THE DIVERSITY AND ORIGINS OF MITOCHONDRION-RELATED ORGANELLES

by

Michelle Marie Leger

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

at

Dalhousie University
Halifax, Nova Scotia
July 2015

© Copyright by Michelle Marie Leger, 2015

For my parents

"About the orchids of that place we knew very little. About flies and beetles almost nothing, fungi nothing, most kinds of organisms nothing. Five thousand kinds of bacteria might be found in a pinch of soil, and about them we knew absolutely nothing. This was wilderness in the sixteenth-century sense, as it must have formed in the minds of the Portuguese explorers, its interior still largely unexplored and filled with strange, myth-engendering plants and animals. [...] And I thought: there is still time to see this land in such a manner."

- Edward Osborne Wilson, 1992, *The Diversity of Life*

“Le biologiste passe, la grenouille reste.”

“The biologist passes, the frog remains.”

- Jean Rostand, 1967, *Inquiétudes d’un biologiste*

Table of Contents

List of Tables

List of Figures

Abstract

Multiple distantly-related eukaryotic lineages have adapted to low-oxygen environments, and possess modified mitochondria, known as mitochondrion-related organelles (MROs). Although relatively few MROs have been investigated in detail, they are known to vary in the types of ATP metabolism they possess, and in the nature of the ancestral mitochondrial functions that they have retained. Here, I expand our knowledge of this diversity, and provide insights into how characteristic anaerobic ATP generation enzymes became widespread among eukaryotes.

MROs known as hydrogenosomes have lost the electron transport chain, and possess a distinctive anaerobic ATP generation pathway that produces hydrogen as an end-product. I describe this type of ATP generation pathway in *Acanthamoeba castellanii*, an amoebozoan previously believed to have typical aerobic mitochondria and I show that this pathway is located in the mitochondria using immunolocalization. This is the first known example of a mitochondrion that possesses both a complete electron transport chain and a complete hydrogenosomal-like ATP generation pathway.

Bacterial cell division is initiated by a tubulin homolog, FtsZ, and three Min proteins that regulate its distribution. In mitochondria, this system has been supplanted by eukaryotic dynamin-related proteins. I show that mitochondrial homologs of FtsZ are widespread among eukaryotes, and that they were likely duplicated in the last eukaryotic common ancestor. I also provide the first evidence for the existence of Min proteins in mitochondria, and show that they are present in four out of six eukaryotic supergroups. The ancestral FtsZ-Min system is more representative of mitochondrial division in diverse eukaryotes than is the dynamin-based system of eukaryotic model organisms.

Finally, I present an *in silico* reconstruction of the biochemical pathways present in the MRO of *Andalucia incarcerata*, a free-living, deep-branching excavate protist, and validate some of these predictions with immunolocalization. *A. incarcerata*'s MRO possesses a hydrogenosomal-like ATP generation pathway that presents a striking example of convergence with those of other, distantly anaerobic eukaryotes. However, it retains a greater complement of mitochondrial metabolic functions and import machinery than the well-described MROs of parasites. My work shows that this larger complement of ancestral mitochondrial functions is a common feature shared by MROs of free-living eukaryotes.

List of Abbreviations Used

24kDa: 24kDa subunit of Electron Transport Chain Complex I (figures)

3HDH: 3-hydroxybutyrylCoA dehydrogenase

51kDa: 51kDa subunit of Electron Transport Chain Complex I (figures)

8, 9, 10, 13: Translocator of the Inner Membrane proteins Tim8, Tim9, Tim10, Tim13, respectively (figures).

AAC: ADP/ATP translocator

ACAD: branched-chain acyl-CoA dehydrogenase

Ace-CoA: acetyl-CoA

ADP: adenosine diphosphate

AGT: alanine – glyoxylate aminotransferase

Ala: alanine

ALT: alanine aminotransferase

AMP: adenosine monophosphate

AS: amoeba saline medium

ASCT: acetate:succinate CoA-transferase

ASCTB: ASCT subtype 1B (figures)

ASCT1B: ASCT subtype 1B (text)

ASCT1C: ASCT subtype 1C

Asp: aspartate

AST: aspartate aminotransferase

ASW: artificial seawater medium

ATM: ABC transporter, Mitochondrial

ATP: adenosine triphosphate

AU: Approximately Unbiased

BαKDH: branched-chain α-ketoglutarate dehydrogenase complex

BCAAT: branched-chain amino acid aminotransferase

BCS: ubiquinol-Cytochrome c reductase Synthesis

BLAST: Basic Local Alignment Search Tool

CAMERA: Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis

cDNA: complementary DNA

CLS: cardiolipin synthase

CoASH: Coenzyme A

CPN: Chaperonin

DLDH: D-lactate dehydrogenase

DNA: deoxyribonucleic acid

E, F, G: [FeFe]-hydrogenase maturases HydE, HydF and HydG, respectively (figures)

EDTA: ethylenediaminetetraacetic acid

EM: Electron microscopy

ERV1: Essential for Respiration and Viability

EST: Expressed Sequence Tag

ETC: Electron Transport Chain

EuPathDB: Eukaryotic Pathogen Database

FDP: flavodiiron protein

FER^{O}: oxidized electron transport ferredoxin

FLD: flavodoxin domain

FRX: Frataxin

GCS: Glycine Cleavage System

Glu: glutamate

Gly: glycine

GPAT: glycerol-3-phosphate acyltransferase

GRX: Glutaredoxin

H, L, P, T: H-, L-, P- and T-protein components of the glycine cleavage system, respectively (figures)

HMM: hidden Markov Model

HPLC: high-performance liquid chromatography

HSP: Heat Shock Protein

HYD: [FeFe]-hydrogenase

FER^R: reduced electron transport ferredoxin

Ile: isoleucine

IM: inner mitochondrial membrane

IND: iron-sulfur protein required for NADH dehydrogenase

ISC: Iron-Sulfur Cluster

ISD: Iron-Sulfur protein biogenesis, Desulfurase-interacting

JAC: J-type Accessory Chaperone

JGI: Joint Genome Institute

Lac: lactate

LB: Luria Bertain liquid medium

LECA: the Last Eukaryotic Common Ancestor

Leu: leucine

LGT: Lateral Gene Transfer

LSU rRNA: large subunit ribosomal RNA

MCE: methylmalonyl-CoA epimerase

MCF: Mitochondrial Carrier Family protein

MCM: methylmalonyl-CoA mutase

MET: Metaxin

MGE: Mitochondrial GrpE

MIA: Mitochondrial intermembrane space Import and Assembly

MPP: Mitochondrial Processing Peptidase

MRO: mitochondrion-related organelle

MS/MS: tandem mass spectrometry

MSDH: methylmalonyl semialdehyde dehydrogenase

mtDNA: mitochondrial DNA

mtTP: mitochondrial targeting peptide

NAD: nicotinamide adenine dinucleotide

NADP: nicotinamide adenine dinucleotide phosphate

NCBI: National Center for Biotechnology Information

NFU: NifU-like

nuDNA: nuclear DNA

Oaa: oxaloacetate

OM: outer mitochondrial membrane

OMAC: Oceanic Mitochondrion-Affiliated Clade

OMO: organelle of mitochondrial origin

PBS: Phosphate-buffered saline

PCC: propionyl-CoA carboxylase

PCR: Polymerase Chain Reaction

PDH: pyruvate dehydrogenase

PFO: pyruvate:ferredoxin oxidoreductase

PGP: phosphatidylglycerolphosphate

PGPS: phosphatidylglycerolphosphate synthase

PHB: Prohibitin

PNT: pyridine nucleotide transhydrogenase.

PTPMT: protein tyrosine phosphatase, mitochondrial

Pyr: pyruvate

RACE: Rapid Amplification of cDNA Ends

RNA: ribonucleic acid

rRNA: ribosomal RNA

SAM: Sorting and Assembly Machinery

SAR: Stramenopiles, Alveolates and Rhizaria

SCS: succinyl-CoA synthetase

SCOT: Succinate:3-oxoacid CoA-transferase

SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Ser: serine

SHMT: serine hydroxymethyltransferase

SOD: superoxide dismutase

SSQ: Stress Seventy subfamily Q

SSU rRNA: small subunit ribosomal RNA

Suc: succinate

Suc-CoA: succinyl-CoA

TAM: Translocator Assembly and Maintenance

TBS: Tris-buffered saline

TCA: Tricarboxylic acid

THF: tetrahydrofolate

TIM: Translocator of the Inner Membrane

TOM: Translocator of the Outer Membrane

TRX: thioredoxin

TRXP: thioredoxin peroxidase

Val: valine

YAH: Yeast Adrenodoxin Homolog

Acknowledgements

My years in Halifax have been strange and wonderful – much more so than I ever could have anticipated, on both counts. CGEB has become a second home for me, and I am beyond grateful to all of its members for creating such a welcoming, stimulating and collegial environment.

First and foremost among those members is my supervisor, Andrew Roger. I cannot thank him enough for giving me a chance as his grad student, sight unseen, and for everything he has done to nurture me throughout my time here. He is an inspiration as a scientist and as a person: in his emphasis on respectful discussion and keeping the focus on the evidence; in the atmosphere of freedom and open collaboration that he cultivates in his lab, and with other labs; and in the patience, generosity and compassion with which he treats his trainees. (Not to mention his deep devotion to Grumpycat!) I am incredibly lucky to have been mentored by him.

The members of my supervisory committee – Alastair Simpson, Ford Doolittle, and John Archibald – have been a constant source of thoughtful advice, constructive criticism, and selfless help. Being able to talk to them outside of formal committee meetings, and being able to see the evolution in their own research over the years, has been particularly encouraging.

I am privileged to have made some good friends to whom I owe a particular and personal debt. Eunsoo Kim became not only my friend, but my role model. Eleni Gentekaki helped me through my hardest times here, and showed me that there is more than one way to keep a spirit of curiosity and exploration in science. Aaron Heiss helped me not to forget (and in some cases showed me firsthand) what an amazing and exceedingly odd place the universe is. Jav Alfaro and Kirsten Kennedy were always there to make me laugh, and I miss feeding the squirrels in Point Pleasant with them. Laura Eme has been a paradoxical source of both insanity and sanity: without her help and encouragement (and cat pictures) I could not have written this thesis! Thank you all.

I have had a great deal of help and advice from members of the Roger lab past and present: thank you Ivica Tamaš, Dan Gaston, Martin Kolisko, Jessica Leigh, Alex Stechmann, Kiyotaka Takishita, Matt Brown, Tommy Harding, Susan Sharpe, Ryoma Kamikawa, Bernard Lemire, Dayana Salas-Leiva, and Jiwon Yang, for their friendship, for teaching me wetwork and bioinformatics, for reading over my work, and for making the Roger lab what it is and was. Special thanks go to Jacquie de Mestral and to Wanda Danilchuk, who kept to Roger lab running in spite of the rest of us; and to Courtney Stairs, my partner in MROmics, who generously gave me comments on this thesis and responded to frantic last-minute questions as it was being written. Laura Hug was always

ready with advice on culturing *Andalucia* or sorting through data. Tasos Tsaousis was a great source of immunolocalization knowledge, encouragement, and late-night hilarity in the lab. Thank you also to other members CGEB, past and present, especially to Bruce Curtis, Dave Spencer and Murray Schnare, and to those who have become friends in addition to helping me with my work: Banoo Malik-Pightling, Julia Hopkins, Gordon Lax, Sergio Muñoz-Gómez, Ugo Cenci, and Yana Eglit.

My collaborators have provided thought-provoking discussion and fruitful partnerships: particular thanks go to Ryan Gawryluk, Mike Gray, Marek Eliáš, Markéta Petrů, Pavel Doležal, and Vojta Žárský. Mary Ann Trevors has been invaluable in helping me with electron microscopy work. I am grateful to Roisin McDevitt, Melissa MacDougall, and Alicia Kirk, who have each gone far out of their way to help me as I wrote this thesis.

Sandie Baldauf and Johanna Fehling introduced me to protists and to molecular evolution, and encouraged me to go to Dalhousie; I am hugely indebted to them for this.

And finally, thank you to my parents for making this possible with their love and their patience.

The work in this thesis was supported by the Fonds National de la Recherche, Luxembourg, and by the Nova Scotia Health Research Foundation.

Chapter 1: Introduction

Over the last twenty years, it has become generally accepted that the last eukaryotic common ancestor (LECA) of all known extant eukaryotes contained a mitochondrion, derived from an α-proteobacterial endosymbiont. The best-known feature of aerobic mitochondria is the production of ATP through oxidative phosphorylation. However, mitochondria are also implicated in a wide range of biochemical pathways, including amino acid catabolism and synthesis, lipid catabolism, glycerophospholipid and sterol biosynthesis, heme biosynthesis, storage of calcium ions, and apoptosis (Scheffler 2007). Other mitochondrial proteins are required for the maintenance of basic mitochondrial functions, such as mitochondrial division and fusion, import and folding of nuclear-encoded mitochondrial proteins, solute import and export, and mitochondrial genome replication, transcription and translation.

Mitochondrion-Related Organelles

Recent years have seen many more studies of mitochondria and their related organelles, from taxa with a range of lifestyles and taxonomic affinities. With the decreasing cost of high-throughput sequencing, a much greater coverage has become possible, providing much broader knowledge of pathways in these organelles.

The first evidence of modified organelles derived from mitochondria was the discovery, in trichomonads, of double membrane-bounded organelles, staining uniformly, and lacking cristae. Initially dubbed paracostal or paraxostylar granules (Daniel, Mattern, Honigberg 1971; Honigberg, Mattern, Daniel 1971), microbody-like particles (Müller 1973), chromatic granules (Filadoro 1970), or mitochondrion-like granules (Sharma, Bourne 1963), these became known as hydrogenosomes, based on their characteristic production of molecular hydrogen (Lindmark, Muller 1973). Subsequently, similar organelles were discovered in rumen ciliates (Yarlett et al. 1981; Embley et al. 1995), in chytrid fungi (Yarlett et al. 1986; Li, Heath, Bauchop 1990; Li, Heath, Cheng 1991; Marvin-Sikkema et al. 1992) and in the diplomonad *Spironucleus salmonicida* (Jerlstrom-Hultqvist et al. 2013) (a recent report was also made of hydrogenosomes in loriciferans, microscopic sediment-dwelling animals (Danovaro et al. 2010), although this was based solely on electron microscopy data and the environment in which the animals were found, without any supporting molecular data).

A link between hydrogenosomes and mitochondria had already been suggested in some publications (Anderson, Beams 1959; Sharma, Bourne 1963) ("One may speculate that these organelles may be modified or specialized mitochondria carrying an array of enzyme systems directly or indirectly related to the general metabolism of the organism." (Anderson, Beams 1959)). At that time, however, no clear link had been positively established. An alternative view was that lineages lacking canonical, cristate mitochondria were ancestrally amitochondriate, having diverged from other eukaryotic lineages prior to the

endosymbiotic event that led to the establishment of mitochondria. Tom Cavalier-Smith proposed the name ‘Archezoa’ for the subkingdom he created for these 'amitochondriate' eukaryotes – metamonads, parabasalids, microsporidia and archamoebae (Cavalier-Smith 1983b). This ‘Archezoa hypothesis’ posited additional traits for the proto-host. In order to engulf the protomitochondrial endosymbiont, argued Cavalier-Smith, it would need to have been capable of phagocytosis (Cavalier-Smith 1983a). Based on extant ‘amitochondriate’ taxa, it likely had a nucleus and an endomembrane system, but not stacked Golgi bodies; and its ribosomes were likely similar in size to prokaryotic ones, smaller than those of typical, mitochondriate eukaryotes (Cavalier-Smith 1987b).

The Archezoa hypothesis was bolstered in light of support by early phylogenies of eukaryotes based on small subunit (SSU) and large subunit (LSU) ribosomal RNA (rRNA) and early protein phylogenies (Cavalier-Smith 1987a; Vossbrinck et al. 1987; Sogin et al. 1989; Leipe et al. 1993; Hashimoto, Hasegawa 1996; Stiller, Hall 1997). These phylogenies placed 'amitochondriate' lineages at the base of a 'crown group' formed by the remaining eukaryotes (Knoll 1992), a finding that would later be found to be artifactual (Siddall, Hong, Desser 1992; Hasegawa, Hashimoto 1993; Stiller, Hall 1999; Philippe, Germot 2000).

It remains possible that the organism that engulfed the protomitochondrial endosymbiont was indeed as posited by the Archezoa hypothesis: a proto-eukaryote with an endomembrane system, capable of phagocytosis, but lacking mitochondria or peroxisomes. However, the second part of the Archezoa hypothesis – that microsporidia, parabasalids, metamonads are relics of this pre-

mitochondrial state – has been definitively refuted. The ancestrally amitochondriate nature of these taxa was challenged by the discovery that their nuclear genomes encoded mitochondrial proteins (Bozner 1996; Bui, Bradley, Johnson 1996; Germot, Philippe, Le Guyader 1996; Horner et al. 1996; Roger, Clark, Doolittle 1996; Germot, Philippe, Le Guyader 1997; Roger 1998). Shortly thereafter, the 'amitochondriate' amoebozoan *Entamoeba histolytica* was found to possess much smaller organelles, also surrounded by a double membrane, to which the mitochondrial protein Cpn60 could be localized (Mai et al. 1999; Tovar, Fischer, Clark 1999). Unlike hydrogenosomes, these organelles did not appear to produce hydrogen; they were named mitosomes (Tovar, Fischer, Clark 1999). Mitosomes were subsequently described in microsporidia (Williams et al. 2002), and in the excavate *Giardia lamblia* (Tovar et al. 2003). The homology between mitochondria, hydrogenosomes and most mitosomes was further cemented by phylogenies placing their iron sulfur cluster assembly proteins in a single clade with α-proteobacterial homologs (Richards, van der Giezen 2006); and by the finding that the latter two import nuclear-encoded proteins using homologs of mitochondrial import proteins, and a similar protein targeting system (Dolezal et al. 2005).

The initial, separate characterization of hydrogenosomes and mitosomes laid the groundwork for a classification of mitochondria and related organelles that was based on their role in energy metabolism. Hydrogenosomes encompassed those organelles that generate energy under anaerobic conditions. It was generally understood that this involved the concomitant production of hydrogen, although this was not always explicitly demonstrated. Meanwhile, those organelles that

did not appear to have a role in energy production were classified as mitosomes (e.g., (van der Giezen, Tovar 2005; Tsaousis et al. 2012a)).

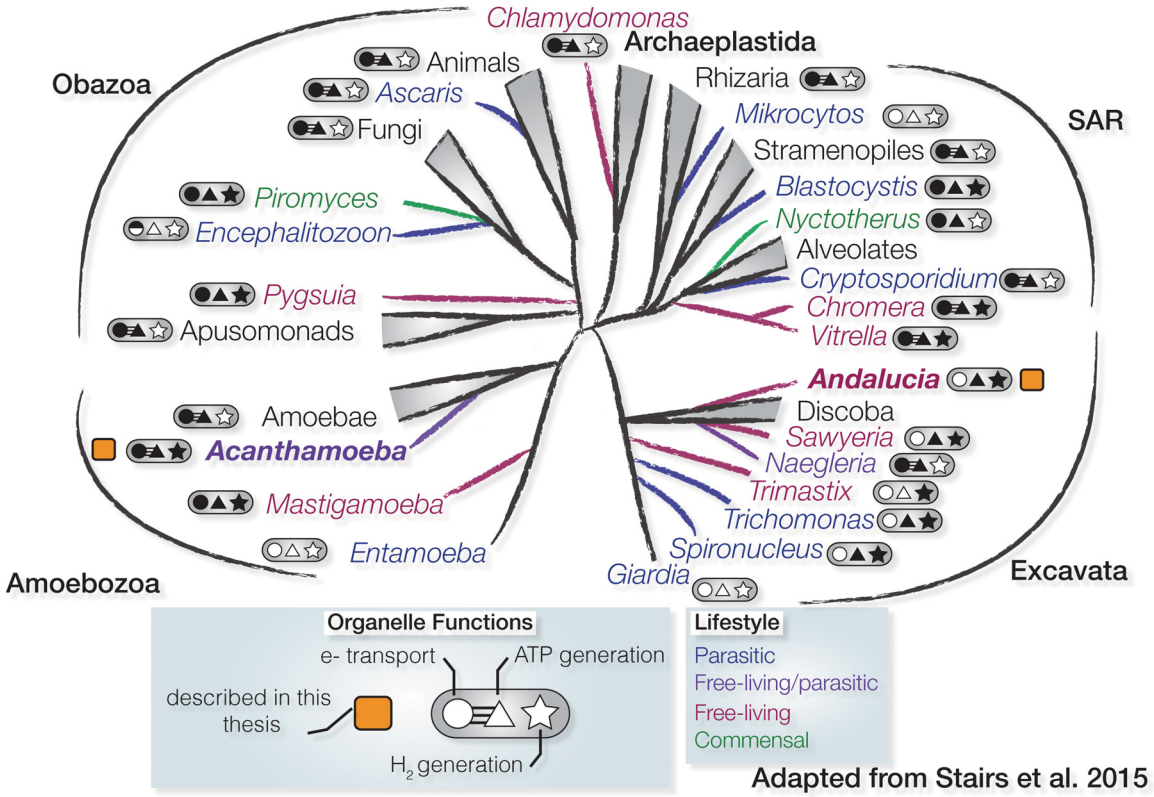


Figure 1.1 Distribution of mitochondrion-related organelles (MROs) across the major supergroups of eukaryotes. Adapted from Stairs, Leger and Roger, 2015. Diversity and origins of anaerobic metabolism in mitochondria and related organelles: a new hypothesis. Phil Trans R Soc B [In Press] Green, commensal organisms; blue, parasitic organisms; purple, facultatively parasitic (but frequently free-living) organisms; pink, free-living organisms. Metabolic functions in each MRO are shown as follows: shaded shapes represent the presence of electron-transporting complexes (circle), ATP synthesis (triangle) and hydrogen production (star). Where at least one proton-pumping complex (Complex I, II, III or IV) and ATP synthase (Complex V) were identified, suggesting that energy is being generated through oxidative phosphorylation, the circle and the triangle are linked. Yellow squares and bold font indicate organisms investigated for this thesis.

Following the establishment of a mitochondrial ancestry for hydrogenosomes and mitosomes, similar organelles were studied in a wider range of taxa (LaGier

et al. 2003; Keithly et al. 2005; Gill et al. 2007; Hampl et al. 2008; Lantsman et al. 2008; Stechmann et al. 2008; de Graaf et al. 2009; Barbera et al. 2010; Burki et al. 2013; Jerlstrom-Hultqvist et al. 2013; Stairs, Leger, Roger 2015 (in press)) (Fig. 1.1).

These studies provided evidence that every major known lineage of eukaryotes had, at some point, possessed mitochondria. They also pointed to a wider range of morphological and biochemical characteristics of these organelles. Some operated only under anaerobic conditions, but, unlike hydrogenosomes and mitosomes, possessed mitochondrial genomes (Akhmanova et al. 1998; Stechmann et al. 2008; Wawrzyniak et al. 2008; de Graaf et al. 2011). Many had retained partial electron transport chains (Lantsman et al. 2008; Stechmann et al. 2008; Stairs et al. 2014), or amino acid metabolism enzymes not previously found in mitosomes or hydrogenosomes (Gill et al. 2007; Zubacova et al. 2013). Some aerobic mitochondria, capable of oxidative phosphorylation, were known to be capable of functioning anaerobically (reviewed in (Müller et al. 2012)). Finally, some aerobic mitochondria were found to possess hydrogenosomal-like anaerobic energy metabolism enzymes (Fritz-Laylin et al. 2010; Fritz-Laylin et al. 2011) (and see Chapter 2 of this thesis). As a result, the need for a term to encompass the full spectrum of these organelles gave rise to the use of terms such as ‘relict mitochondria’ (Keithly et al. 2005), ‘cryptic organelles’ (Williams, Keeling 2003), ‘mitochondrion-like organelles’ (MLOs) (Fenchel, Finlay 1991; Nasirudeen, Tan 2004; Lantsman et al. 2008; Stechmann et al. 2008), ‘mitochondrion-related organelles’ (MROs) (Gill et al. 2007; Barbera et al. 2010; Tsaousis et al. 2012a; Burki et al. 2013; Stairs et al. 2014; Stairs, Leger, Roger

2015 (in press)), or, most recently, 'organelles of mitochondrial origin' (OMOs) (Mentel et al. 2014). A more recent scheme expanded on the idea of classifying MROs according to their role in ATP generation, and addressed the broader spectrum of metabolism types by classifying mitochondria and related organelles into five classes. Under this scheme, Class 1 mitochondria (aerobic mitochondria) generate ATP only through oxidative phosphorylation. Class 2 mitochondria (anaerobic mitochondria) possess an electron transport chain, but are capable of functioning anaerobically using alternative electron acceptors, such as fumarate, while Class 3 mitochondria (H_2-producing mitochondria) possess a bacterial-like anaerobic ATP-generation pathway in addition to the electron transport chain. Class 4 mitochondria (hydrogenosomes) also possess this pathway, but have lost most components of the electron transport chain, such that they are no longer able to generate ATP through oxidative phosphorylation. Class 5 mitochondria (mitosomes) are highly reduced, and lack any role in ATP generation, as well as any of the associated enzymes. Some uncertainties remain in this scheme. For instance, Class 3 makes no distinction between H_2-producing MROs that may also be capable of both oxidative phosphorylation using a complete electron transport chain (e.g., *Naegleria* (Fritz-Laylin et al. 2011), *Acanthamoeba*, Chapter 2 of this thesis), and H_2-producing MROs that retain only a partial electron transport chain (e.g., *Mastigamoeba*, *Pygsuia* (Gill et al. 2007; Stairs et al. 2014)). Furthermore, attempts to generalize the definition of each class to encompass a specific set of additional mitochondrial functions (Maguire, Richards 2014) are hampered by the variation among the proteomes of various MROs.

In general, I employ the term 'MRO' in this thesis because it clearly references the homology between MROs and mitochondria, rather than their mere similarity. The functional continuum and common origins between aerobic mitochondria and the various anaerobically-functioning organelles related to them mean that there should, strictly speaking, be no distinction between them. All of these organelles should therefore properly be referred to as either mitochondria (the 5-Class system admirably does this), or as MROs. Unfortunately, a precedent has been set in the literature that uses 'MROs' to refer to all mitochondrion-related organelles other than strictly aerobic mitochondria. This is in keeping with a function-based definition in use by most biologists, based on well-described aerobic mitochondria, but at odds with a phylogenetically-informed definition. Because of this precedent, I use the term 'MROs' in opposition to 'mitochondria', rather than using a single term for both.

Pathways in MROs

The 5-Class scheme outlined above is useful in categorizing mitochondria and MROs according to their most widely-described characteristic (ATP generation), one that is obviously linked to an organism's overall lifestyle. However, limiting classification of MROs to this feature should not be allowed to obscure the diversity that these organelles encompass in other respects (reviewed in (Maguire, Richards 2014; Stairs, Leger, Roger 2015 (in press)); Fig. 1.2).

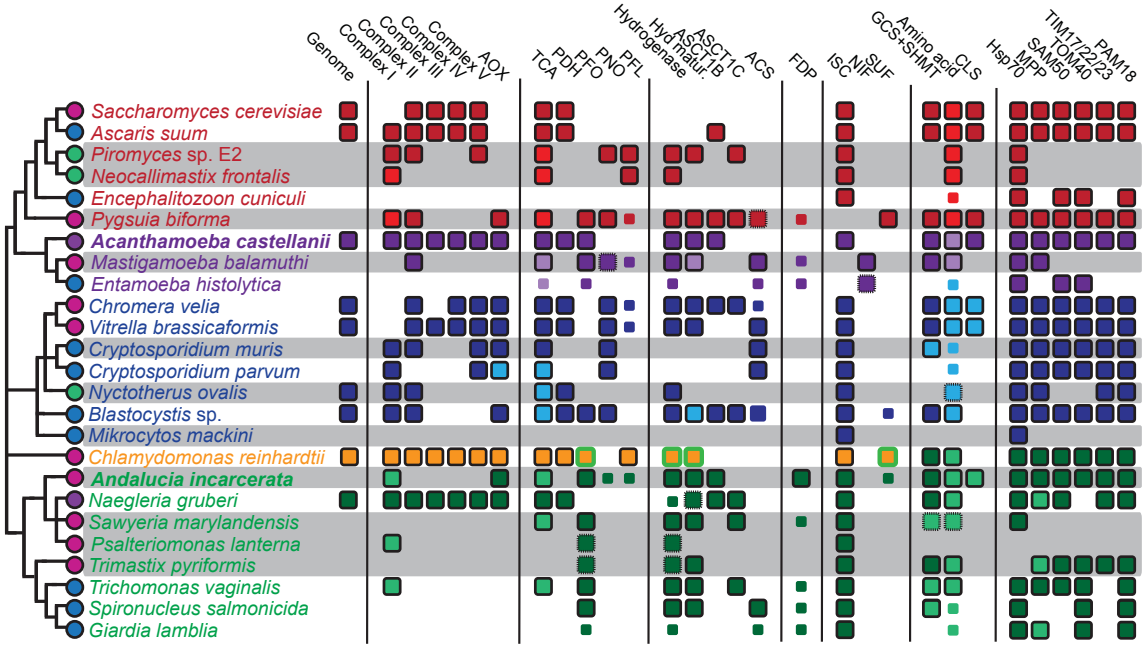


Figure 1.2 Metabolic functions and pathways in selected mitochondrion-related organelles. The schematic to the left indicates the approximate relationships of the organisms to one another. Lineages are indicated in red (Obazoa), purple (Amoebozoa), blue (Stramenopiles, Alveolates and Rhizaria [SAR clade]), and green (Excavata), respectively. Bold font indicates organisms investigated for this thesis. Circle indicate a parasitic (blue), facultatively parasitic (purple), free-living (pink) or commensal (green) lifestyle. Boxes surrounded by a solid black outline indicate a putative MRO localization based on predicted targeting peptides, localization studies or (in the case of mitochondrial protein import machinery components) homology to proteins found exclusively in mitochondria. Dashed outlines indicate incomplete data, making localization predictions impossible. Smaller boxes with no outline indicate a predicted cytosolic location. Green outlines (for *Chlamydomonas*) indicate a plastid location. Lighter boxes indicate an incomplete pathway or number of components, or uncertainty as to how many components are present. Grey backgrounds indicate incomplete genome or transcriptome data. Electron transport: Complex I-V, electron transport chain complexes I-V; AOX, alternative oxidase. ATP generation: TCA, tricarboxylic acid cycle enzymes; PDH, pyruvate dehydrogenase; PFO, pyruvate:ferredoxin oxidoreductase; PNO, pyruvate:NADH oxidoreductase; PFL, pyruvate:formate lyase. Hyd matur., [FeFe]-hydrogenase maturases HydE, HydF and HydG; ASCT, acetate:succinate CoA-transferase; ACS, acetyl-CoA synthetase. Oxygen detoxification: FDP, flavodiiron protein. Iron-sulfur cluster assembly: ISC, mitochondrial iron sulfur cluster assembly system; NIF, nitrogen-fixation system; SUF, sulfur mobilization system. Other ancestrally mitochondrial functions: GCS: glycine cleavage system; SHMT, serine

hydroxymethyltransferase; Amino acid, other amino acid catabolism enzymes (because it is not clear what a 'complete set' of these enzymes would entail, all of the boxes in this section are shaded in a lighter colour); CLS, cardiolipin biosynthesis. Mitochondrial protein import: Hsp70, heat-shock protein Hsp70; MPP, mitochondrial processing peptidase; SAM50, sorting and assembly machinery protein 50; TOM40, translocase of the outer membrane protein 40; TIM17/22/23, translocase of the inner membrane proteins 17, 22 or 23 (treated together because of the difficulty of assigning homologs to one function based on sequence alone); PAM18, presequence translocase-associated motor protein 18. Data from (Neupert 1997; Roberts et al. 2004; Henriquez et al. 2005; Regoes et al. 2005; Ctrnacta et al. 2006; Dolezal et al. 2006; Dagley et al. 2009; de Graaf et al. 2009; Waller et al. 2009; Barbera et al. 2010; Dolezal et al. 2010; Fritz-Laylin et al. 2010; Jedelsky et al. 2011; Schneider et al. 2011; Burki et al. 2013; Heinz, Lithgow 2013; Jerlstrom-Hultqvist et al. 2013; Zubacova et al. 2013; Gawryluk et al. 2014; Maguire, Richards 2014; Stairs et al. 2014; Wojtkowska et al. 2015; Stairs, Leger, Roger 2015 (in press)), or identified by reciprocal best BLAST hits (cutoff e-value $1x10^{-5}$) to data in EuPathDB or GenBank. Hello, Reader! ☺

The first descriptions of MROs were based on biochemical analyses, cell biology or targeted searches for individual genes or pathways, rather than large-scale data (Marvin-Sikkema et al. 1993a; Dyall et al. 2004; Hrdý et al. 2004). More recent large-scale genome, transcriptome or proteome efforts are beginning to shed more light on the roles of MROs in amino acid metabolism (Barbera et al. 2010; Schneider et al. 2011; Stairs et al. 2014). Some weaknesses are still apparent in these efforts. Early Sanger or 454 transcriptomes provided low sequence coverage, resulting in transcriptomes with a high proportion of missing genes, and incomplete sequences that made it impossible to predict the localization of proteins *in silico* (de Graaf et al. 2009; Barbera et al. 2010). Attempts to obtain full-length sequences, or to localize the encoded proteins using immunofluorescence or immunoelectron microscopy techniques, were laborious and time-consuming, and frequently failed. Later transcriptome or genome efforts, while more complete, still depend primarily on *in silico* prediction methods that are optimized on model system opisthokonts or plants.

These may fail in the presence of internal targeting peptides, or N-terminal peptides that are divergent from those of model organisms.

