would be selectively advantageous. PFO will also generate reduced ferredoxin that can be used by a variety of mitochondrial processes (e.g Fe-S cluster biosynthesis; (Lange *et al.* 2000)). However, if the organism acquired hydrogenase, the oxidized ferredoxin could be regenerated for PFO independently of other mitochondrial processes.

At this point, the mitochondrion has a partially functioning electron transport chain (CI and CII), PFO, HYD and partial TCA cycle, mtDNA (encoding components of CI and probably CV), it synthesizes ATP via complex V and ultimately excretes succinate as the end product of metabolism. The next enzyme to be acquired could have been an acetate:succinyl-CoA transferase which could, with acetyl-CoA convert succinate (which is now in abundance) to succinyl-CoA. The succinyl-CoA would be used as a substrate for succinyl-CoA synthetase to generate a trinucleotide (ATP or GTP). This new method of ATP biosynthesis possibly allowed for the eventual loss of ATP synthase (and the corresponding genes encoded on the mtDNA and nuclear genomes). The resulting organelle would then resemble the MROs observed in *Blastocystis* species (Stechmann *et al.* 2008).

5.4 Loss Of CI, mtDNA And Malate Dismutation

Phylogenetic analysis of hydrogenases in Chapter 3 (Figure 3-3) revealed that some species of the *Thermotogales* branch within the larger clade of eukaryotic hydrogenases. Interestingly, studies have shown that the *Thermotoga* hydrogenase can only accept electrons from co-operation of Complex I and ferredoxin and not ferredoxin alone (Schut and Adams 2009). This suggests that the *Thermotoga* hydrogenase interacts with at least some components of Complex I. This is consistent with studies of *Trichomonas vaginalis* hydrogenosomes that showed that two retained components of CI, the 24 kDa and 51 kDa subunits (responsible for oxidation of NADH) interact with ferredoxin and hydrogenase probably to produce H_2 gas (Dyall *et al.* 2004; Hrdy *et al.* 2004). Interestingly, other anaerobic protists including *Sawyeria marylandensis* and *Pygsuia biforma* appear to have retained these same two components of Complex I suggesting that a *Thermotoga*-like coupling of NADH oxidation and H_2 production could also be present in these organisms (Hrdy *et al.* 2004; Barberà *et al.* 2010; Stairs *et al.* 2014).

In light of these observations, it seems likely that once hydrogenase was acquired by an anaerobic protistan lineage, it is possible these CI subunits could have had dual roles. When bound to the membrane-embedded CI subunits, they carried out their normal NADH oxidation reaction and passed electrons to CI subunits that reduced

ubiquinone and pumped proteins in the process. However, when free in the mitochondrial matrix they could alternatively form this specialized Complex I-ferredoxin-hydrogenase complex, ultimately generating hydrogenase gas. It appears that in some lineages, at least, the proton pumping and ubiquinol reducing subunits of CI were completely lost along with ATP synthase, and only the subunits involved in NADH dehydrogenase activity were retained. It is unclear why most of CI and all of ATP synthase would be lost, although it could be related to the fact that if CI was the only ETC complex present to pump protons (and part of the Kreb's cycle running backwards and therefore consuming rather than producing NADH), the proton motive force may have been weak and the ATP yield insufficient to 'cover' the costs of expressing the large numbers of subunits required for complete CI and ATP synthase activity.

In any case, the resulting organelle resembles the organelle I described in Chapter 3 for *Pygsuia biforma* – an organelle lacking a mitochondrial genome, with only CII, two CI subunits, PFO, Hydrogenase, ASCT and SCS. Eventually, the CII (and thus malate dismutation) could be lost if there were an alternative route for generating succinate to be used as a substrate for ASCT. This would produce an organelle resembling the hydrogenosomes of *Trichomonas vaginalis* (although I currently do not know how succinate is formed or imported into the *T. vaginalis* organelle). It is possible that instead of acquiring ASCT, some lineages acquired ACS to generate ATP directly from acetyl-CoA in their MROs, such as *Mastigamoeba balamuthi* (Nyvltova *et al.*, under review) or *Spironucleus salmoncida* (Jerlström-Hultqvist *et al.* 2013). Interestingly, *Mastigamoeba* has a partial malate dismutation pathway; the genome encodes only CII (and not ASCT, SCS or fumarase). This suggests that the *Mastigamoeba* MRO might represent a transition state in between an anaerobically respiring organelle and a *Spironucleus*-like organelle.

5.5 A Note On Fe-S Biosynthesis

It is difficult to speculate on how changes to the Fe-S cluster biosynthesis function of MROs (*i.e.,* SUF in the MROs of *Pygsuia* or NIF in the MROs of *Mastigamoeba*) relate to other anaerobic adaptations of these organelles. In the case of SUF, the ancestor of *Pygsuia's* must have had both SUF and ISC at one point in its history since ISC was directly inherited from the α-proteobacterial ancestor of mitochondria (Lill 2009). SUF systems are more tolerant to oxygen-stress conditions than ISC systems (Outten *et al.* 2004; Wollers *et al.* 2010). Therefore, if the ancestor was already handicapped with oxygen metabolism (*e.g.,* malfunctioning electron transport chain resulting in increased reactive oxygen species) during the transition to anaerobioisis, the SUF system may have functioned better than the ISC system. This would result in a selective advantage for losing the ISC system to use the SUF system exclusively.

5.6 Final Remarks

Above, I have discussed just a few of many possible scenarios that could lead to the metabolic diversity I observe in anaerobic protists today. The importance of continuing to study free-living organisms is two-fold. Firstly, discovering new biochemical modules and combinations of these modules is crucial for understanding the evolution of mitochondria in response to anaerobic conditions. The preceding chapters relied heavily on the idea that anaerobic metabolism was not in the common ancestor of eukaryotes, but was, instead, acquired gradually over evolutionary time independently in different lineages. It is possible that by studying new organisms, the discovery of new metabolic arrangements could strengthen or weaken this hypothesis.

Secondly, studying the anaerobic metabolism of free-living protists is fundamental for understanding metabolism in parasites. By comparing the metabolisms of parasites and their closely related free-living relatives, I can disentangle the exact adaptations that arose in response to anaerobiosis versus those involved with parasitism. This can allow for more targeted research towards developing 'parasitism-

specific' chemotherapeutics and gaining a better understanding of parasite biology in low oxygen environments.

The emerging pattern I observe is that free-living organisms tend to maintain more anabolic pathways than parasites that presumably rely on the host to fulfill these needs. For instance, the organelles of *Pygsuia* have retained a variety of lipid and amino acid metabolic pathways not found in protistan parasites like *Trichomonas, Giardia* or *Entamoeba*. Furthermore, free-living organisms often have different enzymes to perform redundant functions whereas parasites have more streamlined metabolisms. For example *Pygsuia* encodes PFO, pyruvate formate lyase and pyruvate:NADP oxidoreductase for the generation of acetyl-CoA whereas organisms like *Trichomonas*, *Entameoba* or *Giardia* only encode PFO. This observation is slightly harder to reconcile with respect to dependence on the host. It is possible that since free-living organisms can occupy dynamic environments (compared to parasites) there is a selective advantage to having a versatile metabolism. However, this is mostly speculation; more information on the regulation of these various enzymes under different conditions is necessary to determine how they might allow these organisms to adapt to changing environments.

The preceding chapters demonstrate the crucial role of lateral gene transfer in adaptation to hypoxia. Chapter 2 demonstrated how LGT could introduce a non-oxidative acetyl-CoA generation system (PFL) that allows an organism to make acetyl-CoA from pyruvate even under highly reducing conditions. Chapter 3 demonstrated that even near-universally conserved functions of mitochondria such as Fe-S cluster biogenesis can be affected by LGT. Finally Chapter 4 showed that the ability to synthesize specialized cofactors can be introduced by LGT allowing an organism to perform novel functions in hypoxia. As more non-model organisms are studied, I suspect that it will become clearer how LGT serves as an important driving force shaping hypoxia-adaptation in eukaryotes.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105.

