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

© Copyright by Courtney Weir Stairs, 2014
For those who said I could.
Table of Contents

List of Tables ............................................................................................................................... viii
List of Figures ................................................................................................................................. ix
Abstract ........................................................................................................................................... xii
List of Abbreviations Used ........................................................................................................... xiii
Acknowledgements ......................................................................................................................... xvii

Chapter 1  Introduction ...................................................................................................................... 1
  1.1  On The Origin Of Mitochondria ............................................................................................... 1
  1.2  The Archezoa Hypothesis And Amitochondriates .................................................................... 3
  1.3  Diversity And Function Of Mitochondria ................................................................................ 4
    1.3.1  Anaerobic Mitochondria .................................................................................................. 6
    1.3.2  Hydrogen-Producing Mitochondria ............................................................................... 8
        Acanthamoeba ......................................................................................................................... 8
        Naegleria gruberi ...................................................................................................................... 9
        Nyctotherus sp. ....................................................................................................................... 10
        Blastocystis sp. .................................................................................................................... 12
    1.3.3  Hydrogenosomes ........................................................................................................... 13
        Trichomonas vaginalis .............................................................................................................. 13
        Spironucleus salmonicida ....................................................................................................... 15
        Free-living excavates ........................................................................................................... 15
        Mastigamoeba balamuthi ........................................................................................................ 17
    1.3.4  Mitosomes ...................................................................................................................... 18
        Entamoeba ............................................................................................................................ 19
        Mikrocytos makini ................................................................................................................ 19
        Giardia intestinalis ................................................................................................................. 20
        Microsporidians ...................................................................................................................... 21
    1.3.5  Limitations Of Classification .............................................................................................. 22
  1.4  Origin Of Anaerobic Metabolism ............................................................................................. 25
1.4.1 Hydrogen Hypothesis ................................................................. 25
1.4.2 Detecting Lateral Gene Transfer In Eukaryotes ...................... 28
1.4.3 Laterally Acquired Anaerobic Metabolism .............................. 29

1.5 Aims Of This Thesis ........................................................................... 31

Chapter 2 Eukaryotic Pyruvate Formate Lyase And Its Activating Enzyme Were Acquired Laterally From A Firmicute ................................. 33

2.1 Abstract ......................................................................................... 33
2.2 Introduction ................................................................................... 34
2.3 Methods ....................................................................................... 37
  2.3.1 454 EST Project Of Mastigamoeba balamuthi ......................... 37
  2.3.2 Database Searches ................................................................. 37
  2.3.3 Data Set Generation And Phylogenetic Analysis .................. 38
  2.3.4 Congruence Testing And Concatenation ............................... 39
  2.3.5 Topology Testing ................................................................. 40
  2.3.6 Gene Order And Operon Prediction .................................... 41
  2.3.7 Tests For Long-Branch Attraction, Amino Acid Composition Bias ------ 41
  2.3.8 In Silico Subcellular Localization Prediction .................... 42

2.4 Results And Discussion ................................................................. 42
  2.4.1 Identification Of Previously Unidentified Pfl And Pfla In Eukaryotes ----- 42
  2.4.2 Eukaryotic Monophyly Of Pfl And Pfla ................................ 44
  2.4.3 Concatenation Of Pfla And Pfl Increases Monophyly Support And Suggests A Firmicute Ancestry For These Genes ................................. 49
  2.4.4 Atypical Eukaryotic Relationships In Pfl And Pfla Phylogenies Suggest Eukaryote-To-Eukaryote LGT Events ......................................................... 52
  2.4.5 Conservation Of Gene Order: Insights Into The Transfer Mechanism --- 59
  2.4.6 Gene Structures And Organization Of Pfl Genes In Diverse Eukaryotes 60
  2.4.7 In Silico Predictions Of Subcellular Localization .............................................. 61
  2.4.8 The Distribution Of Pyruvate Catalyzing Enzymes In Eukaryotes ........ 61

2.5 Conclusion ..................................................................................... 62
Chapter 3  A Mitochondrial Sulfur Mobilization (SUF) System In The Anaerobe

*Pygsuia biforma* ................................................................. 64

3.1 Abstract .................................................................................. 64

3.2 Introduction ............................................................................... 65

3.3 Methods .................................................................................... 67

3.3.1 *Pygsuia* RNAseq Transcriptome Analysis And Dataset Filtering ........ 67

3.3.2 MRO Protein Prediction .............................................................. 68

3.3.3 *Breviata anathema* Expressed Sequence Tag Project and *Subulatomonas
tetraspora* transcriptome Analysis .............................................. 69

3.3.4 Conserved Functional Domain Search ...................................... 70

3.3.5 Phylogenetic Dataset Construction ........................................... 70

3.3.6 Phylogenetic Analysis .............................................................. 70

3.3.7 Culturing And Microscopy ....................................................... 71

3.3.8 Molecular Biology And Cloning ................................................ 72

3.3.9 Yeast Transformation And Selection .......................................... 73

3.3.10 Antibody Production .............................................................. 74

3.3.11 Protein Expression And Extraction .......................................... 74

3.3.12 Immunoblotting ................................................................. 75

3.4 Results ........................................................................................ 75

3.4.1 Metabolic Pathway Prediction In *Pygsuia biforma* ...................... 75

3.4.2 Pyruvate And Energy Metabolism ............................................. 81

3.4.3 *Pygsuia* MROs Contain Canonical Protein Import/Processing
Machinery ....................................................................................... 92

3.4.4 Fe-S Cluster Biogenesis ............................................................ 94

3.4.5 Solute Transport, Amino Acid And Lipid Metabolism ................. 102

3.4.6 Localization And Morphology Studies Of *Pygsuia biforma* Proteins In
Yeast And *In vivo* ........................................................................... 106

3.4.7 Proteins Predicted In *Breviata anathema* And Not *Pygsuia biforma* -- 110

3.5 Discussion ................................................................................. 110

3.5.1 Energy Metabolism In MROs .................................................... 110
3.5.2 Conservation Of Mitochondrial Protein Import ................................. 112
3.5.3 Acquisition Of SUF-Like Fe-S Cluster Biosynthesis And Loss Of Isc
Machinery In The Pygsuia Lineage ..................................................... 112

3.6 Conclusion ................................................................. 114

Chapter 4 Evolution And Cellular Localization Of Rhodoquinone Biosynthesis In
Pygsuia biforma And Other Anaerobic Eukaryotes ................................ 116

4.1 Abstract .................................................................... 116

4.2 Introduction ................................................................ 117

4.3 Methods ................................................................... 120

4.3.1 Molecular Biology ................................................ 120

4.3.2 Heterologous Expression Of Proteins In E. coli And Immunoblotting... 120

4.3.3 Phylogenetic Dataset Construction And Sequence Analysis .......... 121

4.3.4 Phylogenetic Analysis ............................................. 122

4.3.5 Identification Of Quinone-Utilizing Enzymes In Eukaryotes .......... 123

4.3.6 Antibody Production .............................................. 123

4.3.7 Culturing And Microscopy ...................................... 124

4.4 Results .................................................................. 124

4.4.1 Distribution Of RQUA In Bacteria And Eukaryotes ......... 124

4.4.2 Phylogenetic Analysis Of Bacterial And Eukaryotic RQUA Homologs 124

4.4.3 Primary Sequence Analysis Of Eukaryotic RQUA ................ 128

4.4.4 The Distribution Of Quinone-Utilizing Enzymes In Eukaryotes .... 130

4.4.5 Subcellular Localization Of RQUA In Pygsuia biforma .............. 131

4.5 Discussion ................................................................. 132

4.5.1 Origin Of RQUA And Rhodoquinone Biosynthesis .............. 132

4.5.2 Dispersion Of RQUA By Lateral Gene Transfer (LGT) ............ 133

4.5.3 Eukaryotic RQUA Likely Functions In The Mitochondrion ......... 138

4.5.4 RQUA Function And The ‘Transferability’ Of RQ Biosynthesis
Between Organisms ........................................................................ 139

4.6 Conclusions ............................................................ 141

Chapter 5 Conclusions ..................................................... 142
5.1 A Few Assumptions----------------------------------------------- 142
5.2 Reduction Of The Electron Transport Chain In Response To Hypoxia------ 143
5.3 Changes To Pyruvate And ATP Generation------------------------------- 144
5.4 Loss Of CI, mtDNA And Malate Dismutation------------------------------- 145
5.5 A Note On Fe-S Biosynthesis------------------------------------------ 147
5.6 Final Remarks-------------------------------------------------------- 147
References------------------------------------------------------------------ 149
Appendix A Copyright Permission (Chapter 2)----------------------------- 167
Appendix B Copyright Permission (Chapter 3)----------------------------- 168
List of Tables

Table 2-1 AU Topology Tests for the Single Gene and Concatenated PFL and PFLA Data Sets. .................................................................................................................................40

Table 3-1 Primers used for the amplification of *Pygsuia bifor\textit{ma}* genes encoding SUFCB, NFU1 and ASCT. .................................................................................................................................73

Table 3-2 Proteins involved in amino acid metabolism identified in *Pygsuia bifor\textit{ma}*. .................................................................................................................................77

Table 3-3 Mitochondrial carrier proteins identified in *Pygsuia bifor\textit{ma}*. ORF and protein information as labeled in Table 3-2 .................................................................................................................................78

Table 3-4 Proteins involved in fatty acid, lipid and lipoate metabolism identified in *Pygsuia bifor\textit{ma}*. .................................................................................................................................78

Table 3-5 Proteins involved in pyruvate metabolism identified in *Pygsuia bifor\textit{ma}*. .................................................................................................................................79

Table 3-6 Proteins involved in Fe-S cluster biosynthesis and oxidative stress identified in *Pygsuia bifor\textit{ma}*. .................................................................................................................................79