Anaerobic ATP Generation

The characteristic anaerobic ATP-generating pathway found in hydrogenosomes was first described in, and remains best characterized in, trichomonads (Dyall et al. 2004; Hrdý et al. 2004; Hrdý, Tachezy, Müller 2008; Tsaousis et al. 2012a). It relies on the presence of a suite of proteins not generally found within mitochondria (reviewed in (Müller et al. 2012; Stairs, Leger, Roger 2015 (in press))). Pyruvate is converted to acetyl-CoA and carbon dioxide by one of three enzymes. In typical hydrogenosomes such as those of *Trichomonas* or the diplomonad *Spironucleus* (Steinbuchel, Muller 1986; Dyall et al. 2004; Hrdý et al. 2004; Jerlstrom-Hultqvist et al. 2013), this reaction is catalyzed by pyruvate:ferredoxin oxidoreductase (PFO), while in the MROs of other anaerobic eukaryotes, such as *Blastocystis* sp., *Mastigamoeba balamuthi* or *Cryptosporidium parvum*, pyruvate to acetyl-CoA conversion is instead or additionally carried out by a pyruvate:NADP oxidoreductase (PNO) or a pyruvate:formate lyase (PFL) (Ctrnacta et al. 2006; Lantsman et al. 2008; Stairs, Roger, Hampl 2011). In trichomonad hydrogenosomes, Coenzyme A is subsequently transferred from acetyl-CoA to succinate by an acetate:succinate CoA transferase (ASCT (Tielens et al. 2010)), generating acetate and succinyl-CoA. The subsequent regeneration of succinate, catalyzed by the tricarboxylic acid (TCA) cycle enzyme succinate thiokinase (STK), generates ATP through substrate-level phosphorylation. In the MROs of *Spironucleus salmonicida*, ASCT and STK are not present, and ATP and acetate are instead produced

directly by an acetyl-CoA synthetase (ACS) (Jerlstrom-Hultqvist, Einarsson, Svard 2012; Jerlstrom-Hultqvist et al. 2013). The electrons generated by the conversion of pyruvate to acetyl-CoA are transferred from PFO, via ferredoxin, to [FeFe]-hydrogenase, which catalyses the reductive formation of hydrogen gas from free protons. Three maturases, HydE, HydF and HydG, are involved in the insertion of an iron sulfur cluster into the active site of [FeFe]-hydrogenase (Pütz et al. 2006; Broderick et al. 2014). The 51kDa and 24kDa subunits of mitochondrial electron transport chain Complex I are commonly found in anaerobic eukaryotes in the absence of other Complex I subunits (Stairs, Leger, Roger 2015 (in press)), and can transfer electrons to ferredoxin (Hrdý et al. 2004). [FeFe]-hydrogenase and the Complex I subunits may simultaneously oxidize ferredoxin and NADH, likely using an electron bifurcation mechanism (Müller et al. 2012) that has been described in Thermotogales (Schut, Adams 2009) (in which homologs of the Complex I subunits form a trimeric complex with an [FeFe]-hydrogenase). While this pathway is present in some MROs, the characteristic enzymes involved have also been found in some anaerobic eukaryotes with Class 5 MROs, where they are instead located in the cytosol (Lindmark 1980; Nixon et al. 2003; Emelyanov, Goldberg 2011; Müller et al. 2012; Stairs, Leger, Roger 2015 (in press)).

Class II mitochondria, found in animals such as platyhelminths, nematodes and some bivalves, possess a complete electron transport chain, and can function aerobically. However, under hypoxic conditions, they generate ATP using a modified electron transport chain (Van Hellemond et al. 1995; Tielens, Van Hellemond 1998; Tielens et al. 2010; Müller et al. 2012; Stairs, Leger, Roger

2015 (in press)). In this case, Complex I oxidizes NADH and pumps protons to the intermembrane space (as it does under aerobic conditions), generating an electron potential that drives ATP synthesis by Complex V. However, electrons are transferred from Complex I to a specialized low redox potential quinone, rhodoquinone. Rhodoquinone is reoxidized by succinate dehydrogenase subunits (Complex II) functioning as a fumarate reductase, catalyzing the reduction of fumarate to succinate, the reverse reaction to that catalyzed by succinate dehydrogenase in aerobic mitochondria. This reaction is a step in the conversion of malate to propionate, a process called malate dismutation that also produces ATP by substrate-level phosphorylation (Müller et al. 2012; Stairs, Leger, Roger 2015 (in press)). The enzymes involved in malate dismutation (malate dehydrogenase, succinate dehydrogenase/fumarate reductase, methylmalonate CoA mutase, methylmalonate CoA-epimerase, and propionyl-CoA carboxylase) are normally present in aerobic mitochondria, where they function in the TCA cycle and in fatty acid metabolism respectively. ATP is further produced during the production of acetate by an ASCT/STK system similar to that described above (Tielens et al. 2010).

A similar ATP-production pathway, combining some electron transport chain components, rhodoquinone, and enzymes involved in malate dismutation, has recently been found in a number of obligate anaerobes (Gill et al. 2007; Lantsman et al. 2008; Stechmann et al. 2008; Stairs et al. 2014; Nyvltova et al. 2015) (reviewed in (Stairs, Leger, Roger 2015 (in press))). In these organisms, other components of the electron transport chain and of the TCA cycle have been lost, such that they are no longer able to produce ATP by oxidative

phosphorylation. Furthermore, they combine this system with some or all elements of the hydrogenosomal ATP production pathway described above, suggesting they should be classified as Class 3 mitochondria under the 5-class scheme.

Naegleria gruberi is a hitherto unique (see Chapter 2) case of a eukaryote that appears to combine a complete electron transport chain and TCA cycle (and thus the capacity to perform oxidative phosphorylation) with at least some enzymes characteristic of the hydrogenosomal ATP generation pathway. While it lacks PFO, PNO or PFL, its nuclear genome encodes [FeFe]-hydrogenase and its maturases (Fritz-Laylin et al. 2010; Fritz-Laylin et al. 2011). There is controversy as to whether these enzymes function in the mitochondria or not, as localization studies have failed to confirm the *in silico* predictions of mitochondrial localization (Fritz-Laylin et al. 2010; Tsaousis et al. 2014). While the presence of these enzymes makes *Naegleri*a metabolically unique in any case, their mitochondrial localization would suggest an organelle fitting the criteria of Class 3 mitochondria, but more versatile in its capacity to function as either a mitochondrion or a hydrogenosome.

Iron-Sulfur Cluster Assembly

Iron-sulfur (Fe-S) clusters form a part of the active site of a wide range of proteins involved in diverse cellular processes, including transcription, translation, electron transport, and, in anaerobes, anaerobic ATP generation (Vignais, Billoud 2007; Lill et al. 2012). In the mitochondria of aerobes, assembly and insertion of these clusters into Fe-S proteins is performed by a

suite of enzymes known as the Iron Sulfur Cluster (ISC) system (Lill et al. 2012). Homologs of these enzymes were among the first detected in MROs, and indeed, were used to demonstrate the mitochondrial ancestry of MROs in localization experiments (LaGier et al. 2003; van der Giezen, Cox, Tovar 2004) and phylogenies (Richards, van der Giezen 2006). Iron sulfur cluster assembly by the ISC system is the only essential function of yeast mitochondria (Lill et al. 1999). For this reason, and because ISC proteins were so commonly found in even highly reduced MROs, it was proposed that the retention of reduced MROs was largely driven by the need for a compartment to house this biogenetic function (Embley et al. 2003a; Tovar et al. 2003). However, MROs have now been discovered in which the ISC system has been replaced by one of two Fe-S cluster assembly systems of bacterial origin. The NItrogen Fixation (NIF) system is found in the MROs and cytosols of amoebozoa *Entamoeba histolytica* (Ali et al. 2004; van der Giezen, Cox, Tovar 2004; Maralikova et al. 2010) and its close relative *Mastigamoeba balamuthi* (Gill et al. 2007; Nyvltova et al. 2013) (although immunolocalization (Maralikova et al. 2010) and proteome (Mi-ichi et al. 2009) data are at odds as to whether or not the pathway is truly located in the MROs of *Entamoeba*). Meanwhile, the SUlFur mobilization (SUF) system has replaced the ISC system in the MROs of the breviate *Pygsuia biforma* (Stairs et al. 2014).

Protein Import and Folding

The mammalian mitochondrial proteome comprises over 1000 proteins (Calvo, Mootha 2010), while the largest mitochondrial genome known encodes 64 proteins (Burger et al. 2013). The vast majority of mitochondrial proteins are

instead nucleus-encoded, and imported into mitochondria and MROs with the aid of N-terminal or internal targeting signals. N-terminal targeting peptides are recognized and cleaved by components of the mitochondrial protein import machinery. Once inside the organelle, proteins are bound by mitochondrial chaperones and refolded.

Mitochondrial protein import has been poorly studied in eukaryotes other than yeast and mammals. Recent surveys that include data from more diverse taxa (Eckers et al. 2012; Heinz, Lithgow 2013; Gawryluk et al. 2014; Stairs et al. 2014) suggest that even those with aerobic mitochondria lack some of the components found in the best-studied *Saccharomyces cerevisiae,* although these taxa doubtless have unique mitochondrial protein import components of their own that have not yet been discovered.

Nevertheless, the core components of the mitochondrial protein import machinery are conserved among MROs, even where many of the accessory proteins have been lost (Henriquez et al. 2005; Regoes et al. 2005; Waller et al. 2009; Dolezal et al. 2010; Jedelsky et al. 2011; Schneider et al. 2011; Stairs et al. 2014; Nyvltova et al. 2015) (reviewed in (Maguire, Richards 2014)). The rhizarian *Mikrocytos* is the only organism in which no protein import components other than chaperone Hsp70 been found to date (Burki et al. 2013), although this is most likely an artifact of the relatively few sequences obtained in that study. The chaperones Hsp70 and Cpn60 are ubiquitous among MROs, and so well conserved that they were used as the initial indicators that the 'amitochondriate' eukaryotes had once had mitochondria (Roger, Clark, Doolittle

1996; Germot, Philippe, Le Guyader 1997; Hirt et al. 1997; Roger 1998; Mai et al. 1999). Mitochondrial protein targeting and import is also well conserved, and the location of MRO-targeted proteins can often be predicted by *in silico* prediction methods. The N-terminal targeting peptides of *Giardia* are less well-conserved (Šmíd et al. 2008). This is likely linked to its mitochondrial processing peptidase, which consists of a single beta subunit, rather than a homodimer of alpha and beta subunits as in other eukaryotes (Šmíd et al. 2008). Despite this fact, proteins targeted to *Giardia* MROs can be recognized by the protein import system of *Trichomonas* MROs (Dolezal et al. 2005); possibly this is because *Trichomonas* proteins themselves may not be relying only on N-terminal targeting peptides for import into the MROs (Zimorski et al. 2013).

Other MRO Functions

Relatively little emphasis has been placed on mitochondrial functions such as amino acid or lipid metabolism in the study of MROs. The glycine cleavage system appears to be the best represented amino acid metabolism pathway in MROs (reviewed in (Stairs, Leger, Roger 2015 (in press))). Although it is only partially present in the MROs of *Trichomonas* (Schneider et al. 2011) and *Spironucleus* (Jerlstrom-Hultqvist et al. 2013), and absent from the MROs of microsporidia, *Giardia* (Jedelsky et al. 2011), *Entamoeba* (Loftus et al. 2005), and *Cryptosporidium* (Henriquez et al. 2005), it is complete in those of *Trimastix* (Zubacova et al. 2013), *Mastigamoeba* (Nyvltova et al. 2015) and *Pygsuia* (Stairs et al. 2014). The *Trichomonas* MRO proteome additionally includes several other proteins involved in amino acid metabolism, as does the predicted MRO proteome of *Pygsuia* (Stairs et al. 2014). The latter also includes

other ancestrally mitochondrial functions, such as a complete cardiolipin biosynthesis pathway, and phosphonolipid biosynthesis and fatty acid metabolism enzymes (Stairs et al. 2014).

It appears increasingly likely that if there is any universally conserved component to the MRO proteome (Figure 1.2), that component is limited to elements required for the maintenance of basic functions of the organelle - namely, mitochondrial protein import machinery components required for the import of nuclear-encoded proteins, and chaperonins required for their correct folding (Maguire, Richards 2014).

Mitochondrial Origins in Light of MROs

The idea that mitochondria might be descended from free-living prokaryotes was elaborated in 1918 by Paul Portier, and in 1926 by Ivan Wallin (reviewed in (Wallin 1926)). This followed Constantin Mereschkowsky's earlier elaboration of a similar hypothesis for the origins of plastids (Mereschkowsky 1905), and the notion that eukaryotic cells could be partitioned into smaller, independent elementary units ('bioblasts') (Altman 1890; Wallin 1926). A modern revival of the idea of a symbiotic origin for mitochondria was formulated by Lynn Sagan (later Margulis) in 1967 (Sagan 1967) as part of what would later become known as the serial endosymbiotic theory. Today, the specific phylogenetic affinities and properties of both the α-proteobacterial endosymbiont that gave rise to mitochondria and the host that engulfed it are areas of ongoing debate and investigation.

The Host

An archaeal affinity of the host was initially suggested as archaeal sequence data became available. This revealed proteins involved in transcription, translation and replication machineries that had homologs in eukaryotes, but not in eubacteria, or 'universal' proteins that were more similar to eukaryotic homologs (Auer, Lechner, Bock 1989; Kimura et al. 1989; Puhler et al. 1989; Bult et al. 1996). Additionally, rooting analyses using ancient gene duplications (of ATPase subunits and elongation factors) placed the root of the tree of life between eubacteria, and a clade containing eukaryotes and archaea (Gogarten et al. 1989; Iwabe et al. 1989). These factors prompted Carl Woese to place archaea and eukaryotes sister to one another in the tree of his newly proposed three domains of life (Woese, Kandler, Wheelis 1990). Subsequent work has confirmed that eukaryotic genomes are mosaics of genes of archaeal or bacterial origin, and that the former make up a large part of the informational gene complement of eukaryotes (Rivera et al. 1998; Rivera, Lake 2004; McInerney, O'Connell, Pisani 2014). Woese's tree (the 'three-domains tree') placed eukaryotes and Archaea as distinct clades descended from a common ancestor; however, an earlier model had proposed eukaryotes emerging from within a paraphyletic Archaea (the 'eocyte tree') (Lake et al. 1984). Recent work supports the latter hypothesis (Foster, Cox, Embley 2009; Williams et al. 2012; Spang et al. 2015), and refines the position of eukaryotes. The original eocyte tree placed eukaryotes sister to 'eocytes' (now Crenarcheota) based on an apparent similarity in ribosome structure (Henderson et al. 1984; Lake et al. 1984); in agreement with this, a more recent phylogenomic analysis placed eukaryotes in a clade with the

archaeal superphylum encompassing Thaumarcheota, Aigarchaeota, Crenarcheota and Korarchaeota (the 'TACK' superphylum) (Guy, Ettema 2011; Williams et al. 2012). A recently discovered phylum in this group (the 'Lokiarchaeota') share more proteins with eukaryotes than any other archaeon described to date, and are likely the closest archaeal relatives to eukaryotes (Spang et al. 2015). The bioinformatically reconstructed genomes of these organisms encode eukaryotic-like actin isoforms, and homologs of ESCRT proteins, which in eukaryotes are involved in membrane deformation (Spang et al. 2015). Thus, it is highly possible that they possess the actin-based cytoskeleton and membrane dynamics that would make them capable of phagocytosis. If this is indeed the case, the Lokiarcheota may resemble the hypothetical phagocytosing proto-eukaryote posited by Cavalier-Smith (Cavalier-Smith 1983a)

The Endosymbiont

The eubacterial, and specifically α-proteobacterial nature of the protomitochondrial endosymbiont was originally established by similarities in cytochrome c structures (Dickerson 1980; Gray, Doolittle 1982) and subsequently rRNA and protein phylogenies (Spencer, Schnare, Gray 1984; Yang et al. 1985; Keeling, Doolittle 1997); and, most recently, by a shared complement of proteins linked to cristae formation (Munoz-Gomez et al. 2015). Many attempts have been made to determine the closest living α-proteobacterial relatives to mitochondria. However, mitochondrial genomes are small, A+T-rich, and their protein genes tend to evolve at a rapid rate; all of these features have led to suspicions that they are artificially branching sister to α-proteobacteria

with similar features. These 'suspicious' candidate sister-lineages have included Rickettsiales (Fitzpatrick, Creevey, McInerney 2006) and *Pelagibacter* species (the SAR11 clade) (Georgiades, Raoult 2011; Thrash et al. 2011; Rodriguez-Ezpeleta, Embley 2012). A more recent candidate is a clade of α-proteobacteria identified from metagenomic data, the Oceanic Mitochondrion-Affiliated Clade (OMAC) (Brindefalk et al. 2011).

The question of what drove the initial endosymbiotic event is the subject of active debate. Below, I outline several hypotheses that have been put forward to explain the initial association between host and endosymbiont. I will discuss those hypotheses that focus primarily on the nature of the endosymbiont(s), and the underlying factors that might have made such an association beneficial. I do not address hypotheses that focus on non-mitochondrial features of eukaryotic cells, or the relative timing of when they might have arisen; nor on the long-term consequences of the endosymbiosis on the newly mitochondriate eukaryote.

The Hydrogen Hypothesis

In 1998, Martin and Müller proposed the hydrogen hypothesis (Martin, Müller 1998). They emphasized that no examples are known today where a microbe produces ATP in excess to its own demands, and furthermore is able to export that excess ATP to a new host – two conditions that would have had to be present, had the original endosymbiosis been formed on the basis of ATP transfer from the endosymbiont to the host (Martin, Müller 1998). They outlined a scenario in which the archaeal host was an obligately autotrophic, hydrogen-dependent anaerobe, while the endosymbiont was a facultative anaerobe,

capable of producing hydrogen as a byproduct of ATP production in a manner similar to that of modern hydrogenosomes. Anaerobic ATP production enzymes would therefore have been ancestrally present in eukaryotes, and would subsequently have been lost from the mitochondria of aerobic eukaryotes. They further posited that the host might have been a methanogen, pointing out that close associations have been observed between modern hydrogenosomes and methanogenic endosymbionts of eukaryotes (Finlay, Embley, Fenchel 1993).

Under such a scenario several predictions can be (and, in some cases, have been) made. 1), that enzymes associated with anaerobic ATP production will be found in a variety of modern eukaryotes with an obligately or facultatively anaerobic lifestyle (Martin, Müller 1998). 2), that this complement of enzymes should be more or less similar for all eukaryotes in which they are found (Müller et al. 2012). Furthermore, these enzymes should have been ancestrally present in the mitochondria, and genes encoding any of those enzymes that are cytosolic in modern eukaryotes were transferred to the nucleus from the mitochondrial genome. 3), that eukaryotes should form a clade in phylogenies of these enzymes. 4), that the closest prokaryotic relatives to the protomitochondrial endosymbiont – in other words, α-proteobacteria – have, at least in some cases, retained these enzymes also, and form a clade with eukaryotes in phylogenies. In later formulations of the hydrogen hypothesis, the possibility of lateral gene transfer among bacteria following the endosymbiotic event has been cited as a confounding factor that might negate prediction 4) (Müller et al. 2012). Specifically, it was suggested that eukaryotes might branch more closely with different prokaryotes because of LGT between prokaryotes and gene loss in α-

proteobacteria after the emergence of LECA. This explanation is immune to testing or logic; and it could theoretically be used to explain away any topology or pattern of gene presence/absence not in agreement with *any* favoured hypothesis, not only the hydrogen hypothesis.

The Syntrophic Hypothesis

In the same year as the hydrogen hypothesis appeared, Moreira and López-García presented the syntrophic hypothesis (Moreira, Lopez-Garcia 1998) (later also referred to as the syntrophy or sulfur syntrophy hypothesis). Like the hydrogen hypothesis, this posited that the host was a methanogenic archaeon (although it should be noted that a methanogenic nature is not crucial to the hydrogen hypothesis); unlike the hydrogen hypothesis, however, they posit that the host had already formed a stable endosymbiosis with an anaerobic, sulfate-respiring δ-proteobacterium, prior to a later endosymbiotic event involving the α-proteobacterial endosymbiont that gave rise to mitochondria. Moreira and López-García cite the existence of known symbioses between methanogenic archaea and δ-proteobacteria in support of the plausibility of such a scenario (Zinder, Koch 1984). In this scenario, the initial association between the archaeaon and the δ-proteobacterium would be driven by the archaeon benefitting from the hydrogen produced by the δ-proteobacterium, much as it would benefit from the hydrogen produced by the α-proteobacterium in the hydrogen hypothesis scenario. The syntrophic hypothesis differs from the hydrogen hypothesis in several important respects (Moreira, Lopez-Garcia 1998). It posits more than one single host and δ-proteobacterial endosymbiont,

but rather a eukaryote emerging from a stable consortium of multiple cells, with eukaryotic endomembrane systems emerging as a result of this association. It also posits an anaerobic methanotrophic α-proteobacterium, rather than a fermentative anaerobe. The authors point out that methanotrophy is widespread among α-proteobacteria (Hanson, Hanson 1996), and that many methanotrophic α-proteobacteria in particular are endosymbionts, including of eukaryotes (Hanson, Hanson 1996); and that they form symbiotic associations with δ-proteobacteria (Larkin, Henk 1996).

As with the hydrogen hypothesis, the authors emphasize the need for the α-proteobacterial endosymbiont to have been an anaerobe, and for the initial endosymbiosis to have been based on something other than ATP transfer from the endosymbiont to the host: "No bacterium gives free ATP to the medium" (Moreira, Lopez-Garcia 1998).

Such a scenario would need to make the following predictions: 1) that the closest archaeal relatives of eukaryotes should likely include methanogens. 2) that phylogenies of eukaryotic proteins should recover a clear δ-proteobacterial signal. However, recent work from other groups has identified members of the TACK group as likely sister groups to eukaryotes (Williams et al. 2012; Spang et al. 2015); these do not appear to include methanogens (Petitjean et al. 2015). Furthermore, systematic analyses of eukaryotic gene phylogenies have not recovered either a methanogenic or a δ-protebacterial signal (Rochette, Brochier-Armanet, Gouy 2014).

The Pre-Endosymbiont Hypothesis

Michael Gray's recently formulated pre-endosymbiont hypothesis (Gray 2014) did not arise in response to questions about energy metabolism in the endosymbiosis. Instead, it sought to address the fact that a high proportion of modern mitochondrial proteomes are not of clear α-proteobacterial ancestry (Kurland, Andersson 2000; Schnarrenberger, Martin 2002; Szklarczyk, Huynen 2010). Drawing on the fact that an α-proteobacterial ancestry for the endosymbiont was based primarily on mitochondrial genomes (rather than proteomes), this hypothesis proposes that much of the modern mitochondrial proteome was already encoded in the host cell before the endosymbiotic event. Gray proposes the host already had an extensive degree of intracellular organization that included a discrete metabolic compartment in which various metabolic processes were sequestered. This would explain the non-α-proteobacterial component of modern mitochondrial proteomes; it would also allow for easy retargeting of nuclear-encoded proteins, from an endogenous organelle to the α-proteobacterium. Gray's scenario has both a host and an endosymbiont that may already be aerobic. He makes no speculations as to what the initial benefit of the symbiosis might have been, but he points out that modern aerobic eukaryotes are perfectly capable of forming stable, long-term endosymbioses with α-proteobacteria (Marciano-Cabral 2004). Following the establishment of such a symbiosis, he reasons that the evolution of a means of ATP transport from the endosymbiont to the host would have provided an additional benefit in the form of an energy boost.

The pre-endosymbiont hypothesis does not make clear predictions that can be supported or refuted by work in this thesis. However, it does raise serious questions regarding the import of proteins to the pre-endosymbiont compartment, and to the protomitochondrion. The hypothesis posits that N-terminal targeting peptides were already present on many proteins, allowing them to be targeted to the hypothetical compartment. At least some of the mitochondrial protein import machinery would therefore have been present in the hypothetical compartment; is the modern machinery in this hypothesis a hybrid between this machinery and α-proteobacterial components? Or, if it was completely derived from the hypothetical compartment, how did it all come to be transferred to the protomitochondrion? Iron-sulfur cluster assembly proteins have clear α-proteobacterial affinities in phylogenies (Richards, van der Giezen 2006), yet are encoded in the nucleus. It was clearly not a great obstacle for these proteins to acquire N-terminal targeting peptides, so the presence of existing targeting peptides does not seem to present a great advantage.

The Lateral Gene Transfer (LGT) Model

For some time, responses to the hydrogen hypothesis have suggested alternative explanations for the anaerobic ATP generation enzymes seen in some eukaryotes. These have sought to address the fact that these enzymes, and indeed an anaerobic lifestyle, are very patchily distributed in eukaryotes (it should be noted that the degree to which this is true is disputed by proponents of the hydrogen hypothesis (Müller et al. 2012)), and have invoked lateral gene transfer as an explanation (Andersson et al. 2003; Andersson et al. 2006; Andersson 2009b; Andersson 2009a; Stairs, Roger, Hampl 2011; Takishita et al.

2012; Leger et al. 2013). Recently, Courtney Stairs, Andrew Roger and I formulated some of these alternatives as the LGT model (Stairs, Leger, Roger 2015 (in press)). This model makes no predictions about the original nature of the endosymbiont or the host, nor about the factors driving the endosymbiosis; but it addresses the origins of anaerobic enzymes in modern eukaryotes. We propose that anaerobic ATP generation enzymes were initially acquired in eukaryotes at least once after the establishment of the mitochondrial endosymbiosis, after which they were laterally transferred multiple times between eukaryotic lineages. We also outline a scenario for the gradual loss of the electron transport chain from MROs following the acquisition of various anaerobic ATP generation enzymes, leading to the spectrum of energy metabolism currently observed in MROs of extant organisms.

I propose that this scenario leads to several predictions: 1) that an anaerobic lifestyle, and in particular anaerobic ATP generation enzymes, should be patchily distributed among eukaryotes, often arising in taxa that are related only, or mostly, to aerobes. 2) that the aerobic lineages basal to these anaerobic lineages do not, or do not often, possess anaerobic enzymes like those found in *Naegleria gruberi*. 3) that some heterogeneity is to be expected in the complement of anaerobic ATP generation enzymes borne by individual eukaryotes, explained by multiple possible enzymes carrying out a given function. 4) phylogenies of these enzymes may have one or several eukaryotic clades, but the internal topology within a eukaryotic clade will not reflect the organismal relationships where distantly-related eukaryotes are present. 5) the anaerobic enzymes will be absent from, or sparsely and patchily distributed in, α-proteobacteria. 6) homologs from

α-proteobacteria are not expected to be over-represented relative to other bacterial groups as sister group taxa to eukaryote homologs.

Aims of This Thesis

The overarching aim of this thesis is to explore the functional diversity of mitochondrion-related organelles. I use a bioinformatic approach to reconstruct biochemical pathways, predict the localization of the enzymes involved, and examine their possible origins. A first aim is to expand our knowledge of anaerobic ATP generation into novel eukaryote lineages (Chapters 2, 4). A second aim is to explore other pathways present in MROs that have been less well-studied to date (Chapters 3, 4).

Chapter 2 describes anaerobic ATP generation enzymes – [FeFe]-hydrogenase, [FeFe]-hydrogenase maturases HydE, HydF and HydG, pyruvate:ferredoxin oxidoreductase, and acetate:succinate CoA-transferase - in the amoebozoan *Acanthamoeba castellanii*. Together with enzymes commonly found in mitochondria, these enzymes form a fermentation pathway similar to the one characteristic of hydrogenosomes. *A. castellanii* possesses what previously appeared to be an aerobic mitochondrion; nevertheless, all of the proteins discovered possess predicted mitochondrial targeting peptides. Mass spectrometry and immunolocalization experiments (performed by my collaborator and myself respectively) support these porteins being present in the mitochondrion of *A. castellanii*, suggesting that it is capable of functioning anaerobically. This chapter was previously published, with slight changes, as

Leger, M. M., Gawryluk, R. M. R., Gray, M. W. and Roger, A. J. 2013. *Evidence for a Hydrogenosomal-Type Anaerobic ATP Generation Pathway in Acanthamoeba castellanii*. PLoS One 8, e69532. (doi:10.1371/journal.pone.0069532).

Chapter 3 describes putative mitochondrial homologues of an ancestral bacterial division system in eukaryotes. The presence of mitochondrial FtsZ homologues had previously been demonstrated in stramenopiles, amoebozoa, and rhodophytes. However, three proteins involved in controlling the distribution of FtsZ – MinC, MinD and MinE – were not known to play a part in mitochondrial division, and only plastid homologues of these proteins had been described. In this chapter, I show that homologs of these proteins are widely, though sparsely, distributed in eukaryotes, being present in amoebozoa, an ancyromonad, a breviate, stramenopiles, jakobids, and a heterolobosean, and have mitochondrial targeting peptides. Phylogenetic analyses confirm that they branch separately from plastid-localized homologs, and yeast localization experiments performed by my collaborators confirm their predicted mitochondrial localization. This chapter was previously published, with slight changes, as Leger, M. M., Petrů, M., Žárský, V., Eme, L., Vlček, Č., Harding, T., Lang, B. F., Eliáš, M., Doležal, P. and Roger, A. J. 2015. *An ancestral bacterial division system is widespread in eukaryotic mitochondria*. Proc Natl Acad Sci U S A. 2015 Mar 23. pii: 201421392. [Epub ahead of print]

Chapter 4 is an *in silico* and cell biological analysis of proteins predicted to be present in the mitochondrion-related organelle of *Andalucia incarcerata*, an

anaerobic jakobid, based on Sanger, 454 and RNASeq transcriptome data. Jakobids are free-living, mainly aerobic excavates, and the aerobic jakobids are notable for having apparently primitive, bacterial-like mitochondria and mitochondrial genomes. This chapter is the first molecular study of an MRO in this group, and another well-needed study of the MRO of a free-living protist that can be compared with the better-studied MROs of parasites. Perhaps surprisingly in light of *A. incarcerata*'s close aerobic relatives, I find that the MRO of *A. incarcerata* has not retained any elements of the TCA cycle or electron transport other than those (Complex I subunits, STK) habitually retained in hydrogenosomes. I confirm, using immunoelectron microscopy, that the MRO has retained the ISC system, and that it likely functions in anaerobic ATP generation using [FeFe]-hydrogenases. In general, its ATP generation complement closely resembles that of *Trichomonas*. I also find that the MRO has broad mitochondrial protein import and amino acid metabolism components, and has retained a role in cardiolipin synthesis. In terms of non-ATP generation pathways, it most closely resembles the MRO of *Pygsuia biforma*, another free-living organism, rather than the more reduced MROs of parasites.

Chapter 2: Evidence for a Hydrogenosomal-Type Anaerobic ATP Generation Pathway in *Acanthamoeba castellanii*

This chapter was previously published as: Michelle M. Leger[1], Ryan M. R. Gawryluk[1], Michael W. Gray, Andrew J. Roger. 2013. PLoS One 8, e69532. (doi:10.1371/journal.pone.0069532). Its contents have been amended here following the suggestions of the examination committee.

Abstract

Diverse, distantly-related eukaryotic lineages have adapted to low-oxygen environments and possess mitochondrion-related organelles that have lost the capacity to generate adenosine triphosphate (ATP) through oxidative phosphorylation. A subset of these organelles, hydrogenosomes, has acquired a set of characteristic ATP generation enzymes commonly found in anaerobic bacteria. The recipient of these enzymes could not have survived prior to their acquisition had it not still possessed the electron transport chain present in the ancestral mitochondrion. In the divergence of modern hydrogenosomes from mitochondria, a transitional organelle must therefore have existed that possessed both an electron transport chain and an anaerobic ATP generation pathway. Here, we report a modern analog of this organelle in the habitually aerobic opportunistic pathogen, *Acanthamoeba castellanii*. This organism

[1] These authors contributed equally to this work

possesses a complete set of enzymes comprising a hydrogenosome-like ATP generation pathway, each of which is predicted to be targeted to mitochondria. We have experimentally confirmed the mitochondrial localizations of key components of this pathway using tandem mass spectrometry. This evidence is the first supported by localization and proteome data of a mitochondrion possessing both an electron transport chain and hydrogenosome-like energy metabolism enzymes. Our work provides insight into the first steps that might have occurred in the course of the emergence of modern hydrogenosomes.

Introduction

The capacity to produce adenosine triphosphate (ATP) under low oxygen conditions is found throughout the eukaryote tree, in diverse, distantly-related organisms. Of the lineages of this type that have been studied, most are anaerobic or microaerobic, and possess mitochondrion-related organelles (MROs), which, although derived from mitochondria, have lost the capacity to generate ATP through oxidative phosphorylation (reviewed in (Tsaousis et al. 2012a)). Some of these organelles, known as hydrogenosomes, have adopted a new function in anaerobic ATP generation by acquiring a set of characteristic enzymes that are commonly found in anaerobic bacteria (Lindmark, Muller 1973; Dyall et al. 2004; Hrdý et al. 2004). In other anaerobic/microaerobic eukaryotes with more highly reduced MROs, such as *Giardia intestinalis* and *Entamoeba histolytica*, homologous enzymes are localized in the cytosol, and the MROs of these organisms are not involved in ATP generation (Ellis et al. 1993;

Rodriguez et al. 1996). MROs have long been classified according to their role in energy metabolism, and a recent review (Müller et al. 2012) retains the categories of mitochondria (class 1 under the authors' classification system), hydrogenosomes (class 4) and mitosomes (class 5), while proposing new classes to formally accommodate the more diverse range of MROs now known: anaerobically functioning mitochondria that do not produce hydrogen (class 2) and mitochondria that both possess an electron transport chain and produce hydrogen (class 3).

The best-characterized hydrogenosomes are those of the obligately parasitic parabasalid *Trichomonas vaginalis*. In mitochondria, the first step of pyruvate catabolism involves the decarboxylation of pyruvate by the enzyme complex pyruvate dehydrogenase (PDH, EC 1.2.1.51), which produces acetyl-CoA and CO_2 with the concomitant reduction of NAD^+ to NADH. In the hydrogenosomes of *T. vaginalis*, pyruvate decarboxylation is instead carried out by pyruvate:ferredoxin oxidoreductase (PFO, EC 1.2.7.1) (Steinbuchel, Muller 1986; Hrdý, Müller 1995). In other eukaryotes such as *Cryptosporidium parvum*, *Neocallimastix frontalis* or *Mastigamoeba balamuthi*, a similar function is performed by a pyruvate:NADP oxidoreductase (PNO, EC 1.2.1.51) (Buetow 1989; Ctrnacta et al. 2006) or pyruvate:formate lyase (PFL, EC 2.3.1.54) (Gelius-Dietrich, Henze 2004; Stairs, Roger, Hampl 2011). Electrons produced during pyruvate decarboxylation by PFO are transferred first to a [2Fe-2S] ferredoxin, and then to [FeFe]-hydrogenase (EC 1.12.1.4), which reduces protons, producing molecular hydrogen. Three maturases, HydE, HydF and HydG, are involved in the assembly and insertion of the catalytic iron-sulfur clusters in bacterial

[FeFe]-hydrogenases, and homologs of these enzymes have been reported in some, though not all, eukaryotes possessing this enzyme (Posewitz et al. 2004; Pütz et al. 2006; Hug, Stechmann, Roger 2010). Coenzyme A is transferred from the resulting acetyl-CoA to succinate by an acetate:succinate CoA transferase (ASCT, EC 2.8.3.8) (Tielens et al. 2010), generating acetate and succinyl-CoA. The subsequent regeneration of succinate, catalyzed by the tricarboxylic acid (TCA) cycle enzyme succinate thiokinase (STK, also referred to as succinyl-CoA synthetase, EC 6.2.1.4), generates ATP through substrate-level phosphorylation (Hrdý et al. 2004). ASCTs found in eukaryotes have been classified into three subfamilies (Tielens et al. 2010). Enzymes of subfamily 1A, found in the mitochondria of trypanosomatids (Riviere et al. 2004) and in the hydrogenosomes of rumen ciliates (Boxma et al. 2005), are homologous to the succinyl-CoA:3-oxoacid CoA transferases (SCOT) that are involved in ketone body degradation in metazoan mitochondria. Members of subfamily 1B have been reported in the mitochondria of acetate-producing metazoans such as *Fasciola hepatica* (van Grinsven, van Hellemond, Tielens 2009) and *Artemia franciscana* (Oulton et al. 2003). Based on *in silico* predictions, they are also believed to be present in *Blastocystis* sp. (Tielens et al. 2010) and in the mitochondria of *Naegleria gruberi* (Fritz-Laylin et al. 2010). Enzymes of subfamily 1C have been found in the hydrogenosomes of *T. vaginalis* (van Grinsven et al. 2008), *Blastocystis* sp. (Lantsman et al. 2008; Stechmann et al. 2008), and the chytrid fungus Neocallimastix (Marvin-Sikkema et al. 1993b; Tielens et al. 2010).