Ackrell BAC, Armstrong FA, Cochran B, Sucheta A, Yu T. 1993. Classification of fumarate reductases and succinate dehydrogenases based upon their contrasting behaviour in the reduced benzylviologen/fumarate assay. FEBS Lett. 326:92–94.

Adosraku RK, Smith JD, Nicolaou a, Gibbons W a. 1996. *Tetrahymena thermophila*: analysis of phospholipids and phosphonolipids by high-field 1H-NMR. Biochim. Biophys. Acta 1299:167–174.

Akaike H. 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19:716–723.

Akhmanova A, Voncken FG, Hosea KM, Harhangi H, Keltjens JT, op den Camp HJ, Vogels GD, Hackstein JH. 1999. A hydrogenosome with pyruvate formate-lyase: anaerobic chytrid fungi use an alternative route for pyruvate catabolism. Mol. Microbiol. 32:1103–1114.

Ali V, Shigeta Y, Tokumoto U, Takahashi Y, Nozaki T. 2004. An intestinal parasitic protist, *Entamoeba histolytica,* possesses a non-redundant nitrogen fixation-like system for iron-sulfur cluster assembly under anaerobic conditions. J. Biol. Chem. 279:16863–16874.

Allison DS, Schatz G. 1986. Artificial mitochondrial presequences. Proc. Natl. Acad. Sci. 83:9011–9015.

Altschul S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Amberg DC, Burke D, Strathern JN. 2005. Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual. CSHL Press

Amino H, Osanai A, Miyadera H, Shinjyo N, Tomitsuka E, Taka H, Mineki R, Murayama K, Takamiya S, Aoki T, et al. 2003. Isolation and characterization of the stage-specific cytochrome b small subunit (CybS) of *Ascaris suum* complex II from the aerobic respiratory chain of larval mitochondria. Mol. Biochem. Parasitol. 128:175–186.

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, et al. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Andersson JO, Roger AJ. 2002. Evolutionary analyses of the small subunit of glutamate synthase: gene order conservation, gene fusions, and prokaryote-to-eukaryote lateral gene transfers. Eukaryot. Cell 1:304–310.

Andersson JO, Roger AJ. 2003. Evolution of glutamate dehydrogenase genes: evidence for lateral gene transfer within and between prokaryotes and eukaryotes. BMC Evol. Biol. 3:14.

Andersson JO. 2009. Gene transfer and diversification of microbial eukaryotes. Annu. Rev. Microbiol. 63:177–193.

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UC, Podowski RM, Näslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396:133–140.

Ankarklev J, Jerlström-Hultqvist J, Ringqvist E, Troell K, Svärd SG. 2010. Behind the smile: cell biology and disease mechanisms of *Giardia* species. Nat. Rev. Microbiol. 8:413–422.

Ansell R, Granath K, Hohmann S, Thevelein JM, Adler L. 1997. The two isoenzymes for yeast NAD+-dependent glycerol 3-phosphate dehydrogenase encoded by *GPD1* and *GPD2* have distinct roles in osmoadaptation and redox regulation. EMBO J. 16:2179–2187.

Arkblad EL, Egorov M, Shakhparonov M, Romanova L, Polzikov M, Rydström J. 2002. Expression of proton-pumping nicotinamide nucleotide transhydrogenase in mouse, human brain and *C. elegans*. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 133:13–21.

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86.

Atteia A, van Lis R, Gelius-Dietrich G, Adrait A, Garin J, Joyard J, Rolland N, Martin W. 2006. Pyruvate formate-lyase and a novel route of eukaryotic ATP synthesis in *Chlamydomonas* mitochondria. J. Biol. Chem. 281:9909–9918.

Baba SW, Belogrudov GI, Lee JC, Lee PT, Strahan J, Shepherd JN, Clarke CF. 2004. Yeast *Coq5* C-methyltransferase is required for stability of other polypeptides involved in coenzyme Q biosynthesis. J. Biol. Chem. 279:10052–10059.

Barbera MJ, Ruiz-Trillo I, Leigh J, Hug L, Roger AJ. 2007. The diversity of mitochondria-related organelles amongst eukaryotic microbes. In: Martin WF, Müller M, editors. 1st ed. Origin of mitochondria and hydrogenosomes. Berlin ; New York: Springer. p. 239. B

arberà MJ, Ruiz-Trillo I, Tufts JYA, Bery A, Silberman JD, Roger AJ. 2010. *Sawyeria marylandensis* (Heterolobosea) has a hydrogenosome with novel metabolic properties. Eukaryot. Cell 9:1913–1924.

Baudisch B, Langner U, Garz I, Klösgen RB. 2014. The exception proves the rule? Dual targeting of nuclear-encoded proteins into endosymbiotic organelles. New Phytol. 201:80–90.

Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H. 2010. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. Mol. Biol. Evol. 27:1698–1709.

Bonen L, Cunningham RS, Gray MW, Doolittle WF. 1977. Wheat embryo mitochondrial 18S ribosomal RNA: evidence for its prokaryotic nature. Nucleic Acids Res. 4:663–671.

Boxma B, Graaf R de. 2005. An anaerobic mitochondrion that produces hydrogen. Nature 434:74–79.

Boxma B, Voncken F, Jannink S, van Alen T, Akhmanova A, van Weelden SW, van Hellemond JJ, Ricard G, Huynen M, Tielens AG, et al. 2004. The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate:formate lyase and an alcohol dehydrogenase E. Mol. Microbiol. 51:1389–1399.

Bozner P. 1996. The heat shock response and major heat shock proteins of *Tritrichomonas mobilensis* and *Tritrichomonas augusta*. J. Parasitol. 82:103–111.

Brajcich BC, Iarocci AL, Johnstone L a G, Morgan RK, Lonjers ZT, Hotchko MJ, Muhs JD, Kieffer A, Reynolds BJ, Mandel SM, et al. 2010. Evidence that ubiquinone is a required intermediate for rhodoquinone biosynthesis in *Rhodospirillum rubrum*. J. Bacteriol. 192:436–445.

Bremmer J. 1969. Pyruvate Dehydrogenase, Substrate Specificity and Product Inhibition. Eur. J. Biochem. 8:636–640.

Bricker DK, Taylor EB, Schell JC, Orsak T, Boutron A, Chen Y-C, Cox JE, Cardon CM, Van Vranken JG, Dephoure N, et al. 2012. A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, *Drosophila*, and humans. Science 337:96–100.

Broers CA, Stumm CK, Vogels GD, Brugerolle G. 1990. *Psalteriomonas lanterna* gen. nov., sp. nov., a free-living amoeboflagellate isolated from freshwater anaerobic sediments. Eur. J. Protistol. 25:369–380.

Brown GK, Otero LJ, LeGris M, Brown RM. 1994. Pyruvate dehydrogenase deficiency. J. Med. Genet. 31:875–879.

Brown MW, Sharpe SC, Silberman JD, Heiss AA, Lang BF, Simpson AGB, Roger AJ. 2013. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. Proc. Biol. Sci. 280:20131755.

Brugerolle G, Patterson D. 1997. Ultrastructure of *Trimastix convexa hollande*, an amitochondriate anaerobic flagellate with a previously undescribed organization. Eur. J. Protistol. 33:121–130.

Bui ET, Johnson PJ. 1996. Identification and characterization of [Fe]-hydrogenases in the hydrogenosome of *Trichomonas vaginalis*. Mol. Biochem. Parasitol. 76:305–310.

Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol. Evol. 5:418–438.

Burki F, Corradi N, Sierra R, Pawlowski J, Meyer GR, Abbott CL, Keeling PJ. 2013. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in rhizaria. Curr. Biol. 23:1541–1547.