Table 3-7 Proteins involved in uncategorized or incomplete metabolic processes identified in *Pygsuia bifor\textit{ma}*. .................................................................................................................................79

Table 3-8 Proteins involved in organellar protein import and folding identified in *Pygsuia bifor\textit{ma}*. .................................................................................................................................80

Table 3-9 Proteins involved in electron transport, the Kreb’s cycle and energy metabolism identified in *Pygsuia bifor\textit{ma}*. .................................................................................................................................81

Table 4-1 Accession numbers, sequence completeness and mitochondrial targeting predictions for eukaryotic *rqua* sequences identified from publicly available and in-house sequencing projects .................................................................................................................................122
List of Figures

Figure 1-1: Select metabolic pathways in typical aerobic mitochondria..........................5
Figure 1-2: Schematic representation of mitochondrion-related organelles across the
tree of eukaryotes .............................................................................................................6
Figure 1-3: Select metabolic pathways of Ascaris suum..................................................7
Figure 1-4: Select metabolic pathways of Acanthamoboa castellani mitochondria..........9
Figure 1-5: Select metabolic pathways of Naegleria gruberi mitochondria....................10
Figure 1-6: Select metabolic pathways of Nyctotherus ovalis mitochondria................11
Figure 1-7: Select metabolic pathways of Blastocystis sp. mitochondria.......................13
Figure 1-8: Select metabolic pathways of Trichomonas vaginalis hydrogenosomes........14
Figure 1-9: Select metabolic pathways of Spiroplasma salmonicida hydrogenosomes...15
Figure 1-10: Select metabolic pathways of Trimastix pyroformis hydrogenosomes........17
Figure 1-11: Select metabolic pathways of Mastigamoeba balamuthi
hydrogenosomes .............................................................................................................18
Figure 1-12: Select metabolism of Entamoeba histolytica ............................................19
Figure 1-13: Reduced metabolism of Mikrocytos mackini............................................20
Figure 1-14: Reduced metabolism of Giardia intestinalis.............................................21
Figure 1-15: General metabolism of some microsporidia (Encephalitozoon cuniculi,
Trichipleistophora hominis, Nosema ceranae and Antonospora locustae).....................22
Figure 1-16: Coulson plot of select mitochondrial and MRO-specific genes in select
eukaryotes.........................................................................................................................24
Figure 2-1: The distribution of pyruvate utilizing acetyl-CoA-generating enzymes
across eukaryotic diversity...............................................................................................43
Figure 2-2: Global phylogeny of pyruvate formate lyase in eukaryotes and bacteria......46
Figure 2-3: Global phylogeny of pyruvate formate lyase activating enzyme in
eukaryotes and bacteria.................................................................................................48
Figure 2-4 Phylogeny of concatenated pyruvate formate lyase and its activating
enzyme............................................................................................................................51
Figure 2-5: Atypical eukaryotic phylogeny of concatenated pyruvate formate lyase
and its activating enzyme..............................................................................................53
Figure 2-6: Radial diagrams of different eukaryotic relationships constrained for
topology testing. ........................................................................................................55

Figure 2-7: Removal of fast-evolving sites. .................................................................56

Figure 2-8: Hypothesis for the origin of eukaryotic PFL/PFLAE and its transfer
amongst eukaryotes ....................................................................................................58

Figure 3-1: Predicted MRO metabolism of selected Pygsuia biforma genes
determined by BLAST-mediated homology probing.................................................76

Figure 3-2: Phylogeny of pyruvate ferredoxin oxidoreductase (PFO) .......................83

Figure 3-3: Maximum-likelihood (ML) tree of HYDA (204 sequences, 282 sites). .......84

Figure 3-4: Phylogeny of hydrogenase maturase protein E (90 sequences, 187 sites) ....87

Figure 3-5: Phylogeny of hydrogenase maturase protein F (196 sequences, 232 sites). .85

Figure 3-6: Phylogeny of hydrogenase maturase protein G (87 sequences, 327 sites) ....86

Figure 3-7: Maximum-likelihood (ML) tree of ASCT1C (25 sequences, 478 sites) ........89

Figure 3-8: Maximum-likelihood (ML) tree of ASCT1B (73 sequences, 319 sites) ..........90

Figure 3-9: Maximum-likelihood (ML) tree of RQUA (28 sequences, 193 sites) rooted
with UbiE/COQ5 methyltransferase from Rhodoferax ferrireducens .................92

Figure 3-10: Sequence logo analysis of MTS from Pygsuia biforma matrix proteins. .....93

Figure 3-11: Maximum-likelihood (ML) tree of ISCS (210 sequences, 327 sites) ........95

Figure 3-12: Alignment of mitochondrial, bacterial and Breviate IscS sequences .........96

Figure 3-13: Maximum-likelihood (ML) tree of NFU (123 sequences, 75 sites). ...........97

Figure 3-14: Maximum-likelihood (ML) tree of SUFB (68 sequences, 301 sites) ..........99

Figure 3-15: Maximum-likelihood (ML) tree of SUFC (68 sequences, 195 sites) ..........100

Figure 3-16: Maximum-likelihood (ML) tree of concatenated SUFC and SUFB (68
sequences, 496 sites). .................................................................................................101

Figure 3-17: Alignment of select region of SUFB indicating shared insertion in
Pygsuia and Blastocystis. .........................................................................................102

Figure 3-18: Maximum-likelihood (ML) tree of phosphoenolpyruvate mutase
(PEPM; 65 sequences, 284 sites). ..........................................................................104

Figure 3-19: Maximum-likelihood (ML) tree of concatenated
Phosphonoacetaldehyde aminotransferase ..............................................................105

Figure 3-20: Maximum-likelihood (ML) tree of phosphonopyruvate decarboxylase
(PPDC; 69 sequences, 268 sites). ..........................................................................106
Figure 3-21: Localization of Pygsuia cSUFCB, mSUFCB, mSUFCB-MTS and NFU1 GFP fusion proteins in yeast........................................................................................................107

Figure 3-22: Antibodies raised against SUFCB and ASCT localize to Pygsuia MROs using immunofluorescent confocal microscopy. ..................................................108

Figure 3-23: Antibodies raised against ASCT and SUFCB localize to Pygsuia MROs using immunofluorescence confocal microscopy. ..........................109

Figure 4-1: Structure and electron potential of ubiquinone and rhodoquinone.........118

Figure 4-2: Malate dismutation and energy metabolism in mitochondria and related organelles........................................................................................................................................118

Figure 4-3: Phylogenetic analysis of all eukaryotic and bacterial RQUA homologs (54 sequences and 180 sites)..................................................................................126

Figure 4-4: Bayesian analysis of eukaryotic and bacterial RQUA sequences...........128

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).........................................................................................................129

Figure 4-6: Coulson plot of Q-utilizing and Q biosynthesis enzymes in eukaryotic organisms that encode rqua ........................................................................................................131

Figure 4-7: Antibodies raised against ASCT and SUFCB localize to Pygsuia MROs using immunofluorescence confocal microscopy. ............................................132

Figure 4-8: Two hypotheses for the origin of eukaryotic RQUA. .............................138
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 firmucite 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