Until recently, the only aerobic eukaryotes known to possess both [FeFe]-hydrogenase and PFO were green algae such as *Chlamydomonas reinhardtii* and *Scenedesmus* spp. In these organisms, [FeFe]-hydrogenase and PFO are expressed upon exposure to anoxic conditions and localize to the chloroplast, where they function in both anaerobic energy production and anaerobic photosynthesis (Gaffron, Rubin 1942; Happe, Kaminski 2002; Forestier et al. 2003). In 2010, genes encoding an [FeFe]-hydrogenase and the three [FeFe]-hydrogenase maturases were identified in the nuclear genome of *Naegleria gruberi*, an aerobic heterolobosean; *in silico* predictions suggested that these enzymes might be mitochondrially targeted (Fritz-Laylin et al. 2010). No PFO homologs have been found in the genome of this organism.

Previous studies have attempted to clarify the origin of these enzymes in eukaryotes; these efforts have generally been hampered by the small number of eukaryotic sequences available, and by low resolution in all parts of the tree. Phylogenetic analyses of [FeFe]-hydrogenase sequences have consistently recovered more than one eukaryotic clade, suggesting at least two origins of these enzymes in eukaryotes (Horner et al. 2002; Hackstein 2005; Meyer 2007; Vignais, Billoud 2007; Hug, Stechmann, Roger 2010). A specific relationship between eukaryotic [FeFe]-hydrogenases and their homologs in α-proteobacteria has been rejected in topology tests, providing evidence against a mitochondrial endosymbiotic origin of [FeFe]-hydrogenases in extant eukaryotes (Hug, Stechmann, Roger 2010). Analyses of [FeFe]-hydrogenase maturases recovered robust eukaryotic clades in all cases; however, the internal relationships within these clades were poorly supported and their closest prokaryotic homologs were,

in no case, α-proteobacterial. Similar results were obtained by phylogenetic analyses of PFO (Horner, Hirt, Embley 1999; Rotte et al. 2001; Embley 2006; Hug, Stechmann, Roger 2010). A neighbor-net analysis of ASCT1B and ASCT1C sequences (van Grinsven et al. 2008) recovered eukaryote monophyly for both enzymes, consisting of metazoa in the case of ASCT1B, and fungi and *T. vaginalis* in the case of ASCT1C; at that time these taxa were the only eukaryotes known to possess ASCTs. Again, no clear α-proteobacterial affinity for eukaryotic groups was recovered, and thus there is no clear connection to mitochondrial origins. These observations suggest that lateral gene transfer has played a role in the appearance of these enzymes within eukaryotes; however the number of events involved, and the precise nature of the donor and recipient lineages, remain unclear.

Acanthamoeba castellanii is a free-living soil amoeba, found in a diverse range of marine, freshwater, soil and human-related environments. As an opportunistic pathogen, it is responsible for amoebic keratitis and granulomatous amoebic encephalitis in humans (Marciano-Cabral, Cabral 2003), and under free-living conditions, it grazes on bacterial biofilms (Cometá et al. 2011). Thus it is likely that *A. castellanii* routinely encounters anaerobic or microaerobic conditions. Furthermore, while this amoebozoan has been reported to encyst rapidly when exposed to degassing with N_2 (Turner, Biagini, Lloyd 1997), it is now known to respond well to low-oxygen conditions, replicating faster under these conditions than under aerobic ones (Cometá et al. 2011).

In 2010, Hug *et al.* reported partial sequences of a few enzymes associated with anaerobic ATP generation in publicly available expressed sequence tag (EST) data from *A. castellanii* (Hug, Stechmann, Roger 2010). Here, we report the existence of a complete 'hydrogenosomal' type anaerobic ATP generation pathway, describe the genomic and transcript sequences of all enzymes involved, and show that they all possess classical mitochondrial targeting peptides. We show, by tandem mass spectrometry, that three of these enzymes – PFO, ASCT1B, and the [FeFe]-hydrogenase maturase HydF – are found in the mitochondria of aerobically grown cells. Our findings confirm the presence of a complete hydrogenosome-like ATP generation pathway in *A. castellanii* and strongly suggest that the enzymes are present within the mitochondria of this organism. Our results raise the tantalizing possibility that the mitochondrion of *A. castellanii* is able to act as an organelle with two metabolic modes, producing energy either aerobically, via classical oxidative phosphorylation, or anaerobically, via a hydrogenosomal-type pathway, according to the environmental conditions that the amoeba encounters.

Materials and Methods

EST Assembly

454 ESTs available through the Baylor College of Medicine Human Genome Sequencing Center (http://www.hgsc.bcm.tmc.edu/ftp-archive/AcastellaniNeff/ESTs/) were assembled using CAP3 (Huang, Madan 1999) and Mira (Chevreux, Wetter, Suhai 1999; Chevreux et al. 2004).

Database Searching

Partial EST sequences of [FeFe]-hydrogenase, PFO, HydE and HydG previously identified by Hug *et al.* (Hug, Stechmann, Roger 2010) were retrieved by performing Basic Local Alignment Search Tool (BLAST (Altschul et al. 1997)) searches against the publicly available *A. castellanii* expressed sequence tag (EST) library at TBestDB (O'Brien et al. 2007). Genomic sequence data were obtained by performing subsequent BLASTn searches against scaffolds assembled by the Baylor College of Medicine Human Genome Sequencing Center (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA20303; see Appendix A, Table A.S1 for accession numbers), using the *A. castellanii* ESTs as queries. As no ESTs encoding HydF or an ASCT had been identified, we used a *Clostridium kluyveri* HydF protein sequence (EDK34342) and *Trypanosoma brucei*, *Fasciola hepatica* and *Trichomonas vaginalis* ASCT sequences (EAN79240, ACF06126 and XP_001330176, respectively) as heterologous query sequences for tBLASTn searches. Subsequently, full-length cDNA sequences were obtained by performing BLASTn searches against the 454 EST data, using previously identified EST or genomic sequences as queries. The identities of hits were verified by performing BLASTx searches against GenBank. Start codons and intron positions were verified manually. Predicted mitochondrial targeting peptides were identified using TargetP (Nielsen et al. 1997; Emanuelsson et al. 2000; Emanuelsson et al. 2007).

Cell Culture

Cells were maintained at room temp, but otherwise as described in (Lohan, Gray 2007).

Tandem Mass Spectrometry

Mitochondria were isolated from aerobically grown cells, purified on sucrose gradients, and subfractionated as described in (Gawryluk et al. 2012). A whole mitochondrial fraction (SWM), a soluble protein-enriched fraction (SPE), and a mitochondrial membrane protein-enriched fraction (MPE) were separated on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels; the resulting lanes were excised in approximately equally sized bands and digested, as described in (Gawryluk et al. 2012). These fractions were separated by reverse-phase high-performance liquid chromatography (HPLC) and subjected to tandem mass spectrometry (MS/MS) as described (Gawryluk et al. 2012). An additional whole mitochondrial fraction that had not undergone SDS-PAGE was digested in solution (WM) and separated into fractions using strong cation exchange liquid chromatography; these fractions were resolved by reversed-phase HPLC and subjected to MS/MS. Data were acquired and analyzed against the genomic and 454 EST data using Mascot (Perkins et al. 1999), as described in (Gawryluk, Gray 2010).

Phylogenetic Analyses

For each protein, NCBI databases of proteins and EST sequences were searched using BLASTp or tBLASTx respectively, and added to preexisting datasets used in (Hug, Stechmann, Roger 2010) where applicable. Preliminary alignments were made using MUSCLE (Edgar 2004a; Edgar 2004b), FSA (Bradley et al. 2009) or MAFFT-L-INS-I (Katoh et al. 2002; Katoh et al. 2005; Katoh, Toh 2008) and trimmed using BMGE (Criscuolo, Gribaldo 2010) or a script written by Dr. Daniel Gaston. Preliminary trees were made using FastTree (Price, Dehal,

Arkin 2009), or RAxML (version 7.2.6 (Stamatakis 2006), using the Le and Gascuel [LG] model of amino acid substitution rates (Le, Gascuel 2008) with empirical amino acid frequencies and the gamma model of rate heterogeneity [PROTGAMMALGF]. Based on these initial analyses, long-branching taxa and paralogs were eliminated. In particular, a long-branching clade comprising both eukaryotic and bacterial sequences was identified in [FeFe]-hydrogenase analyses. Final analyses were performed both with and without these sequences. Initially, ASCT1C sequences were analyzed together with distantly homologous ASCT1B sequences, and distantly homologous HydE and HydF sequences were analyzed together; this was done to better identify paralogs in the face of widespread misannotation in GenBank.

Final alignments were made using MAFFT-L-INS-I, verified manually, and trimmed using BMGE. Independent maximum likelihood (ML) trees (200) and 1000 bootstrap replicates were generated in RAxML [PROTGAMMALGF] model, and bootstrap values were mapped onto the best-scoring tree. Bayesian inference posterior probabilities were calculated using PhyloBayes (Lartillot, Lepage, Blanquart 2009) under the [catfix C20] model of evolution (Le, Lartillot, Gascuel 2008).

Topology Tests

We tested support for various grouping topologies using the approximately unbiased (AU) test in CONSEL (Shimodaira, Hasegawa 2001). For each hypothesis tested, five ML trees were generated for a given constraint tree, using the PROTGAMMALGF model and the –g option in RAxML. Subsequent

CONSEL analyses used the 1000 bootstrap trees initially produced to generate p-values for the best trees generated from the constrained RAxML analyses.

Results

Of the 143 nuDNA-encoded proteins identified by MS/MS in (Gawryluk et al. 2012), representing respiratory chain proteins (88 proteins) and non-respiratory chain proteins that contaminated preparations of respiratory complexes (55 proteins), none were known non-mitochondrial contaminants and none were confidently predicted to possess sorting signals that direct proteins to other cellular compartments, such as the endoplasmic reticulum or peroxisomes; accordingly, this protocol was deemed to have produced sufficiently pure mitochondria samples. In order to confirm that the samples used in this particular experiment were highly enriched in mitochondria, we compared the gel electrophoretic profile of the rRNA species recovered from the purified mitochondria to that of total cellular RNA, as a proxy for protein profiles (Appendix A, Fig. A.S1). The RNA profiles were distinct; in particular, we were unable to detect cytosolic LSU or SSU species in our purified mitochondrial samples, indicating a high degree of mitochondrial enrichment in these samples.

The A. castellanii Genome Encodes a Complete Anaerobic ATP Generation Pathway Similar to That Found in T. vaginalis Hydrogenosomes.

Hug *et al.* (Hug, Stechmann, Roger 2010) had identified partial sequences of [FeFe]-hydrogenase, PFO, HydE and HydG in the *A. castellanii* transcriptome (Appendix A, Table A.S1). We have identified corresponding genomic sequences

for all four genes. Searches using a *Clostridium* HydF sequence and ASCT sequences from *Trypanosoma brucei*, *Fasciola hepatica* and *Trichomonas vaginalis* as queries yielded a HydF homolog and two possible candidates for an ASCT, homologous to the *T. brucei* (subfamily 1A) and the *F. hepatica* (subfamily 1B) enzymes. The subfamily 1A enzyme of *T. brucei* is homologous to succinyl-CoA:3-ketoacid-CoA transferase (SCOT), an enzyme that is widespread in mammalian and fungal mitochondria, and that catalyzes the transfer of CoA from succinyl-CoA to a 3-oxoacid. The top BLASTp hits for this candidate were SCOT homologs from *Polysphondylium pallidum* and *Dictyostelium discoideum*, two other, 'cellular slime mold', amoebozoans that do not appear to possess anaerobic ATP generation enzymes. In contrast, the *F. hepatica*-type enzyme has been described in platyhelminths and arthropods, and is homologous to bacterial 4-hydroxybutyrate CoA-transferases; we were unable to find homologs of this enzyme in the *Polysphondylium pallidum* or *Dictyostelium* genome sequences available through dictyBase (Gaudet et al. 2011). Accordingly, we concluded that the *F. hepatica* enzyme hit was the more likely candidate to function in a pathway with the other anaerobic enzymes we had discovered. The *A. castellanii* genome encodes a single adrenodoxin-like [2Fe-2S] ferredoxin, homologous to eukaryotic mitochondrial ferredoxins, which may function as the electron mediator from PFO to [FeFe]-hydrogenase; and STK, which may perform dual functions in the TCA cycle and in anaerobic ATP generation.

Full-length EST sequences for each of these enzymes were retrieved from 454 pyrosequencing data, confirming the locations of spliceosomal introns in the genomic sequences. All of the genomic sequences contained canonical 5′GT-AG3′

spliceosomal introns, refuting the possibility that the genes we identified originated from bacterial contamination of the transcriptomic and genomic data (see Appendix A, Table A.S1 for the list of accession numbers). Interestingly, the genes encoding [FeFe]-hydrogenase and all three maturases are encoded within a single ~50-kb stretch of the genome (Fig. 2.1).

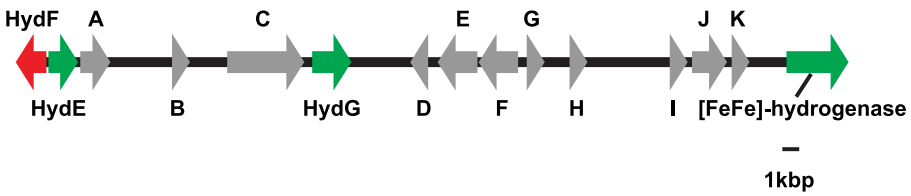


Figure 2.1 Map of the genomic segment encoding [FeFe]-hydrogenase and its associated maturases. The direction of the arrows indicates the transcriptional orientation of the genes. Green, anaerobic metabolism enzyme-encoding gene located on the forward strand. Red, anaerobic metabolism enzyme-encoding gene located on the reverse strand. Gray, predicted gene encoding a product not evidently involved in anaerobic metabolism. Predicted genes annotated according to top BLAST hits in dictyBase: A. Similar to interferon-related protein PC4-like. B. Region with similarity to *Dictyostelium* hypothetical protein, possibly truncated at the 5′ end. C. Similar to importin beta 4. D. Low similarity to *Dictyostelium* vasodilator-stimulated phosphoprotein. E. Similar to RNA-binding region RNP-1 domain-containing protein. F. Similar to molybdenum cofactor synthesis protein 1. G. Similar to molybdenum cofactor synthesis protein 2. H. Region with similarity to sequences annotated as hypothetical protein in GenBank, but without tBLASTx hits in dictyBase. I. Similar to PHD Zn finger-containing protein. J. Low similarity to *Dictyostelium* hypothetical protein. K. Region without significant hits in either GenBank or dictyBase, but which corresponds to EST sequence and apparently contains an intron.

Within this region, 11 other predicted genes were found that had homologs returned by BLAST searches. With the exception of two genes encoding hypothetical proteins, all of the predicted genes had significant tBLASTx hits (with an E-value ≤ 10^{-3}) to one of more of the amoebozoan genomes available through dictyBase. None of these additional genes has any obvious function in anaerobic respiration.

TargetP predicted high probabilities of mitochondrial localization, and identified putative targeting peptide cleavage sites, for all of the anaerobic ATP generation enzymes (Fig. 2.2).

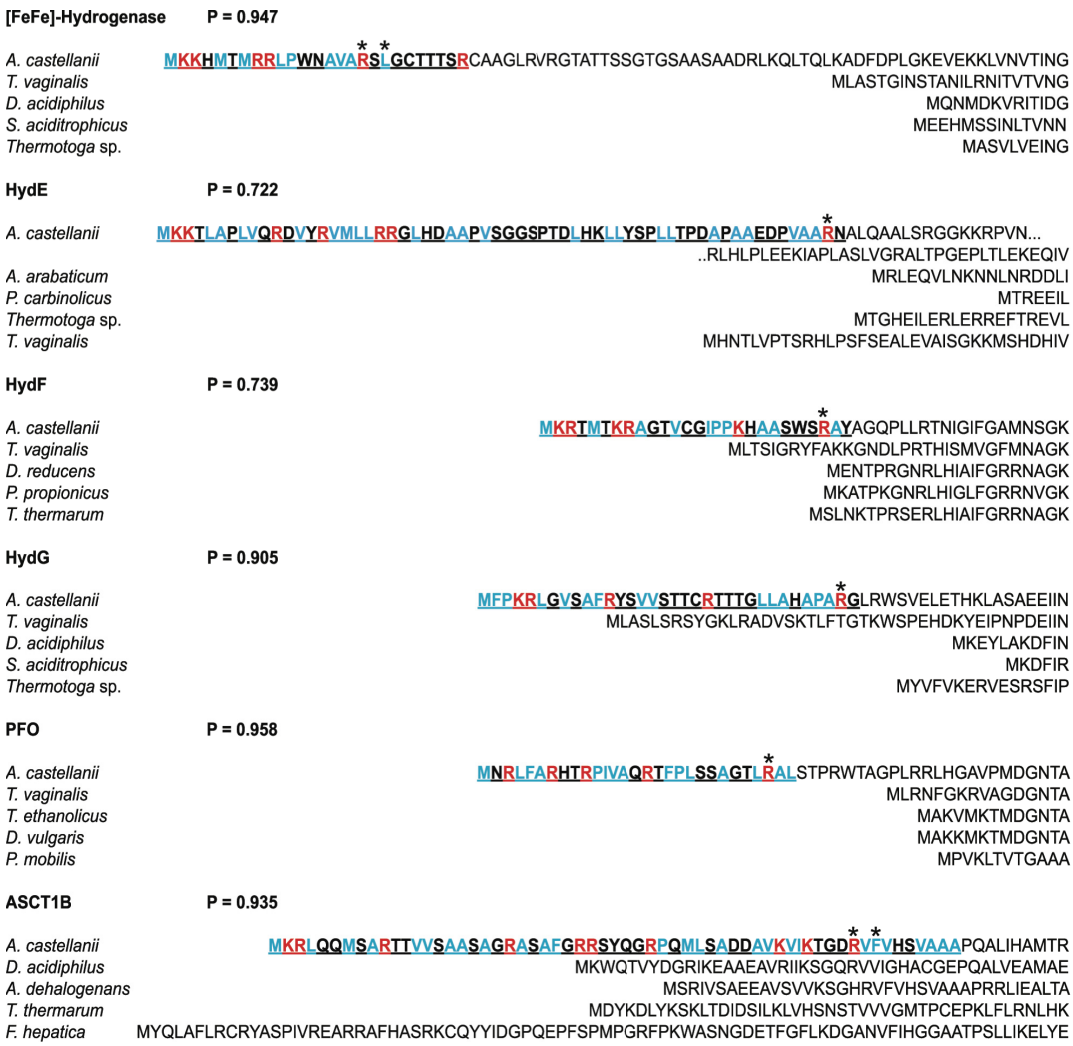


Figure 2.2 N-termini of anaerobic energy generation enzymes in *A. castellanii*, showing TargetP-predicted mitochondrial targeting peptides (mtTPs). Predicted mtTPs are shown underlined and bold. N-termini of bacterial and eukaryotic homologues are shown for comparison; bacterial homologues lack targeting peptides. Positively charged residues in the predicted mtTPs are shown in red; hydrophobic residues are shown in blue. Arginine residues at positions -2, -3 or -10, and hydrophobic residues at position -8, believed to be important in determining the cleavage site (Schneider et al. 1998), are marked with an asterisk. P, TargetP mitochondrial targeting probability.

The predicted mitochondrial targeting peptides (mtTPs) are rich in hydrophobic and positively charged amino acids, consistent with the amphipathic helix structure that mitochondrial targeting peptides are known to adopt (Schneider et al. 1998). The mtTPs have arginine residues at positions -2, -3 or -10 relative to the cleavage site, as well as positively charged residues at position -8 in the latter case. Such residues are believed to be important in determining the site of targeting peptide cleavage (Schneider et al. 1998). The amino acid composition of the predicted targeting peptides is consistent with those predicted by TargetP for other nucleus-encoded proteins known to be mitochondrially targeted in *A. castellanii*, such as mitochondrial malate dehydrogenase, dihydrolipoamide dehydrogenase and isocitrate dehydrogenase (data not shown).

PFO, ASCT1B and the [FeFe]-Hydrogenase Maturase HydF Are Present in the Mitochondrial Proteome

Peptides diagnostic of PFO, ASCT1B and HydF were detected in the mitochondrial protein fractions by tandem mass spectrometry (Table 2.1, Fig. 2.3).

Table 2.1 Anaerobic ATP generation enzymes identified in tandem mass spectrometry experiments. [1] $P < 0.05$ for ion score ≥ 3

	Unique peptides	**Ion score[1]**	**Fractions (see Materials and Methods)**
[FeFe]-hydrogenase	0	N/A	N/A
HydE	0	N/A	N/A
HydG	0	N/A	N/A
HydF	1	59	WM
Pyruvate:ferredoxin oxidoreductase	30	1457	WM, SWM, SPE
Acetate:succinate CoA-transferase type 1B	17	1900	WM, SWM, SPE
Succinyl-CoA 3-oxoacid transferase	14	1631	WM, SWM, SPE
Ferredoxin	3	141	SPE

Pyruvate:Ferredoxin Oxidoreductase

MNRLFARHTRPIVAQRTFPLSSAGTLRAL↓STPRWTAGPLRRLHGAVPMDGNTAAAHVAYGL
SDIHAIYPITPSSQMGELADKWSAEGRLNAFGNTPRVIEMQSEMGAAGTLHGAAVGGALVSTF
TASQGLLLMIPNLYRVAGELMPAVFHVTARAISGQGLSIYGDHSDVMAVKQTGVAMLASASPQ
EAMDLALVAHLSSIRSSVPFVHFFDGFRTSHEINTVEPIKYEDMRKLLDEEALEQFRRRGMNPE
TPNLRGLIDGPEHYFQQVEAANTILDGVLPVVEGYLDEVHKLTGRKYGLFDYHGHPEPRHVIVA
CGSSVSTVEEAVNHRNAQGERVGLIKVRLWRPFSIKHLVDALPKSVEKVAVIDRVRDYLASGG
PLFQEVCTSLMMGGRRDVQLVNGRYGLGSKDFTPGMALAIFDNLKQDQPLHNFVVGIKDDVT
HKSLTVTEEPDTLPAGTKQSIFWGIGGDGTVGANEEAIKLIVENSNMYGQGYFAYSAHKSGGV
TVSHLRFGEKPINSTYQVQNADLIAVHTTPYLKKFPSLLGPLKEGGTVILNSPWNDVAHLDRML
PDFVKRRIARRKARLINVDATAIAHEAGLRGRINMVMQAAFFKASEVLPLDVARAELRRVIDAQ
YARKGRDVLERNYAALDQGLARTVEVAYPDAWAECRDDVADYIVPDPADAPEQLRKVLRPTQ
RMEGDSLPVSAFDPRGAMPSGTSKYEKRGIAPAVAQWTNPDTCTQCNLCSALCPHAAIRPFL
FTQEEAGSAPEGWEGRKAVGKAGKSYQYRVQVSPYDCTGCDVCVKACPTQSLAQVPFVDAL
DRGQARLWDFAAERLPIRSEVYPKESLKGSQFAKPCLEFSGACAGCGETPVVKLLTQLFGDE
LYIANATGCSIVWGGMFPWSAYTTNERGHGPAWGHSLFEDAAEYGFGIRHAVRYRREALRCA
VQRDLTAGAYAAEHPVLAELLRRWDAAYDDRTQSPSLAAKVREYLERLPAPAAAARGPLREL
HAERHMLARKTQWIIGGDGWAYDIGFGGLDHVLASGEKVNVLVLDNEVYANTGGQASKATPR
ASQVKFANAGKTTAKKDLGAMMMQYGNVYVASICLEANPDHAVQALAEAEAFDGPSLVIAYA
PCIAHGIKAGISTEVEEAKRAIKAGYHILYRYNPSLVEQGMNPLSLDSSPPDDQLLQFLRGENR
YEALRQQHPELTDEKQRLLVRDVADRYRHYALLKEQLEPKDDGEEEEEKADEAKEEATA

HydF

MTKRAGTVCGIPPKHAASWSRAY↓AGQPLLRTNIGIFGAMNSGKSTLMNLISQQETSIVDSKP
GTTADTKVALMEMHDLGPVKLFDTPGIDEEGLLGEKKRRKAFDVLKECNAAVVVVNPFNPASL
KAARDVIQEASAKQKKGDDSAQMRVMVVFNVFGQQMAEIRKATNTILDAAEKSLTPDRTSLQI
TSIALDLNSPEAMGRVVKFVTTNTKPHSSNVSLLPSALQLGPDSVVFLNIPMDAETPSGRLLRP
QALVQEELLRQYASTFCYRMDLKKARSPIEEERREEEQRFRSSVDALKSQNKLKLLITDSQAM
DVVHKWTMEPGTAAPSDQGGQSSQETVPLTTFSVMMINYMSGGRLSAFVEGIKRFETLKHGD
KVLICEACNHDRIQDDIGTVQIPAKLRQRFGEGTIGVDHAFGREYQTKILNDYQLVIHCGGCMLD
QQKMAARLSDIEGSGVPITNYGLLLSYLAAKQGLSRVLRPWGL

ASCT1B

MKRLQQMSARTTVVSAASAGRASAFGRRSYQGRPQMLSADDAVKVIKTGDRVFVHSVAAA↓
PQALIHAMTRRAPELRDVEVCHMHIEGDASYADKKYEGSFKNNNFFVGKNVRKGVQEGRLD
YTPVFLSEIPLLFRRGILPLDVALITVSPPDQHGFCSLGTSVDASLAAVQCAKTVIAQVNPHMPR
THGDGFVHESAISFMVDGPAPLIEHKRGKVTETIGKIGKNVAQLVEDGATLQMGIGVIPDAVLAE
LTHHKKLGIHTEMFSDGIIDLVERGVITGENKVIAPRTITVGFCLGTKRLYDFVHENPAVQFRAIE
WVNNPILIKENPKMTAINSAVEVDLSGQICADSIGSKLYSGVGGQMDFMRGAALSHGGKPIIAL
PSTTSRGESKIVPRLKRGAGVVTTRAHVHWVVTEWGAVNLFGKPVKERMKALISIAHPLHRP
WLEKEVERGFWFDVDENSPKIPESAHVDE

SCOT

MRQSTMRGIVLTRGSRGFHTTSLLR↓SKIVGSAEEAVKDVKDGSKLLVGGFGLCGIPEKLIGA
LRTTGVKDLTVVSNNCGVDDFGLGLLLQTRQIKRMISSYVGENAIFEKQYLSGELELELIPQGT
LAERCRAGGAGIPAFYTPTGVGTFLEEGGFPIKYNTDGSVAIASKPREVREFNGRKFIMEEAIT
GDFSLIKGWKADTRGNIVFRYTARNFNPPVATAGKICIAEVEEIVEAGSLHPDEIHLPGIYVNRLI
KGTGYEKRIEKLTLDKGQPAKGDAPKNEAAITREKIARRAALEFQDGMYCNLGIGIPTLASNYI
RSGIHIELQSENGLLGMGPFPKPGNQDPDLINAGKETVTTLPGSSIFSSDQSFAMIRGAHVNLTI
LGGMQVSSNGDLANWVIPGKMVKGPGGAMDLTCSGSRVVVTMEHKDKSGKPKILNRCNLPL
TAQGCVNRIITEMAVFDVDRSGLTLIEVAEGVTVDDVKKYTEPEFKVSPNLKKIAYA

Ferredoxin

MMKRTTTLVTGGRRFAGLQSPVVPFAWAGRQATSSSASTLRSARLY↓SADADSKKTVHVTFI
DKDGTEIPLEAPVGKSVLELAHDNKIDLEGACEASLACSTCHVILDKEYYDKLPAPVEEEEDML
DLAFGLTETSRLGCQIIISPELEGIRLKLPPATRNMMVDGYKPPHH

Figure 2.3 Peptides identified in tandem mass spectrometry experiments mapped to anaerobic ATP generation enzyme sequences. Peptides identified in mitochondrial fractions are shown in red. mtTP cleavage sites predicted by TargetP are shown with arrows. Note that the cleavage site predicted for ASCT1B is likely incorrect, as it falls within a peptide identified in the mitochondrial fraction.

Thirty unique PFO-specific peptides were identified in the WM, SWM and SPE fractions, with a high ion score (1457), evidence that PFO is present in *A. castellanii* mitochondria even under aerobic conditions. No PFO-specific peptides were identified in the MPE fraction; these findings are consistent with PFO in *A. castellanii* being a soluble matrix protein, in contrast to that of *T. vaginalis*, which is bound to the hydrogenosomal membrane (Williams, Lowe, Leadlay 1987). ASCT1B was similarly well represented in the mitochondrial proteome (ion score: 1900, 17 unique peptides).

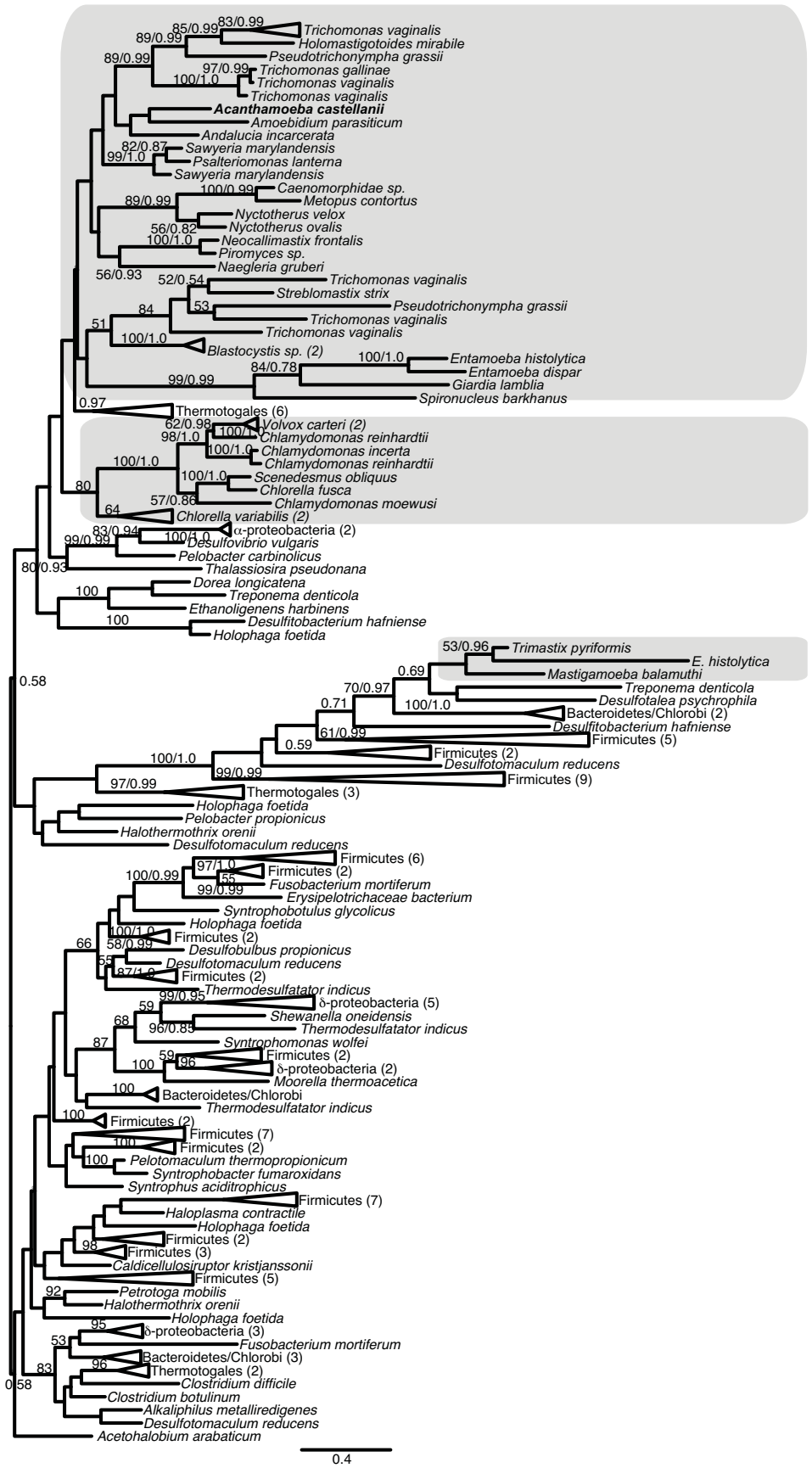
A single peptide from the [FeFe]-hydrogenase maturase HydF was identified, as were three peptides from ferredoxin. No peptides corresponding to [FeFe]-hydrogenase, or to the two other [FeFe]-hydrogenase maturases, were recovered.

In addition, we performed immunogold labeling experiments on *A. castellanii* cells that had been exposed to anaerobic conditions for 6 or 24 hr, using an antibody raised against *A. castellanii* [FeFe]-hydrogenase (Appendix A, Supporting Methods). Antibody staining in these cells was elevated in mitochondria (approx. 2.9-fold higher than in the cytosol, and approx. 1.7-fold higher than in the nucleus), consistent with the presence of a predicted mitochondrial targeting peptide for [FeFe]-hydrogenase (Appendix A, Figures A.S2, A.S3, A.S4), although the high levels of antibody required and the degree of cross-reaction in other cellular locations suggest that optimal conditions for high expression remain to be established for [FeFe]-hydrogenase in *A. castellanii*.

Evolutionary Histories of Anaerobic ATP Generation Enzymes.

Previous analyses of [FeFe]-hydrogenase phylogenies have failed to recover eukaryotes as a monophyletic clade (Horner, Foster, Embley 2000; Horner et al. 2002; Voncken et al. 2002; Hug, Stechmann, Roger 2010). Our results (Fig. 2.4), which include sequences from a larger number of eukaryotic taxa than were previously available, are consistent with these findings, in that we recover at least three distinct eukaryotic clades.

Figure 2.4 Phylogeny of [FeFe]-hydrogenase in eukaryotes and bacteria. **(Following page)** The topology shown is the maximum likelihood (ML) tree generated by RAxML analyses; 279 sites were examined across 175 taxa. Bootstrap support values ≥50% and Bayesian posterior probabilities ≥0.5 are shown. Eukaryotes are shaded gray.