Butler JA, Mishur RJ, Bokov AF, Hakala KW, Weintraub ST, Rea SL. 2012. Profiling the anaerobic response of *C. elegans* using GC-MS. PLoS One 7:e46140.

Calvo SE, Mootha VK. 2010. The mitochondrial proteome and human disease. Annu. Rev. Genomics Hum. Genet. 11:25–44.

Castro-Guerrero NA, Jasso-Chávez R, Moreno-Sánchez R. 2005. Physiological role of rhodoquinone in *Euglena gracilis* mitochondria. Biochim. Biophys. Acta 1710:113–121.

Cavalier-Smith T, Chao EE. 1996. Molecular phylogeny of the free-living archezoan *Trepomonas agilis* and the nature of the first eukaryote. J. Mol. Evol. 43:551–562.

Cavalier-Smith T. 1983. A 6-kingdom classification and a unified phylogeny. Endocytobiol. II:1027–1034.

Cavalier-Smith T. 1989. Archaebacteria and Archezoa. Nature 339(6220):100-101.
Chabrière E, Charon MH, Volbeda A, Pieulle L, Hatchikian EC, Fontecilla-Camps JC. 1999. Crystal structures of the key anaerobic enzyme pyruvate:ferredoxin oxidoreductase, free and in complex with pyruvate. Nat. Struct. Biol. 6:182–190.
Chavez LA, Balamuth W, Gong T. 1986. A Light and Electron Microscopical Study of a New, Polymorphic Free-Living Amoeba, *Phreatamoeba balamuthi* n. g., n. sp. J. Eukaryot. Microbiol. 33:397–404.
Claros MG, Vincens P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. Fed. Eur. Biochem. Soc. J. 241:779–786.
Cometa I, Schatz S, Trzyna W, Rogerson A. 2011. Tolerance of naked amoebae to low oxygen levels with an emphasis on the genus *Acanthamoeba*. Acta Protozool. 50:33–40.
Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, Desany B, Egholm M, Hutchison S, Pettis JS, Lipkin WI, et al. 2009. Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. PLoS Pathog. 5:e1000466.
Corradi N, Akiyoshi DE, Morrison HG, Feng X, Weiss LM, Tzipori S, Keeling PJ. 2007. Patterns of genome evolution among the microsporidian parasites Encephalitozoon cuniculi, *Antonospora locustae* and *Enterocytozoon bieneusi.* PLoS One 2:e1277.
Crane BR, Siegel LM, Getzoff ED. 1995. Sulfite reductase structure at 1.6 A: evolution and catalysis for reduction of inorganic anions. Science 270:59–67.
Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. 10:210.
Denoeud F, Roussel M, Noel B, Wawrzyniak I, Silva C Da, Diogon M, Viscogliosi E, Brochier-armanet C, Couloux A, Poulain J, et al. 2011. Genome sequence of the stramenopile *Blastocystis* , a human anaerobic parasite. Genome Biol. 12:R29.
Derelle E, Ferraz C, Rombauts S, Rouzé P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynié S, Cooke R, et al. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc. Natl. Acad. Sci. U. S. A. 103:11647–11652.
Dolezal P, Likic V, Tachezy J, Lithgow T. 2006. Evolution of the molecular machines for protein import into mitochondria. Science 313:314–318.
Dolgikh V V, Senderskiy I V, Pavlova O a, Naumov AM, Beznoussenko G V. 2011. Immunolocalization of an alternative respiratory chain in *Antonospora (Paranosema) locustae* spores: mitosomes retain their role in microsporidial energy metabolism. Eukaryot. Cell 10:588–593.
Doolittle WF. 1998a. A paradigm gets shifty. Nature 392:15–16.
Doolittle WF. 1998b. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. 14:307–311.
De Duve C. 2007. The origin of eukaryotes: a reappraisal. Nat. Rev. Genet. 8:395–403.
Dyall SD, Yan W, Delgadillo-correa MG, Lunceford A, Loo JA, Clarke CF, Johnson PJ. 2004. Non-mitochondrial complex I proteins in a hydrogenosomal oxidoreductase complex. 28:1103–1107.
Eddy SR. 1998. Profile hidden Markov models. Bioinformatics 14:755–763.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Eisen J a. 2000. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. Curr. Opin. Genet. Dev. 10:606–611.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300:1005–1016.

Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. Nature 440:623–630.

Emelyanov V V, Goldberg A V. 2011. Fermentation enzymes of *Giardia intestinalis,* pyruvate:ferredoxin oxidoreductase and hydrogenase, do not localize to its mitosomes. Microbiology 157:1602–1611.

Emelyanov V V. 2003. Common evolutionary origin of mitochondrial and rickettsial respiratory chains. Arch. Biochem. Biophys. 420:130–141.

Engel P, Kramer R, Undenl AG. 1992. Anaerobic Fumarate Transport in *Escherichia coli* by an fir-Dependent Dicarboxylate Uptake System Which Is Different from the Aerobic Dicarboxylate Uptake System.

Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package).

Ferla MP, Thrash JC, Giovannoni SJ, Patrick WM. 2013. New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. PLoS One 8:e83383.

Field HI, Coulson RMR, Field MC. 2013. An automated graphics tool for comparative genomics: the Coulson plot generator. BMC Bioinformatics 14:141.

Finlay BJ, Embley TM, Fenchel T. 1993. A new polymorphic methanogen, closely related to *Methanocorpusculum parvum*, living in stable symbiosis within the anaerobic ciliate *Trimyema* sp. J. Gen. Microbiol. 139:371–378.

Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, et al. 2006. Pfam: clans, web tools and services. Nucleic Acids Res. 34:D247–D251.

Föll RL, Pleyers A, Lewandovski GJ, Wermter C, Hegemann V, Paul RJ. 1999. Anaerobiosis in the nematode *Caenorhabditis elegans*. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 124:269–280.

Frey M, Rothe M, Wagner A, Knappe J. 1994. Adenosylmethionine-dependent synthesis of the glycyl radical in pyruvate formate-lyase by abstraction of the glycine C-2 pro-S hydrogen atom. Studies of [2H]glycine-substituted enzyme and peptides homologous to the glycine 734 site. J. Biol. Chem. 269:12432.

Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, et al. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. Cell 140:631–642.

García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28:2678–2679.

Gawryluk RMR, Chisholm KA, Pinto DM, Gray MW. 2014. Compositional complexity of the mitochondrial proteome of a unicellular eukaryote (*Acanthamoeba castellanii*,

supergroup Amoebozoa) rivals that of animals, fungi, and plants. J. Proteomics 109C:400–416.

Geli V, Glick B. 1990. Mitochondrial protein import. J. Bioenerg. Biomembr. 22:725–751.

Gelius-Dietrich G, Henze K. 2004. Pyruvate formate lyase (PFL) and PFL activating enzyme in the chytrid fungus *Neocallimastix frontalis*: a free-radical enzyme system conserved across divergent eukaryotic lineages. J. Eukaryot. Microbiol. 51:456–463.

Germot A, Philippe H, Le Guyader H. 1996. Presence of a mitochondrial-type 70-kDa heat shock protein in *Trichomonas vaginalis* suggests a very early mitochondrial endosymbiosis in eukaryotes. Proc. Natl. Acad. Sci. 93:14614–14617.

Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2:31–34.

Van der Giezen M, Cox S, Tovar J. 2004. The iron-sulfur cluster assembly genes iscS and iscU of *Entamoeba histolytica* were acquired by horizontal gene transfer. BMC Evol. Biol. 4:7.

Van der Giezen M, Rechinger KB, Svendsen I, Durand R, Hirt RP, Fevre M, Embley TM, Prins RA. 1997. A mitochondrial-like targeting signal on the hydrogenosomal malic enzyme from the anaerobic fungus *Neocallimastix frontalis*: support for the hypothesis that hydrogenosomes are modified mitochondria. Mol. Microbiol. 23:11–21.