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ace</td>
<td>Acetate</td>
</tr>
<tr>
<td>ACoA</td>
<td>Acetyl-Coa</td>
</tr>
<tr>
<td>ACS</td>
<td>Acetyl-Coa Synthetase</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>AK</td>
<td>Adenylate Kinase</td>
</tr>
<tr>
<td>Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>AlaAT</td>
<td>Alanine Aminotransferase</td>
</tr>
<tr>
<td>AOX</td>
<td>Alternative Oxidase</td>
</tr>
<tr>
<td>ASCT</td>
<td>Acetate:Succinyl-Coa Transferase</td>
</tr>
<tr>
<td>AspAT</td>
<td>Aspartate Aminotransferase</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td>BCAT</td>
<td>Brached Chain Amino Acid Aminotransferase</td>
</tr>
<tr>
<td>BI</td>
<td>Bayesian Inference</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>bp</td>
<td>Base Pairs</td>
</tr>
<tr>
<td>BV</td>
<td>Bootstrap Value</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CDP-DAG</td>
<td>Cytidine Phosphate Diacylglycerol Glycerol</td>
</tr>
<tr>
<td>CI</td>
<td>Complex I, NADH:Ubiquinone Oxidoreductase, NADH Dehydrogenase</td>
</tr>
<tr>
<td>CIA</td>
<td>Cytosolic Fe-S Cluster Assembley Machinery</td>
</tr>
<tr>
<td>CII</td>
<td>Complex II, Succinate Dehydrogenase; Succinate:Ubiquinone Oxidoreductase</td>
</tr>
<tr>
<td>CIIaf</td>
<td>Succinate Dehydrogenase Assembly Factor 2,</td>
</tr>
<tr>
<td>CIII</td>
<td>Complex III, Ubiquinol:Cytochrome C Oxidoreductase</td>
</tr>
<tr>
<td>Cit</td>
<td>Citrate</td>
</tr>
<tr>
<td>CIV</td>
<td>Complex IV, Cytochrome C Oxidase</td>
</tr>
<tr>
<td>CL</td>
<td>Cardiolipin</td>
</tr>
<tr>
<td>CLS</td>
<td>Cardiolipin Synthase</td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>Cpn</td>
<td>Mitochondrial Chaperonin</td>
</tr>
<tr>
<td>CS</td>
<td>Citrate Synthase-Like</td>
</tr>
<tr>
<td>CV</td>
<td>Complex V, ATP Synthase</td>
</tr>
<tr>
<td>Cym1</td>
<td>Presequence Protease</td>
</tr>
<tr>
<td>CYSJ</td>
<td>Sulfite Reductase Flavin Reductase Alpha Domain</td>
</tr>
<tr>
<td>DAPI</td>
<td>4′,6-Diamidino-2-Phenylindole</td>
</tr>
<tr>
<td>DHAP</td>
<td>Dihydroxyacetone Phosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>EGT</td>
<td>Endosymbiotic Gene Transfer</td>
</tr>
<tr>
<td>EST</td>
<td>Expressed Sequence Tag</td>
</tr>
<tr>
<td>ETC</td>
<td>Electron Transport Chain</td>
</tr>
<tr>
<td>ETF</td>
<td>Electron Transferring Flavoprotein</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ETFDH</td>
<td>ETF Dehydrogenase (ETF:Q Oxidoreductase)</td>
</tr>
<tr>
<td>FADH₂</td>
<td>Flavin Adenine Dinucleotide</td>
</tr>
<tr>
<td>Fe-S</td>
<td>Iron-Sulfur</td>
</tr>
<tr>
<td>FH</td>
<td>Fumarase, Fumarate Hydratase</td>
</tr>
<tr>
<td>FRD</td>
<td>Fumarate Reductase</td>
</tr>
<tr>
<td>Fum</td>
<td>Fumarate</td>
</tr>
<tr>
<td>G3PDH</td>
<td>Glycerol-3-Phosphate Dehydrogenase</td>
</tr>
<tr>
<td>GCS</td>
<td>Glycine Cleavage System</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>GK</td>
<td>Glycerol Kinase</td>
</tr>
<tr>
<td>Gly3p</td>
<td>Glycerol-3-Phosphate</td>
</tr>
<tr>
<td>GrpE-Mge1</td>
<td>Mitochondrial GrpE Chaperon-Mge1</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-Transferase</td>
</tr>
<tr>
<td>HADA</td>
<td>Trifunctional Enzyme Alpha Subunit</td>
</tr>
<tr>
<td>HADB</td>
<td>Trifunctional Enzyme Beta Subunit</td>
</tr>
<tr>
<td>HAR</td>
<td>Hydroxylamine Reductase</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>HPM</td>
<td>Hydrogen Producing Mitochondria</td>
</tr>
<tr>
<td>HTD2</td>
<td>3-Hydroxyacyl Thioester Dehydratase 2</td>
</tr>
<tr>
<td>HYDA</td>
<td>[Fefe]-Hydrogenase</td>
</tr>
<tr>
<td>HYDE-G</td>
<td>HYDA Maturase Proteins E-G</td>
</tr>
<tr>
<td>lcp55</td>
<td>Mitochondrial Aminopeptidase</td>
</tr>
<tr>
<td>IMM</td>
<td>Inner Mitochondrial Membrane</td>
</tr>
<tr>
<td>IMS</td>
<td>Intermembrane Space,</td>
</tr>
<tr>
<td>Ind</td>
<td>Indole</td>
</tr>
<tr>
<td>IND1</td>
<td>Fe-S Cluster Binding Protein, Ind1-Like</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl B-D-1-Thiogalactopyranoside</td>
</tr>
<tr>
<td>ISC</td>
<td>Iron-Sulfur Cluster</td>
</tr>
<tr>
<td>Kar-FabG</td>
<td>Ketoacyl Reductase (Fabg-Like)</td>
</tr>
<tr>
<td>KBL</td>
<td>2-Amino-3-Ketobutyrate Coa Ligase</td>
</tr>
<tr>
<td>Lac</td>
<td>Lactate</td>
</tr>
<tr>
<td>LDH</td>
<td>Lactate Dehydrogenase</td>
</tr>
<tr>
<td>Leu</td>
<td>Leucine</td>
</tr>
<tr>
<td>LGT</td>
<td>Lateral Gene Transfer</td>
</tr>
<tr>
<td>LplA-Lip3</td>
<td>Lipoate-Protein Ligase</td>
</tr>
<tr>
<td>Mal</td>
<td>Malate</td>
</tr>
<tr>
<td>MCF</td>
<td>Mitochondrial Carrier Protein</td>
</tr>
<tr>
<td>ME1</td>
<td>Malic Enzyme NAD-Dependent</td>
</tr>
<tr>
<td>MECR</td>
<td>Trans-2-Enoyl-Coa Reductase</td>
</tr>
<tr>
<td>Met</td>
<td>Metaxin</td>
</tr>
<tr>
<td>MIA40</td>
<td>Mitochondrial Intermembrane Space Assemble Protein</td>
</tr>
<tr>
<td>MIP</td>
<td>Mitochondrial Intermediate Peptidase</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>MM</td>
<td>Mitochondrial Matrix</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MMC</td>
<td>Methylmalonyl-Coa</td>
</tr>
<tr>
<td>MPC</td>
<td>Mitochondrial Pyruvate Carrier</td>
</tr>
<tr>
<td>MPP</td>
<td>Mitochondrial Processing Peptidase</td>
</tr>
<tr>
<td>MRO</td>
<td>Mitochondrion Related Organelle</td>
</tr>
<tr>
<td>Msf1/Ups2</td>
<td>Slowmo Protein - Mitochondrial Protein Sorting</td>
</tr>
<tr>
<td>mtDNA</td>
<td>Mitochondrial DNA</td>
</tr>
<tr>
<td>mtDnaJ</td>
<td>Mitochondrial Dnaj Chaperone</td>
</tr>
<tr>
<td>mtHsp70</td>
<td>Mitochondrial Heat Shock Protein 70</td>
</tr>
<tr>
<td>MTS</td>
<td>Mitochondrial Targeting Signal</td>
</tr>
<tr>
<td>NA</td>
<td>Numerical Aperature</td>
</tr>
<tr>
<td>NADH</td>
<td>Nicotinamide Adenine Dinucleotide</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Centre For Biotechnology Information</td>
</tr>
<tr>
<td>NFU1</td>
<td>NIFU-Like Protein</td>
</tr>
<tr>
<td>NIF</td>
<td>Nitrogen-Fixation</td>
</tr>
<tr>
<td>NPQDH</td>
<td>NAD(P)H Dehydrogenase (Quinone)</td>
</tr>
<tr>
<td>NUO</td>
<td>NADH:Ubiquionone Oxidoreductase (NADH Dehydrogenase)</td>
</tr>
<tr>
<td>Oaa</td>
<td>Oxaloacetate</td>
</tr>
<tr>
<td>OMM</td>
<td>Outer Mitochondrial Membrane</td>
</tr>
<tr>
<td>ORF</td>
<td>Open Reading Frame</td>
</tr>
<tr>
<td>Pam18</td>
<td>Mitochondrial Pam18-Like Protein</td>
</tr>
<tr>
<td>Pb</td>
<td>Pygsuia biforma</td>
</tr>
<tr>
<td>PCC</td>
<td>Propionyl-Coa Carboxylase</td>
</tr>
<tr>
<td>PDC</td>
<td>Pyruvate Dehydrogenase Complex</td>
</tr>
<tr>
<td>PE</td>
<td>Phosphatidyl Ethanolamine</td>
</tr>
<tr>
<td>PEP</td>
<td>Phosphoenolpyruvate</td>
</tr>
<tr>
<td>PEPM</td>
<td>Phosphoenolpyruvate Mutase</td>
</tr>
<tr>
<td>PFL</td>
<td>Pyruvate Formate Lyase</td>
</tr>
<tr>
<td>PFLA</td>
<td>PFL Activating Enzyme</td>
</tr>
<tr>
<td>PFO</td>
<td>Pyruvate:Ferredoxin Oxidoreductase</td>
</tr>
<tr>
<td>PG</td>
<td>Phosphatidylglycerol</td>
</tr>
<tr>
<td>PGP</td>
<td>Phosphatidylglycerol Phosphate</td>
</tr>
<tr>
<td>PGPS</td>
<td>CDP-Diacylglycerol--Glycerol-3-Phosphate 3-Phosphatidyltransferase</td>
</tr>
<tr>
<td>Pim1</td>
<td>Mitochondrial Serine Protease - Lon Protease, Pim1-Like</td>
</tr>
<tr>
<td>PNO</td>
<td>Pyruvate:NADPH Oxidoreductase</td>
</tr>
<tr>
<td>PP</td>
<td>Posterior Probabilities</td>
</tr>
<tr>
<td>PPA</td>
<td>Phosphonoacetaldehyde</td>
</tr>
<tr>
<td>PPC</td>
<td>Propc Carboxylase</td>
</tr>
<tr>
<td>PPD</td>
<td>Phosphonopyruvate Decarboxylase</td>
</tr>
<tr>
<td>PPyr</td>
<td>Phosphonopyruvate</td>
</tr>
<tr>
<td>Pro</td>
<td>Prohibitin</td>
</tr>
<tr>
<td>ProP</td>
<td>Propionyl-Coa</td>
</tr>
<tr>
<td>PRX</td>
<td>Peroredoxin</td>
</tr>
<tr>
<td>PS</td>
<td>Phosphatidyl Serine</td>
</tr>
<tr>
<td>PSD1</td>
<td>Phosphotidylserine Decarboxylase 1</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PTPMT1</td>
<td>Protein-Tyrosine Phosphatase Mitochondrial</td>
</tr>
<tr>
<td>rDNA</td>
<td>Ribosomal DNA</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>RQ</td>
<td>Rhodoquinone</td>
</tr>
<tr>
<td>RQUA</td>
<td>Rhodoquinone Biosynthesis Protein RQUA</td>
</tr>
<tr>
<td>SAM</td>
<td>S-Adenosyl Methionine</td>
</tr>
<tr>
<td>Sam50</td>
<td>Sorting And Assembly Machinery 50</td>
</tr>
<tr>
<td>SAR</td>
<td>Stromatopile-Alveolae-Rhizaria</td>
</tr>
<tr>
<td>SCS</td>
<td>Succinyl-Coa Synthetase, Succinate Thiokinase</td>
</tr>
<tr>
<td>SDHA</td>
<td>Succinate Dehydrogenase Flavoprotein Subunit</td>
</tr>
<tr>
<td>SDHB</td>
<td>Sdh Fe-S Subunit</td>
</tr>
<tr>
<td>SDHC</td>
<td>SDH Large Cytochrome B Subunit</td>
</tr>
<tr>
<td>SDHD</td>
<td>SDH Small Cytochrome B Subunit</td>
</tr>
<tr>
<td>SDR</td>
<td>Short Chain Dehydrogenase</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>Ser</td>
<td>Serine</td>
</tr>
<tr>
<td>SHMT</td>
<td>Serine Hydroxymethyl Transferase</td>
</tr>
<tr>
<td>SOD</td>
<td>Superoxide Dismutase</td>
</tr>
<tr>
<td>SSU</td>
<td>Small Subunit</td>
</tr>
<tr>
<td>Suc</td>
<td>Succinate</td>
</tr>
<tr>
<td>SUF</td>
<td>Sulfur Mobilization</td>
</tr>
<tr>
<td>SUFCB</td>
<td>Sulfur Assimilation CB-D Fusion, Cytosolic</td>
</tr>
<tr>
<td>SW802</td>
<td>Salt Water 802 Medium</td>
</tr>
<tr>
<td>Tam41</td>
<td>Translocator Assembly And Maintenance Protein 41</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris-Buffered Saline</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxylic Acid Or Citric Acid</td>
</tr>
<tr>
<td>TDH</td>
<td>Threonine Dehydrogenase</td>
</tr>
<tr>
<td>Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>TIM</td>
<td>Translocator Of The Inner Mitochondrial Membrane</td>
</tr>
<tr>
<td>TnaA</td>
<td>Tryptophanase</td>
</tr>
<tr>
<td>Trp</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>UBI</td>
<td>Ubiquionone Biosynthesis Genes</td>
</tr>
<tr>
<td>UPGMA</td>
<td>Unweighted Pair Group Method With Arithmetic Mean</td>
</tr>
<tr>
<td>UQ</td>
<td>Ubiquinone</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>YPAD</td>
<td>Yeast Extract, Peptone, Adenine And Dextrose Medium</td>
</tr>
</tbody>
</table>
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, Mereschkowsky 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 mitochondrial targeted proteins to one of four sub-organelar 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 FADH₂) that are used by the electron transport chain (ETC) to generate a proton gradient which fuels ATP synthesis ultimately reducing O₂ to H₂O (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).