Trichomonas vaginalis
Holomastigotoides mirabile
Pseudotrichonympha grassii
Trichomonas gallinae
Trichomonas vaginalis
Trichomonas vaginalis
Acanthamoeba castellanii
Amoebidium parasiticum
Andalucia incarcerata
Sawyeria marylandensis
Psalteriomonas lanterna
Sawyeria marylandensis
Caenomorphidae sp.
Metopus contortus
Nyctotherus velox
Nyctotherus ovalis
Neocallimastix frontalis
Piromyces sp.
Naegleria gruberi
Trichomonas vaginalis
Streblomastix strix
Pseudotrichonympha grassii
Trichomonas vaginalis
Trichomonas vaginalis
Blastocystis sp. (2)
Entamoeba histolytica
Entamoeba dispar
Giardia lamblia
Spironucleus barkhanus
Thermotogales (6)
Volvox carteri (2)
Chlamydomonas reinhardtii
Chlamydomonas incerta
Chlamydomonas reinhardtii
Scenedesmus obliquus
Chlorella fusca
Chlamydomonas moewusi
Chlorella variabilis (2)
α-proteobacteria (2)
Desulfovibrio vulgaris
Pelobacter carbinolicus
Thalassiosira pseudonana
Dorea longicatena
Treponema denticola
Ethanoligenens harbinens
Desulfitobacterium hafniense
Holophaga foetida
Trimastix pyriformis
E. histolytica
Mastigamoeba balamuthi
Treponema denticola
Desulfotalea psychrophila
Bacteroidetes/Chlorobi (2)
Desulfitobacterium hafniense
Firmicutes (5)
Firmicutes (2)
Desulfotomaculum reducens
Firmicutes (9)
Thermotogales (3)
Holophaga foetida
Pelobacter propionicus
Halothermothrix orenii
Desulfotomaculum reducens
Firmicutes (6)
Firmicutes (2)
Fusobacterium mortiferum
Erysipelotrichaceae bacterium
Syntrophobotulus glycolicus
Holophaga foetida
Firmicutes (2)
Desulfobulbus propionicus
Desulfotomaculum reducens
Firmicutes (2)
Thermodesulfatator indicus
δ-proteobacteria (5)
Shewanella oneidensis
Thermodesulfatator indicus
Syntrophomonas wolfei
Firmicutes (2)
δ-proteobacteria (2)
Moorella thermoacetica
Bacteroidetes/Chlorobi
Thermodesulfatator indicus
Firmicutes (2)
Firmicutes (7)
Firmicutes (2)
Pelotomaculum thermopropionicum
Syntrophobacter fumaroxidans
Syntrophus aciditrophicus
Firmicutes (7)
Haloplasma contractile
Holophaga foetida
Firmicutes (2)
Firmicutes (3)
Caldicellulosiruptor kristjanssonii
Firmicutes (5)
Petrotoga mobilis
Halothermothrix orenii
Holophaga foetida
δ-proteobacteria (3)
Fusobacterium mortiferum
Bacteroidetes/Chlorobi (3)
Thermotogales (2)
Clostridium difficile
Clostridium botulinum
Alkaliphilus metalliredigenes
Desulfotomaculum reducens
Acetohalobium arabaticum
0.4

While our trees suffer from the same poor resolution (i.e., low bootstrap support for many branches) that has been reported in previous analyses, it is possible to conduct approximately unbiased (AU) topology tests to determine whether the data have sufficient information to reject alternative phylogenetic hypotheses using an alpha-level of 0.05 as the significance threshold. The hypotheses tested are shown in Table 2.2 and include tests for the monophyly of eukaryote sequences as a whole, the grouping of *A. castellanii* sequences with homologs from other amoebozoans, and tests for grouping of eukaryotic and/or *A. castellanii* sequences with α-proteobacterial sequences (as expected if they were of mitochondrial origin).

Table 2.2 Approximately unbiased (AU) tests of alternate topologies. [1] α-proteobacteria [2] Euk clade 1: opisthokonts + *Blastocystis* [3] Euk clade 2: *Thecamonas* + *Salpingoeca* + *Monosiga* [4] β-proteobacteria

Hypothesis tested	**AU test P-value**
[FeFe]-hydrogenase	
ML tree	0.952
Eukaryote monophyly	0.094
Acanthamoeba + *Mastigamoeba* monophyly	3e-41
Acanthamoeba + short-branching *Entamoeba* monophyly	4e-06
Acanthamoeba + long-branching *Entamoeba* monophyly	0.004
(*Acanthamoeba*, α-prot.S2), other eukaryotes, (bacteria)	0.015
[FeFe]-hydrogenase (long branches removed)	
ML tree	0.921
Eukaryote monophyly	0.547

Hypothesis tested	AU test P-value
Acanthamoeba + *Entamoeba* monophyly	0.003
(*Acanthamoeba*, α-prot.S2), other eukaryotes, (bacteria)	2e-05
PFO	
ML tree	0.707
Eukaryote monophyly	0.674
Acanthamoeba + *Mastigamoeba* monophyly	0.181
Acanthamoeba + *Entamoeba* monophyly	0.059
Acanthamoeba + *Mastigamoeba* + *Entamoeba* monophyly	0.045
(*Acanthamoeba*, α-prot.S2), other eukaryotes, (bacteria)	2e-37
ASCT	
ML tree	0.803
Eukaryote monophyly	0.002
(*Acanthamoeba* , Euk clade 1[2]), other eukaryotes (bacteria)	0.625
(*Acanthamoeba* , *Malawimonas*, *Capsaspora*, Euk clade 1[2]), other euks, (bacteria)	0.452
(Euk clade 1[2], α-prot.S2), other eukaryotes *Pseudovibrio Commensalibact.*, (bacteria)	0.036
(Euk clade 2[3], α-prot.S2), other eukaryotes, *Pseudovibrio*, *Commensalibact.*, (bacteria)	0.032
(*Acanthamoeba*, α-prot.S2), other eukaryotes, *Pseudovibrio*, *Commensalibacter*, (bacteria)	0.032
HydE	
ML tree	0.848
Eukaryote monophyly	0.866

Hypothesis tested	AU test P-value
Acanthamoeba + *Mastigamoeba* monophyly	0.948
((Eukaryotes, α-prot.S2), β-prot.s), *Spironucleus*, (bacteria)	0.605
HydF	
ML tree	0.757
Eukaryote monophyly	0.758
Acanthamoeba + *Mastigamoeba* monophyly	0.678
HydG	
ML tree	0.985
Eukaryote monophyly	0.944
Acanthamoeba + *Mastigamoeba* monophyly	0.880

These tests show, for example, that monophyly of the *A. castellanii* sequence with other amoebozoan homologs, such as the sequence from *Mastigamoeba balamuthi* or the *Entamoeba histolytica* sequence, can be rejected, whereas eukaryote monophyly itself cannot be rejected. Preliminary RAxML and FastTree analyses recovered one unusually long-branching eukaryotic/bacterial clade, corresponding to the major [FeFe]-hydrogenase Clade B described in Hug et al. (Hug, Stechmann, Roger 2010), which includes the *M. balamuthi*, both *Trimastix pyriformis*, and one of the *E. histolytica* enzymes. Separate analyses were performed excluding the taxa in this clade (Appendix A, Fig. A.S5). Removing this clade did not alter the overall topology enough to recover

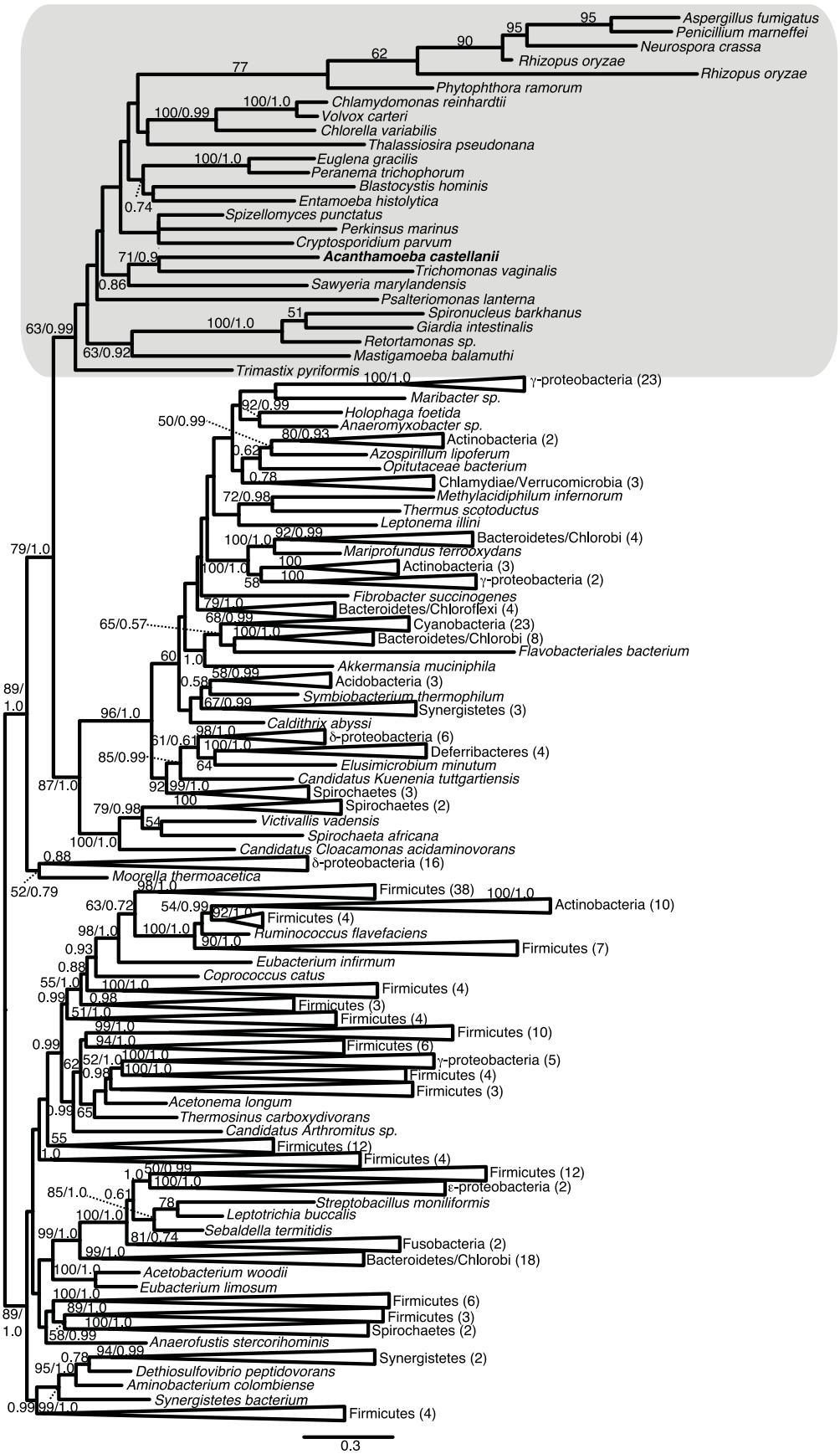
eukaryote monophyly; again, however, in AU tests using this dataset, eukaryote monophyly was not rejected.

The only previous phylogenetic analyses of [FeFe]-hydrogenase maturases (Hug, Stechmann, Roger 2010) were notable in that they recovered eukaryote monophyly for all three enzymes, despite this not having been the case for [FeFe]-hydrogenase itself – even accounting for the lack of known sequences of these enzymes in some eukaryotes. This observation also holds true for our analyses (Appendix A, Fig.s S6 – S8), despite the additional eukaryotic sequence data that have become available in the interim. Although *Spironucleus vortens* groups with α- and β-proteobacteria in the HydE tree (Appendix A, Fig. A.S6), this position has low bootstrap support and, as with the other two maturases, topology tests (Table 2.2) do not reject eukaryote monophyly for HydE. In contrast with the topology tests for [FeFe]-hydrogenase, PFO and ASCT1B, a specific grouping of eukaryotes and α-proteobacteria (as expected if the enzymes were of mitochondrial origin) is not rejected by topology tests (Table 2.2). Within the main eukaryotic clade, low support precludes drawing conclusions about internal relationships, including that of *A. castellanii*; monophyly of *A. castellanii* and *M. balamuthi* is not rejected by topology tests for any of the maturases.

Previous phylogenetic analyses of PFO reached different conclusions as to the recovery of eukaryote monophyly (Horner, Hirt, Embley 1999; Hug, Stechmann, Roger 2010). Our analyses excluded a long-branching *Monocercomonoides* sequence that may have distorted the topology recovered by Hug et al. (Hug,

Stechmann, Roger 2010); consequently, we recover eukaryotic monophyly (Fig. 2.5), a finding consistent with that of Horner and colleagues (Horner, Hirt, Embley 1999).

Figure 2.5 Phylogeny of PFO in eukaryotes and bacteria. **(Following page)** The topology shown is the ML tree estimated by RAxML; 954 sites were examined across 335 taxa. Bootstrap support 50% and posterior probabilities ≥0.5 are shown. Eukaryotes are shaded gray.


Aspergillus fumigatus
Penicillium marneffei
Neurospora crassa
Rhizopus oryzae
Rhizopus oryzae
Phytophthora ramorum
Chlamydomonas reinhardtii
Volvox carteri
Chlorella variabilis
Thalassiosira pseudonana
Euglena gracilis
Peranema trichophorum
Blastocystis hominis
Entamoeba histolytica
Spizellomyces punctatus
Perkinsus marinus
Cryptosporidium parvum
Acanthamoeba castellanii
Trichomonas vaginalis
Sawyeria marylandensis
Psalteriomonas lanterna
Spironucleus barkhanus
Giardia intestinalis
Retortamonas sp.
Mastigamoeba balamuthi
Trimastix pyriformis
γ-proteobacteria (23)
Maribacter sp.
Holophaga foetida
Anaeromyxobacter sp.
Actinobacteria (2)
Azospirillum lipoferum
Opitutaceae bacterium
Chlamydiae/Verrucomicrobia (3)
Methylacidiphilum infernorum
Thermus scotoductus
Leptonema illini
Bacteroidetes/Chlorobi (4)
Mariprofundus ferrooxydans
Actinobacteria (3)
γ-proteobacteria (2)
Fibrobacter succinogenes
Bacteroidetes/Chloroflexi (4)
Cyanobacteria (23)
Bacteroidetes/Chlorobi (8)
Flavobacteriales bacterium
Akkermansia muciniphila
Acidobacteria (3)
Symbiobacterium thermophilum
Synergistetes (3)
Caldithrix abyssi
δ-proteobacteria (6)
Deferribacteres (4)
Elusimicrobium minutum
Candidatus Kuenenia tuttgartiensis
Spirochaetes (3)
Spirochaetes (2)
Victivallis vadensis
Spirochaeta africana
Candidatus Cloacamonas acidaminovorans
δ-proteobacteria (16)
Moorella thermoacetica
Firmicutes (38)
Actinobacteria (10)
Firmicutes (4)
Ruminococcus flavefaciens
Firmicutes (7)
Eubacterium infirmum
Coprococcus catus
Firmicutes (4)
Firmicutes (3)
Firmicutes (4)
Firmicutes (10)
Firmicutes (6)
γ-proteobacteria (5)
Firmicutes (4)
Firmicutes (3)
Acetonema longum
Thermosinus carboxydivorans
Candidatus Arthromitus sp.
Firmicutes (12)
Firmicutes (4)
Firmicutes (12)
ε-proteobacteria (2)
Streptobacillus moniliformis
Leptotrichia buccalis
Sebaldella termitidis
Fusobacteria (2)
Bacteroidetes/Chlorobi (18)
Acetobacterium woodii
Eubacterium limosum
Firmicutes (6)
Firmicutes (3)
Spirochaetes (2)
Anaerofustis stercorihominis
Synergistetes (2)
Dethiosulfovibrio peptidovorans
Aminobacterium colombiense
Synergistetes bacterium
Firmicutes (4)
0.3

As in the case of [FeFe]-hydrogenase, a hypothetical *Acanthamoeba* + *Entamoeba* clade was rejected by topology tests; but an *Acanthamoeba* + *Mastigamoeba* clade was not. However, large groupings within the eukaryote clade have poor support, preventing the inference of clear internal relationships.

A previous neighbor-net analysis of ASCT1B and ASCT1C sequences (van Grinsven et al. 2008) recovered a monophyletic cluster of animal sequences, representing the only eukaryotes known to possess ASCT1B-like sequences at that time. Our analyses include additional animal sequences, as well as sequences from a number of other lineages. Monophyly of all of these eukaryote sequences is rejected in topology tests (Table 2.2), as is the grouping of α-proteobacteria with any of the major eukaryote groups. However, both the grouping of *A. castellanii* with the aerobic flagellate *Malawimonas jakobiformis* as well as the branches separating this clade from away from opisthokonts and *Blastocystis* have low bootstrap support (Appendix A, Fig. A.S9), and an alternate position of this organism, grouping with opisthokonts and *Blastocystis*, is not rejected.

Discussion

Acanthamoeba castellanii is known to inhabit aerobic environments and to produce energy via oxidative phosphorylation; it possesses a mitochondrion with a functional TCA cycle and electron transport chain similar to those found in other aerobic eukaryotes (Edwards, Lloyd 1978; Gawryluk et al. 2012). Nevertheless, it inhabits a wide range of soil, aquatic, and man-made

environments, and it is likely that it encounters low-oxygen conditions with some frequency. An organism with such a lifestyle would likely derive significant survival benefit from being able to function under a wide range of conditions, including periods of anoxia. Here, we present the first case of a complete hydrogenosome-like ATP generation pathway with predicted mitochondrial targeting in a habitually aerobic eukaryote, and show that several of its key enzymes are present in the mitochondria. Our work provides the first evidence supported by localization data for a mitochondrion possessing the metabolic components of a 'hybrid' organelle, which may enable it to adopt functions in oxidative phosphorylation or anaerobic metabolism according to the conditions that it encounters. The existence of such organelles in an extant organism immediately suggests a possible sequence of events in the first steps leading to the emergence of hydrogenosomes. Upon acquiring an anaerobic energy-generating pathway, a previously obligate aerobe would be able to thrive in a more diverse range of habitats, as well as surviving temporal fluctuations in oxygen levels. Subsequently, descendents of such a cell inhabiting exclusively low-oxygen environments, with a reduced need to perform oxidative phosphorylation, might lose components of the electron transport chain, as well as other mitochondrial functions (Fig. 2.6).

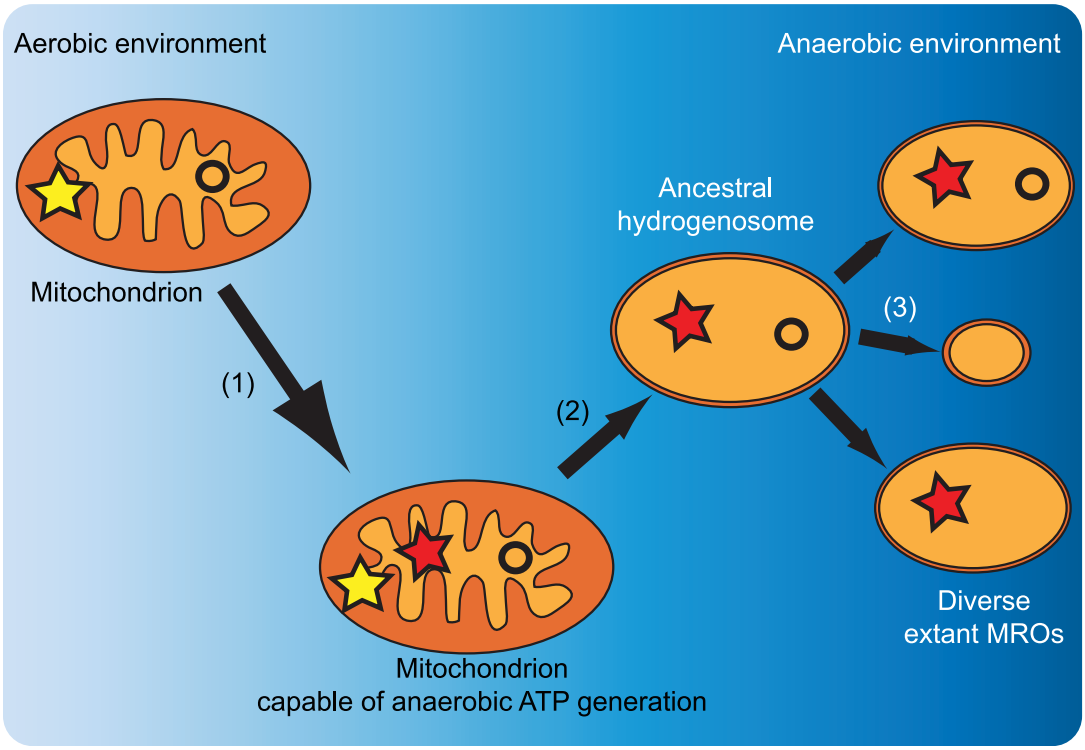


Figure 2.6 The origins of mitochondrion-related organelles. A hypothetical scenario for the acquisition of anaerobic ATP generation enzymes and the subsequent emergence of extant mitochondrion-related organelles (MROs). (1) Acquisition of anaerobic energy generation enzymes. (2) Loss of the capacity for oxidative phosphorylation. (3) Loss of diverse mitochondrial functions. Yellow stars represent the electron transport chain, while red stars represent the hydrogenosomal anaerobic ATP generation pathway. Circles represent the mitochondrial genomes.

We have confirmed the presence in *A. castellanii* of a complete anaerobic ATP generation pathway similar to that found in hydrogenosomes. All of these enzymes are predicted to have mitochondrial localization, and we confirm this localization for the characteristic hydrogenosomal energy enzymes PFO, ASCT1B, HydF and [FeFe]-hydrogenase. It should be noted that we cannot exclude the possibility of a dual mitochondrial and cytosolic localization of these enzymes; nevertheless, the presence in mitochondrial fractions of PFO, ASCT1B and HydF, and the elevated localization of [FeFe]-hydrogenase within the mitochondria, suggests that these organelles can function as a site of anaerobic

respiration in *A. castellanii*. Further research should elucidate the conditions under which this pathway is induced in *A. castellanii*, and the interplay between anaerobic respiration and encystation as different – or perhaps complementary – survival modes in this organism. In addition, the detection of PFO, ASCT1B and HydF peptides in mitochondrial fractions purified from aerobically grown cells raises the possibility that some or all of these enzymes may be upregulated in response to environmental factors other than oxygen concentration.

The most intriguing question raised by this study concerns the origin of these genes. The physical proximity of [FeFe]-hydrogenase and its three maturases in the genome might suggest the lateral transfer of a single bacterial operon as an acquisition mechanism. However, the presence of so many interspersing genes and the absence of a clear bacterial donor candidate argue against a single, recent transfer from bacteria. Topology tests reject the grouping of *A. castellanii* and α-proteobacteria (with remaining eukaryotes unconstrained) for [FeFe]-hydrogenase, PFO and ASCT1B. This result is consistent with previous studies, which recover at least two independent origins for [FeFe]-hydrogenase (Horner et al. 2002; Voncken et al. 2002; Meyer 2007; Vignais, Billoud 2007; Hug, Stechmann, Roger 2010), and specifically reject α-proteobacterial ancestry for [FeFe]-hydrogenase in topology tests (Hug, Stechmann, Roger 2010).

Eukaryote monophyly is recovered for two of the maturases and for PFO, and is not rejected in topology tests for [FeFe]-hydrogenase and the remaining maturase. This finding might seem consistent with the hydrogen hypothesis, which holds that the original endosymbiont that gave rise to mitochondrion-

related organelles within eukaryotes was a facultative anaerobe possessing an [FeFe]-hydrogenase, retained by a methanogenic, hydrogen-dependent host for the hydrogen it generated as a waste product (Martin, Müller 1998). Nevertheless, the lack of a clear affinity to α-proteobacterial homologues for these enzymes, and their distribution within eukaryotes – in particular their absence from so many taxa closely related to anaerobes – weakens such a conclusion. All or most of these genes might still have been present in the protomitochondrial endosymbiont as a result of a lateral gene transfer (LGT) event from a different prokaryote, or might have been acquired by the ancestral eukaryote by other means (Embley 2006); these scenarios would be more consistent with the very small number of contemporary α-proteobacteria reported to possess homologs of these genes. However, if this is the case, then (1) the patchy distribution of anaerobic ATP generation enzymes among eukaryotes in general, (2) the rejection of monophyly of *A. castellanii* with two other amoebozoans for [FeFe]-hydrogenase, and (3) the absence of genes for these enzymes in the genomes of other members of Amoebozoa, remain to be explained.

Intriguingly, the existence of monophyletic eukaryotic clades with unusual internal topology has been reported for other enzymes with patchy distributions among eukaryotes. The authors of these studies proposed multiple lateral transfers between eukaryotes as a hypothesis to explain patchy distributions and unexpected phylogenetic relationships for other enzymes found in both aerobic and anaerobic protists (Andersson et al. 2006; Andersson et al. 2007; Andersson 2009b; Andersson 2009a; Stairs, Roger, Hampl 2011; Takishita et al. 2012;

Stairs et al. 2014). This mode of acquisition would provide an attractive alternative scenario for the acquisition of anaerobic metabolism genes; the transfer of genes between eukaryotes would remove the need for the acquisition of eukaryotic regulatory sequences for the enzymes in the recipient, and would account for the distribution of anaerobic ATP generation enzymes in extant eukaryotes. Furthermore, it would provide an elegant explanation for the common pool of anaerobic ATP generation enzymes found in anaerobic eukaryotes (Müller et al. 2012). Frustratingly, the lack of phylogenetic resolution in much of the tree, including for internal eukaryote relationships and for the position of *A. castellanii* itself, does not allow us to draw strong inferences that would help us to distinguish among competing hypotheses as to the origin of anaerobic enzymes in *A. castellanii.*

Our work cements the possibility, initially raised by Hug et al. (Hug, Stechmann, Roger 2010), that anaerobic ATP generation enzymes might be more widespread among eukaryotes than previously thought, not being limited to anaerobic or microaerobic lineages. It also highlights the importance of exploratory sequencing efforts focusing on a wide range of organisms. So far, anaerobic energy enzymes have been described in two non-photosynthetic organisms, *A. castellanii* and *N. gruberi.* As an opportunistic human pathogen that also harbours pathogenic bacteria, and a close relative of an opportunistic human pathogen, respectively, both of these organisms have links to human health, making them attractive targets for sequencing efforts. However, the more important link between them is likely their lifestyle, which exposes them to a wide range of habitats that vary in oxygen concentration, a lifestyle likely made

possible by the anaerobic energy enzymes they have acquired. Free-living soil protists are relatively poorly studied, and investigations into the metabolic complements of a wider range of such organisms will reveal whether these hybrid organelles are found more commonly in nature than previously suspected.

Acknowledgements

We thank Dr. Brendan Loftus for early sharing of his *Acanthamoeba castellanii* EST data; Prof. John Walker for providing OverExpress™ C41(DE) cells; Dr. Daniel Gaston for the use of his phylogenetic alignment trimming script; Mary Ann Trevors and Dr. Gary Faulkner for help with immunoelectron microscopy experiments; Dr. Edward Susko for advice on statistical analyses; and Drs. David Spencer, Murray Schnare and Laura Eme for helpful discussion.

Chapter 3: An Ancestral Bacterial Division System Is Widespread in Eukaryotic Mitochondria

This chapter was published as: Michelle M. Leger, Markéta Petrů, Vojtěch Žárský, Laura Eme, Čestmír Vlček, Tommy Harding, B. Franz Lang, Marek Eliáš, Pavel Doležal, and Andrew J. Roger. 2015. Proc Natl Acad Sci U S A. 2015 Mar 23. pii: 201421392. [Epub ahead of print]. Its contents have been amended here following the suggestions of the examination committee.

Abstract

Bacterial division initiates at the site of a contractile Z-ring composed of polymerized FtsZ. The location of the Z-ring in the cell is controlled by a system of three mutually antagonistic proteins, MinC, MinD and MinE. Plastid division is also known to be dependent on homologues of these proteins, derived from the ancestral cyanobacterial endosymbiont that gave rise to plastids. In contrast, the mitochondria of model systems such as *Saccharomyces cerevisiae*, mammals and *Arabidopsis thaliana* appear to have replaced the ancestral α-proteobacterial Min-based division machinery with host-derived dynamin-related proteins that form outer contractile rings. Here, we show that the mitochondrial division system of these organisms is the exception, rather than the rule for eukaryotes. We describe endosymbiont-derived, bacterial-like division systems comprising FtsZ and Min proteins in diverse less-studied

eukaryote protistan lineages, including jakobid and heterolobosean excavates, a malawimonad, stramenopiles, amoebozoans, a breviate, and an apusomonad. For two of these taxa, the amoebozoan *Dictyostelium purpureum* and the jakobid *Andalucia incarcerata*, we confirm a mitochondrial localization of these proteins by their heterologous expression in *Saccharomyces cerevisiae*. The discovery of a proteobacterial-like division system in mitochondria of diverse eukaryotic lineages suggests it was the ancestral feature of all eukaryotic mitochondria and has been supplanted by a host-derived system multiple times in distinct eukaryote lineages.

Introduction

During bacterial division, septum formation is mediated by the Z-ring, a contractile ring structure made up of the polymerized tubulin homologue FtsZ (reviewed in (de Boer 2010; Meier, Goley 2014)). The site at which FtsZ polymerizes is determined by the Min system (reviewed in (Lutkenhaus 2007; de Boer 2010; Lutkenhaus, Pichoff, Du 2012; Natale, Pazos, Vicente 2013)), comprising three proteins, MinC, MinD and MinE. Adenosine triphosphate (ATP)-bound, dimerized MinD binds the inner cell membrane at the poles of the cell, forming aggregates. MinD aggregates bind and activate dimerized MinC (Ghosal et al. 2014), which then inhibits local FtsZ polymerization (Fig. 3.1A). Concomitantly, dimerized MinE forms a spiral ring whose constant polymerization and depolymerization causes it to oscillate across the cell (Hale, Meinhardt, de Boer 2001; Loose et al. 2008). Where MinE comes into contact

with MinD, it causes the release of ATP, and the subsequent liberation of MinD from the membrane (Hu, Lutkenhaus 2001; Park et al. 2012). In this way, MinD and MinC cannot inhibit FtsZ polymerization near the midpoint of the cell. The polymerizing Z-ring is stabilized and tethered to the membrane by FtsA, ZipA, and the nonessential ZapA and (in some organisms) ZapB (Hale, de Boer 1997; Wang et al. 1997; Hale, de Boer 1999; Pichoff, Lutkenhaus 2002; Galli, Gerdes 2010). Maturation of the Z-ring into a complete septal ring continues with the subsequent recruitment by FtsA of further components of the divisome (i.e., FtsB, FtsE, FtsI/PBP3, FtsK, FtsL, FtsN, FtsQ, FtsW and FtsX), which proceed to stabilize FtsZ and contribute to peptidoglycan synthesis (reviewed in (Lutkenhaus, Pichoff, Du 2012; Egan, Vollmer 2013; Natale, Pazos, Vicente 2013; den Blaauwen, Andreu, Monasterio 2014)) (Fig. 3.1B) prior to Z-ring constriction and the completion of septum formation (Fig. 3.1C).

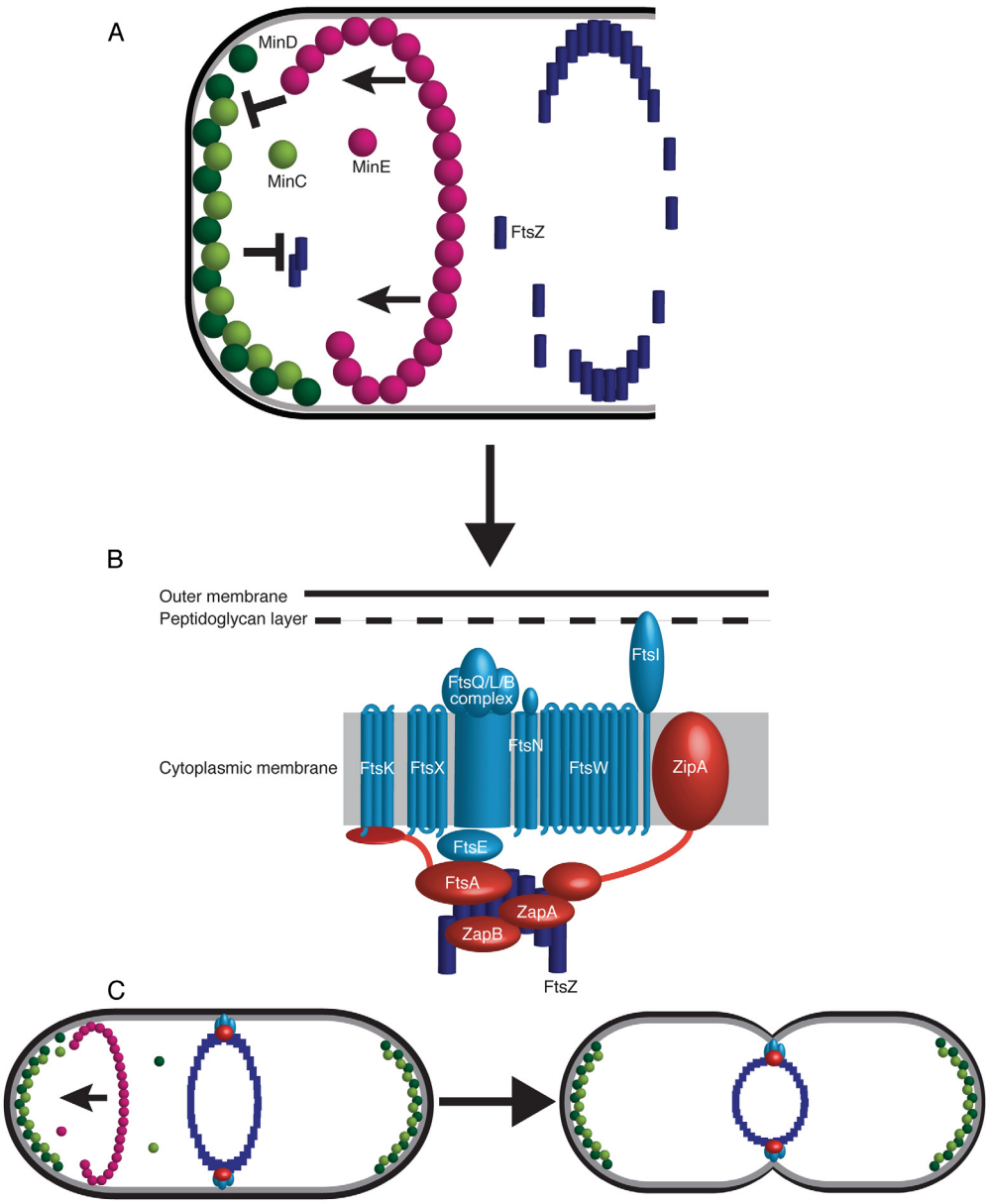


Figure 3.1 Partial schematic overview of division machinery in *Escherichia coli*. A, roles of Min proteins during FtsZ polymerization. B, subsequent recruitment of early and late stage proteins involved in Z-ring stabilization and attachment to the cell membrane. C, overview of septation initiation at the cell level. Dark blue rectangles, FtsZ; light green circles, MinC; dark green circles, MinD; magenta circles, MinE; red shapes, early stage cell division proteins; light blue shapes, late stage cell division proteins. For the sake of clarity, not all proteins known to localize to the midcell during division are shown. In particular, this schematic focuses on proteins known to localize to the cytoplasmic membrane, and excludes most proteins localizing primarily to the peptidoglycan layer and the outer membrane. Based on reviews (de Boer 2010; Egan, Vollmer 2013; den Blaauwen, Andreu, Monasterio 2014).