Gijzen HJ, Broers CA, Barughare M, Stumm CK. 1991. Methanogenic bacteria as endosymbionts of the ciliate *Nyctotherus ovalis* in the cockroach hindgut. Appl. Envir. Microbiol. 57:1630–1634.

Gilbert C, Schaack S, 2nd JKP, Brindley PJ, Feschotte C. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature 464:1347–1350.

Gill EE, Diaz-Trivino S, Barbera MJ, Silberman JD, Stechmann A, Gaston D, Tamas I, Roger AJ, Diaz-Triviño S, Barberà MJ. 2007. Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*. Mol. Microbiol. 66:1306–1320.

Goldberg A V, Molik S, Tsaousis AD, Neumann K, Kuhnke G, Delbac F, Vivares CP, Hirt RP, Lill R, Embley TM. 2008. Localization and functionality of microsporidian iron-sulphur cluster assembly proteins. Nature 452:624–628.

González-Mariscal I, García-Testón E, Padilla S, Martín-Montalvo A, Pomares Viciana T, Vazquez-Fonseca L, Gandolfo Domínguez P, Santos-Ocaña C. 2014. The regulation of coenzyme q biosynthesis in eukaryotic cells: all that yeast can tell us. Mol. Syndromol. 5:107–118.

Goodgame RW. 1996. Understanding Intestinal Spore-Forming Protozoa: Cryptosporidia, Microsporidia, Cyclospora. Ann. Intern. Med. 124:429.

Graaf RM De, Duarte I, Alen TA Van, Kuiper JWP, Schotanus K, Rosenberg J, Huynen MA, Hackstein JHP, de Graaf RM, van Alen T a. 2009. The hydrogenosomes of *Psalteriomonas lanterna*. BMC Evol. Biol. 9:287.

De Graaf RM, Ricard G, van Alen TA, Duarte I, Dutilh BE, Burgtorf C, Kuiper JWP, van der Staay GWM, Tielens AGM, Huynen MA, et al. 2011. The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. Mol. Biol. Evol. 28:2379–2391.

Grant J, Lahr DJG, Rey FE, Burleigh JG, Knight R, Molestina RE, Katz LA. 2012. Gene discovery from a pilot study of the transcriptomes from three diverse microbial eukaryotes: *Corallomyxa tenera, Chilodonella uncinata*, and *Subulatomonas tetraspora*. Protist Genomics 1:3–18. Gray MW, Burger G, Lang BF. 2001. The origin and early evolution of mitochondria. Genome Biol 2:1–5.

Gray MW, Burger G, Lang FB. 1999. Mitochondrial Evolution. Science 283(5407):1476–1481.

Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. Annu. Rev. Genet. 38:477–524.

Van Grinsven KWA, Rosnowsky S, van Weelden SWH, Pütz S, van der Giezen M, Martin W, van Hellemond JJ, Tielens AGM, Henze K. 2008. Acetate:succinate CoA-transferase in the hydrogenosomes of *Trichomonas vaginalis*: identification and characterization. J. Biol. Chem. 283:1411–1418.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." Proc. Natl. Acad. Sci. U. S. A. 106:3859–3864.

Hampl V, Stairs CW, Roger AJ. 2011. The tangled past of eukaryotic enzymes involved in anaerobic metabolism. Mob. Genet. Elements 1:71–74.

Heinz E, Lithgow T. 2012. Back to basics: A revealing secondary reduction of the mitochondrial protein import pathway in diverse intracellular parasites. Biochim. Biophys. Acta.

Heinz E, Williams TA, Nakjang S, Noël CJ, Swan DC, Goldberg A V, Harris SR, Weinmaier T, Markert S, Becher D, et al. 2012. The genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into microsporidian genome dynamics and reductive evolution. PLoS Pathog. 8:e1002979.

Heiss AA, Walker G, Simpson AGB. 2013. The flagellar apparatus of *Breviata anathema*, a eukaryote without a clear supergroup affinity. Eur. J. Protistol.

Hemschemeier A, Jacobs J, Happe T. 2008. Biochemical and physiological characterization of the pyruvate formate-lyase Pfl1 of *Chlamydomonas reinhardtii*, a typically bacterial enzyme in a eukaryotic alga. Eukaryot. Cell 7:518–526.

Herzig S, Raemy E, Montessuit S, Veuthey J-L, Zamboni N, Westermann B, Kunji ERS, Martinou J-C. 2012. Identification and functional expression of the mitochondrial pyruvate carrier. Science 337:93–96.

Hiltunen JK, Autio KJ, Schonauer MS, Kursu V a S, Dieckmann CL, Kastaniotis AJ. 2010. Mitochondrial fatty acid synthesis and respiration. Biochim. Biophys. Acta 1797:1195–1202.

Hine PM, Bower SM, Meyer GR, Cochennec-Laureau N, Berthe FCJ. 2001. Ultrastructure of *Mikrocytos mackini*, the cause of Denman Island disease in oysters *Crassostrea* spp. and Ostr*e*a spp. in British Columbia, Canada. Dis. Aquat. Organ. 45:215–227.

Hirt RP, Logsdon JM, Healy B, Dorey MW, Doolittle WF, Embley TM. 1999. Microsporidia are related to Fungi: Evidence from the largest subunit of RNA polymerase II and other proteins. Proc. Natl. Acad. Sci. 96:580–585.

Van Hoek AHAM, van Alen TA, Sprakel VSI, Leunissen JAM, Brigge T, Vogels GD, Hackstein JHP. 2000. Multiple Acquisition of Methanogenic Archaeal Symbionts by Anaerobic Ciliates. Mol. Biol. Evol. 17:251–258.

Hoffmeister M, van der Klei A, Rotte C, van Grinsven KWA, van Hellemond JJ, Henze K, Tielens AGM, Martin W. 2004. *Euglena gracilis* rhodoquinone:ubiquinone ratio and mitochondrial proteome differ under aerobic and anaerobic conditions. J. Biol. Chem. 279:22422–22429.

Horner DS, Hirt RP, Embley TM. 1999. A single eubacterial origin of eukaryotic pyruvate: ferredoxin oxidoreductase genes: implications for the evolution of anaerobic eukaryotes. Mol. Biol. Evol. 16:1280–1291.

Horner DS, Hirt RP, Kilvington S, Lloyd D, Embley TM. 1996. Molecular data suggest an early acquisition of the mitochondrion endosymbiont. Proc. Biol. Sci. 263:1053–1059.

Hrdy I, Hirt RP, Dolezal P, Bardonová L, Foster PG, Tachezy J, Embley TM. 2004. *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature 432:618–622.

Hrdy I, Muller M. 1995. Primary structure and eubacterial relationships of the pyruvate:Ferredoxin oxidoreductase of the amitochondriate eukaryote *Trichomonas vaginalis*. J. Mol. Evol. 41.

Hug LA, Stechmann A, Roger AJ. 2010. Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes. Mol. Biol. Evol. 27:311–324.

Hurt EC, Schatz G. 1987. A cytosolic protein contains a cryptic mitochondrial targeting signal. Nature 325:499–503.

Iwata F, Shinjyo N, Amino H, Sakamoto K, Islam MK, Tsuji N, Kita K. 2008. Change of subunit composition of mitochondrial complex II (succinate-ubiquinone reductase/quinol-fumarate reductase) in *Ascaris suum* during the migration in the experimental host. Parasitol. Int. 57:54–61.

Jedelský PL, Doležal P, Rada P, Pyrih J, Smíd O, Hrdý I, Sedinová M, Marcinčiková M, Voleman L, Perry AJ, et al. 2011. The Minimal Proteome in the Reduced Mitochondrion of the Parasitic Protist *Giardia intestinalis*. PLoS One 6:e17285.