![Aerobic mitochondria](image)

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 (QH₂) 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 QH₂. 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 CII 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 (RQH₂). 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 RQH₂. These organisms also generate ATP via propionyl-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 propionyl-CoA carboxylase (PPC) to generate ATP and propionyl-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 (RQH₂). Malate dismutation: Malate (Mal) is converted
to fumarate (Fum) via fumarate hydratase [6]. Electrons from rhodoquinol and transferred to fum to generate succinate (suc) via CI. 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 RQH₂. 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:ferrodoxin 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 ferrodoxin 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 Acanthamoeba 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 Cl generating ubiquinone (QH2). 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).

*Naegleria gruberi*

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 (QH2). Electrons from ubiquinol and transferred to fum to generate succinate (suc) via CI. 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 RQH2 (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...
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 Cl 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 H₂ 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 Cl generating rhodoquinol (RQH₂). Electrons from rhodoquinol and transferred to Fum to generate Succ 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).
**Blastocystis sp.**

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 ferredoxin are used to produce hydrogen via hydrogenase [11]. A proton gradient is generated via CI generating rhodoquinol (RQH₂). 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 *Acanthamoeba* 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 ASC (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.

*Trichomonas vaginalis*

![Diagram of metabolic pathways](image)

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.

![Spironucleus salmonicida](image)

**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 (Pyr) 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.
**Trimastix pyroformis**

![Diagram of Trimastix pyroformis metabolic pathways]

Figure 1-10: Select metabolic pathways of *Trimastix pyroformis* hydrogenosomes. Numbering and labels are shown as 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 amoeboid 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 *Spirocrus* 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) (Nyvltová, 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; Nyvtová 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 (Nyvtová et al. 2013).

![Mastigamoeba balamuthi](image)

**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.](image)

**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 cytoplastsms like other intracellular parasites (Schmitz-Esser et al. 2004; Tsaousis et al. 2008).

![Mikrocytos mackini](image)

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).
**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.*, *Encephalitozoon cuniculi* (Katinka *et al.* 2001) and *Trachipleistophora hominis* (Williams *et al.* 2002)), and insects (*e.g.*, *Nozema 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 (*e.g.* *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 (Tsousis *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-phosphate 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 (Encephalitozoon 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 projects to identify homologs in different organisms. Each pie represents the super family of organisms surveyed: Obazoza (Ec, Encephalitozoon cuniculi; Sc, Saccharomyces cerevisiae; Pb, Pygsula biforma; Nf, Neocallostax frontalis; Ps, Piromyces sp.), Amoebozoa (Ac, Acanthamoeba castellani; Dd, Dictyostelium discoideum; Eh, Entamoeba histolytica; Mb, Mastigamoeba balalmuthi), SAR clade (Bh, Blastocystis hominis; Mm, Mircycytos mackini; No, Nyctotherus ovalis) and Excavata (Tp, Trimastix pyroformis; Gi, Giardia intestinalis; Ss, Spironucleus salmonicida; 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 microsporidian genomes. Abbreviations not mentioned in text: GCS, glycine cleavage system; SHMT, serine hydroxymethyltransferase; PEMP, phosphoenolpyruvate
mutase; PPD, phenopyruvate 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-Garcia 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 $\text{H}_2$-dependent methanogenic archaeon. In an anaerobic environment, the waste products of the symbiont ($\text{H}_2$, $\text{CO}_2$ 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 $\gamma$- or $\delta$-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 $\alpha$-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 $\alpha$-proteobacteria but instead often $\delta$–/$\varepsilon$– 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 $\alpha$-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 $\alpha$-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 ‘$\alpha$-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 $\alpha$-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 \( \alpha \)-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 \( \alpha \)-proteobacteria (Müller et al. 2012). Conveniently, this argument neglects the observation that many typically ‘aerobic’ genes (e.g., pyruvate dehydrogenase) shared by \( \alpha \)-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 \( \alpha \)-proteobacterial affinity) (Emelyanov 2003). The foregoing convoluted arguments explaining why anaerobic enzymes in eukaryotes do not show clear \( \alpha \)-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


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 archaeoplastidan, 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 (Lehtio 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 CO₂ 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 (Gélius-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 Archaeoplastida. 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$_2$H$_2$PO$_4$, 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 (BLASTN), 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: ATGTCCGGCTCTATCCGG and reverse primer MbPFL_909R: CTATGAGCAGCAGGCACGC). 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 diagnosed conserved active site motif (CC, Lehtiö and Goldman 2004), glycy radical domain ([LIVM]R[LIVM]$_2$GY, 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 Cysl 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 PFIA 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.) were generated using the adjusted bootstrap probability (aBPn) program under the LG + 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. 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.

<table>
<thead>
<tr>
<th>Phylogeny</th>
<th>Monophyly Test</th>
<th>AU Test P value</th>
<th>AU Test P value of Partitioned Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PFL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^2) and γ-proteobacteria</td>
<td>0.0004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^3) and α-proteobacteria</td>
<td>0.0024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^5) and α-clade</td>
<td>0.0130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^5) and cyanobacteria</td>
<td>0.1580</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PFLA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^2) and γ-proteobacteria</td>
<td>0.8010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^3) and α-proteobacteria</td>
<td>0.0010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^5) and α-clade</td>
<td>0.4660</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^5) and cyanobacteria</td>
<td>0.0740</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Concatenated tree(^2) (PFL–PFLA)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^2) and α-proteobacteria</td>
<td>2 × 10(^{-22})</td>
<td>5 × 10(^{-38})</td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^3) and α-clade</td>
<td>0.0130</td>
<td>0.0040</td>
<td></td>
</tr>
<tr>
<td>Eukaryotes(^5) and cyanobacteria</td>
<td>1 × 10(^{-5})</td>
<td>0.0090</td>
<td></td>
</tr>
<tr>
<td><strong>Concatenated tree (PFL–PFLA; exclusively eukaryotes(^5))</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical eukaryotic phylogeny</td>
<td>0.0020</td>
<td>0.0030</td>
<td></td>
</tr>
<tr>
<td>Chlorophytes and prasinophytes</td>
<td>0.0100</td>
<td>0.0080</td>
<td></td>
</tr>
<tr>
<td>Chlorophytes, prasinophytes, and Mastigamoeba</td>
<td>0.0380</td>
<td>0.0330</td>
<td></td>
</tr>
<tr>
<td>Fungi and prasinophytes</td>
<td>0.5230</td>
<td>0.4660</td>
<td></td>
</tr>
<tr>
<td>Opisthokonts</td>
<td>0.6610</td>
<td>0.6650</td>
<td></td>
</tr>
</tbody>
</table>

\(^2\)ML tree search with the indicated constraint.