Plastids are known to possess FtsZ (Osteryoung, Vierling 1995; Fraunholz, Moerschel, Maier 1998; Sato et al. 2005), MinD (Wakasugi et al. 1997; Colletti et al. 2000), MinE (Wakasugi et al. 1997; Itoh et al. 2001) and in some cases MinC (Miyagishima et al. 2014) homologues of cyanobacterial endosymbiotic origin; in some cases, the latter are encoded on the plastid genome (Wakasugi et al. 1997; Douglas, Penny 1999). In contrast, only two examples of putative mitochondrial Min proteins have been reported, in the stramenopiles *Nannochloropsis oceanica* and *Ectocarpus siliculosus* (Vieler et al. 2012). Indeed, although eukaryotic mitochondria are derived from an α-proteobacterial endosymbiont, the ancestral bacterial division machinery has been partly or wholly replaced by eukaryote-specific proteins in model system eukaryotes where mitochondrial division has been studied. Whereas Amoebozoa (Gilson et al. 2003), stramenopiles (Beech et al. 2000; Kiefel, Gilson, Beech 2004) and the red alga *Cyanidioschyzon merolae* (Takahara et al. 2000; Takahara et al. 2001) have retained experimentally-confirmed mitochondrial FtsZ, animals and fungi (opisthokonts) and plants examined to date lack this protein. In the latter taxa, an outer contractile ring is instead formed by Dnm1p/Drp1, a eukaryote-specific dynamin GTPase (Bleazard et al. 1999; Labrousse et al. 1999; Smirnova et al. 2001). This protein is implicated in mitochondrial division in organisms across the eukaryotic tree, including *Arabidopsis thaliana* (Arimura, Tsutsumi 2002; Arimura et al. 2004; Mano et al. 2004), the parabasalid *Trichomonas vaginalis* (Wexler-Cohen et al. 2014), *Dictyostelium discoideum* (Wienke et al. 1999), and *Cyanidioschyzon merolae* (Nishida et al. 2003), suggesting that the outer contractile ring is a widespread eukaryotic feature. In *T. vaginalis* and *A.*

thaliana, the nature of the inner contractile ring is not yet understood, although the presence of two Dnm1/Drp1 homologues in *A. thaliana* (Arimura et al. 2004) raises the possibility that they form an outer and an inner contractile ring, respectively. Recent work (Purkanti, Thattai 2015) reconstructing the evolution of eukaryotic dynamins suggests that the ancestral mitochondrial dynamin was a bifunctional protein that also mediated vesicle scission. This protein underwent duplication events, followed by subfunctionalization, independently in at least three lineages (opisthokonts, land plants and alveolates). Nevertheless, the ancestral bifunctional form appears to have been retained in amoebozoa such as *Dictyostelium discoideum*, the red alga *Cyanidioschyzon merolae*, and stramenopiles (and possibly additional eukaryotes that have currently less well-characterized dynamins). The distribution of ancestral-like bifunctional mitochondrial/vesicle fission dynamins thus appears to mirror that of mitochondrial FtsZ (Purkanti, Thattai 2015).

Here, we hypothesize that the complete loss of the α-proteobacterial division system is the exception, rather than the rule for eukaryotes. We show that mitochondria-targeted homologues of bacterial Min proteins are patchily but widely distributed among diverse eukaryote lineages; and we further demonstrate that Min proteins from two of these lineages, the amoebozoan *Dictyostelium purpureum* and the jakobid excavate *Andalucia incarcerata*, localize to mitochondria when expressed in yeast.

Materials and Methods

Database Searches

Publicly available databases and sequencing projects were searched using the Basic Local Alignment Search Tools (BLAST) tools blastp and tblastn (Altschul et al. 1997). A large number of databases containing eukaryotic sequences were screened with these tools using query sequences from *Dictyostelium purpureum* (XP_003286111, XP_003292258, XP_003293637, XP_642499), *Ectocarpus siliculosus* (CBJ32744, CBJ31561, CBJ28079, CBJ48312) *Andalucia incarcerata*, *Pseudomonas fluorescens* (AEV64338, AEV64339, AEV64340, AEV64767) and *Anabaena* sp. 90 (YP_006998153, AFW94434, YP_006996248, YP_006996249). The databases searched included the Nucleotide collection (nr/nt), NCBI Genomes, Whole-Genome Shotgun contigs, Expressed Sequence Tags, High-throughput Genomic Sequences and Transcriptome Shotgun Assembly divisions of GenBank (Benson et al. 2014)(last accessed February 9, 2015); the Broad Institute project databases (Wilding et al. 2002) (last accessed April 23, 2014); the Joint Genome Institute (JGI) genome databases (Grigoriev et al. 2012; Nordberg et al. 2014) (last accessed February 9, 2015); dictyBase, 2013 release (Kreppel et al. 2004; Basu et al. 2013; Fey et al. 2013); the EnsemblProtists database (Flicek et al. 2014)(last accessed February 9, 2015); the Eukaryotic Pathogen Database Resources (EuPathDB) (Aurrecoechea et al. 2007)(last accessed February 9, 2015); and the Marine Microbial Eukaryote Transcriptome Sequencing Project (Keeling et al. 2014) (last accessed June 3, 2014), via the Community cyberinfrastructure for Advanced Microbial Ecology

Research and Analysis (CAMERA) portal (Sun et al. 2011) (for a full list of sequences identified, see Appendix B Table B.S1, Accession Numbers). In addition, we searched our own unpublished genome or transcriptome assemblies from several protist taxa of key evolutionary interest: two jakobids (*Andalucia incarcerata* and *Andalucia godoyi*), the heterolobosean *Pharygmononas kirbyi*, and *Malawimonas californiana*. Potential homologues identified were screened manually in order to exclude contaminants from bacterial or other eukaryotic sources, by searching for introns and excluding sequences with a notably high degree of similarity to bacterial or distantly-related eukaryotic homologues. Subcellular localization and targeting peptides were predicted using TargetP, using 'plant' parameters for plastid-bearing taxa, and 'non-plant' parameters for taxa lacking plastids (Emanuelsson et al. 2000; Emanuelsson et al. 2007).

Sequence Generation

Pharyngomonas kirbyi strain AS12B (Park et al. 2011) was cultivated at 37 °C in 10% salt medium (NaCl 1.6M, KCl 34.0 mM, $MgCl_2$ 44.2 mM, $CaCl_2$ 4.0 mM, $MgSO_4$ 4.5 mM) supplemented with *Citrobacter* sp. as a food source prior to RNA isolation. RNA was extracted using TRIzol (Life Technologies) following the manufacturer's instructions and stored at -80 °C. The RNA sample was treated with Turbo DNAse (Life Technologies) prior to conversion to cDNA using the GeneRacer kit with SuperScript III reverse transcriptase (Life Technologies) and stored at -20 °C. Primers were designed to amplify genes of interest using available sequences. Primer sequences were as follows: MinCF: 5'-ATGTCACGTCGATGGTTAGT-3'; MinCR: 5'-TAATACAAAAAAAAAACA-3'; MinDF: 5'-ATGTATCGATCAACGAGTTC-3'; MinDR: 5'-

TTAGTTCCTGCTAAATAATC-3'. Polymerase Chain Reactions (PCR) were done using the Phusion high-fidelity DNA polymerase (New England BioLabs) where the initial denaturation at 98 °C for 30 s was followed by 30 cycles of DNA denaturation at 98 °C for 10 s, primer annealing at 40 °C for 30 s and strand elongation at 72 °C for 60 s, with a final extension at 72 °C for 10 s. PCR products were purified by gel extraction using the Nucleospin Extract II kit (Macherey-Nagel) and directly sequenced using the PCR primers.

Phylogenetic Analyses

For each protein, alignments were generated from datasets including all known eukaryotic homologues, and bacterial homologues harvested from NCBI using MUSCLE v.3.8.31 (Edgar 2004a) or MAFFT-L-INSI v7.149b (Katoh et al. 2002; Katoh et al. 2005; Katoh, Standley 2013), and trimmed using BMGE 1.1 (Criscuolo, Gribaldo 2010) (-m BLOSUM30, all other parameters default). Preliminary phylogenies were generated using FastTree, and datasets were manually refined. Twenty independent Maximum Likelihood (ML) tree estimates and 200 bootstrap replicates were generated using RAxML v.8.0.23 (Stamatakis 2014) under the PROTGAMMALG4X (Le, Dang, Gascuel 2012) model of amino acid substitution. Bayesian inference posterior probabilities were calculated using PhyloBayes v.3.3f (Lartillot, Lepage, Blanquart 2009) under the catfix C20 model of evolution. We tested whether specific phylogenetic hypotheses were rejected by the data using the Approximately Unbiased test implemented in CONSEL v.1.20 (Shimodaira, Hasegawa 2001) (Appendix B Table B.S2). Maximum likelihood trees given specific constraints (i.e., corresponding to specific hypotheses) were generated using RAxML. In addition,

the 200 trees from bootstrap replicates were included in the hypothesis testing analyses performed with CONSEL.

Yeast Culture, Transformation and Microscopy

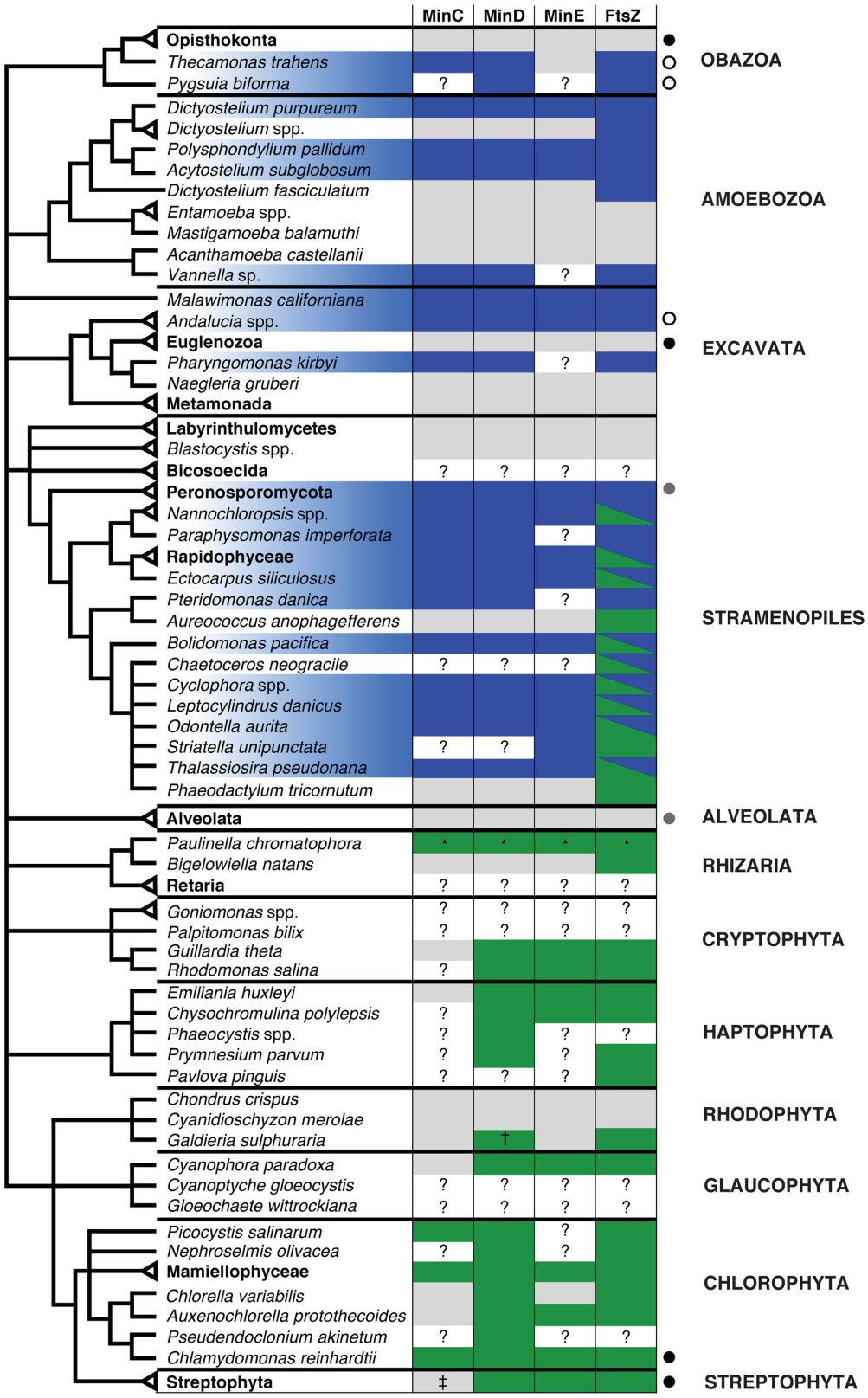
Saccharomyces cerevisiae strain YPH499 was grown at 30°C on YPD medium or selective medium without uracil after lithium-acetate transformation. For ectopic expression of AiMinC, D and E, the complete AiMinC, D and E open reading frames (ORFs) were amplified by PCR from *A. incarcerata* cDNA. For ectopic expression of DpMinC, D and E, the complete DpMinC, D and E open reading frames were amplified by PCR from synthesized DNA fragments, containing *E. coli*-codon-optimized sequences. The resulting PCR products were cloned separately into pUG35 using XbaI/ClaI restriction sites (AiMinD, DpMinC and E) or BamHI/HindIII restriction sites (AiMinC and E, DpMinD) allowing the expression of green fluorescent protein (GFP) on the C-terminus of each protein. For fluorescence microscopy, cells were incubated with MitoTracker Red CMXRos (1:10 000) for 10 min, washed once in PBS and mounted in 2% low-melting agarose. Cells were viewed using an Olympus IX81 microscope and a Hamamatsu Orca-AG digital camera using the cell^R imaging program at 100x magnification.

Results and Discussion

We identified sequences encoding at least one Min protein from a number of eukaryotic taxa, (Fig. 3.2, Appendix B Table B.S1), including ancestrally plastid-lacking lineages such as the apusomonad *Thecamonas trahens*, the breviate

Pygsuia biforma, the jakobid excavates *Andalucia godoyi* and *A. incarcerata*, the malawimonad *Malawimonas californiana*, and several amoebozoan lineages such as *Dictyostelium purpureum*.

Figure 3.2 Presence and absence of bacterial Min proteins and FtsZ in selected eukaryotic taxa. **(Following page)** Blue, predicted mitochondrial proteins; green, predicted plastid proteins; grey, no protein found encoded in complete genome data; ?, no protein found encoded in transcriptome or incomplete genome data; *, chromatophore protein; †, predicted pseudogene. ‡, with the exception of *Physcomitrella patens*. Boxes shaded half blue, half green represent multiple paralogues, predicted to be mitochondrial and plastid respectively. In cases where only a transcriptome or incomplete genome is available, it should be noted that the presence of a plastid protein does not exclude the possibility of one or more mitochondrial paralogues also being present; and vice versa. Eukaryotic taxa possessing predicted mitochondrial Min proteins are shaded in blue. Mitochondrial or plastid predictions are based on phylogenetic affinity with previously localized proteins, predicted subcellular localization, and localization in yeast (*Andalucia incarcerata, Dictyostelium discoideum*). Black circles indicate taxa in which reticulate mitochondria have previously been described; grey circles indicate groups for which reticulate mitochondria have been described in at least one member; black-bordered white circles indicate taxa in which only single or unbranched mitochondria have been described. The schematic phylogeny reflects the current understanding of relationships based on multiple phylogenomic analyses. For a more complete table, see Appendix B Table B.S1.


MinC
MinD
MinE
FtsZ
Opisthokonta
Thecamonas trahens
Pygsuia biforma
OBAZOA
Dictyostelium purpureum
Dictyostelium spp.
Polysphondylium pallidum
Acytostelium subglobosum
Dictyostelium fasciculatum
Entamoeba spp.
Mastigamoeba balamuthi
Acanthamoeba castellanii
Vannella sp.
AMOEBOZOA
Malawimonas californiana
Andalucia spp.
Euglenozoa
Pharyngomonas kirbyi
Naegleria gruberi
Metamonada
EXCAVATA
Labyrinthulomycetes
Blastocystis spp.
Bicosoecida
Peronosporomycota
Nannochloropsis spp.
Paraphysomonas imperforata
Rapidophyceae
Ectocarpus siliculosus
Pteridomonas danica
Aureococcus anophagefferens
Bolidomonas pacifica
Chaetoceros neogracile
Cyclophora spp.
Leptocylindrus danicus
Odontella aurita
Striatella unipunctata
Thalassiosira pseudonana
Phaeodactylum tricornutum
STRAMENOPILES
Alveolata
ALVEOLATA
Paulinella chromatophora
Bigelowiella natans
Retaria
RHIZARIA
Goniomonas spp.
Palpitomonas bilix
Guillardia theta
Rhodomonas salina
CRYPTOPHYTA
Emiliania huxleyi
Chysochromulina polylepsis
Phaeocystis spp.
Prymnesium parvum
Pavlova pinguis
HAPTOPHYTA
Chondrus crispus
Cyanidioschyzon merolae
Galdieria sulphuraria
RHODOPHYTA
Cyanophora paradoxa
Cyanoptyche gloeocystis
Gloeochaete wittrockiana
GLAUCOPHYTA
Picocystis salinarum
Nephroselmis olivacea
Mamiellophyceae
Chlorella variabilis
Auxenochlorella protothecoides
Pseudendoclonium akinetum
Chlamydomonas reinhardtii
CHLOROPHYTA
Streptophyta
STREPTOPHYTA

Three previously reported FtsZ sequences identified in haptophytes (*Gephyrocapsa oceanica* and *Pleurochrysis carterae*) and a glaucophyte (*Cyanophora paradoxa*) (Kiefel, Gilson, Beech 2004) were excluded as probable α-proteobacterial contaminants, based on their position in preliminary phylogenies, their high degree of similarity to α-proteobacterial sequences, and, in the case of *C. paradoxa*, our inability to recover the reported mitochondrial FtsZ sequence from the genome sequence (Price et al. 2012). All complete genomes encoding at least one Min protein also encoded at least one FtsZ homolog; however, the reverse was not true. Min proteins were retained not only in lineages with typical aerobic mitochondria, but also in lineages possessing mitochondrion-related organelles (MROs) such as *A. incarcerata* (Simpson, Patterson 2001) and *Pygsuia biforma (Stairs et al. 2014).*

Most of these Min and FtsZ homologues possess predicted mitochondrial targeting peptides (Appendix B Table B.S1). In order to confirm these predictions, we expressed GFP-tagged homologues of Min proteins in *Saccharomyces cerevisiae*, in conjunction with the mitochondrial stain MitoTracker Red CMXRos (Fig.s 3.3A, 3.3B). We chose Min proteins from two representative taxa lacking plastids: the amoebozoan *Dictyostelium purpureum* (Fig. 3.3A) and the jakobid excavate *Andalucia incarcerata* (Fig. 3B). In both cases, the GFP signal co-localized with the Mitotracker signal, supporting the predicted targeting of *A. incarcerata* and *D. discoideum* Min proteins to the inside of the mitochondria.

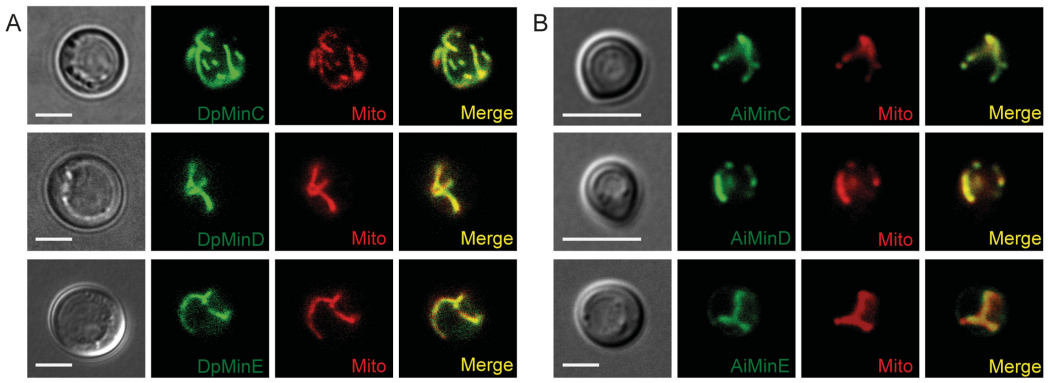


Figure 3.3 Min proteins from *D. purpureum* (A) and *A. incarcerata* (B) expressed in *S. cerevisiae*.Differential Interference Contrast (DIC) images of *S. cerevisiae* cells expressing Min fusion proteins (*left*); in green, MinC, MinD or MinE expressed with the C-terminal GFP tag in *S. cerevisiae*; in red, mitochondria labeled with MitoTracker Red CMXRos (Mito); merged images (Merge) show mitochondrial localization of all Min proteins. Scale bars indicate 5 μm.

Single-protein phylogenies of MinC, D, E and FtsZ recover all predicted mitochondrial homologues as well-resolved clades, distinct from known and predicted plastid sequences (Fig. 3.4 and 3.5; Appendix B Fig. B.S1 and BS2). For MinD (Fig. 3.4) and FtsZ (Fig. 3.5 and Appendix B Fig. B.S3), the hypothesis that the plastid and mitochondrial homologues group in a monophyletic clade was rejected by AU tests (Appendix B Table B.S2); however, this hypothesis could not be rejected for the more divergent MinC (Appendix B Fig. B.S1) and MinE (Appendix B Fig. B.S2). In all three Min phylogenies, mitochondrial homologues emerged within proteobacterial sequences; although, because the resolution within that clade was too poor to identify the closest homologues (Fig. 4, Appendix B Fig. B.S1 and Appendix B Fig. B.S2), we cannot exclude the possibility that these proteins originate from a group other than the α-proteobacteria.

Figure 3.4 Unrooted Maximum Likelihood (ML) tree of MinD sequences. **(Following page)** Phylogenetic analyses were performed on 328 sequences and 226 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, cyanobacteria green, proteobacteria orange, and α-proteobacteria magenta.

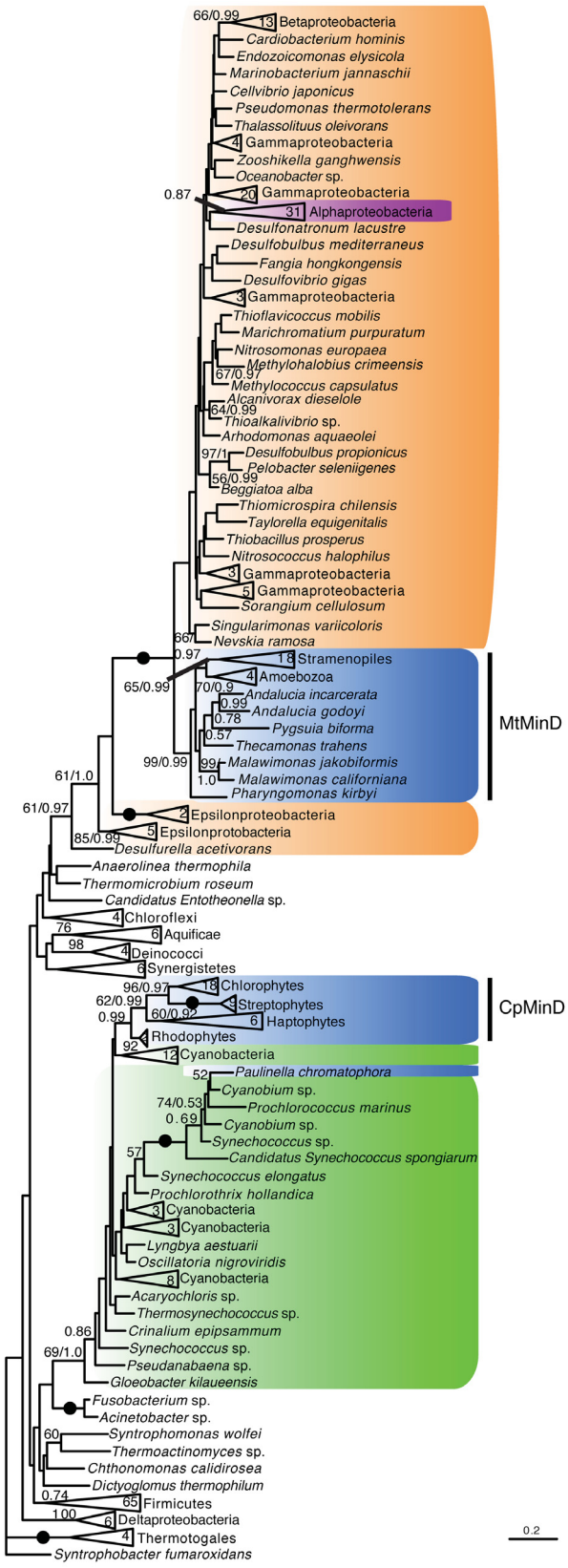
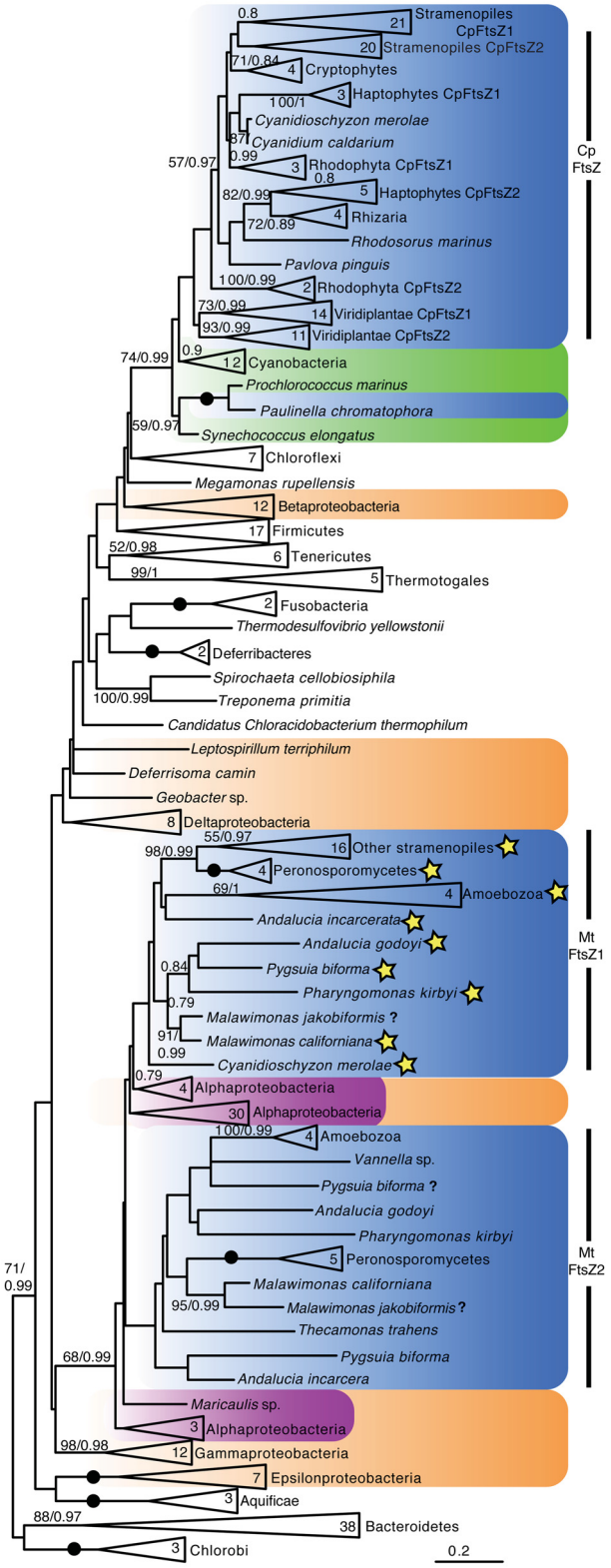

66/0.99
13 Betaproteobacteria
Cardiobacterium hominis
Endozoicomonas elysicola
Marinobacterium jannaschii
Cellvibrio japonicus
Pseudomonas thermotolerans
Thalassolituus oleivorans
4 Gammaproteobacteria
Zooshikella ganghwensis
Oceanobacter sp.
0.87
20 Gammaproteobacteria
31 Alphaproteobacteria
Desulfonatronum lacustre
Desulfobulbus mediterraneus
Fangia hongkongensis
Desulfovibrio gigas
3 Gammaproteobacteria
Thioflavicoccus mobilis
Marichromatium purpuratum
Nitrosomonas europaea
Methylohalobius crimeensis
67/0.97
Methylococcus capsulatus
Alcanivorax dieselole
64/0.99
Thioalkalivibrio sp.
Arhodomonas aquaeolei
97/1
Desulfobulbus propionicus
Pelobacter seleniigenes
56/0.99
Beggiatoa alba
Thiomicrospira chilensis
Taylorella equigenitalis
Thiobacillus prosperus
Nitrosococcus halophilus
3 Gammaproteobacteria
5 Gammaproteobacteria
Sorangium cellulosum
Singularimonas variicoloris
66/
Nevskia ramosa
0.97
18 Stramenopiles
4 Amoebozoa
65/0.99
70/0.9
Andalucia incarcerata
0.99
Andalucia godoyi
0.78
Pygsuia biforma
0.57
Thecamonas trahens
99/0.99
99
Malawimonas jakobiformis
1.0
Malawimonas californiana
Pharyngomonas kirbyi
MtMinD
61/1.0
61/0.97
2 Epsilonproteobacteria
5 Epsilonprotobacteria
85/0.99
Desulfurella acetivorans
Anaerolinea thermophila
Thermomicrobium roseum
Candidatus Entotheonella sp.
4 Chloroflexi
76
6 Aquificae
98
4 Deinococci
6 Synergistetes
96/0.97
18 Chlorophytes
62/0.99
4 Streptophytes
0.99
60/0.92
6 Haptophytes
CpMinD
Rhodophytes
92
12 Cyanobacteria
52
Paulinella chromatophora
Cyanobium sp.
74/0.53
Prochlorococcus marinus
0.69
Cyanobium sp.
Synechococcus sp.
57
Candidatus Synechococcus spongiarum
Synechococcus elongatus
Prochlorothrix hollandica
3 Cyanobacteria
3 Cyanobacteria
Lyngbya aestuarii
Oscillatoria nigroviridis
8 Cyanobacteria
Acaryochloris sp.
Thermosynechococcus sp.
Crinalium epipsammum
0.86
Synechococcus sp.
69/1.0
Pseudanabaena sp.
Gloeobacter kilaueensis
Fusobacterium sp.
Acinetobacter sp.
60
Syntrophomonas wolfei
Thermoactinomyces sp.
Chthonomonas calidirosea
Dictyoglomus thermophilum
0.74
65 Firmicutes
100
6 Deltaproteobacteria
4 Thermotogales
Syntrophobacter fumaroxidans
0.2

Figure 3.5 Unrooted Maximum Likelihood (ML) tree of FtsZ sequences. **(Following page)** Phylogenetic analyses were performed on 327 sequences and 257 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, cyanobacteria green, proteobacteria orange, and α-proteobacteria magenta. Eukaryotic paralogues lacking the variable C-terminal spacer region are indicated by stars, while those with incomplete sequence at the C-terminus are indicated by question marks. The exception to this pattern is a *Corethron hystrix* sequence that, despite branching with other stramenopiles in the MtFtsZ1 clade, possesses a C-terminal variable region (see Appendix B Fig. B.S3).


Stramenopiles CpFtsZ1
Stramenopiles CpFtsZ2
Cryptophytes
Haptophytes CpFtsZ1
Cyanidioschyzon merolae
Cyanidium caldarium
Rhodophyta CpFtsZ1
Haptophytes CpFtsZ2
Rhizaria
Rhodosorus marinus
Pavlova pinguis
Rhodophyta CpFtsZ2
Viridiplantae CpFtsZ1
Viridiplantae CpFtsZ2
Cp FtsZ
Cyanobacteria
Prochlorococcus marinus
Paulinella chromatophora
Synechococcus elongatus
Chloroflexi
Megamonas rupellensis
Betaproteobacteria
Firmicutes
Tenericutes
Thermotogales
Fusobacteria
Thermodesulfovibrio yellowstonii
Deferribacteres
Spirochaeta cellobiosiphila
Treponema primitia
Candidatus Chloracidobacterium thermophilum
Leptospirillum terriphilum
Deferrisoma camin
Geobacter sp.
Deltaproteobacteria
Other stramenopiles
Peronosporomycetes
Amoebozoa
Andalucia incarcerata
Andalucia godoyi
Pygsuia biforma
Pharyngomonas kirbyi
Malawimonas jakobiformis ?
Malawimonas californiana
Cyanidioschyzon merolae
Mt FtsZ1
Alphaproteobacteria
Alphaproteobacteria
Amoebozoa
Vannella sp.
Pygsuia biforma ?
Andalucia godoyi
Pharyngomonas kirbyi
Peronosporomycetes
Malawimonas californiana
Malawimonas jakobiformis ?
Thecamonas trahens
Pygsuia biforma
Andalucia incarcera
Mt FtsZ2
Maricaulis sp.
Alphaproteobacteria
Gammaproteobacteria
Epsilonproteobacteria
Aquificae
Bacteroidetes
Chlorobi
0.2

An early study by Miyagishima and colleagues (Miyagishima et al. 2004) reported the presence of two copies of plastid-targeted FtsZ in photosynthetic eukaryotes, as well as two copies of predicted mitochondrial FtsZ in *C. merolae* and *D. discoideum*. The authors hypothesized that duplication of FtsZ occurred early during primary plastid endosymbiosis, and that a similar process might also have accompanied the establishment of the protomitochondrial endosymbiont. Our broader taxonomic sampling allowed us to confirm the presence of two types of mitochondrial FtsZ homologue in the majority of the eukaryotic taxa examined. They form two distinct phylogenetic clades, each of which contains one homologue from each eukaryote. In addition, although these clades lack strong statistical support, one encompasses copies retaining a variable C-terminal spacer domain that is also found in bacterial homologues (TerBush, Yoshida, Osteryoung 2013), while sequences from the other clade lack this domain (Fig. 5 and Appendix B Fig. B.S3). A robust grouping of α-proteobacterial and both putative mitochondrial FtsZ paralogues was recovered. We subsampled these sequences, using γ-proteobacterial sequences as an outgroup, and re-analyzed them in an attempt to better resolve this clade. We also excluded amoebozoan sequences because of their unusually high AT content and long branches in preliminary trees (Appendix B Fig. B.S3). Unfortunately, we were unable to obtain better resolution of the branching order amongst α-proteobacterial and eukaryotic clades. Nevertheless, the C-terminally-truncated FtsZ proteins were identified only in eukaryotes, and we did not identify more than one FtsZ homologue in α-proteobacteria. We therefore conclude that the duplication event that gave rise to the FtsZ paralogues found in extant

eukaryotes likely occurred early in eukaryotic evolution, rather than earlier, in the α-proteobacterial lineage that gave rise to the mitochondrion.