Jerlström-Hultqvist J, Einarsson E, Xu F, Hjort K, Ek B, Steinhauf D, Hultenby K, Bergquist J, Andersson JO, Svärd SG. 2013. Hydrogenosomes in the diplomonad *Spironucleus salmonicida*. Nat. Commun. 4:2493.

Jiang D, Zhao L, Clish CB, Clapham DE. 2013. Letm1, the mitochondrial Ca2+/H+ antiporter, is essential for normal glucose metabolism and alters brain function in Wolf-Hirschhorn syndrome. Proc. Natl. Acad. Sci. U. S. A. 110:E2249–E2254.

Kamikawa R, Sanchez-Perez GF, Sako Y, Roger AJ, Inagaki Y. 2009. Expanded phylogenies of canonical and non-canonical types of methionine adenosyltransferase reveal a complex history of these gene families in eukaryotes. Mol. Phylogenet. Evol. 53:565–570.

Kanehisa M. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28:27–30.

Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, et al. 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi.* Nature 414:450–453.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9:286–298.

Katz JE, Dlakić M, Clarke S. 2003. Automated identification of putative methyltransferases from genomic open reading frames. Mol. Cell. Proteomics 2:525–540.

Keeling PJ, Corradi N, Morrison HG, Haag KL, Ebert D, Weiss LM, Akiyoshi DE, Tzipori S. 2010. The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. Genome Biol. Evol. 2:304–309.

Keeling PJ, Fast NM. 2002. Microsporidia: biology and evolution of highly reduced intracellular parasites. Annu. Rev. Microbiol. 56:93–116.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. 9:605–618.

Kerbey BAL, Randle PJ, Cooper RH, Whitehouse S, Pask HT, Denton RM. 1976. Regulation of Pyruvate Dehydrogenase in Rat Heart. 154:327–348.

Kita K, Takamiya S, Furushima R, Ma Y, Suzuki H, Ozawa T, Oya H. 1988. Electron-transfer complexes of *Ascaris suum* muscle mitochondria. III. Composition and fumarate reductase activity of complex II. Biochim. Biophys. Acta - Bioenerg. 935:130–140.

Van Klink F, Alizadeh H, Stewart GL, Pidherney MS, Silvany RE, He Y, McCulley JP, Niederkorn JY. 1992. Characterization and pathogenic potential of a soil isolate and an ocular isolate of *Acanthamoeba castellanii* in relation to *Acanthamoeba keratitis.* Curr. Eye Res. 11:1207–1220.

Kobayashi M, Matsuo Y, Takimoto A, Suzuki S, Maruo F, Shoun H. 1996. Denitrification, a Novel Type of Respiratory Metabolism in Fungal Mitochondrion. J. Biol. Chem. 271:16263–16267.

Kumar S, Skjaeveland A, Orr RJS, Enger P, Ruden T, Mevik B-H, Burki F, Botnen A, Shalchian-Tabrizi K. 2009. AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. BMC Bioinformatics 10:357.

Kunin V, Ouzounis CA. 2003. The balance of driving forces during genome evolution in prokaryotes. Genome Res. 13:1589–1594.

Lagier-Tourenne C, Tazir M, López LC, Quinzii CM, Assoum M, Drouot N, Busso C, Makri S, Ali-Pacha L, Benhassine T, et al. 2008. ADCK3, an ancestral kinase, is mutated in a form of recessive ataxia associated with coenzyme Q10 deficiency. Am. J. Hum. Genet. 82:661–672.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Lane CE, Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. Trends Ecol. Evol. (Personal Ed. 23:268–275.

Lange H, Kaut a., Kispal G, Lill R. 2000. A mitochondrial ferredoxin is essential for biogenesis of cellular iron-sulfur proteins. Proc. Natl. Acad. Sci. 97:1050–1055.

Lanier W, Moustafa A, Bhattacharya D, Comeron JM. 2008. EST analysis of *Ostreococcus lucimarinus,* the most compact eukaryotic genome, shows an excess of introns in highly expressed genes. PLoS One 3:e2171.

Lantsman Y, Tan KS, Morada M, Yarlett N. 2008. Biochemical characterization of a mitochondrial-like organelle from *Blastocystis* sp. subtype 7. Microbiology 154:2757–2766.

Lapointe J, Wang Y, Bigras E, Hekimi S. 2012. The submitochondrial distribution of ubiquinone affects respiration in long-lived Mclk1+/- mice. J. Cell Biol. 199:215–224.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25:2286–2288.

Le SQ, Dang CC, Gascuel O. 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. Mol. Biol. Evol. 29:2921–2936.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25:1307–1320.

Le SQ, Lartillot N, Gascuel O. 2008. Phylogenetic mixture models for proteins. Philos. Trans. R. Soc. London.Series B, Biol. Sci. 363:3965–3976.

Leger MM, Gawryluk RMR, Gray MW, Roger AJ. 2013. Evidence for a Hydrogenosomal-Type Anaerobic ATP Generation Pathway in Acanthamoeba castellanii. PLoS One 8:e69532.

Lehtiö L, Goldman A, Lehtio L. 2004. The pyruvate formate lyase family: sequences, structures and activation. Protein Eng. Des. Sel. 17:545–552.

Leigh JW, Susko E, Baumgartner M, Roger AJ. 2008. Testing congruence in phylogenomic analysis. Syst. Biol. 57:104–115.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Liang C-R, Rosenberg H. 1968. On the distribution and biosynthesis of 2-aminoethylphosphonate in two terrestrial molluscs. Comp. Biochem. Physiol. 25:673–681.

Lill R. 2009. Function and biogenesis of iron-sulphur proteins. Nature 460:831–838.

Liu Z, Li X, Zhao P, Gui J, Zheng W, Zhang Y. 2011. Tracing the evolution of the mitochondrial protein import machinery. Comput. Biol. Chem. 35:336–340.

Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, et al. 2005. The genome of the protist parasite *Entamoeba histolytica*. Nature 433:865–868.

Lonjers ZT, Dickson EL, Chu T-PT, Kreutz JE, Neacsu F a, Anders KR, Shepherd JN. 2012. Identification of a new gene required for the biosynthesis of rhodoquinone in *Rhodospirillum rubrum*. J. Bacteriol. 194:965–971.

López-García P, Moreira D. 1999. Metabolic symbiosis at the origin of eukaryotes. Trends Biochem. Sci. 24:88–93.

Lucattini R, Likic VA, Lithgow T. 2004. Bacterial proteins predisposed for targeting to mitochondria. Mol. Biol. Evol. 21:652–658.

Ma H, Kunes S, Schatz PJ, Botstein D. 1987. Plasmid construction by homologous recombination in yeast. Gene 58:201–216.

Ma YC, Funk M, Dunham WR, Komuniecki R. 1993. Purification and characterization of electron-transfer flavoprotein: rhodoquinone oxidoreductase from anaerobic mitochondria of the adult parasitic nematode, *Ascaris suum*. J. Biol. Chem. 268:20360–20365.

Maniatis T, Fritsh EF, Sambrook J. 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor: Cold Spring Harbor Laboratory

Maralikova B, Ali V, Nakada-Tsukui K, Nozaki T, van der Giezen M, Henze K, Tovar J. 2010. Bacterial-type oxygen detoxification and iron-sulfur cluster assembly in amoebal relict mitochondria. Cell. Microbiol. 12:331–342.

Markham G, Hafner E, Tabor C, Tabor H. 1980. S-Adenosylmethionine synthetase from *Escherichia coli.* J. Biol. Chem. 255:9082–9092.

Martin W, Müller M. 1998. The hydrogen hypothesis for the first eukaryote. Nature 392:37–41.

Martin WF. 2011. Early evolution without a tree of life. Biol. Direct 6:36.