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

\(^5\)For concatenation, the following data were removed: *Brugia malayi*, *Litopenaeus 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 cyS domain typically associated with pyruvate:NADP+ oxidoreductases (PNO) or −SR for those sequences with cyS1 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, BVc = 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, BVc = 97, PP = 1.0; Figure 2-3). To test whether eukaryotic monophylly 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 monophylly 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 \( \gamma \)-proteobacterial clade relative to other sequences (this is the deepest significant split in the tree), removal of the \( \gamma \)-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 \( \alpha \)-proteobacteria (mitochondria) or cyanobacteria (chloroplasts) respectively were significantly rejected by the data (Table 2-1). Eukaryotes + \( \alpha \)-proteobacteria monophyly was rejected for both PFL (AU \( P \) value \( [P] = 0.002 \)) and PFLA (\( P = 0.001 \)). As I noted that the \( \alpha \)-proteobacteria often grouped with the \( \beta \)-proteobacteria and actinobacteria in optimal trees I tested whether the eukaryotic affinity for this group as a whole (\( \alpha \)-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.) 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 *fimbicute-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 Pfi 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<sub>c</sub> = 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–fimicute–<i>Bacteroides</i> grouping with strong support values (BV = 84, BV<sub>c</sub> = 96, PP = 1.0). All alternative topologies were rejected by AU tests (Table 2-1), including cyanobacteria + eukaryote monophyly (P<sub>unpart</sub> = 0.00005 and P<sub>part</sub> = 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 (BVr) 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, BVc = 98, and PP = 0.93) and firmicute–
Bacteroides affinity (BV = 89, BVc = 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.](image)

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, BVc = 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 (*Neocalamastix 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 (Hampf *et al.* 2009; Nozaki *et al.* 2009). Finally, the haptophyte and diatoms branch together with strong support (BV = 100, BVc = 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 ichthyosporean) and monophyly of the bikonts with the haptophyte and stramenopile (diatom) to the immediate exclusion of the archaeaplastid 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_{unpart} = 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.

![Radial diagrams of different eukaryotic relationships](image)

**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* ($P_{PHL} = 0.0001$, $P_{PFLA} = 0.00384$) 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)

![Graph A](image)

**Support for eukaryotic monophyly with removal of fast evolving sites**

![Graph B](image)

**Support for eukaryote-firmicute affinity with removal of fast evolving sites**

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 monophly (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 archaeoplastidan 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.

![Diagram](image)

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
*pfila* 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 *pfila* 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 *pfala* 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 *pfala* genes. Within the *Micromonas* species, there was intron position conservation in *pfala* 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; Gélius-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* ($P_{Target} = 0.319$ 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 mitochondrial 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*


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$_2$O. 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.
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 biformal, 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 biformal 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 pneumoniae 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 Pygsuia model versus
being from another source. These seven exceptions were nevertheless verified to be from Pygsuaia 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 pcca2, 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 parologue of four other Pygsuaia 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 terasporsa 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 terasporsa, using the Pygsuia homologues as queries. Due to the low sequence coverage and protein complement identified in the Breviata anathema and Subulatomonas terasporsa 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 < $10^{-5}$ 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 previously 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 NH₄Cl, 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 SUF CB, N FU1 and AS CT.

<table>
<thead>
<tr>
<th>Primer Description</th>
<th>Direction</th>
<th>Sequence</th>
<th>Restriction Enzyme Site</th>
<th>Destination Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. biforma msufcb</em> 1F; HindIII</td>
<td>Forward</td>
<td>GGGAAAGCTTATGCCTAGGGGCTTTAAC</td>
<td>HindIII</td>
<td>pTS395 (C-term GFP)</td>
</tr>
<tr>
<td><em>P. biforma msufcb</em> 2034R; HindIII</td>
<td>Reverse</td>
<td>GGGCAAGCTTGGAGAGGCGCCCTAAC</td>
<td>HindIII</td>
<td>pTS395 (C-term GFP)</td>
</tr>
<tr>
<td><em>P. biforma msufcb</em> MTS 111R; HindIII</td>
<td>Forward</td>
<td>GGGAAAGCTTATGCCTAGGGGCTTTAAC</td>
<td>HindIII</td>
<td>pTS395 (C-term GFP)</td>
</tr>
<tr>
<td><em>P. biforma msufcb</em> MTS 2034R; HindIII</td>
<td>Reverse</td>
<td>GGGCAAGCTTGGAGAGGCGCCCTAAC</td>
<td>HindIII</td>
<td>pTS395 (C-term GFP)</td>
</tr>
<tr>
<td><em>P. biforma asct</em> 1F; HindIII</td>
<td>Forward</td>
<td>CCCAAAGCTTATGCCTAGGGGCTTTAAC</td>
<td>HindIII</td>
<td>pTS395 (C-term GFP)</td>
</tr>
<tr>
<td><em>P. biforma asct</em> 2034R; HindIII</td>
<td>Reverse</td>
<td>CCCAAAGCTTATGCCTAGGGGCTTTAAC</td>
<td>HindIII</td>
<td>pTS395 (C-term GFP)</td>
</tr>
<tr>
<td><em>P. biforma nfu1</em> 1F; HindIII</td>
<td>Forward</td>
<td>GGGAAAGCTTATGCCTAGGGGCTTTAAC</td>
<td>HindIII</td>
<td>pTS395 (C-term GFP)</td>
</tr>
<tr>
<td><em>P. biforma nfu1</em> 398R; HindIII</td>
<td>Reverse</td>
<td>GGGAAAGCTTATGCCTAGGGGCTTTAAC</td>
<td>HindIII</td>
<td>pTS395 (C-term GFP)</td>
</tr>
<tr>
<td>pTS395_GFPR</td>
<td>Reverse</td>
<td>CTTCCTTTATCAT</td>
<td>Ndel</td>
<td>pET16b (His Tag)</td>
</tr>
<tr>
<td><em>P. biforma asct</em> 1F; Ndel</td>
<td>Forward</td>
<td>GGGCCATATATGCGACTCTGTGCTGGTAC</td>
<td>Ndel</td>
<td>pET16b (His Tag)</td>
</tr>
<tr>
<td><em>P. biforma asct</em> 891R; Ndel</td>
<td>Reverse</td>
<td>GGGCCATATATGCGACTCTGTGCTGGTAC</td>
<td>Ndel</td>
<td>pET16b (His Tag)</td>
</tr>
<tr>
<td><em>P. biforma msufcb</em> 1F; Xhol</td>
<td>Forward</td>
<td>GGGTCGAGATGCTAGGGGCTTTAAC</td>
<td>Xhol</td>
<td>pGEX-4T-1 (GST Tag)</td>
</tr>
<tr>
<td><em>P. biforma msufcb</em> 2034R; Xhol</td>
<td>Reverse</td>
<td>GGGTCGAGATGCTAGGGGCTTTAAC</td>
<td>Xhol</td>
<td>pGEX-4T-1 (GST Tag)</td>
</tr>
</tbody>
</table>

### 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 microscopy (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 (4000 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 chemiluminescence 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, IIS chaperone/lowmo 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: Fe⁴⁺ carrier], [Rhodoquinone (dark pink): UQ: ubiquinone; RQ: rhodoquinone; RQUA: RQ methyltransferase], [Lipoate metabolism (light pink): LIP: lipoate; PLPA: 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; CII, succinate DH; AF, CII assembly factor; F, flavin; SCS, succinyl-CoA synthetase; ASC; acetate: succinyl-CoA transferase(s); DHAP; dihydroxyacetone phosphate; Gly3p, glycerol-3-phosphate; G3PD, Gly3p DH; ETF: electron transferring flavoprotein; AOX: alternative oxidase; Q/QH2, 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, hydroxacyl DH, *acyl-carrier protein is used in place of CoA during biosynthesis], [Amino acid metabolism (light brown): Thr, threonine; TDH, Thr DH; KBL, keto-butyrte lyase; Trp, tryptophan; TRN, tryptophanase; Ind, indole; Gly, glycine, Ala, alanine; ALAT, aminotransferase (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: peroxiredoxin; Prx5: peroxidase], [Lipid metabolism (orange): CDP-DAG, cytidine diphosphate diacylglycerol; PGPS, CDP-DAG-glycerol-3-phosphate 3-phosphatidyltransferase; PGP, phosphatidylglycerol phosphatase; 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>
<thead>
<tr>
<th>Protein Name</th>
<th>Short Name</th>
<th>Complete</th>
<th>ORF Coverage</th>
<th>Predicted Localization</th>
<th>Mitoprot Prediction</th>
<th>TargetP prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine Aminotransferase</td>
<td>AlaAT</td>
<td>C</td>
<td>285.74</td>
<td>MM</td>
<td>0.7871</td>
<td>0.749</td>
</tr>
<tr>
<td>Aspartate Aminotransferase</td>
<td>AspAT</td>
<td>C</td>
<td>166.84</td>
<td>MM</td>
<td>0.9165</td>
<td>0.558</td>
</tr>
<tr>
<td>Brached Chain Amino Acid Aminotransferase</td>
<td>BCAT</td>
<td>C</td>
<td>406.11</td>
<td>MM</td>
<td>0.5406</td>
<td>0.785</td>
</tr>
<tr>
<td>Lipoamide protein</td>
<td>GcaH</td>
<td>C</td>
<td>2,775.97</td>
<td>MM</td>
<td>0.9729</td>
<td>0.86</td>
</tr>
<tr>
<td>Dihydrolipoamide dehydrogenase</td>
<td>GcsL</td>
<td>C</td>
<td>236.52</td>
<td>MM</td>
<td>0.9925</td>
<td>0.823</td>
</tr>
<tr>
<td>Glycine decarboxylase 1</td>
<td>GcsP1</td>
<td>C</td>
<td>149.69</td>
<td>MM</td>
<td>0.9896</td>
<td>0.91</td>
</tr>
<tr>
<td>Glycine decarboxylase 2</td>
<td>GcsP2</td>
<td>C</td>
<td>353.92</td>
<td>MM</td>
<td>0.999</td>
<td>0.941</td>
</tr>
<tr>
<td>Aminomethyl transferase</td>
<td>GcsT</td>
<td>C</td>
<td>1,275.96</td>
<td>MM</td>
<td>0.977</td>
<td>0.927</td>
</tr>
<tr>
<td>2-amino-3-ketobuterate CoA ligase</td>
<td>KBL</td>
<td>C</td>
<td>1,815.70</td>
<td>MM</td>
<td>0.9487</td>
<td>0.91</td>
</tr>
<tr>
<td>Serine Hydroxymethyl Transferase</td>
<td>SHMT</td>
<td>C</td>
<td>141.86</td>
<td>MM</td>
<td>0.9383</td>
<td>0.87</td>
</tr>
<tr>
<td>Threonine Dehydrogenase</td>
<td>TDH</td>
<td>C</td>
<td>491.41</td>
<td>MM</td>
<td>0.9447</td>
<td>0.917</td>
</tr>
<tr>
<td>Tryptophanase</td>
<td>TnaA</td>
<td>C</td>
<td>331.1</td>
<td>MM</td>
<td>0.686</td>
<td>0.605</td>
</tr>
</tbody>
</table>
Table 3-3 Mitochondrial carrier proteins identified in *Pygsua biforma*. ORF and protein information as labeled in Table 3-2