Altogether, these lines of evidence are consistent with the hypothesis that the nuclear-encoded mitochondrial Min and FtsZ homologues of eukaryotes originated by endosymbiotic gene transfer from the ancestral mitochondrial endosymbiont.

Although found in diverse eukaryotes, the Min proteins are sparsely distributed, a pattern that can only partly be reconciled with taxonomic representation in the available data. A striking example of gene loss is seen in the Mycetozoa (*Dictyostelium* spp., *Acytostelium subglobosum*, and *Polysphondylium pallidum*). Here, *Dictyostelium discoideum, D. citrinum, D. intermedium and D. firmibasis* have retained only FtsZ, while their sister taxon *D. purpureum* and the more basal taxa *Polysphondylium pallidum, P. violaceum* and *Acytostelium globosum* have additionally maintained all three Min proteins. Meanwhile, the yet more distantly related *D. fasciculatum* (Heidel et al. 2011; Romeralo et al. 2011) appears to have independently lost the Min proteins, and, like *D. discoideum,* only possesses FtsZ. This overall pattern raises the question of why Min proteins were retained in some taxa, yet lost in others. No obvious correlation was found with mitochondrial cristae morphology, as Min proteins were found in organisms possessing discoid (e.g., *Pharyngomonas kirbyi* (Harding et al. 2013)) or tubular (e.g., *A. godoyi* (Lara, Chatzinotas, Simpson 2006)) cristae, as well as in lineages with MROs that apparently lack cristae entirely (e.g., *A. incarcerata* and *P. biforma*) (Simpson, Patterson 2001; Stairs et

al. 2014). Nor is there any obvious difference in either overall mitochondrial morphology or lifestyle between lineages that possess Min proteins, and lineages that do not. Kiefel et al. (Kiefel, Gilson, Beech 2004) have raised the possibility that FtsZ is lost in lineages with reticulate mitochondria, and hence the placement of the division site may not affect mitochondrial function. This hypothesis remains a plausible explanation that might apply to Min proteins; *Andalucia godoyi* and *Pygsuia biforma* each possess a single mitochondrion or mitochondrion-related organelle (Lara, Chatzinotas, Simpson 2006; Brown et al. 2013), and *Thecamonas trahens* is predicted to have discrete, non-branching mitochondria based on 3-D reconstructions (Heiss, Walker, Simpson 2013) (Fig. 3.2). Meanwhile, a number of the lineages lacking Min proteins are known to possess reticulate mitochondria in at least one tissue and during at least one life stage including opisthokonts (Barbera et al. 2010), plants (Bendich, Gauriloff 1984; Segui-Simarro, Coronado, Staehelin 2008), the euglenozoan *Euglena gracilis* (Buetow 1989), and apicomplexa (Hogan et al. 1960; van Dooren et al. 2005) (Fig. 3.2). One exception seems to be *Phytophthora cinnamomi*, an organism described in the literature as having 3-4 reticulate mitochondria per cell (Hardham 1987), and in which we found Min homologues. Unfortunately, there are relatively few taxa in our survey for which detailed microscopy data are available that would permit conclusions to be drawn about the three-dimensional structure of their mitochondria. Furthermore, many organisms known to possess reticulate mitochondria may also possess unbranched mitochondria in some tissues, during some parts of their life cycle (Segui-Simarro, Coronado, Staehelin 2008), or alongside reticulate mitochondria, as in

P. cinnamomi (Hardham 1987). It is therefore also possible that the presence or absence of Min proteins reflect some unknown transient mitochondrial morphological features specific to replication. Clearly, genetic and functional studies of mitochondrial Min systems are greatly needed to understand their precise roles.

Further questions are raised by the apparent absence of homologues of all other components of the bacterial divisome from the surveyed eukaryotes, including ZipA, ZapA, FtsA, FtsB, FtsE, FtsI, FtsK, FtsL, FtsN, FtsQ, FtsW and FtsX. BLAST searches of databases using α-proteobacterial and *E. coli* homologues of these proteins as queries yielded no candidate homologues. The bacterial divisome components recruited late in the division process (FtsB, FtsE, FtsI, FtsL, FtsN, FtsQ, FtsW and FtsX) are primarily involved in facilitating peptidoglycan synthesis, and so their apparent absence is perhaps not surprising, given the lack of a peptidoglycan wall in any mitochondria. But it is not clear how the Z-ring remains stabilized and anchored to the membrane in the absence of FtsA, ZipA or ZapA. ZED, a coiled-coil domain protein with 25.8% sequence identity to ZapA, is reported to be involved in mitochondrial Z-ring formation in the red alga *Cyanidioschyzon merolae* (Yoshida et al. 2009). However, we were unable to identify any homologues of this protein in other eukaryotes. The two distinct FtsZ paralogues may form an alternating copolymer that forms the Z-ring; or the Z-ring might be composed of a single paralog, while the second paralogue might instead be involved in attachment of the Z-ring to the membrane. In either case, the anchoring mechanism of FtsZ remains a mystery.

Recent work (Friedman et al. 2011) implicates the endoplasmic reticulum (ER) in the control of the mitochondrial division site location and subsequent Dnm1p recruitment in yeast. This type of external division site control contrasts with that of the Min protein system, which regulates division site location from the mitochondrial matrix. The contrast between these control mechanisms raises the questions of when the role of the ER in mitochondrial division may have emerged; whether any taxa possess both Min proteins and Dnm1p/Drp1; and how these organisms (if they exist) recruit Dnm1p/Drp1 in the absence of ER-mediated division site control. Therefore, an important avenue of further study is the taxonomic distribution of mitochondrial Dnm1p/Drp1, and its functional interplay with FtsZ. Study of this distribution is hampered by the fact that multiple paralogues of dynamins have different functions within eukaryotic cells (Purkanti, Thattai 2015), including vesicular trafficking in yeast (Miyagishima, Nishida, Kuroiwa 2003), and unknown functions in less-studied organisms such as *T. vaginalis* (Wexler-Cohen et al. 2014). These proteins lack N-terminal targeting peptides, and so in the absence of localization data, a mitochondrial function cannot clearly be ascribed to any one of them based on sequence data alone. In any case, investigations into the molecular mechanisms governing the coordination of the various kinds of inner and outer contractile rings are critically needed in diverse eukaryote lineages to fully understand what are features of the division system of the last eukaryotic common ancestor and what are more recent lineage-specific innovations.

Accession Numbers

Accession numbers of Min protein and FtsZ homologues identified in this study and used in analyses are listed in SI Dataset S01, available online at http://www.pnas.org/content/early/2015/03/19/1421392112.abstract?tab=ds

Acknowledgements

The authors thank Dr. Michael W. Gray for planting the seeds of the collaboration leading to this paper. M.M.L. is supported by the National Research Fund, Luxembourg (FNR), and by the Nova Scotia Health Research Foundation (NSHRF). L.E. was supported by a Centre for Comparative Genomics and Evolutionary Bioinformatrics postdoctoral fellowship from the Tula Foundation. T.H. was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). This work was supported by a Regional Partnerships Program Grant (FRN#62809) from the Canadian Institutes of Health Research (CIHR) and the NSHRF awarded to A.J.R, a Czech Science Foundation grant (13-24983S) awarded to M.E., a Czech Science Foundation grant (13-29423S) awarded to P.D, and the European Regional Development Fund awarded to the Biomedicine Center of the Academy of Sciences and Charles University (CZ.1.05/1.1.00/02.0109). M.M.L. would like to thank Yana Eglit for her help in taming Adobe Illustrator. The authors would like to acknowledge the *Acytostelium* genome consortium for kindly providing *A. subglobosum* gene sequences.

Chapter 4: Novel Mitochondrion-Related Organelles in the Free-Living Jakobid *Andalucia incarcerata*

Abstract

Mitochondrion-related organelles (MROs) have arisen independently in a wide range of anaerobic protist lineages. Only a few of these organelles and their functions have been investigated in detail, and most of what is known about MROs comes from studies of parasitic organisms such as the parabasalid *Trichomonas vaginalis*. Here, we describe the MRO of a free-living anaerobic jakobid excavate, *Andalucia incarcerata*. We report an RNAseq-based reconstruction of *A. incarcerata*'s MRO proteome, with an associated biochemical map of the pathways predicted to be present in this organelle. The pyruvate metabolism and oxidative stress response pathways are strikingly similar to those found in the MROs of other anaerobic protists, such as *Trichomonas*. This elegant example of convergent evolution is suggestive of an anaerobic biochemical 'module' of prokaryotic origins that has been laterally transferred among eukaryotes, enabling them to adapt rapidly to anaerobiosis. We also identified genes corresponding to a variety of mitochondrial processes not found in *Trichomonas,* including intermembrane space components of the mitochondrial protein import apparatus, and enzymes involved in amino acid metabolism and cardiolipin biosynthesis. In this respect, the MROs of *A.*

incarcerata more closely resemble those of the much more distantly related *Pygsuia biforma*, likely reflecting their shared lifestyle as free-living anaerobes.

Introduction

It is now accepted that all known extant eukaryote lineages contain, or once contained, a mitochondrion of some description, originating from an α-proteobacterial endosymbiont present in the last eukaryotic common ancestor. Highly derived mitochondrion-related organelles (MROs) are found in diverse anaerobic protist lineages (reviewed in (Tsaousis et al. 2012a; Makiuchi, Nozaki 2014)), and these have been classified primarily based on their roles in ATP generation. Traditionally, MROs that produce energy anaerobically have been classified as hydrogenosomes (Lindmark, Muller 1973; Bui, Johnson 1996; Dyall et al. 2004; Hrdý et al. 2004), while more reduced organelles that lack a role in energy metabolism have been classified as mitosomes (Tovar, Fischer, Clark 1999; Jedelsky et al. 2011); a more recent classification scheme aims to encompass the broader spectrum of energy functions since discovered in MROs (Müller et al. 2012). Only the first three classes have retained an electron transport chain. Class 1 mitochondria, typical mitochondria, are only capable of generating ATP through oxidative phosphorylation. Class 2 mitochondria are capable of functioning anaerobically using alternative electron acceptors, such as fumarate, while class 3 mitochondria also possess a bacterial-like anaerobic ATP-generation pathway. Class 4 mitochondria, also known as hydrogenosomes, also possess this pathway, but have lost most components of the electron transport

chain, such that they are no longer able to generate ATP through aerobic phosphorylation, The more highly reduced class 5 mitochondria, also known as mitosomes, lack any role in ATP generation and any of the associated enzymes. Based on the distribution of MROs within eukaryotic phyla (many of which contain predominantly mitochondriate taxa), it has been proposed that this adaptation to anaerobic conditions has occurred multiple times, including at least four independent events within ciliates alone (Embley et al. 1995). Nevertheless, striking biochemical similarities are apparent between these taxa in the form of enzymes shared between distantly-related anaerobic and microaerophilic eukaryotes (Andersson et al. 2003; Andersson et al. 2006; Stairs, Roger, Hampl 2011; Müller et al. 2012; Tsaousis et al. 2012a).

Hydrogenosomes (class 4 mitochondria) were first described in trichomonads (Lindmark, Muller 1973), and have since been discovered in many anaerobic protistan lineages, including the excavates *Sawyeria marylandensis* (Barbera et al. 2010), *Psalteriomonas lanterna* (de Graaf et al. 2009), and *Spironucleus salmonicida* (Jerlstrom-Hultqvist et al. 2013); chytridiomycete fungi such as *Nyctotherus ovalis* (Bowman et al. 1992); and some anaerobic ciliates (Yarlett et al. 1981; Akhmanova et al. 1998). They contain a characteristic anaerobic ATP-generating pathway (Dyall et al. 2004; Hrdý et al. 2004; Hrdý, Tachezy, Müller 2008; Tsaousis et al. 2012a), which relies on the presence of two key proteins not found within eukaryotic mitochondria: pyruvate:ferredoxin oxidoreductase (PFO) and [FeFe]-hydrogenase. PFO decarboxylates pyruvate, producing acetyl-CoA and carbon dioxide. In some anaerobic eukaryotes, such as *Blastocystis* sp.,

Mastigamoeba balamuthi or *Cryptosporidium parvum*, pyruvate to acetyl-CoA conversion is carried out by a pyruvate:NADP oxidoreductase (PNO) or a pyruvate:formate lyase (PFL) (Ctrnacta et al. 2006; Lantsman et al. 2008; Stairs, Roger, Hampl 2011). Coenzyme A is subsequently transferred from acetyl-CoA to succinate by an acetate:succinate CoA transferase (ASCT (Tielens et al. 2010)), generating acetate and succinyl-CoA. The subsequent regeneration of succinate, catalyzed by the tricarboxylic acid (TCA) cycle enzyme succinate thiokinase (STK), generates ATP through substrate-level phosphorylation. The electrons generated in the decarboxylation of pyruvate are transferred from PFO, via ferredoxin, to [FeFe]-hydrogenase, which catalyses the reductive formation of hydrogen gas from free protons. Three maturases, HydE, HydF and HydG, are involved in the insertion of an iron sulfur cluster into the active site of [FeFe]-hydrogenase. The 51kDa and 24kDa subunits of mitochondrial electron transport chain Complex I are commonly found in anaerobic eukaryotes in the absence of other Complex I subunits, and are capable of transferring electrons to ferredoxin (Hrdý et al. 2004). The origin(s) of PFO and [FeFe]-hydrogenase and its maturases in diverse anaerobic eukaryotes are currently under debate. Phylogenetic analyses, which have frequently suffered from poor resolution, have not provided definitive evidence for lateral transfers of these protein genes from either anaerobic bacteria or between eukaryotes, nor have they shown proof of a mitochondrial origin (Vignais, Billoud, Meyer 2001; Davidson et al. 2002; Horner et al. 2002; Hug, Stechmann, Roger 2010; Leger et al. 2013; Stairs et al. 2014).

With the advent of high-throughput sequencing, MROs from an increasing number of anaerobic lineages have been characterized, with an emphasis on the totality of predicted functions of the organelles, and not merely on energy metabolism. These studies have revealed a great diversity of ancestrally mitochondrial and newly acquired biochemical pathways in MROs (Akhmanova et al. 1998; van Hoek et al. 2000; Boxma et al. 2005; Lantsman et al. 2008; Stechmann et al. 2008; Denoeud et al. 2011; Tsaousis et al. 2011; Nyvltova et al. 2013; Stairs et al. 2014; Nyvltova et al. 2015).

Here, we examine the mitochondrion-related organelle of the anaerobic protist *Andalucia incarcerata*. *A. incarcerata* is a member of the Jakobida (Lara, Chatzinotas, Simpson 2006), a group of flagellated protists that contain typical aerobic mitochondria with the largest, most eubacterial-like genomes found to date (Lang et al. 1997; Burger et al. 2013; Kamikawa et al. 2014), eubacterial-like RNA polymerases, and Shine-Dalgarno-like motifs upstream of translation sites (Lang et al. 1997). As the only known anaerobic member of this group, *A. incarcerata* represents a unique opportunity to examine organellar adaptation to an anaerobic environment. The original identification by electron microscopy described uniformly-staining MROs lacking cristae (Simpson, Patterson 2001); however, the biochemical nature of the MROs has not been studied until now.

A. incarcerata appears to represent another independent derivation of an anaerobic MRO, specifically from a lineage containing mitochondria with ancestral features. The biochemical processes present within this cellular compartment could therefore provide valuable information as to the

mechanisms by which MROs evolve from mitochondria. Furthermore, together with other recent studies (Gill et al. 2007; de Graaf et al. 2009; Barbera et al. 2010; Nyvltova et al. 2013; Stairs et al. 2014; Nyvltova et al. 2015), *Andalucia incarcerata* presents an opportunity to contrast this process in a free-living organism with organellar reduction events that have occurred in the better-studied parasite lineages (Loftus et al. 2005; Aguilera, Barry, Tovar 2008; Jedelsky et al. 2011; Schneider et al. 2011; Jerlstrom-Hultqvist et al. 2013).

Materials and Methods

Cell Culture

A. incarcerata MB1 had previously been isolated from Mahone Bay, Nova Scotia, Canada (Simpson, Perley, Lara 2008). Cells were maintained at 21°C in a medium composed of 50:50 Page's modified Neff's Amoeba Saline (AS; 2 mM Na_2HPO_4, 2 mM KH_2PO_4, 0.03 mM $MgSO_4 \cdot 7H_2O$, 0.05 mM $CaCl_2 \cdot 2H_2O$, 4.1 mM NaCl):Artificial seawater (ASW; 0.42 M NaCl, 9 mM KCl, 9.25 mM $CaCl_2 \cdot 2H_2O$, 23 mM $MgCl \cdot 6H_2O$, 25.5 mM $MgSO_4 \cdot 7H_2O$, 2.14 mM $NaHCO_3$) plus 3% Luria-Bertani (LB) broth (10g tryptone, 5g yeast extract, 10g NaCl in 1l). Cultures were maintained in 50ml Falcon tubes filled to approx. 45ml.

Expressed Sequence Tag (EST) Library Creation

Total RNA was extracted from high-density cultures of *A. incarcerata* using TRIzol® Reagent (Invitrogen) following the manufacturer's instructions for cells grown in suspension. A 2 µg aliquot of high quality total RNA was sent to

Amplicon Express (Pullman, USA) for EST library creation. RNA underwent two rounds of poly-A selection, followed by reverse-transcription to complementary DNA (cDNA). cDNA reads were size-selected, and cloned into pBluescript II SK+.

The completed EST library was sent to the National Research Council (NRC), Halifax location for Sanger sequencing. A total of 3456 clones were sequenced and processed using Phred and Phrap (Ewing, Green 1998; Ewing et al. 1998).

The library was then amplified, nebulized and subjected to 454 GS-FLX Titanium pyrosequencing at the McGill University Genome Quebec Innovation Centre, producing 11905 reads.

A mixed Sanger/454 assembly was produced using Mira (Chevreux, Wetter, Suhai 1999).

Illumina RNASeq

Total RNA was extracted from high-density cultures of *A. incarcerata* as described above, and sent to Macrogen (Seoul, Korea) for two rounds of poly-A selection, followed by Illumina HiSeq2000 paired-end sequencing. 63.4 million raw reads were produced, and assembled *de novo* using Trinity (Grabherr et al. 2011). Contigs were compared with the existing Sanger/454 transcriptome to verify the accuracy of the transcripts.

Identification and Analysis of Putative Organellar Protein Genes

A conservative set of sequences encoding possible MRO proteins was identified using CBOrg, a comparative BLAST tool for predicting MRO proteomes based on

Homo sapiens, *Saccharomyces cerevisiae*, *Tetrahymena thermophila* and *Trichomonas vaginalis* proteomes (Gaston, Tsaousis, Roger 2009). The sequences in this set were re-examined using Basic Local Alignment Search Tool (BLAST (Altschul et al. 1997)) searches against the non-redundant nucleotide and protein databases in GenBank to confirm homology to known mitochondrial proteins of other eukaryotes, and to eliminate likely bacterial contaminants. Additionally, hidden Markov Model (HMM) searches were used to identify proteins involved in protein import and translocation, as these are often divergent. This was done using HMMER 3.0 (http://www.hmmer.org), with HMMs constructed from multiple sequence alignments of homologous sequences provided by Dr. Trevor Lithgow and aligned using MUSCLE v3.8.31 (Edgar 2004b; Edgar 2010). The presence or absence of a full 5' end of the gene was determined through translation of the sequence and identification of a putative initiator methionine and an in-frame stop codon upstream. BLAST homology was used to confirm the presence of a complete 5' end compared to other known sequences for the gene in question.

In Silico Detection of N-Terminal Targeting Peptides

For sequences with a full 5' end of the coding sequence, the translated protein sequence was screened for the presence of a potential N-terminal targeting peptide using the internet-based targeting peptide prediction programs TargetP (Emanuelsson et al. 2000; Emanuelsson et al. 2007) and Mitoprot (Claros 1995). Targeting peptides were designated as probable if organelle localization

probability was above 50%, and if an N-terminal extension was present relative to the closest bacterial homologs.

Generating Complete Sequence Data

Total RNA was extracted as described above. From a total RNA extract, messenger RNA (mRNA) was enriched using the Poly(A) Purist mRNA purification kit (Ambion Inc, TX). cDNA was generated from mRNA using the GeneRacer kit (Invitrogen Corp., USA). Total DNA was extracted by centrifuging high-density cultures, resuspending the pellets in extraction buffer (1 M Tris-HCl pH 7.5, 5 M NaCl, 0.5 M EDTA, 20% sodium dodecyl sulfate (SDS), incubating at 50°C for 10 min., and centrifuging for 10 min. at 10000g. The resulting supernatant was subjected to two rounds of phenol:chloroform:isoamyl alcohol (25:24:1) (Fluka, Germany) extraction, followed by a third extraction in chloroform:isoamyl alcohol (24:1), and precipitation through addition of 1/10th volume 3 M sodium acetate and 6/10th volume of isopropanol

Rapid Amplification of cDNA Ends (RACE) PCR was used to obtain missing sequence data from MinC, which was not completely represented in the transcriptomic data, and standard PCR was used to obtain full-length sequence data from specific genes of interest to verify sequence information, or to search for introns in genomic DNA (for HydE, HydF, HydG, all [FeFe]-hydrogenases, all PFOs, ASCT1B, acyl-CoA synthetase, Grx5, SufCB, and ferredoxin-fused flavodiiron protein). Gene-specific primers for genes of interest were designed based on sequence information from the *A. incarcerata* transcriptome.

Western blotting

[FeFe]-hydrogenase 1 and IscS, both predicted to be localized to the MROs based on N-terminal targeting peptides, were amplified by PCR from *A. incarcerata* cDNA using primers with N-terminal restriction sites (NdeI and XhoI for IscS, and XhoI and BamHI for [FeFe]-hydrogenase 1). Each gene was cloned into pET-16b (Novagen) downstream of a 6xHis tag for recombinant expression in OverExpress™ C41(DE) cells, kindly provided by Prof. John Walker (Medical Research Council Mitochondrial Biology Unit, Cambridge, UK). The recombinant protein was expressed in inclusion bodies, which were purified using BugBuster reagent (Novagen).

Western blotting

Antibodies were tested against inclusion bodies from C41(DE) cells expressing recombinant *A. castellanii* IscS or [FeFe]-hydrogenase 1 from the pET-16b vector, or the empty pET-16b vector (as a negative control). The antibodies used were a polyclonal anti-*Giardia* IscS antibody (Tovar et al. 2003) kindly provided by Dr. Jan Tachezy (Charles University in Prague, Czech Republic), and a custom polyclonal anti-[FeFe]-hydrogenase antibody previously raised against a *Mastigamoeba balamuthi* [FeFe]-hydrogenase peptide (CPFGAVMTRSFMLDVMRAMRDSRSAGSKVVAMVAPAVAGHMGNAPIWSICE ALKRAGFDEALEVSIGADTTTENEAHEFEERFGEGAPKGKFSFMTTSCCPAYVA CVRKHVPEIEDAVSHTRSPMHYTAKLAKERWPGCTTVFVGPCTAKLHEASIDE YTDFAITVVEALSLLRGRGVALDNSQ; Abgent). A monoclonal antibody raised against the polyhistidine tag (anti-His; Thermo Scientific) was used as a positive control. Proteins were separated by sodium dodecyl sulfate-polyacrylamide gel

electrophoresis, blotted onto polyvinylidene difluoride membranes, and blocked overnight at 4°C in 5% milk. The blots were then incubated with primary antibody at a ratio of 1:5000 (anti-His), 1:500 (anti-IscS) or 1:250 (anti-[FeFe]-hydrogenase) in 1% milk for 1 hr at room temp, washed, and incubated with horseradish peroxidase-conjugated secondary anti-mouse (for anti-His) or anti-rabbit (for anti-IscS or anti-[FeFe]-hydrogenase) IgG antibodies (Sigma), 1:30000 or 1:50000 respectively, for 1 hr at room temp. The blots were then washed again, incubated with electrochemiluminescence western blotting reagents (Amersham), and photographed using either film or a FluorChem E imaging system (Protein Simple).

Immuno-Electron Microscopy (Immuno-EM)

Cells were fixed in 4% paraformaldehyde and 0.5% glutaraldehyde in 0.1 sodium cacodylate buffer, rinsed in 0.1M sodium cacodylate buffer, subjected to stepwise dehydration in ethanol, and embedded in LR White Resin. 100nm sections were cut using a LKB Huxley ultramicrotome with a Diatome diamond knife, and placed on 200 mesh nickel grids. Sections were blocked in PBS containing 0.8 % Bovine Serum Albumin and 0.01 % Tween 20 for 30 min, then incubated with primary antibody diluted in blocking solution for 3 hours at room temp or overnight at 4°C. The primary antibodies used were a polyclonal anti-*Giardia* IscS antibody raised against the recombinant protein (Tovar et al. 2003), or a custom polyclonal antibody raised against a *Mastigamoeba* [FeFe]-hydrogenase peptide (Abgent). The grids were then washed for 10 min each of four times in blocking solution before incubation with gold-conjugated anti-rabbit (IscS) or

anti-rat ([FeFe]-hydrogenase) secondary antibody (Sigma or Electron Microscopy Sciences) diluted 1:200 in blocking solution. Labeled sections underwent one 10-minute wash in blocking solution, three 10-minute washes in PBS, and two 30-second washes in dH_2O. Following antibody labeling, grids were stained using 2% aqueous uranyl acetate and rinsed, stained with lead citrate, rinsed and air-dried. The sections were viewed using a JEOL JEM 1230 transmission electron microscope at 80 kV, and images were captured using a Hamamatsu ORCA-HR digital camera.

The areas of the nucleus, mitochondria and cytosol of each cell cross-section were measured using ImageJ for 8 cells (IscS localization) or 13 cells ([FeFe]-hydrogenase localization), and the gold particles in each part of the cell were counted by eye.

Phylogenetic Analyses

For each protein, all known eukaryotic and bacterial homologues in NCBI were collected and then multiple alignments were generated using PROBCONS v.1.2 (Do et al. 2005) or MAFFT-L-INSI v7.149b (Katoh et al. 2002; Katoh et al. 2005; Katoh, Standley 2013), and trimmed of ambiguously aligned sites with BMGE 1.1 (Criscuolo, Gribaldo 2010) (-m BLOSUM30, all other parameters default). Preliminary phylogenies were generated using FastTree, and datasets were manually curated. Twenty independent Maximum Likelihood (ML) tree estimates and 200 bootstrap replicates were generated using RAxML v.8.0.23 (Stamatakis 2014) under the PROTGAMMALG4X (Le, Dang, Gascuel 2012) model of amino acid substitution. Bayesian inference posterior probabilities

were calculated using PhyloBayes v.3.3f (Lartillot, Lepage, Blanquart 2009) under the catfix C20 model of evolution. We tested whether specific phylogenetic hypotheses were rejected by the data using the Approximately Unbiased test implemented in CONSEL v.1.20 (Shimodaira, Hasegawa 2001). Maximum likelihood trees given specific constraints (i.e. corresponding to specific hypotheses) were generated using RAxML, and the 200 trees from bootstrap replicates were included in analyses, as required by CONSEL. PFO and [FeFe]-hydrogenase phylogenies were performed both including paralogs (sulfite reductases, and periplasmic [FeFe]-hydrogenases respectively).

Results and Discussion

The Andalucia MROs Likely Lack an Organellar Genome

The closest known relative of *A. incarcerata*, *Andalucia godoyi*, possesses the most gene-rich mitochondrial genome known so far (Burger et al. 2013). However, no orthologs of the mitochondrial genome-encoded *A. godoyi* genes could be recovered from the *A. incarcerata* transcriptome; in the absence of these genes, or any sequences encoding mitochondrial transcriptional or translational apparatus components, it seems likely that *A. incarcerata* MROs lack an organellar genome.

Identification of Organellar Proteins In Silico

CBOrg and comparative BLAST analysis identified 159 proteins of putative mitochondrial or hydrogenosomal origin. Of these, 112 sequences contain N-terminal targeting peptides (Appendix C, Table C.S1). It should be noted that the

majority of the proteins discussed above have been assigned as organellar based mainly on BLAST homology to typically mitochondrial proteins. Given the identification of both subunits of the mitochondrial processing peptidase, it seems likely that typical targeting pathways are active within the organelle, and thus identification of N-terminal targeting peptides will provide further evidence for the localization of these proteins. Most of the proteins predicted to be localized to the MROs have N-terminal extensions relative to bacterial homologs; these extensions are generally predicted to be N-terminal targeting peptides by TargetP and Mitoprot. Exceptions are homologs of known membrane proteins, such as many components of the mitochondrial protein import system. It is also reasonable to expect that a significant portion of the proteins actually targeted to the organelle contain cryptic internal targeting signals (Neupert 1997) and as such will not have been identified in this survey. Fig. 4.1 shows a biochemical map of selected pathways predicted to be present in the MRO.

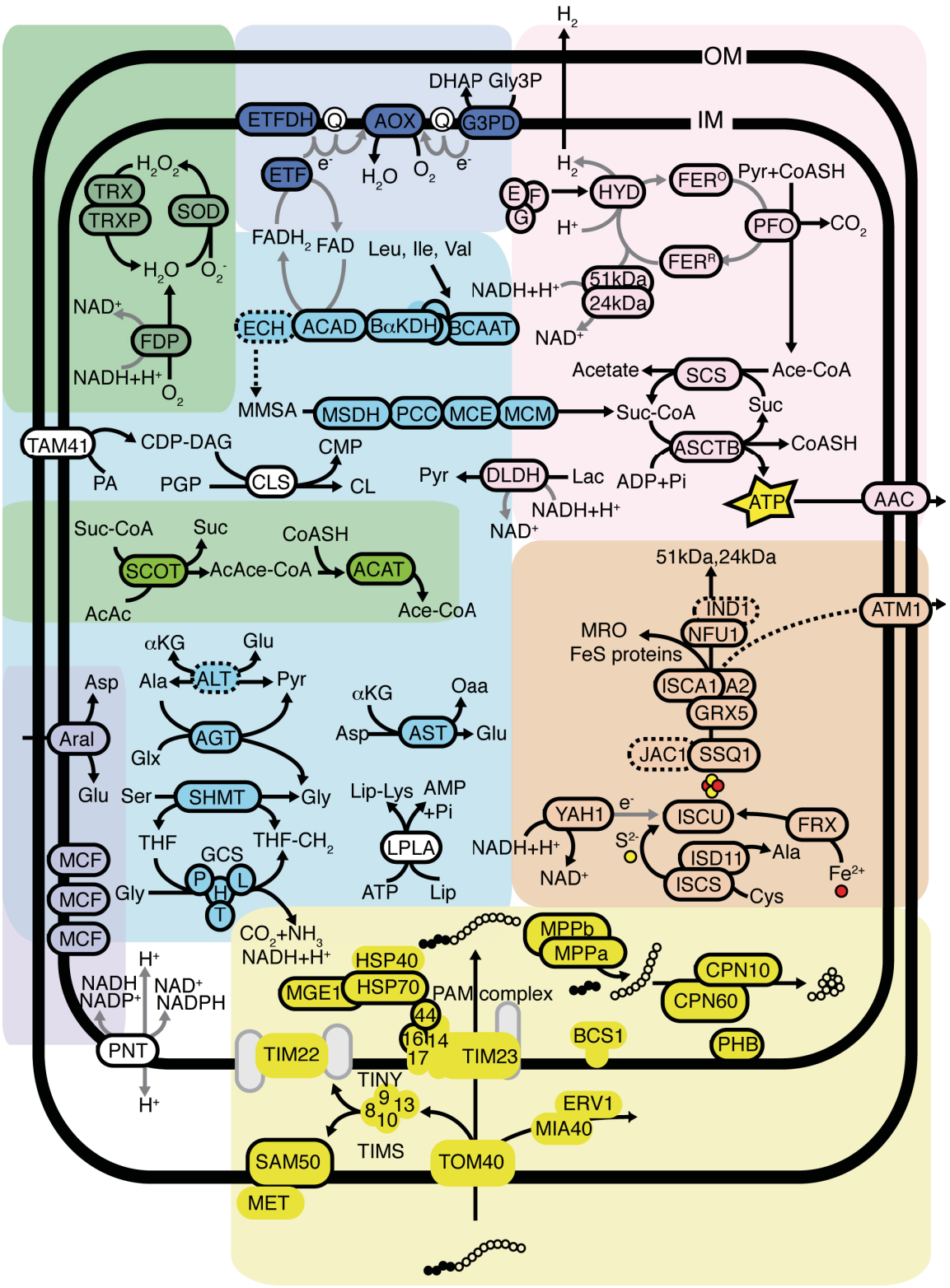


Figure 4.1 A hypothetical biochemical map of the mitochondrion-related organelle in *A. incarcerata*. All proteins depicted are based on Blastcompare analyses of the EST data intended to identify mitochondrion or hydrogenosome-localized proteins. Solid

circles indicate the protein contains a predicted N-terminal targeting peptide. Dashed circles correspond to genes with an incomplete coding region at the 5' end, and thus the presence/absence of a targeting peptide could not be evaluated. Grey circles correspond to components of pathways or solute carriers which were not represented in the EST survey, but which are anticipated to be present in the organelle based on the presence of other components within complexes and pathways. Pathways and associated abbreviations of their enzyme components are depicted as follows:

General: OM: outer mitochondrial membrane; IM: inner mitochondrial membrane; AMP: adenosine monophosphate; ADP: adenosine diphosphate; ATP: adenosine triphosphate; NAD: nicotinamide adenine dinucleotide; NADP: nicotinamide adenine dinucleotide phosphate; AcAc: acetoacetate; Ace-CoA: acetyl-CoA; AcAce-CoA: acetoacetyl-CoA; CoASH: Coenzyme A; MMSA: (S)-methylmalonate-semialdehyde; Suc: succinate; Suc-CoA: succinyl-CoA; Ala: alanine; Asp: aspartate; Glu: glutamate; Glx: glyoxylate; Gly: glycine; Ile: isoleucine; Leu: leucine; Lys: lysine; Ser: serine; Val: valine; Lac: lactate; Lip: lipoate; Oaa: oxaloacetate; Pyr: pyruvate; Q: quinone; THF: tetrahydrofolate. Yellow circles: sulphur moiety of FeS clusters; red circles: iron moiety of FeS clusters; contiguous white circles: imported protein; contiguous black circles: mitochondrial targeting peptide of imported protein. Solid black arrows indicate biochemical pathways; dashed black arrows indicate poorly-understood or unclear pathways; solid grey arrows indicate electron transfer.