Marvin-Sikkema FD, Gomes TMP, Grivet JP, Gottschal JC, Prins RA. 1993. Characterization of hydrogenosomes and their role in glucose metabolism of *Neocallimastix* sp. L2. Arch. Microbiol. 160:388–396.

Mentel M, Martin W. 2010. Anaerobic animals from an ancient, anoxic ecological niche. BMC Biol. 8:32.

Mentel M, Röttger M, Leys S, Tielens AGM, Martin WF. 2014. Of early animals, anaerobic mitochondria, and a modern sponge. Bioessays 36:924–932.

Mereschkowsky C. 1905. Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. Biol. Zentrablatt 25:593–604.

Metcalf WW, van der Donk W a. 2009. Biosynthesis of phosphonic and phosphinic acid natural products. Annu. Rev. Biochem. 78:65–94.

Mi-ichi F, Abu Yousuf M, Nakada-Tsukui K, Nozaki T. 2009. Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway. Proc. Natl. Acad. Sci. U. S. A. 106:21731–21736.

Milenkovic D, Ramming T, Müller JM, Wenz L-S, Gebert N, Schulze-Specking A, Stojanovski D, Rospert S, Chacinska A. 2009. Identification of the signal directing Tim9 and Tim10 into the intermembrane space of mitochondria. Mol. Biol. Cell 20:2530–2539.

Mills DB, Canfield DE. 2014. Oxygen and animal evolution: Did a rise of atmospheric oxygen “trigger” the origin of animals? BioEssays 36:1145–1155.

Mills DB, Ward LM, Jones C, Sweeten B, Forth M, Treusch AH, Canfield DE. 2014. Oxygen requirements of the earliest animals. Proc. Natl. Acad. Sci. U. S. A. 111:4168–4172.

Mogi T, Kita K. 2010. Diversity in mitochondrial metabolic pathways in parasitic protists *Plasmodium* and *Cryptosporidium*. Parasitol. Int. 59:305–312.

Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, et al. 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. Science 317:1921–1926.

Mulder DW, Boyd ES, Sarma R, Lange RK, Endrizzi JA, Broderick JB, Peters JW. 2010. Stepwise [FeFe]-hydrogenase H-cluster assembly revealed in the structure of HydA(DeltaEFG). Nature 465:248–251.

Müller M, Lindmark DG. 1978. Respiration of hydrogenosomes of *Tritrichomonas foetus*. II. Effect of CoA on pyruvate oxidation. J. Biol. Chem. 253:1215–1218.

Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu R-Y, van der Giezen M, Tielens AGM, Martin WF. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol. Mol. Biol. Rev. 76:444–495.

Nass MM, Nass S. 1963. Intramitochondrial fibers with DNA characteristics. J. Cell Biol. 19:593–611.

Nozaki H, Maruyama S, Matsuzaki M, Nakada T, Kato S, Misawa K. 2009. Phylogenetic positions of Glaucophyta, green plants (Archaeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes. Mol. Phylogenet. Evol. 53:872–880.

Nyvltová E, Suták R, Harant K, Sedinová M, Hrdy I, Paces J, Vlcek C, Tachezy J. 2013. NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of Mastigamoeba balamuthi. Proc. Natl. Acad. Sci. U. S. A.

O'Kelly CJ, Farmer MA, Nerad TA. 1999. Ultrastructure of *Trimastix pyriformis* (Klebs) Bernard et al.: similarities of *Trimastix* species with retortamonad and jakobid flagellates. Protist 150:149–162.

O'Kelly CJ, Silberman JD, Amaral Zettler LA, Nerad TA, Sogin ML. 2003. *Monopylocystis visvesvarai* n. gen., n. sp. and *Sawyeria marylandensis* n. gen., n. sp.: two new amitochondrial heterolobosean amoebae from anoxic environments. Protist 154:281–290.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304.

Ochoa S, Mehler AH, Kornberg A. 1948. Biosynthesis of dicarboxylic acids by carbon dioxide fixation; isolation and properties of an enzyme from pigeon liver catalyzing the reversible oxidative decarboxylation of 1-malic acid. J. Biol. Chem. 174:979–1000.

Olsen GJ, Woese CR, Overbeek R. 1994. The winds of (evolutionary) change: breathing new life into microbiology. J. Bacteriol. 176:1–6.

Olson JW, Agar JN, Johnson MK, Maier RJ. 2000. Characterization of the NifU and NifS Fe−S Cluster Formation Proteins Essential for Viability in *Helicobacter pylori*. Biochemistry 39:16213–16219.

Outten FW, Djaman O, Storz G. 2004. A suf operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli.* Mol. Microbiol. 52:861–872.

Padilla-López S, Jiménez-Hidalgo M, Martín-Montalvo A, Clarke CF, Navas P, Santos-Ocaña C. 2009. Genetic evidence for the requirement of the endocytic pathway in the uptake of coenzyme Q6 in *Saccharomyces cerevisiae.* Biochim. Biophys. Acta 1788:1238–1248.

Patel MS, Korotchkina LG. 2006. Regulation of the pyruvate dehydrogenase complex. Biochem. Soc. Trans. 34:217–222.

Pertea M, Ayanbule K, Smedinghoff M, Salzberg SL. 2009. OperonDB: a comprehensive database of predicted operons in microbial genomes. Nucleic Acids Res. 37:D479–D482.

Petrin D, Delgaty K, Bhatt R, Garber G. 1998. Clinical and microbiological aspects of *Trichomonas vaginalis*. Clin. Microbiol. Rev. 11:300–317.

Petrossian TC, Clarke SG. 2009. Multiple Motif Scanning to identify methyltransferases from the yeast proteome. Mol. Cell. Proteomics 8:1516–1526.

Philippe H. 1993. MUST, a computer package of Management Utilities for Sequences and Trees. Nucleic Acids Res. 21:5264–5272.

Pisani D, Cotton JA, McInerney JO. 2007. Supertrees disentangle the chimerical origin of eukaryotic genomes. Mol. Biol. Evol. 24:1752–1760.

Poole AM, Penny D. 2007. Evaluating hypotheses for the origin of eukaryotes. Bioessays 29:74–84.

Portier P. 1918. Les symbiotes. Paris: Masson

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.

Py B, Gerez C, Angelini S, Planel R, Vinella D, Loiseau L, Talla E, Brochier-Armanet C, Garcia Serres R, Latour J-M, et al. 2012. Molecular organization, biochemical function, cellular role and evolution of NfuA, an atypical Fe-S carrier. Mol. Microbiol. 86:155–171.

Quinzii CM, Hirano M. 2010. Coenzyme Q and mitochondrial disease. Dev. Disabil. Res. Rev. 16:183–188.

Raynaud C, Sarçabal P, Meynial-Salles I, Croux C, Soucaille P, Sarc P. 2003. Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum.* Proc. Natl. Acad. Sci. U. S. A. 100:5010–5015.

Reeves RE, Warren LG, Susskind B, Lo HS. 1977. An energy-conserving pyruvate-to-acetate pathway in *Entamoeba histolytica.* Pyruvate synthase and a new acetate thiokinase. J. Biol. Chem. 252:726–731.

Rew R. 1974. Enzyme localization in the anaerobic mitochondria of Ascaris lumbricoides. J. Cell Biol. 63:125–135.

Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. 2006. Evolution of Filamentous Plant Pathogens: Gene Exchange across Eukaryotic Kingdoms. Curr. Biol. 16:1857–1864.

Richards TA, Soanes DM, Jones MDM, Vasieva O, Leonard G, Paszkiewicz K, Foster PG, Hall N, Talbot NJ. 2011. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proc. Natl. Acad. Sci. U. S. A. 108:15258–15263.

Riordan CE, Ault JG, Langreth SG, Keithly JS. 2003. *Cryptosporidium parvum* Cpn60 targets a relict organelle. Curr. Genet. 44:138–147.