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

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

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Short Name</th>
<th>Complete</th>
<th>ORF Coverage</th>
<th>Predicted Localization</th>
<th>Mitoprot Prediction</th>
<th>TargetP prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trifunctional enzyme alpha subunit</td>
<td>HADA</td>
<td>C</td>
<td>40.43</td>
<td>MM</td>
<td>0.4216</td>
<td>0.435</td>
</tr>
<tr>
<td>Trifunctional enzyme beta subunit</td>
<td>HADB</td>
<td>C</td>
<td>3,022.23</td>
<td>MM</td>
<td>0.281</td>
<td>0.552</td>
</tr>
<tr>
<td>Hydroxylamine reductase</td>
<td>HAR</td>
<td>C</td>
<td>213.7</td>
<td>MM</td>
<td>0.9889</td>
<td>0.951</td>
</tr>
<tr>
<td>3-hydroxyacyl thioester dehydratase 2</td>
<td>HDT2</td>
<td>C</td>
<td>158.12</td>
<td>MM</td>
<td>0.8448</td>
<td>0.923</td>
</tr>
<tr>
<td>Ketoacyl reductase</td>
<td>Kar</td>
<td>C</td>
<td>45.16</td>
<td>MM</td>
<td>0.0676</td>
<td>0.018</td>
</tr>
<tr>
<td>Ketoacyl reductase (FabG-like)</td>
<td>Kar-FabG</td>
<td>C</td>
<td>128.59</td>
<td>MM</td>
<td>0.3941</td>
<td>0.069</td>
</tr>
<tr>
<td>Trans-2-enoyl-CoA reductase</td>
<td>MECR</td>
<td>C</td>
<td>176.87</td>
<td>MM</td>
<td>0.9576</td>
<td>0.793</td>
</tr>
<tr>
<td>Cardiolipin Synthase</td>
<td>CLS</td>
<td>C</td>
<td>32.38</td>
<td>MM</td>
<td>0.9912</td>
<td>0.872</td>
</tr>
<tr>
<td>Phosphoenolpyruvate mutase</td>
<td>PEPM</td>
<td>C</td>
<td>8,644.50</td>
<td>MM</td>
<td>0.9937</td>
<td>0.73</td>
</tr>
<tr>
<td>CDP-diacylglycerol–glycerol-3-phosphate 3-</td>
<td>PGPS</td>
<td>C</td>
<td>31.19</td>
<td>MM</td>
<td>0.6864</td>
<td>0.828</td>
</tr>
<tr>
<td>phosphatidyltransferase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphatidylserine decarboxylase 1</td>
<td>Psd1</td>
<td>C</td>
<td>68.02</td>
<td>IMS</td>
<td>0.8706</td>
<td>0.881</td>
</tr>
<tr>
<td>Protein-tyrosine phosphatase mitochondrial</td>
<td>TPTMFT1</td>
<td>C</td>
<td>83.82</td>
<td>MM</td>
<td>0.9123</td>
<td>0.163</td>
</tr>
<tr>
<td>Lipoate-protein ligase</td>
<td>LplA-Lip3</td>
<td>C</td>
<td>246.7</td>
<td>MM</td>
<td>0.8355</td>
<td>0.655</td>
</tr>
</tbody>
</table>
Table 3-5 Proteins involved in pyruvate metabolism identified in *Pygsua biforma*. ORF and protein information as labeled in Table 3-2