Pyruvate metabolism (pink): 24kDa: 24kDa subunit of Electron Transport Chain Complex I; 51kDa: 51kDa subunit of Electron Transport Chain Complex I; AAC: ADP/ATP translocator; ASCTB: acetate:succinate CoA-transferase, subtype 1B; E, F, G: [FeFe]-hydrogenase maturases HydE, HydF and HydG, respectively; FER^{O}: oxidized electron transport ferredoxin; HYD: FER^{R}: reduced electron transport ferredoxin; HYD: [FeFe]-hydrogenase; PFO: pyruvate:ferredoxin oxidoreductase; SCS: succinyl-CoA synthetase.

Carriers (lilac): Aral: Aralar solute carrier; MCF: Mitochondrial Carrier Family protein.

Ketone body degradation (light green): ACAT: acetyl-CoA C-acetyltransferase; SCOT: succinate:3-oxoacid CoA-transferase.

ROS and oxygen stress (dark green): FDP: flavodiiron protein; SOD: superoxide dismutase; TRX: thioredoxin; TRXP: thioredoxin peroxidase.

Mitochondrial protein import and folding (yellow): BCS: ubiquinol-Cytochrome c reductase Synthesis; ERV: Essential for Respiration and Viability; CPN: Chaperonin; HSP: Heat Shock Protein; MET: Metaxin; MGE: Mitochondrial GrpE; MIA: Mitochondrial intermembrane space Import and Assembly; MPP: Mitochondrial Processing Peptidase; PHB: Prohibitin; SAM: Sorting and Assembly Machinery; TIM: Translocator of the Inner Membrane; TOM: Translocator of the Outer Membrane; 8, 9, 10, 13: Translocator of the Inner Membrane proteins Tim8, Tim9, Tim10, Tim13, respectively.

Iron-sulfur cluster assembly (orange): ATM: ABC transporter, Mitochondrial; FRX: Frataxin; GRX: Glutaredoxin; IND: iron-sulfur protein required for NADH dehydrogenase; ISC: Iron-Sulfur Cluster; ISD: Iron-Sulfur protein biogenesis, Desulfurase-interacting; JAC: J-type Accessory Chaperone; NFU: NifU-like; SSQ: Stress Seventy subfamily Q; YAH: Yeast Adrenodoxin Homolog.

Amino acid metabolism (light blue): AGT: alanine – glyoxylate aminotransferase; ALT: alanine aminotransferase; AST: aspartate aminotransferase; DLDH: D-lactate dehydrogenase; GCS: Glycine Cleavage System; H, L, P, T: H-, L-, P- and T-protein components of the glycine cleavage system, respectively; SHMT: serine hydroxymethyltransferase; ACAD: branched-chain acyl-CoA dehydrogenase; BαKDH:

branched-chain α-ketoglutarate dehydrogenase complex; BCAAT: branched-chain amino acid aminotransferase; ECH: enoyl-CoA hydratase; MCE: methylmalonyl-CoA epimerase; MCM: methylmalonyl-CoA mutase; MSDH: methylmalonyl semialdehyde dehydrogenase; PCC: propionyl-CoA carboxylase.
Electron transport (dark blue): AOX: alternative oxidase; ETF: electron transport flavoprotein; ETFDH: electron transport flavoprotein dehydrogenase; G3PD: glycerol-3-phosphate dehydrogenase**.**
Other (white): CDP-DAG: cytidine diphosphate diacylglcerol; CMP: cytidine monophosphate; CL: cardiolipin; CLS: cardiolipin synthase; LPLA: lipoate protein ligase; PA: phosphatidic acid; PNT: pyridine nucleotide transhydrogenase; TAM: translocator assembly and maintenance.

ATP Generation

Conspicuously absent from the EST library are any subunits of complex I other than the 51kDa and 24kDa subunits, or components of the electron transport chain downstream of complex I, as well as any subunits of the pyruvate dehydrogenase (PDH) complex, all typically present in and associated with aerobic ATP generation in classic mitochondria. In contrast, the identification of PFO, [FeFe]-hydrogenase and its maturases, ferredoxin, ASCT, STK, and the two complex I subunits constitutes the bulk of the typical hydrogenosomal ATP generation pathway previously described in *T. vaginalis* (Dyall et al. 2004; Hrdý et al. 2004; Carlton et al. 2007; Hrdý, Tachezy, Müller 2008). Like *T. vaginalis*, *A. incarcerata* possesses multiple MRO-targeted homologs of both PFO and [FeFe]-hydrogenase; but both of its copies of ASCT belong to the subtype 1B family, which is distinct from the subtype 1C enzyme found in *T. vaginalis* (Tielens et al. 2010), providing an example of convergent biochemical adaptation between the two species. The *A. incarcerata* transcriptome also encodes PFL and PNO, although these enzymes are not predicted to localize to the MROs.

In addition to the two MRO-targeted [FeFe]-hydrogenase homologs, the transcriptome encodes two further [FeFe]-hydrogenases that lack N-terminal

mitochondrial targeting peptides. Each of these is fused to a C-terminal CysJ similar to those of pyruvate:NADP oxidoreductase (PNO) and NADPH cytochrome p450 reductase. This type of [FeFe]-hydrogenase has previously been described in *T. vaginalis*, (Carlton et al. 2007), and in the breviate *Pygsuia biforma* (Stairs et al. 2014); in both of these organisms, the CysJ-fused [FeF]-hydrogenases are also predicted to be cytosolic. We confirmed [FeFe]-hydrogenase localization in the MROs using immunogold labeling (Fig. 4.2; Appendix C, Fig. C.S1). We used a heterologous polyclonal primary antibody raised in rat against *Mastigamoeba* [FeFe]-hydrogenase peptide. We probed labeled cells with a secondary anti-rat antibody conjugated to gold particles, and measured the distribution of these particles in *Andalucia incarcerata* cells. We found a sevenfold higher density of gold particles in the MROs compared with the cytosol. We also observed cross-reaction of the antibody with engulfed bacteria, and with the nucleus (although staining in the nucleus was significantly lower than that in the MROs). The former likely results from cross-reaction with bacterial homologs, while the latter might be attributed to the presence of Nuclear prelamin A Recognition Factor-Like protein, a distant homolog of [FeFe]-hydrogenase. The significantly higher staining within MROs suggests that the bulk of [FeFe]-hydrogenases in *A. incarcerata* are located in the MROs, confirming a likely role of this organelle in ATP production.

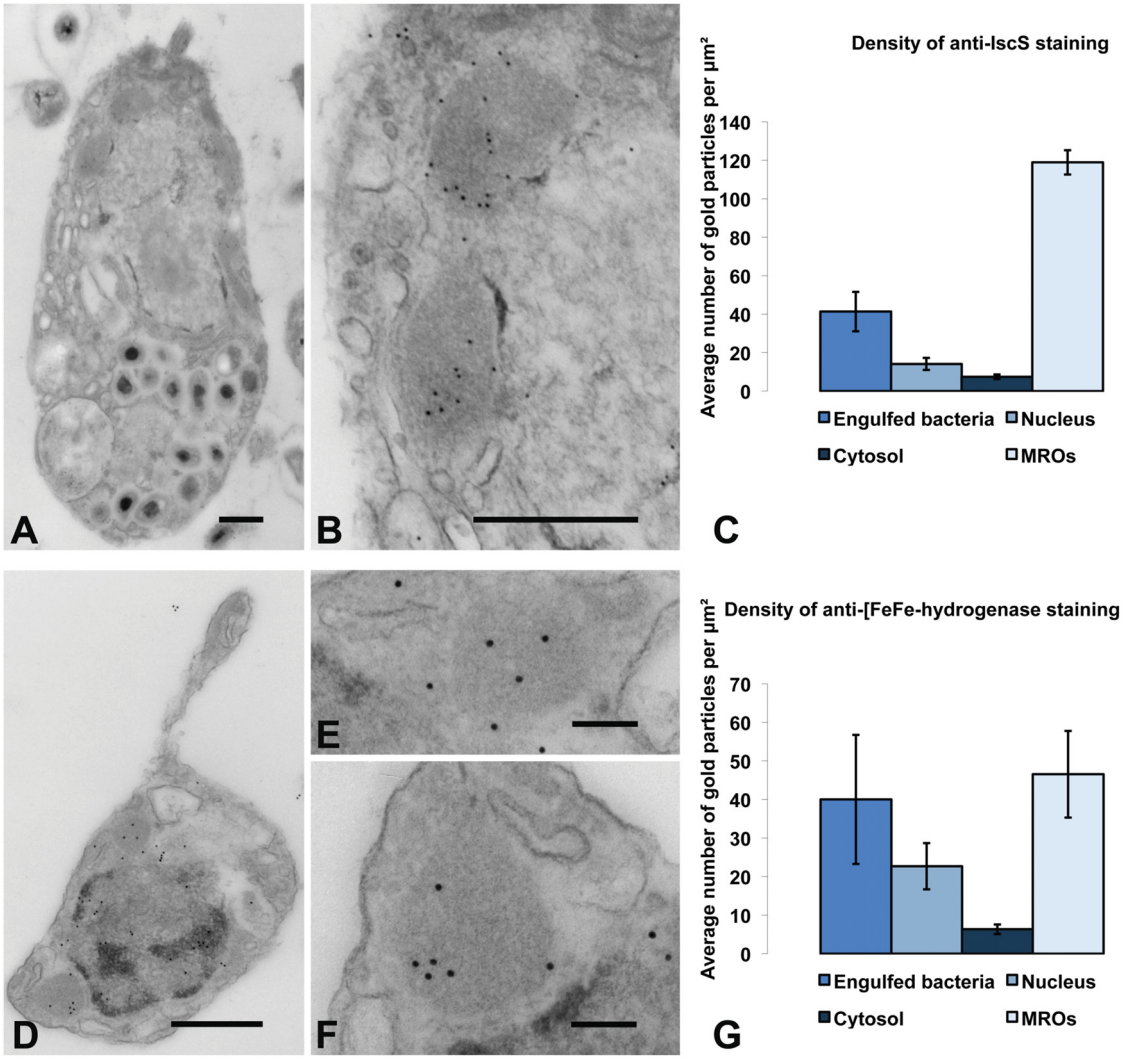


Figure 4.2 Immunogold localization of IscS and [FeFe]-hydrogenase in *A. incarcerata* cells. A. A representative whole cell fixed for immunogold staining; scale bar, 500 nm; this image has been cropped in order to show only the whole cell from which the insets shown were derived. B. Magnified section from the cell pictured in (A), showing gold particles corresponding to anti-*Giardia* IscS localization; scale bar, 500 nm. C. Mean density of immunogold labeling in engulfed bacteria, the cytosol, the nucleus, and MROs (7 cells), ± standard error of the mean. D. Representative whole cell fixed for immunogold staining; scale bar, 500 nm; this image has been cropped in order to show only the whole cell from which the insets shown were derived. E, F. Magnified sections from the cell pictured in (A), showing gold particles corresponding to anti-*Mastigamoeba* [FeFe]-hydrogenase localization; scale bar, 500 nm. G. Mean density of immunogold labeling in engulfed bacteria, the cytosol, the nucleus, and MROs (13 cells), ± standard error of the mean. Brightness and contrast have been adjusted in each image to enhance visibility of the mitochondria and gold particles.

Phylogenies recover eukaryote monophyly for PFO, and all of the [FeFe]-hydrogenase maturases (Fig. 4.3, Fig. 4.5, Appendix C, Fig.s C.S2, C.S4, C.S5). Two major clades of [FeFe]-hydrogenase are recovered, one of which is a long-branching clade that includes bacterial H_2-evolving periplasmic [FeFe]-hydrogenases (Vignais, Billoud 2007), as well as *Mastigamoeba*, *Entamoeba* and *Trimastix* homologs. As expected, monophyly of all eukaryotic hydrogenases, both periplasmic-like and non-periplasmic, was rejected in AU tests (Table 4.1) (Hug, Stechmann, Roger 2010; Leger et al. 2013). However, as in previous studies, monophyly of non-periplasmic [FeFe]-hydrogenases was not rejected in AU tests (Fig. 4.4, Table 4.1, Appendix C, Fig. C.S3) (Leger et al. 2013). The origins of these anaerobic energy enzymes remain frustratingly unclear, as there is no clear prokaryotic sister clade to any of these groups. In contrast to previous studies, a specific grouping of α-proteobacteria and eukaryotes was not rejected in AU tests; nevertheless, the best constrained trees generated by RAxML for these hypotheses placed α-proteobacteria branching from within eukaryotes (data not shown), and the reverse hypothesis was rejected in most cases. As in previous studies (Hug, Stechmann, Roger 2010; Leger et al. 2013), the hypothetical grouping of α-proteobacteria with eukaryotes was clearly rejected for PFO, whether sulfite reductases were included in the analyses or not.

Table 4.1 Approximately unbiased tests of alternate topologies * Rejected with P-val < 0.05 ** Rejected with P-val < 0.01

Hypothesis tested	**AU test P-value**
[FeFe]-hydrogenase	
ML tree	0.657
Monophyly of Clade 1 and Clade 2 eukaryotic hydrogenases; *Thalassiosira* unconstrained	7e-37**
Monophyly of Clade 1 eukaryotes and Clade 1 α-proteobacteria; *Thalassiosira* unconstrained; eukaryotic periplasmic hydrogenases unconstrained; all other α-proteobacteria unconstrained	0.414
Clade 1 α-proteobacteria basal to Clade 1 eukaryotes; *Thalassiosira* unconstrained; Clade 2 eukaryotes unconstrained; all other α-proteobacteria unconstrained	0.560
[FeFe]-hydrogenase (non-periplasmic)	
ML tree	0.559
Eukaryote monophyly; *Thalassiosira* unconstrained	0.634
Monophyly of eukaryotes and Clade 1 α-proteobacteria; *Thalassiosira* unconstrained; all other α-proteobacteria unconstrained	0.688
Monophyly of eukaryotes and α-proteobacteria; *Thalassiosira* unconstrained; all other α-proteobacteria unconstrained	9e-05**
Clade 1 α-proteobacteria basal to eukaryotes; *Thalassiosira* unconstrained; all other α-proteobacteria unconstrained	1e-04**
Clade 2 α-proteobacteria basal to eukaryotes; *Thalassiosira* unconstrained; all other α-proteobacteria unconstrained	8e-05**
PFO	
ML tree	1.000
Monophyly of eukaryotes and α-proteobacteria	3e-04**
α-proteobacteria basal to eukaryotes	2e-04**
PFO+sulfite reductase	
ML tree	0.999
Monophyly of eukaryotes and α-proteobacteria	0.008**
α-proteobacteria basal to eukaryotes	0.027*
ASCT1B	
ML tree	0.739
Eukaryote monophyly	7e-05**
Monophyly of Clade 1 eukaryotes and Clade 2 eukaryotes; other	0.718

Hypothesis tested	AU test P-value
eukaryotes unconstrained	
Monophyly of Clade 1, 2 and 3 eukaryotes; other eukaryotes unconstrained	0.582
Monophyly of Clade 2 and Clade 3 eukaryotes; other eukaryotes unconstrained	0.557
Monophyly of Clade 2 and Clade 4 eukaryotes; other eukaryotes unconstrained	2e-04**
Monophyly of Clade 1, 2 and 4 eukaryotes; other eukaryotes unconstrained	2e-63**
Monophyly of Clade 1 eukaryotes and α-proteobacteria; other eukaryotes unconstrained	0.537
Monophyly of Clade 2 eukaryotes and α-proteobacteria; other eukaryotes unconstrained	0.547
Flavodiiron protein	
ML tree	1.000
Eukaryote monophyly	0.004**

Figure 4.3 Unrooted Maximum Likelihood (ML) tree of PFO sequences. **(Following page)** Phylogenetic analyses were performed on 212 sequences and 958 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and α-proteobacteria magenta. Eukaryotic PFO and PNO sequences are indicated.

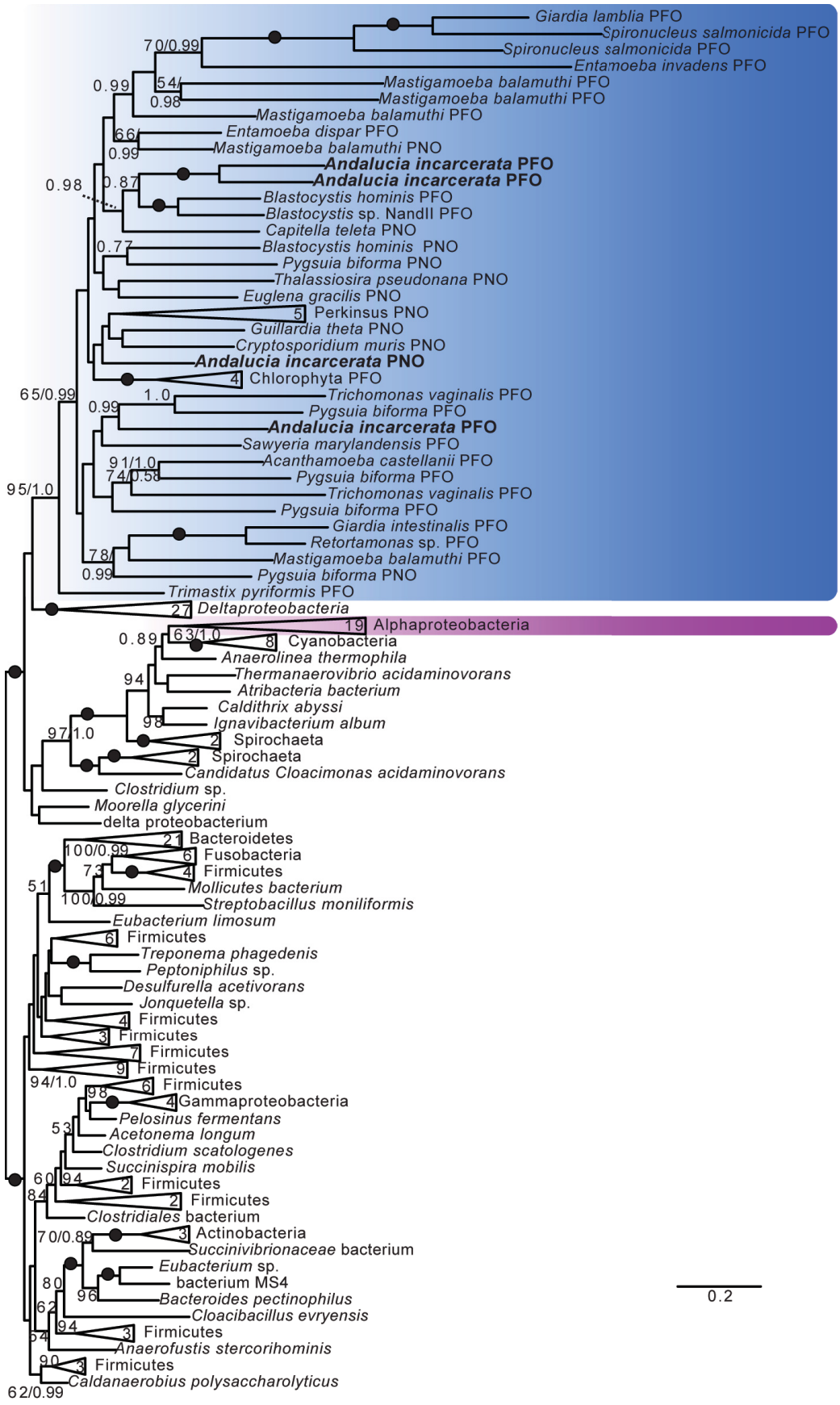

Giardia lamblia PFO
Spironucleus salmonicida PFO
Spironucleus salmonicida PFO
Entamoeba invadens PFO
Mastigamoeba balamuthi PFO
Mastigamoeba balamuthi PFO
Mastigamoeba balamuthi PFO
Entamoeba dispar PFO
Mastigamoeba balamuthi PNO
Andalucia incarcerata PFO
Andalucia incarcerata PFO
Blastocystis hominis PFO
Blastocystis sp. NandII PFO
Capitella teleta PNO
Blastocystis hominis PNO
Pygsuia biforma PNO
Thalassiosira pseudonana PNO
Euglena gracilis PNO
Perkinsus PNO
Guillardia theta PNO
Cryptosporidium muris PNO
Andalucia incarcerata PNO
Chlorophyta PFO
Trichomonas vaginalis PFO
Pygsuia biforma PFO
Andalucia incarcerata PFO
Sawyeria marylandensis PFO
Acanthamoeba castellanii PFO
Pygsuia biforma PFO
Trichomonas vaginalis PFO
Pygsuia biforma PFO
Giardia intestinalis PFO
Retortamonas sp. PFO
Mastigamoeba balamuthi PFO
Pygsuia biforma PNO
Trimastix pyriformis PFO
Deltaproteobacteria
Alphaproteobacteria
Cyanobacteria
Anaerolinea thermophila
Thermanaerovibrio acidaminovorans
Atribacteria bacterium
Caldithrix abyssi
Ignavibacterium album
Spirochaeta
Spirochaeta
Candidatus Cloacimonas acidaminovorans
Clostridium sp.
Moorella glycerini
delta proteobacterium
Bacteroidetes
Fusobacteria
Firmicutes
Mollicutes bacterium
Streptobacillus moniliformis
Eubacterium limosum
Firmicutes
Treponema phagedenis
Peptoniphilus sp.
Desulfurella acetivorans
Jonquetella sp.
Firmicutes
Firmicutes
Firmicutes
Firmicutes
Firmicutes
Gammaproteobacteria
Pelosinus fermentans
Acetonema longum
Clostridium scatologenes
Succinispira mobilis
Firmicutes
Firmicutes
Clostridiales bacterium
Actinobacteria
Succinivibrionaceae bacterium
Eubacterium sp.
bacterium MS4
Bacteroides pectinophilus
Cloacibacillus evryensis
Firmicutes
Anaerofustis stercorihominis
Firmicutes
Caldanaerobius polysaccharolyticus
0.2

Figure 4.4 Unrooted Maximum Likelihood (ML) tree of non-periplasmic-like [FeFe]-hydrogenase sequences. **(Following page)** Phylogenetic analyses were performed on 238 sequences and 346 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and α-proteobacteria magenta. [FeFe]-hydrogenases with C-terminal CysJ or flavodoxin (FLD) domains are indicated.

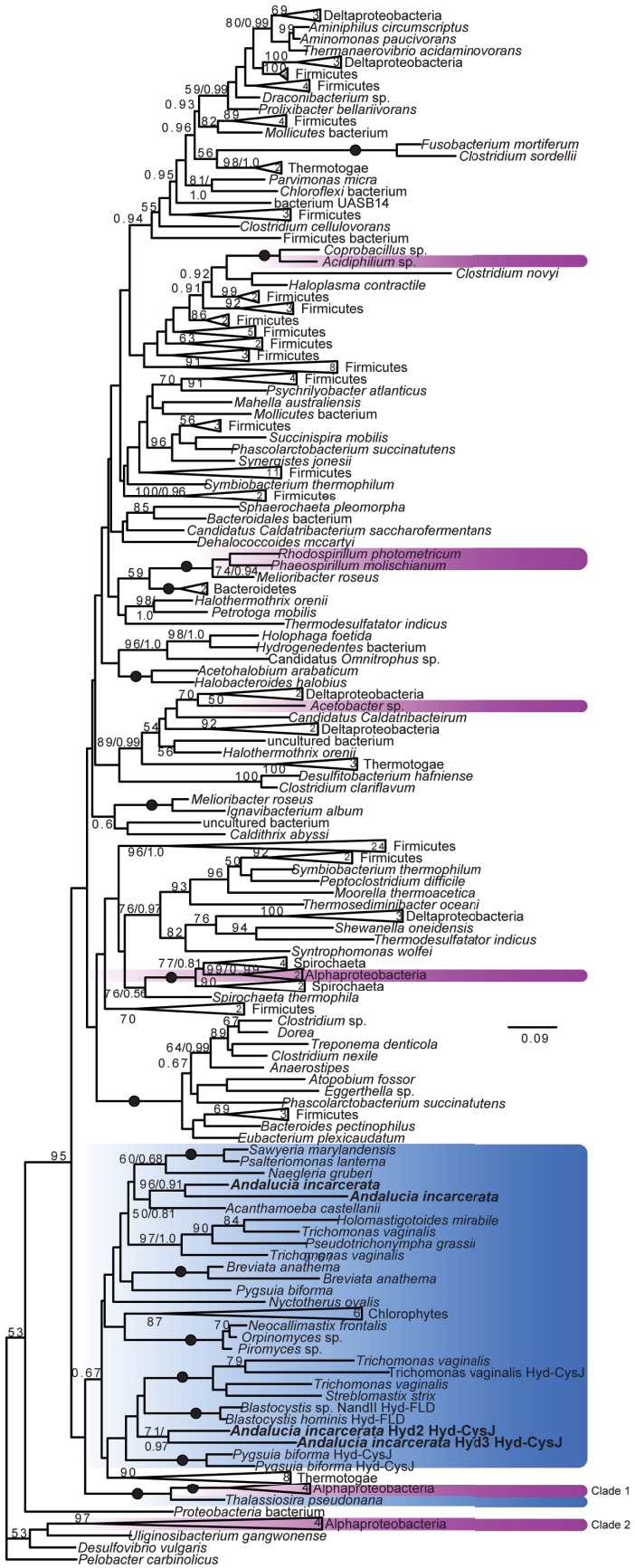
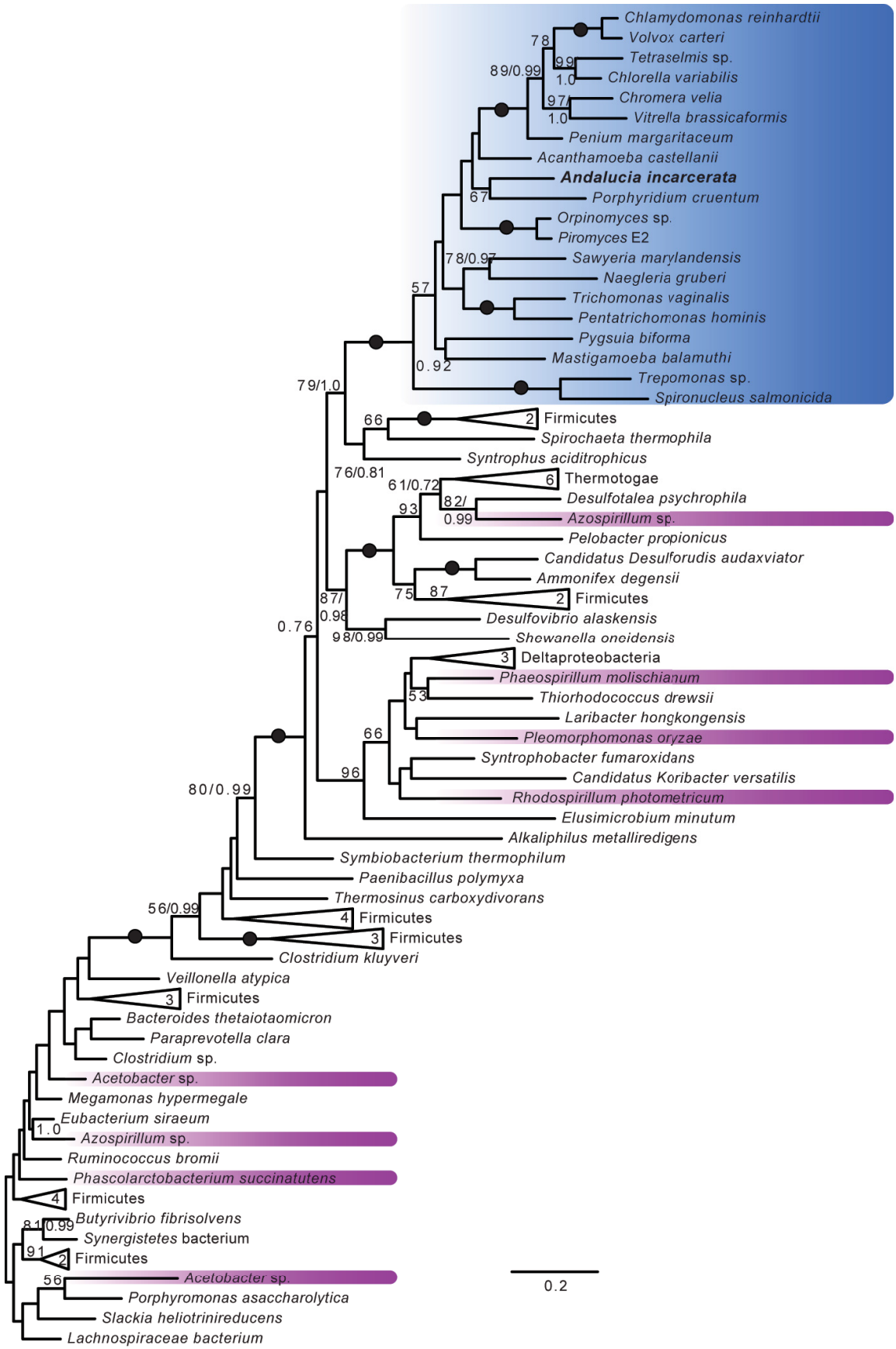

Deltaproteobacteria
Aminiphilus circumscriptus
Aminomonas paucivorans
Thermanaerovibrio acidaminovorans
Deltaproteobacteria
Firmicutes
Firmicutes
Draconibacterium sp.
Prolixibacter bellariivorans
Firmicutes
Mollicutes bacterium
Fusobacterium mortiferum
Clostridium sordellii
Thermotogae
Parvimonas micra
Chloroflexi bacterium
bacterium UASB14
Firmicutes
Clostridium cellulovorans
Firmicutes bacterium
Coprobacillus sp.
Acidiphilium sp.
Clostridium novyi
Haloplasma contractile
Firmicutes
Firmicutes
Firmicutes
Firmicutes
Firmicutes
Firmicutes
Firmicutes
Psychrilyobacter atlanticus
Mahella australiensis
Mollicutes bacterium
Firmicutes
Succinispira mobilis
Phascolarctobacterium succinatutens
Synergistes jonesii
Firmicutes
Symbiobacterium thermophilum
Firmicutes
Sphaerochaeta pleomorpha
Bacteroidales bacterium
Candidatus Caldatribacterium saccharofermentans
Dehalococcoides mccartyi
Rhodospirillum photometricum
Phaeospirillum molischianum
Melioribacter roseus
Bacteroidetes
Halothermothrix orenii
Petrotoga mobilis
Thermodesulfatator indicus
Holophaga foetida
Hydrogenedentes bacterium
Candidatus Omnitrophus sp.
Acetohalobium arabaticum
Halobacteroides halobius
Deltaproteobacteria
Acetobacter sp.
Candidatus Caldatribacteirum
Deltaproteobacteria
uncultured bacterium
Halothermothrix orenii
Thermotogae
Desulfitobacterium hafniense
Clostridium clariflavum
Melioribacter roseus
Ignavibacterium album
uncultured bacterium
Caldithrix abyssi
Firmicutes
Firmicutes
Symbiobacterium thermophilum
Peptoclostridium difficile
Moorella thermoacetica
Thermosediminibacter oceani
Deltaproteobacteria
Shewanella oneidensis
Thermodesulfatator indicus
Syntrophomonas wolfei
Spirochaeta
Alphaproteobacteria
Spirochaeta
Spirochaeta thermophila
Firmicutes
Clostridium sp.
Dorea
Treponema denticola
Clostridium nexile
Anaerostipes
Atopobium fossor
Eggerthella sp.
Phascolarctobacterium succinatutens
Firmicutes
Bacteroides pectinophilus
Eubacterium plexicaudatum
0.09
Sawyeria marylandensis
Psalteriomonas lanterna
Naegleria gruberi
Andalucia incarcerata
Andalucia incarcerata
Acanthamoeba castellanii
Holomastigotoides mirabile
Trichomonas vaginalis
Pseudotrichonympha grassii
Trichomonas vaginalis
Breviata anathema
Breviata anathema
Pygsuia biforma
Nyctotherus ovalis
Chlorophytes
Neocallimastix frontalis
Orpinomyces sp.
Piromyces sp.
Trichomonas vaginalis
Trichomonas vaginalis Hyd-CysJ
Trichomonas vaginalis
Streblomastix strix
Blastocystis sp. NandII Hyd-FLD
Blastocystis hominis Hyd-FLD
Andalucia incarcerata Hyd2 Hyd-CysJ
Andalucia incarcerata Hyd3 Hyd-CysJ
Pygsuia biforma Hyd-CysJ
Pygsuia biforma Hyd-CysJ
Thermotogae
Alphaproteobacteria
Clade 1
Thalassiosira pseudonana
Proteobacteria bacterium
Alphaproteobacteria
Clade 2
Uliginosibacterium gangwonense
Desulfovibrio vulgaris
Pelobacter carbinolicus

Figure 4.5 Unrooted Maximum Likelihood (ML) tree of HydG sequences. **(Following page)** Phylogenetic analyses were performed on 87 sequences and 401 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and α-proteobacteria magenta.


Chlamydomonas reinhardtii
Volvox carteri
Tetraselmis sp.
Chlorella variabilis
Chromera velia
Vitrella brassicaformis
Penium margaritaceum
Acanthamoeba castellanii
Andalucia incarcerata
Porphyridium cruentum
Orpinomyces sp.
Piromyces E2
Sawyeria marylandensis
Naegleria gruberi
Trichomonas vaginalis
Pentatrichomonas hominis
Pygsuia biforma
Mastigamoeba balamuthi
Trepomonas sp.
Spironucleus salmonicida
2 Firmicutes
Spirochaeta thermophila
Syntrophus aciditrophicus
6 Thermotogae
Desulfotalea psychrophila
Azospirillum sp.
Pelobacter propionicus
Candidatus Desulforudis audaxviator
Ammonifex degensii
2 Firmicutes
Desulfovibrio alaskensis
Shewanella oneidensis
3 Deltaproteobacteria
Phaeospirillum molischianum
Thiorhodococcus drewsii
Laribacter hongkongensis
Pleomorphomonas oryzae
Syntrophobacter fumaroxidans
Candidatus Koribacter versatilis
Rhodospirillum photometricum
Elusimicrobium minutum
Alkaliphilus metalliredigens
Symbiobacterium thermophilum
Paenibacillus polymyxa
Thermosinus carboxydivorans
4 Firmicutes
3 Firmicutes
Clostridium kluyveri
Veillonella atypica
3 Firmicutes
Bacteroides thetaiotaomicron
Paraprevotella clara
Clostridium sp.
Acetobacter sp.
Megamonas hypermegale
Eubacterium siraeum
Azospirillum sp.
Ruminococcus bromii
Phascolarctobacterium succinatutens
4 Firmicutes
Butyrivibrio fibrisolvens
Synergistetes bacterium
2 Firmicutes
Acetobacter sp.
Porphyromonas asaccharolytica
Slackia heliotrinireducens
Lachnospiraceae bacterium
0.2

The overall lack of resolution in these trees, and the lack of a clear prokaryotic sister group to eukaryotes that could point to these enzymes being either of mitochondrial origin, or laterally transferred from a specific prokaryotic group, is unfortunately typical of anaerobic energy generation enzyme phylogenies (Horner, Foster, Embley 2000; Horner et al. 2002; Hug, Stechmann, Roger 2010; Stairs et al. 2014). The prevalence of these enzymes among eukaryotes, and eukaryotic monophyly in some phylogenies, have been cited as evidence for their presence in the original mitochondrial endosymbiont (Horner, Hirt, Embley 1999; Horner, Foster, Embley 2000; Horner et al. 2002; Müller et al. 2012). The suggestion of non-periplasmic [FeFe]-hydrogenase eukaryotic monophyly, and the fact that α-proteobacteria + eukaryote monophyly cannot be rejected for non-periplasmic [FeFe]-hydrogenases, could be argued to be further evidence for this hypothesis. While these findings should not be ignored, they should however be treated with caution. If the monophyly of eukaryotes in anaerobic energy generation enzyme phylogenies is indeed a result of their presence in the protomitochondrial endosymbiont, then this should be balanced against the relative scarcity of [FeFe]-hydrogenases in α-proteobacteria; the extreme scarcity of the three [FeFe]-hydrogenase maturases in the same group; and the rejection of an α-proteobacterial affinity for PFO. Ancestral presence of modern anaerobic energy generation enzymes in the protomitochondrial endosymbiont requires massive loss from, or replacement of, these enzymes in α-proteobacteria; or lateral transfer of at least some of them into the lineage giving rise to the endosymbiont shortly before the endosymbiotic event; and certainly, large-scale subsequent loss in diverse eukaryote lineages. A more parsimonious

explanation might be lateral transfer between anaerobic eukaryotes, which presents an attractive explanation for an increasingly common pattern of genes shared only between anaerobic eukaryotes (Andersson et al. 2003; Andersson et al. 2006; Stairs, Roger, Hampl 2011; Tsaousis et al. 2012b; Stairs et al. 2014). It should be noted that the latter hypothesis does not exclude the possibility of anaerobic energy generation enzymes originally being present in the protomitochondrial endosymbiont, but suggests that if they were originally present, they have since been lost or replaced in extant eukaryotes (Tsaousis et al. 2012a).