Robinson KM. 1997. The COQ5 Gene Encodes a Yeast Mitochondrial Protein Necessary for Ubiquinone Biosynthesis and the Assembly of the Respiratory Chain. J. Biol. Chem. 272:9175–9181.

Roche B, Aussel L, Ezraty B, Mandin P, Py B, Barras F. 2013. Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity. Biochim. Biophys. Acta 1827:455–469.

Rodionov DA, Gelfand MS. 2005. Identification of a bacterial regulatory system for ribonucleotide reductases by phylogenetic profiling. Trends Genet. 21:385–389.

Rodríguez-Ezpeleta N, Embley TM. 2012. The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria. PLoS One 7:e30520.

Roger AJ, Clark CG, Doolittle WF. 1996. A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. Proc. Natl. Acad. Sci. U. S. A. 93:14618–14622.

Roger AJ, Svard SG, Tovar J, Clark CG, Smith MW, Gillin FD, Sogin ML. 1998. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc. Natl. Acad. Sci. U. S. A. 95:229–234.

Rotte C, Stejskal F, Zhu G, Keithly JS, Martin W. 2001. Pyruvate : NADP Oxidoreductase from the Mitochondrion of *Euglena gracilis* and from the Apicomplexan *Cryptosporidium parvum:* A Biochemical Relic Linking Pyruvate Metabolism in Mitochondriate and Amitochondriate Protists. Mol. Biol. Evol. 18:710–720.

Sagan L. 1967. On the origin of mitosing cells. J. Theor. Biol. 14:255–274.

Sakai C, Tomitsuka E, Esumi H, Harada S, Kita K. 2012. Mitochondrial fumarate reductase as a target of chemotherapy: From parasites to cancer cells. Biochim. Biophys. Acta - Gen. Subj. 1820:643–651.

Sanchez LB, Galperin MY, Muller M. 2000. Acetyl-CoA synthetase from the amitochondriate eukaryote *Giardia lamblia* belongs to the newly recognized superfamily of acyl-CoA synthetases (Nucleoside diphosphate-forming). J. Biol. Chem. 275:5794–5803.

Sarkar M, Hamilton CJ, Fairlamb AH. 2003. Properties of phosphoenolpyruvate mutase, the first enzyme in the aminoethylphosphonate biosynthetic pathway in *Trypanosoma cruzi*. J. Biol. Chem. 278:22703–22708.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504.

Schmitz-Esser S, Linka N, Collingro A, Beier CL, Neuhaus HE, Wagner M, Horn M. 2004. ATP/ADP Translocases: A Common Feature of Obligate Intracellular Amoebal Symbionts Related to Chlamydiae and Rickettsiae. J. Bacteriol. 186:683–691.

Schneider RE, Brown MT, Shiflett AM, Dyall SD, Hayes RD, Xie Y, Loo J a, Johnson PJ. 2011. The *Trichomonas vaginalis* hydrogenosome proteome is highly reduced relative to mitochondria, yet complex compared with mitosomes. Int. J. Parasitol. 41:1421–1434.

Schut GJ, Adams MWW. 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. J. Bacteriol. 191:4451–4457.

Schwartz W. 1970. Lynn Margulis, Origin of Eukaryotic Cells. Evidence and Research Implications for a Theory of the Origin and Evolution of Microbial, Plant, and Animal

Cells on the Precambrian Earth. XXII u. 349 S., 89 Abb., 49 Tab. New Haven-London 1970: Yale University. New Haven, Conneticut

Searcy DEG. 2003. Metabolic integration during the evolutionary origin of mitochondria. 13:229–238.

Shi X, Gu H, Susko E, Field C. 2005. The comparison of the confidence regions in phylogeny. Mol. Biol. Evol. 22:2285–2296.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–1247.

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

Sicheritz-Pontén T, Kurland CG, Andersson SGEE. 1998. A phylogenetic analysis of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. Biochim. Biophys. Acta - Bioenerg. 1365:545–551.

Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, Meyer HE, Schönfisch B, Perschil I, Chacinska A, Guiard B, et al. 2003. The proteome of *Saccharomyces cerevisiae* mitochondria. Proc. Natl. Acad. Sci. U. S. A. 100:13207–13212.

Simpson AGB, Bernard C, Patterson DJ. 2000. The ultrastructure of *Trimastix marina* Kent 1880 (Eukaryota), an excavate flagellate. Eur. J. Protistol. 36:229–251.

Smith AC, Blackshaw JA, Robinson AJ. 2012. MitoMiner: a data warehouse for mitochondrial proteomics data. Nucleic Acids Res. 40:D1160–D1167.

Smith DGS, Gawryluk RMR, Spencer DF, Pearlman RE, Siu KWM, Gray MW. 2007. Exploring the mitochondrial proteome of the ciliate protozoon *Tetrahymena thermophila*: direct analysis by tandem mass spectrometry. J. Mol. Biol. 374:837–863.

Sogin ML, Gunderson JH, Elwood HJ, Alonso RA, Peattie DA. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia.* Science 243:75–77.

Soltys BJ, Gupta RS. 1994. Presence and cellular distribution of a 60-kDa protein related to mitochondrial hsp60 in *Giardia lamblia*. J. Parasitol. 80:580–590.

Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, van der Giezen M, Roger AJ. 2014. A SUF Fe-S Cluster Biogenesis System in the Mitochondrion-Related Organelles of the Anaerobic Protist *Pygsuia*. Curr. Biol. 24:1176–1186.

Stairs CW, Roger AJ, Hampl V. 2011. Eukaryotic pyruvate formate lyase and its activating enzyme were acquired laterally from a Firmicute. Mol. Biol. Evol. 28:2087–2099.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stechmann A, Baumgartner M, Silberman JD, Roger AJ. 2006. The glycolytic pathway of *Trimastix pyriformis* is an evolutionary mosaic. BMC Evol. Biol. 6:101.

Stechmann A, Hamblin K, Pérez-Brocal V, Gaston D, Richmond GS, van der Giezen M, Clark CG, Roger AJ, Perez-Brocal V, Giezen M Van Der. 2008. Organelles in

Blastocystis that blur the distinction between mitochondria and hydrogenosomes. Curr. Biol. 18:580–585.

Stehling O, Lill R. 2013. The role of mitochondria in cellular iron-sulfur protein biogenesis: mechanisms, connected processes, and diseases. Cold Spring Harb. Perspect. Biol. 5.

Steinbüchel A, Müller M. 1986. Anaerobic pyruvate metabolism of *Tritrichomonas foetus* and *Trichomonas vaginalis* hydrogenosomes. Mol. Biochem. Parasitol. 20:57–65.

Sucheta A, Ackrell BACA, Cochran B, Armstrong FAA. 1992. Diode-like behaviour of a mitochondrial electron-transport enzyme. Nature 356:361–362.

Sukumaran J, Holder MT. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Susko E. 2009. Bootstrap support is not first-order correct. Syst. Biol. 58:211–223.

Susko E. 2010. First-order correct bootstrap support adjustments for splits that allow hypothesis testing when using maximum likelihood estimation. Mol. Biol. Evol. 27:1621–1629.

Tachezy J, Sanchez LB, Muller M. 2001. Mitochondrial Type Iron-Sulfur Cluster Assembly in the Amitochondriate Eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as Indicated by the Phylogeny of IscS. Mol. Biol. Evol. 18:1919–1928.

Takahashi Y, Tokumoto U. 2002. A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids. J. Biol. Chem. 277:28380–28383.

Takamiya S, Matsui T, Taka H, Murayama K, Matsuda M, Aoki T. 1999. Free-living nematodes *Caenorhabditis elegans* possess in their mitochondria an additional rhodoquinone, an essential component of the eukaryotic fumarate reductase system. Arch. Biochem. Biophys. 371:284–289.