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

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

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Short Name</th>
<th>Complete</th>
<th>ORF Coverage</th>
<th>Predicted Localization</th>
<th>Mitoprot Prediction</th>
<th>TargetP prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe-S cluster binding protein, Ind1-like</td>
<td>Ind1</td>
<td>C</td>
<td>243.4</td>
<td>MM</td>
<td>0.9623</td>
<td>0.948</td>
</tr>
<tr>
<td>Sulfur Assimilation CB-D fusion</td>
<td>mSUFCB</td>
<td>C</td>
<td>140.88</td>
<td>MM</td>
<td>0.993</td>
<td>0.951</td>
</tr>
<tr>
<td>Sulfur Assimilation CB-D fusion, cytosolic</td>
<td>cSuFcb</td>
<td>C</td>
<td>509.78</td>
<td>Cytosolic</td>
<td>0.0218</td>
<td>0.07</td>
</tr>
<tr>
<td>NIFU-like protein</td>
<td>Nfu1</td>
<td>C</td>
<td>1,596.36</td>
<td>MM</td>
<td>0.9993</td>
<td>0.952</td>
</tr>
<tr>
<td>Peroxisome, Mitochondrial-like</td>
<td>Pyx2</td>
<td>C</td>
<td>6,623.76</td>
<td>MM</td>
<td>0.2333</td>
<td>0.111</td>
</tr>
<tr>
<td>Glutathione amide-dependent peroxidase</td>
<td>Pyx5</td>
<td>C</td>
<td>2,638.06</td>
<td>MM</td>
<td>0.9955</td>
<td>0.865</td>
</tr>
<tr>
<td>Superoxide dismutase 1 (Fe-Mn)</td>
<td>SOD1</td>
<td>C</td>
<td>1,460.03</td>
<td>MM</td>
<td>0.871</td>
<td>0.489</td>
</tr>
<tr>
<td>Superoxide dismutase 2 (Fe-Mn)</td>
<td>SOD2</td>
<td>C</td>
<td>420.1</td>
<td>MM</td>
<td>0.6243</td>
<td>0.784</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Short Name</th>
<th>Complete</th>
<th>ORF Coverage</th>
<th>Predicted Localization</th>
<th>Mitoprot Prediction</th>
<th>TargetP prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acyl-CoA synthetase-like protein</td>
<td>AcSyn</td>
<td>C</td>
<td>45.77</td>
<td>MM</td>
<td>0.994</td>
<td>0.835</td>
</tr>
<tr>
<td>Adenylate kinase</td>
<td>AK</td>
<td>C</td>
<td>27.47</td>
<td>MM</td>
<td>0.9937</td>
<td>0.927</td>
</tr>
<tr>
<td>Fatty acid synthase</td>
<td>FoIC</td>
<td>C</td>
<td>8.54</td>
<td>MM</td>
<td>0.9358</td>
<td>0.814</td>
</tr>
<tr>
<td>Glycerol kinase</td>
<td>GK</td>
<td>C</td>
<td>40.89</td>
<td>MM</td>
<td>0.9771</td>
<td>0.715</td>
</tr>
<tr>
<td>Mitochondrial aminopeptidase</td>
<td>Ip55</td>
<td>C</td>
<td>905.83</td>
<td>MM</td>
<td>0.904</td>
<td>0.778</td>
</tr>
<tr>
<td>NAD(P)H dehydrogenase (quinone)</td>
<td>NPQD</td>
<td>C</td>
<td>86.51</td>
<td>MM</td>
<td>0.9797</td>
<td>0.826</td>
</tr>
<tr>
<td>Short chain dehydrogenase</td>
<td>SDR</td>
<td>C</td>
<td>654</td>
<td>MM</td>
<td>0.9335</td>
<td>0.733</td>
</tr>
<tr>
<td>Propinoyl-CoA carboxylase alpha subunit</td>
<td>PCCa1</td>
<td>3’ end IC</td>
<td>9.37</td>
<td>MM</td>
<td>0.7924</td>
<td>0.542</td>
</tr>
<tr>
<td>Propinoyl-CoA carboxylase beta subunit</td>
<td>PCCb</td>
<td>Complete</td>
<td>167.25</td>
<td>MM</td>
<td>0.4932</td>
<td>0.59</td>
</tr>
<tr>
<td>Propinoyl-CoA carboxylase alpha subunit</td>
<td>PCCa2</td>
<td>5’ end IC</td>
<td>18.4</td>
<td>Unpredictable</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Short Name</th>
<th>Complete</th>
<th>ORF Coverage</th>
<th>Predicted Localization</th>
<th>Mitoprot Prediction</th>
<th>TargetP prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrial Chaperin 10</td>
<td>Cpn10</td>
<td>C</td>
<td>651.22</td>
<td>MM</td>
<td>0.8906</td>
<td>0.455</td>
</tr>
<tr>
<td>Mitochondrial Chaperin 60</td>
<td>Cpn60</td>
<td>C</td>
<td>142.62</td>
<td>MM</td>
<td>0.854</td>
<td>0.761</td>
</tr>
<tr>
<td>Presequence Protease</td>
<td>Cym1</td>
<td>C</td>
<td>17.06</td>
<td>IMS</td>
<td>0.7599</td>
<td>0.578</td>
</tr>
<tr>
<td>Mitochondrial GrrE Chaperin-Mge1</td>
<td>Grr-E-Mge1</td>
<td>C</td>
<td>45.56</td>
<td>MM</td>
<td>0.9077</td>
<td>0.920</td>
</tr>
<tr>
<td>Metaxin</td>
<td>Met</td>
<td>C</td>
<td>662.31</td>
<td>OM</td>
<td>0.0609</td>
<td>0.061</td>
</tr>
<tr>
<td>Mitochondrial Intermediate Peptidase</td>
<td>MIP</td>
<td>C</td>
<td>27.8</td>
<td>MM</td>
<td>0.8053</td>
<td>0.853</td>
</tr>
<tr>
<td>Mitochondrial processing peptidase alpha subunit</td>
<td>MPPalpha</td>
<td>C</td>
<td>32.97</td>
<td>MM</td>
<td>0.9929</td>
<td>0.872</td>
</tr>
<tr>
<td>Mitochondrial processing peptidase beta subunit</td>
<td>MPPbeta</td>
<td>C</td>
<td>54.95</td>
<td>MM</td>
<td>0.9157</td>
<td>0.817</td>
</tr>
<tr>
<td>Slowmo protein - Mitochondrial protein sorting</td>
<td>Mfs1/Ups2</td>
<td>C</td>
<td>320.77</td>
<td>IMS</td>
<td>0.3486</td>
<td>0.173</td>
</tr>
<tr>
<td>Mitochondrial DnaJ Chaperone</td>
<td>mtDnaJ</td>
<td>C</td>
<td>180.49</td>
<td>MM</td>
<td>0.9951</td>
<td>0.878</td>
</tr>
<tr>
<td>Mitochondrial Heat Shock Protein 70</td>
<td>mtHsp70</td>
<td>C</td>
<td>225.85</td>
<td>MM</td>
<td>0.9864</td>
<td>0.848</td>
</tr>
<tr>
<td>Mitochondrial Pam18-like protein</td>
<td>Pam18</td>
<td>C</td>
<td>259.97</td>
<td>IM</td>
<td>0.8841</td>
<td></td>
</tr>
<tr>
<td>Mitochondrial Serine Protease - Lon protease, Pim1-like</td>
<td>Pim1</td>
<td>C</td>
<td>47.85</td>
<td>MM</td>
<td>0.968</td>
<td>0.808</td>
</tr>
<tr>
<td>Prohibitin</td>
<td>Pro</td>
<td>C</td>
<td>184.88</td>
<td>IM</td>
<td>0.708</td>
<td>0.105</td>
</tr>
<tr>
<td>Sorting and Assembly Machinery 50</td>
<td>Sam50</td>
<td>C</td>
<td>106.88</td>
<td>OM</td>
<td>0.118</td>
<td>0.04</td>
</tr>
<tr>
<td>Translocator assembly and maintenance protein 41</td>
<td>Tam41</td>
<td>C</td>
<td>56.82</td>
<td>IM</td>
<td>0.2363</td>
<td>0.084</td>
</tr>
<tr>
<td>Translocator of the inner mitochondrial membrane 10</td>
<td>Tim10</td>
<td>C</td>
<td>36.04</td>
<td>IMS</td>
<td>0.0948</td>
<td>0.174</td>
</tr>
<tr>
<td>Translocator of the inner mitochondrial membrane 13</td>
<td>Tim13</td>
<td>C</td>
<td>180.33</td>
<td>IMS</td>
<td>0.0553</td>
<td>0.072</td>
</tr>
<tr>
<td>Translocator of the inner mitochondrial membrane 16</td>
<td>Tim16/ Tim16</td>
<td>C</td>
<td>127.12</td>
<td>IM</td>
<td>0.7396</td>
<td>0.076</td>
</tr>
<tr>
<td>Translocator of the inner mitochondrial membrane 17</td>
<td>Tim17</td>
<td>C</td>
<td>52.97</td>
<td>IM</td>
<td>0.0405</td>
<td>0.047</td>
</tr>
<tr>
<td>Translocator of the inner mitochondrial membrane 22</td>
<td>Tim22</td>
<td>C</td>
<td>41.39</td>
<td>IM</td>
<td>0.2897</td>
<td>0.322</td>
</tr>
<tr>
<td>Translocator of the inner mitochondrial membrane 44</td>
<td>Tim44</td>
<td>C</td>
<td>59.15</td>
<td>IM</td>
<td>0.9991</td>
<td>0.864</td>
</tr>
<tr>
<td>Translocator of the inner mitochondrial membrane 50</td>
<td>Tim50</td>
<td>C</td>
<td>3’ end IC</td>
<td>8.73</td>
<td>0.9745</td>
<td>0.89</td>
</tr>
<tr>
<td>Translocator of the inner mitochondrial membrane 8</td>
<td>Tim8</td>
<td>C</td>
<td>248.55</td>
<td>IMS</td>
<td>0.0182</td>
<td>0.067</td>
</tr>
<tr>
<td>Translocator of the inner mitochondrial membrane 9</td>
<td>Tim9</td>
<td>C</td>
<td>219.19</td>
<td>IMS</td>
<td>0.0261</td>
<td>0.106</td>
</tr>
<tr>
<td>Translocator of the outer mitochondrial membrane 40</td>
<td>Tom40</td>
<td>C</td>
<td>131.2</td>
<td>OM</td>
<td>0.0984</td>
<td>0.112</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Short Name</th>
<th>Complete</th>
<th>ORF Coverage</th>
<th>Predicted Localization</th>
<th>Mitoprot Prediction</th>
<th>TargetP prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative oxidase</td>
<td>AOX</td>
<td>C</td>
<td>181.43</td>
<td>IM</td>
<td>0.9774</td>
<td>np</td>
</tr>
<tr>
<td>Acetate:Succinate CoA transferase Type 1B</td>
<td>ASCT 1B</td>
<td>C</td>
<td>706.49</td>
<td>MM</td>
<td>0.7717</td>
<td>0.502</td>
</tr>
<tr>
<td>Acetate:Succinate CoA transferase Type 1C</td>
<td>ASCT 1C</td>
<td>C</td>
<td>643.23</td>
<td>MM</td>
<td>0.9837</td>
<td>0.709</td>
</tr>
<tr>
<td>Succinate Dehydrogenase Assembly Factor 2, Citrate Synthase-like</td>
<td>CItaf</td>
<td>C</td>
<td>212.73</td>
<td>MM</td>
<td>0.846</td>
<td>0.821</td>
</tr>
<tr>
<td>Electron transfer flavoprotein alpha subunit</td>
<td>ETFa</td>
<td>C</td>
<td>48.65</td>
<td>IM</td>
<td>0.9486</td>
<td>0.194</td>
</tr>
<tr>
<td>Electron transfer flavoprotein beta subunit</td>
<td>ETFb</td>
<td>C</td>
<td>34.13</td>
<td>IM</td>
<td>0.8149</td>
<td>0.336</td>
</tr>
<tr>
<td>ETF-ubiquinone oxidoreductase</td>
<td>ETFDH</td>
<td>C</td>
<td>1,478</td>
<td>IM</td>
<td>0.6233</td>
<td>0.647</td>
</tr>
<tr>
<td>Fumarase</td>
<td>Furn</td>
<td>C</td>
<td>81.44</td>
<td>MM</td>
<td>0.96</td>
<td>0.571</td>
</tr>
<tr>
<td>Glycerol-3-phosphate dehydrogenase, membrane bound</td>
<td>G3PDH</td>
<td>C</td>
<td>226.4</td>
<td>IM</td>
<td>0.8846</td>
<td>0.164</td>
</tr>
<tr>
<td>NADH dehydrogenase 24 kDa Subunit, NuoE</td>
<td>NuoE</td>
<td>C</td>
<td>740.81</td>
<td>MM</td>
<td>0.9925</td>
<td>0.908</td>
</tr>
<tr>
<td>NADH dehydrogenase 51 kDa Subunit, NuoF</td>
<td>NuoF</td>
<td>C</td>
<td>446.61</td>
<td>MM</td>
<td>0.9489</td>
<td>0.899</td>
</tr>
<tr>
<td>Succinyl-CoA synthetase, alpha subunit</td>
<td>SCSa</td>
<td>C</td>
<td>735.35</td>
<td>MM</td>
<td>0.8893</td>
<td>0.825</td>
</tr>
<tr>
<td>Succinyl-CoA synthetase, beta subunit</td>
<td>SCSb</td>
<td>C</td>
<td>836.31</td>
<td>MM</td>
<td>0.5658</td>
<td>0.743</td>
</tr>
<tr>
<td>Succinate Dehydrogenase Flavoprotein Subunit</td>
<td>SdhA</td>
<td>C</td>
<td>20.83</td>
<td>MM</td>
<td>0.8078</td>
<td>0.275</td>
</tr>
<tr>
<td>SDH Fe-S Subunit</td>
<td>SdhB</td>
<td>C</td>
<td>335.88</td>
<td>MM</td>
<td>0.8877</td>
<td>0.947</td>
</tr>
<tr>
<td>SDH Large Cytochrome b Subunit</td>
<td>SdhC</td>
<td>C</td>
<td>97.24</td>
<td>IM</td>
<td>0.8974</td>
<td>0.864</td>
</tr>
<tr>
<td>SDH Small Cytochrome b Subunit</td>
<td>SdhD</td>
<td>C</td>
<td>1,693.76</td>
<td>IM</td>
<td>0.9861</td>
<td>0.918</td>
</tr>
<tr>
<td>Rhodoquinone biosynthesis protein RQUA</td>
<td>RquA</td>
<td>C</td>
<td>849.93</td>
<td>MM</td>
<td>0.9919</td>
<td>0.958</td>
</tr>
</tbody>
</table>

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. bifurca*, 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. bifurca* 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, *Pygsulia* 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 *Pygsulia*.