Takishita K, Chikaraishi Y, Leger MM, Kim E, Yabuki A, Ohkouchi N, Roger AJ. 2012. Lateral transfer of tetrahymanol-synthesizing genes has allowed multiple diverse eukaryote lineages to independently adapt to environments without oxygen. Biol. Direct 7:5.

Thauer RK, Kirchniawy FH, Jungermann KA. 1972. Properties and Function of the Pyruvate-Formate-Lyase Reaction in Clostridiae. Eur. J. Biochem. 27:282–290.

Thrash JC, Boyd A, Huggett MJ, Grote J, Carini P, Yoder RJ, Robbertse B, Spatafora JW, Rappé MS, Giovannoni SJ. 2011. Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. Sci. Rep. 1:13.

Tian H-F, Feng J-M, Wen J-F. 2012. The evolution of cardiolipin biosynthesis and maturation pathways and its implications for the evolution of eukaryotes. BMC Evol. Biol. 12:32.

Tielens AGM, Van Hellemond JJ, Hellemond JJ Van. 1998. The electron transport chain in anaerobically functioning eukaryotes. Biochim. Biophys. Acta 1365:71–78.

Tielens AGM. 1994. Energy generation in parasitic helminths. Parasitol. Today 10:346–352.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. 5:123–135.

Tovar J, Fischer A, Clark CG. 1999. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. Mol. Microbiol. 32:1013–1021.

Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, van der Giezen M, Hernández M, Müller M, Lucocq JM. 2003. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. Nature 426:172–176.

Tsaousis AD, Kunji ER, Goldberg A V, Lucocq JM, Hirt RP, Embley TM. 2008. A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. Nature 453:553–556.

Tsaousis AD, Nyvltová E, Sutak R, Hrdy I, Tachezy J. 2014. A nonmitochondrial hydrogen production in *Naegleria gruberi*. Genome Biol. Evol. 6:792–799.

Tsaousis AD, Ollagnier de Choudens S, Gentekaki E, Long S, Gaston D, Stechmann A, Vinella D, Py B, Fontecave M, Barras F, et al. 2012. Evolution of Fe/S cluster biogenesis in the anaerobic parasite *Blastocystis*. Proc. Natl. Acad. Sci. U. S. A. 109:10426–10431.

Ueda M, Fujimoto M, Arimura S-I, Tsutsumi N, Kadowaki K-I. 2008. Presence of a latent mitochondrial targeting signal in gene on mitochondrial genome. Mol. Biol. Evol. 25:1791–1793.

Vanlerberghe GC, McIntosh L. 1997. ALTERNATIVE OXIDASE: From Gene to Function. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48:703–734.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. 2001. The sequence of the human genome. Science 291:1304–1351.

Vossbrinck CR, Maddox J V, Friedman S, Debrunner-Vossbrinck BA, Woese CR. 1987. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. Nature 326:411–414.

Wagner AF, Frey M, Neugebauer FA, Schafer W, Knappe J. 1992. The free radical in pyruvate formate-lyase is located on glycine-734. Proc. Natl. Acad. Sci. 89:996–1000.

Wallin IE. 1927. Symbionticism and the Origin of Species. Рипол Классик

Wang Y, Hekimi S. 2012. Molecular genetics of ubiquinone biosynthesis in animals. Crit. Rev. Biochem. Mol. Biol.:1–20.

Williams BAP, Elliot C, Burri L, Kido Y, Kita K, Moore AL, Keeling PJ. 2010. A broad distribution of the alternative oxidase in microsporidian parasites. PLoS Pathog. 6:e1000761.

Williams BAP, Hirt RP, Lucocq JM, Embley TM. 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. Nature 418:865–869.

Wilson EB. 1925. The cell in development and heredity 3rd edition. Macmillan

Wilson R, Williamson D. 1997. Extrachromosomal DNA in the Apicomplexa. Microbiol. Mol. Biol. Rev. 61:1–16.

Wollers S, Layer G, Garcia-Serres R, Signor L, Clemancey M, Latour J-M, Fontecave M, Ollagnier de Choudens S. 2010. Iron-sulfur (Fe-S) cluster assembly: the SufBCD complex is a new type of Fe-S scaffold with a flavin redox cofactor. J. Biol. Chem. 285:23331–23341.

Wu J, Hu Z, DeLisi C. 2006. Gene annotation and network inference by phylogenetic profiling. BMC Bioinformatics 7:80.

Xu F, Jerlström-Hultqvist J, Einarsson E, Astvaldsson A, Svärd SG, Andersson JO. 2014. The genome of *Spironucleus salmonicida* highlights a fish pathogen adapted to fluctuating environments.Heitman J, editor. PLoS Genet. 10:e1004053.

Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR. 1985. Mitochondrial origins. Proc. Natl. Acad. Sci. U. S. A. 82:4443–4447.

Zeghouf M, Defaye G, Fontecave M, Coves J. 1998. The flavoprotein component of the *Escherichia coli* sulfite reductase can act as a cytochrome P450c17 reductase. Biochem. Biophys. Res. Commun. 246:602–605.

Zierdt CH. 1983. Blastocystis hominis, a protozoan parasite and intestinal pathogen of human beings. Clin. Microbiol. Newsl. 5:57–59.

Zillig W, Klenk H-P, Palm P, Pühler G, Gropp F, Garrett R a., Leffers H. 1989. The phylogenetic relations of DNA-dependent RNA polymerases of archaebacteria, eukaryotes, and eubacteria. Can. J. Microbiol. 35:73–80.

Zubáčová Z, Novák L, Bublíková J, Vacek V, Fousek J, Rídl J, Tachezy J, Doležal P, Vlček C, Hampl V, et al. 2013. The mitochondrion-like organelle of *Trimastix pyriformis* contains the complete glycine cleavage system. PLoS One 8:e55417.

Appendix A Copyright Permission (Chapter 2)

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Oct 07, 2014

This is a License Agreement between Courtney W Stairs ("You") and Oxford University Press ("Oxford University Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3483720652958
License date	Oct 07, 2014
Order Content Publisher	Oxford University Press
Order Content Publication	Molecular Biology and Evolution
Order Content Title	Eukaryotic Pyruvate Formate Lyase and Its Activating Enzyme Were Acquired Laterally from a Firmicute:
Order Content Author	Courtney W. Stairs, Andrew J. Roger, Vladimir Hampl
Order Content Date	07/01/2011
Type of Use	Thesis/Dissertation
Institution name	None
Title of your work	Functions and Origins of Mitochondrion-Related Organelles in Anaerobic Protists
Publisher of your work	n/a
Expected publication date	Nov 2014
Permissions cost	0.00 USD
Value added tax	0.00 USD
TotalTotal	**0.00 USD**
TotalTotal	**0.00 USD**

Appendix B Copyright Permission (Chapter 3)

ELSEVIER LICENSE
TERMS AND CONDITIONS

Oct 07, 2014

This is a License Agreement between Courtney W Stairs ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Courtney W Stairs
Customer address	5580 College Street Rm 8D1 Halifax, NS B3H2Z8
License number	3483711193264
License date	Oct 07, 2014
Licensed content publisher	Elsevier
Licensed content publication	Current Biology
Licensed content title	A SUF Fe-S Cluster Biogenesis System in the Mitochondrion-Related Organelles of the Anaerobic Protist Pygsuia
Licensed content author	Courtney W. Stairs,Laura Eme,Matthew W. Brown,Cornelis Mutsaers,Edward Susko,Graham Dellaire,Darren M. Soanes,Mark van der Giezen,Andrew J. Roger
Licensed content date	2 June 2014
Licensed content volume number	24
Licensed content issue number	11
Number of pages	11
Start Page	1176
End Page	1186
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Functions and Origins of Mitochondrion-Related Organelles in Anaerobic Protists