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. bifurca* and one (incomplete) copy in *B. anathema* (*Ba*-HYDA). In addition, I found three other putative HYDA-like proteins in *P. bifurca* 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 breviate 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 posterior 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 Pygsuisa 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 maturation 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 (SCS/a/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 *Pygsuaia* 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:ubiquinone 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 Pygsua 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. Pygsua 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; Nyvtová *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, *Pygsusia* 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 *Pygsusia* encodes a protein containing an ISCS-like domain fused C-terminal to a 4-thiouridine biosynthesis protein (THII; PblSCS-THII). A partial sequence encoding ISCS was identified in *B. anathema* without the THII domain. Interestingly, *Leishmania* species also encode the *Pygsusia*-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 through LGT from a prokaryotic source, although the donor lineage is currently unclear. Furthermore, in contrast to the *Breviata* sequence, the *Pygsusia* 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 Breviata 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 Pygsua, 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 FADH₂-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 CXₙCX₂C 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\(^{++}\)/H\(^{+}\) 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 phospholipids. The latter pathway is rare in eukaryotes having only been documented in molluscs, trypanosomes and ciliates (Liang and Rosenberg 1968; Adorsaku *et al.* 1996; Sarkar *et al.* 2003; Metcalf and van der Donk 2009). In *Tetrahymena*, phoshoenol 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 organelar in *Tetrahymena*. In *Pygusia*, 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 *Pygusia* 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 *Pygusia* 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 *Pygusia* 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 (PEM; 65 sequences, 284 sites). Support values and taxa labeling as in Figure 3-2.
Figure 3-19: Maximum-likelihood (ML) tree of concatenated Phosphonoacetaldehyde 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 tax label as in Figure 3-2.

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

In the absence of a genetic system in Pygsua, Pygsua 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.](image)

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 be 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 Breviata 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 transcriptome 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 Pygsua, 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, Pygsua 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 Pygsua 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 Pygsua 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 Pygsua 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 Pygsuaia is similar to other non-Opisthokonts (e.g., Dictyostelium) with the exception of the outer membrane complex (Pygsuaia only encodes TOM40 and SAM50). Pygsuaia 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, Pygsuaia 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 Pygsuaia 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 Pygsuaia 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 SUFBC2D 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 SUFBC2D 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 SUFBC2D 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 Pysuiia 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 enzyme 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 (Cl) 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$_2$O (See Chapter 1, Figure 1-1). In the process Cl, CIII, and CIV generate a proton gradient across the inner mitochondrial membrane that fuels oxidative phosphorylation of ADP to ATP by the F1-F$_0$ 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$_2$O. However, under anaerobic conditions, the Ascaris respiratory chain shifts to use the electron carrier rhodoquinone, (RQ, Figure 4-1) and only Cl, CII and CV (Figure 4-2). In these conditions, Cl 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 Caenorhabditis 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; Hrdý 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'-CCCCGATCTGCAATGCGGTGTGCAACACC-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 pretein 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 *qua* sequences identified from publicly available and in-house sequencing projects.

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Organism</th>
<th>Accession</th>
<th>database</th>
<th>Complete&lt;sup&gt;a&lt;/sup&gt;</th>
<th>MitoProt TargetP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archamoeba</td>
<td>Mastigamoeba balamuthi</td>
<td>KM1886976</td>
<td>Genbank</td>
<td>0.8324</td>
<td>0.918</td>
</tr>
<tr>
<td>Discosea</td>
<td>Paramoeba atlantica</td>
<td>0201483178</td>
<td>CAMERA</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Discosea</td>
<td>Neoparamoeba australiana</td>
<td>0201515656</td>
<td>CAMERA</td>
<td>0.709</td>
<td>n.d.</td>
</tr>
<tr>
<td>Tubulinia</td>
<td>Copromyxyla protea</td>
<td>this study</td>
<td>this study</td>
<td>0.8457</td>
<td>0.879</td>
</tr>
<tr>
<td>Heterokonta</td>
<td>Blastocystis_hominis</td>
<td>300176487</td>
<td>Genbank nr</td>
<td>0.9884</td>
<td>0.888</td>
</tr>
<tr>
<td>Heterokonta</td>
<td>Blastocystis sp.</td>
<td>this study</td>
<td>this study</td>
<td>0.9231</td>
<td>0.885</td>
</tr>
<tr>
<td>Heterokonta</td>
<td>Atthaea septentrionalis</td>
<td>0198293608</td>
<td>CAMERA</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Heterokonta</td>
<td>Atthaea sp.</td>
<td>371497227</td>
<td>Genbank EST</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Heterokonta</td>
<td>Staurosira complex</td>
<td>0202500160</td>
<td>CAMERA</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Ciliophora</td>
<td>Condylomysa magnus</td>
<td>KM186977</td>
<td>this study</td>
<td>0.9182</td>
<td>0.878</td>
</tr>
<tr>
<td>Ciliophora</td>
<td>Fabrea fabrea salina</td>
<td>0202430720</td>
<td>CAMERA</td>
<td>0.7987</td>
<td>0.596</td>
</tr>
<tr>
<td>Mollusca</td>
<td>Aplysia californica&lt;sup&gt;a&lt;/sup&gt;</td>
<td>613494758</td>
<td>Genbank TSA</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Chaenoflagellata</td>
<td>Monosiga ovata</td>
<td>163101368</td>
<td>Genbank EST</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Breviata</td>
<td>Pygsula biforma</td>
<td>KM186978</td>
<td>Genbank SRA</td>
<td>0.9991</td>
<td>0.958</td>
</tr>
<tr>
<td>Euglenida</td>
<td>Euglena gracilis</td>
<td>EC670846</td>
<td>Genbank EST</td>
<td>0.9697</td>
<td>0.886</td>
</tr>
<tr>
<td>Euglenida</td>
<td>Eutreptiella gymnastica</td>
<td>0113734402</td>
<td>CAMERA</td>
<td>0.9955</td>
<td>0.922</td>
</tr>
<tr>
<td>Euglenida</td>
<td>Eutreptiella-like</td>
<td>0174352920</td>
<td>CAMERA</td>
<td>0.9986</td>
<td>0.947</td>
</tr>
<tr>
<td>Foraminifera</td>
<td>Reticulomyxa filosa</td>
<td>569420904</td>
<td>Genbank nr</td>
<td>0.6108</td>
<td>0.82</td>
</tr>
<tr>
<td>Foraminifera</td>
<td>Ammonia sp.</td>
<td>0197027732</td>
<td>CAMERA</td>
<td>0.5539</td>
<td>0.554</td>
</tr>
<tr>
<td>Foraminifera</td>
<td>Elphidium marginatum</td>
<td>0202694266</td>
<td>CAMERA</td>
<td>0.9908</td>
<td>0.881</td>
</tr>
</tbody>
</table>

<sup>a</sup>Potential contaminant  
<sup>b</sup>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 QUA evolved from a methyltransferase (UBIE) involved in ubiquinone biosynthesis (Lonjers et al. 2012). Indeed, initial phylogenies of QUA and UBIE sequences revealed QUA 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 discoideum and Chlamydomonas reinhardtii 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 bifurca 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 Phylogenetic 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/maenquinone biosynthesis C-methyltransferase 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 Condyllostoma), foraminiferans (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, Paramoeba 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 ‘supergroup’ 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 \textit{rqua} relative to fumarate reductase (\textit{frd})/succinate dehydrogenase (\textit{sdh}) subunits A-D are indicated with arrows next to bacteria where applicable. Non- \textit{frd}/\textit{sdh} genes are shown with light blue arrows and \textit{rqua} shown in purple. Select UBIE outgroup was used to root the tree (grey sequences).

To determine if \textit{rqua} is genetically linked to other potential RQ biosynthesis genes, the genomic context of \textit{rqua} in the various bacterial genomes was investigated manually. The \textit{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, \textit{rqua} is encoded physically close (\textit{i.e.}, within one gene or 1 kilobase pair) to the genes annotated as anaerobic fumarate reductase (\textit{frd})/succinate dehydrogenase (\textit{sdh}) subunits (A-D) (Figure 4-3). Furthermore, in 12 of these bacteria (\textit{Competibacter denitrificans, Candidatus Contendobacter odensis, Rhodomicrobium udaipurense, Rhodomicrobium udaipurense, Rhodomicrobium vanniellii, Burkholderiales, Caenispirillum salinarum, Novispirillum itersonii, Methylcella silvestris, Thauera sp., Aromatoleum aromaticum, Laribacter hongkongensis, Pseudogulbenkiania ferrooxidans, and Chromobacterium violaceum}) there is second \textit{frd}/\textit{sdh} operon located elsewhere in the genome. The genetic linkage I observed between \textit{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; [VI[FG]-
[LIVCS]-[DENLV]-[VALMIT]-[GLYCF]-[CSTAYFP]-[GA]-[PTSNKRMV]-[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](image)

129
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 \textit{rqua}-containing eukaryotic genomes and transcriptomes. FIGURE using homologs of these enzymes from \textit{Dictyostelium discoideum} and \textit{Arabidopsis thaliana} as query sequences. Interestingly, most MRO-containing protists (\textit{Pygsuia}, \textit{Mastigamoeba} and \textit{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; Chromalveolata, 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 indceted 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.

![Image](image.png)

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-PbRQUA antibodies recognize purified recombinant PbRQUA. Protein extracts were isolated from whole cell extracts of *E. coli* transformed with empty plasmid pGEX4T-1 (1) or pGEX-PbRQUA (2) and cultured on conditions to induce expression of the encoded protein and GST-PbRQUA purified using glutathione magnetic beads (3) were resolved by SDS-PAGE and probed by immunoblotting using anti-PbRQUA 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–proteobacteria (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’.  

136
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 consistent with the observation that, like the closely-related ubiqinone 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 CIIls 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 H₂ gas (Dyall et al. 2004; Hrdy et al. 2004). Interestingly, other anaerobic protists including Sawyeria marylandensis and Pygsuia bifiform appear to have retained these same two components of Complex I suggesting that a Thermotoga-like coupling of NADH oxidation and H₂ 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 anaerobiosis, 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


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.


161


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


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


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 Fimbicute: |
| 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 |
| Total | 0.00 USD |
| Total | 0.00 USD |
Appendix B  Copyright Permission (Chapter 3)

ELSEVIER LICENSE TERMS AND CONDITIONS

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.

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

168