In either case, these analyses highlight the importance of two factors. The first is the importance of gathering more data, shown by the increase in the representation of both eukaryotic and α-proteobacterial taxa in our trees relative to even recent studies (Hug, Stechmann, Roger 2010; Leger et al. 2013; Stairs et al. 2014); further data may completely alter the taxonomic representation in, and topologies of, these phylogenies. The second is that the periplasmic hydrogenases (and their eukaryotic homologs), while they may be functionally equivalent to the non-periplasmic hydrogenases, are paralogs, and that caution should be used when grouping them together in analyses, as has been done in the past (Hug, Stechmann, Roger 2010). AU tests performed on groupings of α-proteobacteria and eukaryotes, for instance, will likely generate spurious results if periplasmic and non-periplasmic hydrogenases are treated as a single group, as these enzymes likely have quite different origins in eukaryotes.

Iron-Sulfur Cluster Assembly

Iron-sulfur (Fe-S) cluster biosynthesis is known to be the only truly essential function of yeast mitochondria (Lill et al. 1999). The mitochondrial ISC iron sulfur cluster assembly pathway is a crucial and conserved function of most MROs described to date (LaGier et al. 2003); it has been suggested that the need to provide an environment for this process explains the retention of the most reduced MROs (Embley et al. 2003b; Tovar et al. 2003). Nevertheless, several examples of MROs are known in which the mitochondrial ISC system has been replaced by bacterial NIF (*Entamoeba histolytica* (Ali et al. 2004; van der Giezen, Cox, Tovar 2004; Maralikova et al. 2010), *Mastigamoeba balamuthi* (Gill et al. 2007; Nyvltova et al. 2013)) or SUF (*Pygsuia biforma* (Stairs et al. 2014)) systems.

The main components of the iron sulfur cluster assembly pathway were identified in the EST survey of *A. incarcerata* (Fig. 4.1, brown). Most notable are IscS, a cysteine desulfurase that generates sulfur destined for cluster assembly and the scaffold proteins IscU, IscA1 and IscA2 (Lill et al. 2012). Accessory proteins found include the mitochondrial transporter Atm1, as well as the iron donor frataxin, and ferredoxin Yah1, which maintain a redox balance, completing the complement of Fe-S assembly proteins (for the complete set of proteins found in this pathway, see Fig. 4.1). We confirmed the localization of this typical mitochondrial pathway to the putative MROs using immunogold labeling (Fig. 4.2). We used a heterologous polyclonal primary antibody raised in rabbit against *Giardia* IscS (Tovar et al. 2003). We probed labeled cells with a

secondary anti-rabbit antibody conjugated to gold particles, and measured the distribution of these particles in *Andalucia incarcerata* cells. As expected, we found a greater (specifically, sixteenfold higher) density of gold particles in the MROs compared with the cytosol; consistent with the presence of IscS homologs in bacteria, we also observed cross-reaction of the antibody with engulfed bacteria.

Protein Transport and Folding

Major components of the complexes involved in translocation across the outer and inner mitochondrial membranes, and insertion into the outer mitochondrial membrane were represented in the transcriptome data. Many accessory subunits of these complexes that are present in *Saccharomyces cerevisiae* were not recovered, and most of the apparently missing components are also absent from *T. vaginalis* (Schneider et al. 2011).

Notably absent from our dataset are accessory components of the Translocase of the Outer Membrane (TOM) and the Translocase of the Inner Membrane 22 (Tim22) complexes and the Inner Membrane Proteases (Imp1/2). Less surprisingly, the mitochondrial oxidase assembly protein Oxa1 is also absent; this protein mediates the membrane insertion of mitochondrially-encoded Cox1p and Cox3p (Hell, Neupert, Stuart 2001), neither of which are found in *A. incarcerata*. All of these import proteins are also absent from *T. vaginalis*. *A. incarcerata* differs from *T. vaginalis* in that it possesses homologs of the intermembrane space chaperone proteins known as Tiny Tims – Tim8, Tim9,

Tim10 and Tim13 –, as well as the mitochondrial intermembrane space import and assembly (MIA) proteins, Mia40 and Erv1.

Many of the matrix proteins involved in folding the imported proteins are present (Fig. 4.1, yellow). They include mtHsp70 and Cpn60, which are typical mitochondrial markers involved in protein transport and refolding, and both the alpha- and beta-subunits of the mitochondrial processing peptidase (MPP), the non-catalytic subunit of an enzyme that cleaves N-terminal targeting peptides after protein import into the organellar matrix.

As the majority of the proteins involved in protein import are highly divergent between taxa, *A. incarcerata* homologs of some proteins may simply not have been retrieved by our methods. However, mitochondrial protein import has been poorly studied in eukaryotes other than yeast and mammals, and recent surveys that include data from more diverse taxa (Eckers et al. 2012; Heinz, Lithgow 2013; Gawryluk et al. 2014; Stairs et al. 2014) suggest that the apparently 'reduced' complement found in *A. incarcerata* may in fact be more representative of eukaryotes overall. Despite uncertainly arising from the incomplete nature of transcriptome data, particularly in *Sawyeria* (Barbera et al. 2010) and *Psalteriomonas* (de Graaf et al. 2009), a trend is apparent between the more reduced protein import machinery of the MROs of parasites such as diplomonads, *Entamoeba* and microsporidia (Dagley et al. 2009; Waller et al. 2009; Dolezal et al. 2010; Jedelsky et al. 2011; Schneider et al. 2011; Jerlstrom-Hultqvist et al. 2013), and the more complex protein import machinery of typical mitochondria (Table 4.2). As with the free-living *Pygsuia (Stairs et al. 2014)*, the

protein import machinery of *A. incarcerata* more closely resembles that of typical mitochondria, suggesting that a reduction in MRO protein import apparatus reflects a parasitic lifestyle, rather than being a necessary accompaniment to the derivation of MROs from aerobic mitochondria.

The only other jakobid in which the mitochondrial protein import apparatus has been studied in some detail is *Reclinomonas americana*. This organism is notable in that, in addition to components of the canonical mitochondrial protein import pathway, it includes subunits of the bacterial SecY and twin-arginine transport (TAT) complexes, which are involved in protein insertion into and secretion across the inner membrane (Tong et al. 2011). Subsequently, these proteins have been found to be mitochondrially encoded in other jakobids, including *Andalucia godoyi* (Burger et al. 2013), and in the discobid *Tsukubamonas globosa* (Kamikawa et al. 2014). No nuclear-encoded homologs of these components have so far been found in eukaryotes, and it appears that they were lost from *A. incarcerata* along with the mitochondrial genome.

Table 4.2 Amino acid metabolism in MROs. 1 Source: (Jedelsky et al. 2011) 2 Source: (Jerlstrom-Hultqvist et al. 2013) 3 Source: (Schneider et al. 2011) 4 Source: (Stairs et al. 2014) +: Predicted to be present in the MRO based on proteomic data or bioinformatic predictions. +? Incomplete sequence at the N-terminus, making localization prediction uncertain. Orange, FeS cluster assembly; Bright yellow, glycine cleavage system; pale yellow, glycine, serine and threonine metabolism; pink, alanine metabolism; brown, tyrosine metabolism; pale blue, lysine and tryptophan metabolism; green, arginine metabolism; purple, glutamate and proline metabolism; dark blue, leucine, isoleucine and valine (branched-chain amino acid) metabolism

	Giardia[1]	*Spironucleus*[2]	*Trichomonas*[3]	*Pygsuia* [4]	*Andalucia*
Iron sulfur cluster assembly (for comparison)	ISC	ISC	ISC	SUF	ISC
GCS H protein		?	+	+	+
GCS L protein				+	+
GCS P protein				+	+
GCS T protein					
Serine hydroxymethyl-transferase		+	+	+	+
Alanine-glyoxylate aminotransferase					+
Glyoxylate reductase					
Glycerate kinase					
Phosphoserine aminotransferase			+		
Serine/threonine dehydratase					
L-Threonine dehydrogenase				+	+?
Glycine-C-acetyltransferase				+	+?
Threonine synthase			+		
Alanine aminotransferase			+?	+	+?
Tyrosine aminotransferase			+		+
2-ketoglutarate DH complex					+
Dihydro-lipoamide S-succinyltransferase					+
Tryptophanase				+	
Methionine gamma-lyase			+		
S-adenosyl-methionine					+

	Giardia[1]	*Spironucleus*[2]	*Trichomonas*[3]	*Pygsuia* [4]	*Andalucia*
synthetase					
Aspartate aminotransferase			+	+	+
Arginine deiminase			+	+	
GABA amino-transferase					+
Succinate semialdehyde dehydrogenase					+
Glutamate dehydrogenase			+		+
Branched-chain amino acid aminotransferase				+	+
Br.ch. Amino acid a-ketoacid dehydrogenase				+	+
Dihydrolipoyl transacylase			+		+
Isovaleryl-CoA dehydrogenase					
Butyryl-CoA dehydrogenase					
Short/branched-chain Acyl-CoA dehydrogenase					+
3-methylcrotonyl-CoA carboxylase					
Enoyl-Co hydratase					+?
3-hydroxy-isobutyryl-CoA hydrolase					+
3-hydroxyacyl-CoA dehydrogenase					
acetyl-CoA C-acyltransferase					+
Hydroxymethyl-glutaryl-CoA lyase					+?
Propionyl-coA carboxylase				+	+
Methylmalonate semialdehyde dehydrogenase				+	+
Methylmalonyl-CoA epimerase				+	+

Amino Acid Metabolism

Genes encoding all components of the glycine cleavage system were identified from the transcripts, including the L, P, T and H proteins and serine hydroxymethyl transferase (SHMT) (Figure 4.1, blue; Table 4.3). Targeting peptides have been identified for all of these proteins, providing evidence that this pathway is active within the organellar compartment. Whereas only partial glycine cleavage systems are present in the MROs of the parasites *Trichomonas vaginalis* (Schneider et al. 2011) and *Spironucleus salmonicida* (Jerlstrom-Hultqvist et al. 2013), a complete glycine cleavage system is predicted to function in the MROs of *Trimastix pyriformis* (Zubacova et al. 2013), *Mastigamoeba balamuthi* (Nyvltova et al. 2015) and *Pygsuia biforma* (Stairs et al. 2014). The presence of a glycine cleavage system may therefore be the rule, rather than the exception, in the MROs of free-living anaerobes.

Table 4.3 Mitochondrial protein import proteins in various taxa. Dark blue: free-living mitochondrion; pale blue: free-living MRO; dark pink: parasite, mitochondrion; light pink, parasite, MRO; purple: facultative parasite, mitochondrion. Data from (Neupert 1997; Hoogenraad, Ward, Ryan 2002; Henriquez et al. 2005; Regoes et al. 2005; Williams, Keeling 2005; Dolezal et al. 2006; Atteia et al. 2009; Dagley et al. 2009; Waller et al. 2009; Barbera et al. 2010; Dolezal et al. 2010; Jedelsky et al. 2011; Liu et al. 2011; Schneider et al. 2011; Eckers et al. 2012; Heinz, Lithgow 2013; Jerlstrom-Hultqvist et al. 2013; Zubacova et al. 2013; Gawryluk et al. 2014; Murcha et al. 2014a; Murcha et al. 2014b; Stairs et al. 2014; Murcha et al. 2015; Wojtkowska et al. 2015), this study.[1] Transcriptome data, incomplete; particularly in the cases of *Sawyeria marylandensis* and *Trimastix pyriformis* [2] pATOM36 [3] Tom9 [4] pATOM [5] Metaxin [6] Tim8/13 [7] Tim17/22/23

	Homo sapiens	*Saccharomyces cerevisiae*	*Encephalitozoon cuniculi*	*Antonospora locustae*	*Pygsuia biforma*[1]	*Dictyostelium discoideum*	*Entamoeba histolytica*	*Acanthamoeba castellanii*	*Arabidopsis thaliana*	*Chlamydomonas reinhardtii*	*Tetrahymena thermophila*	*Cryptosporidium parvum*	*Trypanosoma brucei*	*Andalucia incarcerata*[1]	*Sawyeria marylandensis*[1]	*Trimastix pyriformis*[1]	*Trichomonas vaginalis*	*Spironucleus salmonicida*	*Giardia lamblia*	*Rickettsia prowazekii*	*Escherichia coli*
Tom20	+	+	-	-	?	+	-	+	+	+	-	-	+ 2	?	?	?	-	-	-	-	-
Tom70	+	+	+	+	?	+	-	+	-	-	-	-	-	?	?	?	-	-	-	-	-
Tom22	+	+	-	-	?	+	-	+	+ 3	+ 3	+	-	-	?	?	?	-	-	-	-	-
Tom40	+	+	+	+	+	+	+	+	+	+	+	+	+ 4	+	?	+	+	+	+	-	-
Tom5	+	+	-	-	?	-	-	-	+	+	-	-	-	?	?	?	-	-	-	-	-
Tom6	+	+	-	-	?	-	-	-	+	+	-	-	-	?	?	?	-	-	-	-	-
Tom7	+	+	-	-	?	+	-	+	+	+	+	-	-	?	?	?	-	-	-	-	-
Sam35	+ 5	+	-	-	?	-	-	+	-	+	-	-	+ 5	?	?	?	-	-	-	-	-
Sam37	+ 5	+	-	-	+ 5	-	-	+	+ 5	-	-	-	+ 5	+ 5	?	?	-	-	-	-	-
Sam50	+	+	+	+	+	+	+	+	+	+	-	-	+	+	?	+	+	-	-	+	+
Mdm10	-	+	-	-	?	-	-	+	-	-	-	-	-	?	?	?	-	-	-	-	-
Mdm12	-	+	-	-	?	-	-	+	-	-	-	-	-	+	?	?	-	-	-	-	-
Mmm1	-	+	-	-	?	-	-	+	-	-	-	-	-	?	?	?	-	-	-	-	-
Mim1	-	+	-	-	?	-	-	-	-	-	-	-	-	?	?	?	-	-	-	-	-

	Homo sapiens	*Saccharomyces cerevisiae*	*Encephalitozoon cuniculi*	*Antonospora locustae*	*Pygsuia biforma*[1]	*Dictyostelium discoideum*	*Entamoeba histolytica*	*Acanthamoeba castellanii*	*Arabidopsis thaliana*	*Chlamydomonas reinhardtii*	*Tetrahymena thermophila*	*Cryptosporidium parvum*	*Trypanosoma brucei*	*Andalucia incarcerata*[1]	*Sawyeria marylandensis*[1]	*Trimastix pyriformis*[1]	*Trichomonas vaginalis*	*Spironucleus salmonicida*	*Giardia lamblia*	*Rickettsia prowazekii*	*Escherichia coli*
Tim8	+	+	-	-	+	-	-	+	+	+	+	-	+ 6	+	?	?	-	-	-	-	-
Tim9	+	+	-	-	+	-	-	+	+	+	+	-	+	+	?	?	-	-	-	-	-
Tim10	+	+	-	-	+	+	-	+	+	+	+	-	+	+	?	?	-	-	-	-	-
Tim13	+	+	-	-	+	+	-	+	+	+	-	-	-	+	?	?	-	-	-	-	-
Mia40	+	+	-	-	+	-	-	-	+	-	-	-	-	+	?	?	-	-	-	-	-
Erv1	+	+	+	-	?	+	-	+	+	+	+	+	+	+	?	?	-	-	-	-	-
Tim54	-	+	-	-	?	-	-	-	-	-	-	-	-	?	?	?	-	-	-	-	-
Tim22	+	+	+	+ 7	+	+	-	+	+	+	+	-	-	+	?	?	+ 7	-	-	-	-
Tim18	-	+	-	-	?	-	-	-	-	-	-	-	-	?	?	?	-	-	-	-	-
Tim12	-	+	-	-	?	-	-	-	-	-	-	-	-	?	?	?	-	-	-	-	-
Tim50	+	+	-	-	+	+	-	+	+	+	+	-	+	?	?	?	-	-	-	-	-
Tim21	+	+	-	-	?	+	-	+	+	+	-	-	-	?	?	?	-	-	-	-	-
Tim23	+	+	-	-	+	+	-	+	+	+	+	+	-	+	+	?	-	-	-	-	-
Tim17	+	+	-	+	+	+	-	+	+	+	+	+	+	+	?	+	-	-	-	-	-
mtHsp70	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	?	+	+	+	+	+
Mge1	+	+	-	-	+	+	-	-	+	+	+	+	+	+	?	?	+	+	+	+	+
Pam16	+	+	-	-	+	+	-	+	+	+	+	-	+	+	?	?	+	-	+	-	-
Pam17	-	+	-	-	?	-	-	-	-	-	-	-	-	?	?	?	-	-	-	-	-
Pam18	+	+	+	+	+	+	-	+	+	+	+	+	+	+	?	+	+	+	+	+	+
Tim44	+	+	-	-	+	+	-	+	+	+	+	+	+	?	+	?	+	-	-	+	-
Imp1	+	+	-	-	+	-	-	+	+	+	+	+	-	?	?	?	-	-	-	+	+

	Homo sapiens	*Saccharomyces cerevisiae*	*Encephalitozoon cuniculi*	*Antonospora locustae*	*Pygsuia biforma*[1]	*Dictyostelium discoideum*	*Entamoeba histolytica*	*Acanthamoeba castellanii*	*Arabidopsis thaliana*	*Chlamydomonas reinhardtii*	*Tetrahymena thermophila*	*Cryptosporidium parvum*	*Trypanosoma brucei*	*Andalucia incarcerata*[1]	*Sawyeria marylandensis*[1]	*Trimastix pyriformis*[1]	*Trichomonas vaginalis*	*Spironucleus salmonicida*	*Giardia lamblia*	*Rickettsia prowazekii*	*Escherichia coli*
Imp2	+	+	-	+	?	-	-	+	+	+	-	-	+	?	?	?	-	-	-	-	-
Oxa1	+	+	-	-	?	+	-	+	+	+	+	-	+	?	?	?	-	-	-	+	+
MPPa	+	+	-	-	+	+	-	+	+	+	+	+	+	+	?	+	+	-	-	-	-
MPPb	+	+	-	-	+	+	-	+	+	+	+	+	+	+	?	?	+	-	+	-	-

All of the key enzymes in the initial stages of branched-chain amino acid degradation, as well as most downstream enzymes for the further breakdown of leucine, isoleucine, and valine were identified in *Andalucia incarcerata* (Fig. 4.1, blue). While enzymes carrying out the early steps of branched-chain amino acid degradation have been described in the *T. vaginalis* genome (Carlton et al. 2007), this pathway is not hypothesized to be present in the hydrogenosome (Carlton et al. 2007; Schneider et al. 2011), despite mitochondrial localization being typical in aerobic organisms.

In addition to these proteins, we identified enzymes involved in alanine, arginine, glutamate, lysine, tryptophan, methionine, and tyrosine metabolism (Table 4.3). These are largely absent from the reduced amino acid metabolism complement of *Trichomonas* MROs (Schneider et al. 2011) (and completely absent from the much more reduced proteome of *Giardia* MROs (Jedelsky et al.

2011)). A number of these enzymes, including branched-chain amino acid degradation enzymes, also appear to be expressed in the MROs of *Pygsuia* (Stairs et al. 2014). This is consistent with the hypothesis that the MROs of free-living organisms may retain more mitochondrial metabolic functions than those of parasites; nevertheless, *A. incarcerata* MROs appear to possess a larger complement of amino acid metabolic enzymes than *Pygsuia* MROs, notably those involved in valine, isoleucine and leucine degradation, and in arginine and glutamate metabolism (Table 4.3).

Other Functions and Pathways

As previously reported, the *A. incarcerata* transcriptome encodes two homologs of the bacterial and mitochondrial division protein FtsZ, and homologs of three proteins, MinC, MinD and MinE, that, in bacteria, control the placement of FtsZ ((Leger et al. 2015), Chapter 3).

The phospholipid cardiolipin is a key component of both prokaryotic and mitochondrial inner membranes; in aerobic eukaryotes, it is synthesized in mitochondria. Cardiolipin synthesis proteins were until recently believed to be absent from MRO-bearing protists (Tian, Feng, Wen 2012), but a CDP-diacylglycerol-3-a bioinformatic survey of the *Pygsuia* MRO proteome uncovered the first evidence of a complete cardiolipin biosynthesis pathway in an MRO (Stairs et al. 2014). Our survey uncovered two enzymes in this pathway, a eukaryotic transferase-type cardiolipin synthesis and the CDP-diacylglycerol synthase Tam41 (Tamura et al. 2013). We found no evidence of phosphatidylglycerol phosphate (PGP) synthase or PGP phosphatase, which are

involved in phosphatidylglycerol formation from CDP-diacylglycerol. In the absence of these enzymes, phosphatidylglycerol might be taken up from bacterial food sources; alternatively, homologs of these proteins present in *A. incarcerata* might simply not have been recovered in our transcriptome.

Recently, a novel, fused SufCB enzyme was identified in *Blastocystis* sp. - the first evidence for a Suf system enzyme in a eukaryote. Despite the fact that the IscS system is present in the MROs of *Blastocystis* sp., this enzyme was also shown to be expressed in these organelles (Tsaousis et al. 2012b). Homologs of this enzyme were subsequently found to be encoded in the transcriptome of an anaerobic breviate, *Pygsuia biforma*, which possesses both an MRO-targeted and a cytosolic copy (Tielens et al. 2010). Based on phylogenetic studies, the enzyme was hypothesized to have been laterally acquired from an archaeon related to the Methanomicrobiales, and subsequently transferred laterally from one eukaryote to the other (Tsaousis et al. 2012b; Stairs et al. 2014). We recovered a sequence encoding a homolog of this enzyme from our transcriptome data. The *A. incarcerata* enzyme displays the same fusion of the prokaryotic SufC and SufB proteins as those found in *Blastocystis* sp. and *P. biforma*; however, the enzyme lacks a targeting peptide, suggesting that it is located in the cytosol (Appendix C, Table C.S1). While the genomic sequence of the enzyme does not include introns, it is also present in transcriptome data derived from a closely related taxon, provisionally named *A. trypanoides* (T. Panek and I. Cepicka, personal communication), suggesting that it is not a prokaryotic contaminant.

Like MROs of *Trichomonas* (Schneider et al. 2011) and *Pygsuia* (Stairs et al. 2014), but unlike those of *Giardia* (Brown, Upcroft, Upcroft 1995), *A. incarcerata* organelles have retained mitochondrial reactive oxygen stress response proteins (superoxide dismutase, thioredoxin, and thioredoxin peroxidase; Fig. 4.1). In addition, *A. incarcerata* possesses homologs of a flavodiiron protein commonly found in anaerobic prokaryotes, as well as some anaerobic protists (Andersson et al. 2003); previously described in *Giardia* and *Trichomonas* as having a high affinity for O_2, this enzyme has been hypothesized to be involved in oxygen detoxification in eukaryotes (Di Matteo et al. 2008; Smutna et al. 2009; Vicente et al. 2009; Vicente et al. 2012). *A. incarcerata* possesses two paralogs of this protein that differ in their predicted localization and phylogenetic affinities. The first is predicted to be cytosolic, and groups together with *Entamoeba*, *Mastigamoeba* and *Sawyeria* homologs. The second possesses a mitochondrial targeting peptide, and forms a clade with *Pygsuia*, *Breviata*, *Giardia*, *Trichomonas* and *Spironucleus*. Consistent with earlier results suggesting at least two eukaryotic acquisitions from prokaryotes (Andersson et al. 2003; Andersson et al. 2006), AU tests reject the hypothesis of eukaryote monophyly for this protein (P-val < 0.01) (Table 4.1; Appendix C, Fig. C.S7). Lateral transfer of the protein between eukaryotes likely followed the initial acquisitions; it is particularly striking that *A. incarcerata* has acquired paralogs from two distinct sources. Interestingly, the MRO-targeted *A. incarcerata* protein is a fusion protein, possessing a C-terminal ferredoxin domain. This arrangement, which was confirmed by PCR amplification, appears to be unique to *A. incarcerata*. Flavodiiron proteins typically accept electrons

from rubredoxins (which in turn accept electrons from NADH via a flavoprotein) (Di Matteo et al. 2008; Leger et al. 2015). Rubredoxins do not appear to be present in *A. incarcerata*'s MRO; the fused ferredoxin likely fulfils this role for the *A. incarcerata* organellar flavodiiron protein.

Conclusions and Significance

The rise of low-cost, high-coverage sequencing in recent years is making it possible to examine mitochondria and their related organelles on a much broader taxonomic and biochemical scale than was previously possible. It is now possible to investigate the true diversity and flexibility of these organelles, rather than focusing mainly on energy generating functions in a limited number of taxa. The examination of MROs in free-living anaerobes and microaerophiles is allowing us to tease apart adaptations to anoxia from byproducts of a parasitic lifestyle. The presence of a hydrogenosome-like organelle in a lineage closely related to organisms recognized for the unusual, primitive features of their mitochondria presents a particularly interesting opportunity. The complete absence of aerobic energy generation machinery and mitochondrial genome-encoded genes highlights the dramatic changes that can occur in the course of MRO derivation, and the provides proof of the independent origins of MROs in *A. incarcerata.*

A comparison of the MROs of *A. incarcerata* with those of other free-living and parasitic taxa reveals significant similarities as well as differences. Like the well-characterized MROs of *Trichomonas*, *A. incarcerata* MROs house an oxygen

scavenging system for maintaining an anaerobic environment, and a anaerobic ATP generation pathway involving PFO, [FeFe]-hydrogenase, [FeFe]-hydrogenase maturases, Complex I subunits and ASCT; the differences in ASCT subtypes between the two organisms suggest at least some degree of convergent adaptation. However, in its complement of typical mitochondrial enzymes, such as those involved in amino acid catabolism, cardiolipin synthesis and mitochondrial protein import, the MROs of *A. incarcerata* more closely resemble those of the much more distantly related *Pygsuia biforma*. These similarities may result from a shorter amount of time elapsed since the process of MRO emergence in these taxa relative to diplomonads, but more likely reflect their shared lifestyle as free-living anaerobes; free-living organisms may have less need of these catabolic pathways, as they can derive more metabolic end-products from the host. As a transcriptomic survey, this study is necessarily incomplete, and it is likely that further examination of *A. incarcerata*'s genome will clarify the role of this organelle within the cell, as well as its relationship to classical mitochondria. Such an in-depth examination will undoubtedly yield useful insights into the order and timing of protein and pathway loss in mitochondrion-related organelles. In coming years, it will be exciting to examine the mitochondrial proteomes of jakobids such as *Andalucia godoyi* in order to draw inferences about the origins of *A. incarcerata* MRO proteins, particularly those involved in anaerobic energy generation and oxygen detoxification.

Chapter 5: Conclusions

Overall Lifestyle and Mitochondrion-Related Organelles

Mitochondrion-related organelles were first described in trichomonads, and most of the early work on these and other MROs has focused on parasitic organisms (Rotte et al. 2001; LaGier et al. 2003; Nixon et al. 2003; Tovar et al. 2003; Ali et al. 2004; Dyall et al. 2004; Hrdý et al. 2004; Roberts et al. 2004; Chan et al. 2005; Henriquez et al. 2005; van der Giezen, Leon-Avila, Tovar 2005; Hrdý, Tachezy, Müller 2008). This has made it difficult to tease apart distinctive MRO properties that are adaptations to lower-oxygen conditions from those that were adaptations to a parasitic lifestyle.

In recent years, the advent of high throughput and lower cost sequencing methods has permitted the gathering of large-scale transcriptomic and/or genomic data from diverse eukaryotic microbes, including some free-living anaerobes. Although taxonomic sampling is still relatively limited, some general trends are apparent. Firstly, obligate parasites appear more likely to have highly reduced organelles, with no role in ATP generation (mitosomes or class 5 mitochondria; e.g., *Entamoeba*, *Giardia*, microsporidia, and *Mikrocytos*). Secondly, even the less-reduced MROs of other parasitic organisms appear to have a more streamlined complement of both metabolic genes and organellar protein import machinery (see Tables 4.2 and 4.3 respectively, Chapter 4). The reduced amino acid metabolism complement of parasite MROs may be

accounted for by their ability to derive metabolic intermediates from the host. Thirdly, the glycine cleavage system is retained in more MROs than any other amino acid metabolism enzymes, but it is restricted to MROs that also generate ATP. It is possible that in these organelles, the GCS has an important role in maintaining the redox balance in the MROs, by contributing to the NADH pool. This may be a particular advantage to MROs that produce ATP using Complex I or its 51kDa and 24kDa subunits, as these subunits use NADH as a substrate (Ryoma Kamikawa, personal communication).

Our understanding of MRO pathways will be greatly expanded as sequence data from more MRO-harbouring organisms become available, but it will also depend on broadening our understanding of aerobic mitochondria in a wider range of eukaryotes. The predicted MRO proteomes reconstructed *in silico* so far rely mainly on lists of pathways characterized in mitochondria of model organisms. As a result, even proteins that were likely present in LECA are difficult to detect if they are absent from model organisms. Chapter 3 illustrates this: the discovery of Min proteins in the *Andalucia incarcerata* proteome was fortuitous, and ultimately shed light of mitochondrial functions in a wide range of eukaryotes. Further study will hopefully elucidate not only the metabolic functions in MROs, but also their mechanisms of division, inheritance, cristae formation (where applicable), and whether they undergo fusion.

Meanwhile, the predicted MRO proteomes of *Trichomonas* (Schneider et al. 2011), *Giardia* (Jedelsky et al. 2011), *Blastocystis* (Tsaousis et al. 2012b) and *Pygsuia* (Stairs et al. 2014) exposed a set of proteins not generally found in

aerobic eukaryotes, but shared with other anaerobic eukaryotes. These include the oxygen-scavenging flavodiiron protein and SufCB (Chapter 4), but also proteins with poorly understood functions, such as a putative acyl-CoA synthetase that *Trichomonas* and *Giardia* also share with *Andalucia* (Appendix C, Table C.S1), and a number of proteins with unknown function (Jedelsky et al. 2011; Schneider et al. 2011). Further study of these proteins will likely uncover a wider complement of unusual enzymes shared among anaerobic eukaryotes and clarify their origins.

Origins of Anaerobic ATP Generation Metabolism

In Chapter 1, I discussed two main hypotheses for the origins of mitochondria and their anaerobic ATP-generating metabolism: the hydrogen hypothesis, and the LGT hypothesis. In the following section, I discuss the implications of the results presented in this thesis for these hypotheses.

Distribution Of Anaerobic ATP Generation Enzymes

The hydrogen hypothesis (see Chapter 1) predicts that: 1) enzymes associated with anaerobic ATP production will be found in a variety of modern eukaryotes with an obligately or facultatively anaerobic lifestyle and 2), this complement of enzymes should be more or less similar for all eukaryotes in which they are found. With respect to these predictions, my results have added to the wide range of modern eukaryotes now known to possess enzymes of anaerobic ATP production. Unfortunately, the LGT hypothesis is equally consistent with the presence of anaerobic ATP production in disparate eukaryote lineages. While the

patchy distribution of these enzymes within the eukaryote super-groups seems to match the LGT hypothesis better, taxonomic sampling is still relatively sparse, and will need to be increased to find out whether their distribution is truly patchy, as it currently appears.

The second prediction of the hydrogen hypothesis is that, because LECA possessed a particular collection of metabolic enzymes, modern anaerobic eukaryotes should each possess a subset of these enzymes (Müller et al. 2012). The LGT hypothesis, meanwhile, is also consistent with some commonalities in the ATP-generation complement of anaerobic eukaryotes. But, in contrast to the hydrogen hypothesis, it is consistent with a sparser distribution of any given enzyme within eukaryotes, and with unusual topologies of gene phylogenies that reflect lateral transfer rather than vertical inheritance.

From the table in Figure 1.2 (Chapter 1), it is possible to reconstruct the ATP-generation complement of LECA if the most straightforward version of the hydrogen hypothesis were correct (i.e., a version that did not invoke LGT in eukaryotes subsequent to LECA). It would have needed to include PDH, a full TCA cycle and electron transport chain. In addition, it would have possessed at least two types of [FeFe]-hydrogenase (the more common enzyme found in, e.g., *Trichomonas*, and periplasmic-type related H_2-evolving hydrogenases found in *Trimastix*, *Entamoeba* and *Mastigamoeba*; see Chapter 4); the hydrogenase maturases; PFO; PNO; PFL; ASCT1B; ASCT1C; and ACS. This is notinconceivable: even given the relatively small subset of organelles described so far, several organisms encode the majority of these proteins. But it does seem

unlikely: none are known to encode all of them, and the subset of these proteins that remains localized to MROs is smaller still. In some organisms these enzymes are instead localized in the cytosol, or in the plastids of some green algae. In light of this fact, and the number of losses that would have had to occur across the tree of eukaryotes to explain their presence in only a handful of distantly related organisms, lateral gene transfer between eukaryotes seems a more parsimonious explanation for the distribution of these enzymes.

Phylogenetic Evidence

Eukaryotes frequently form a single clade in phylogenies of anaerobic ATP generation enzymes; an exception is ASCT1B (Chapter 4). This is consistent with predictions made by the hydrogen hypothesis and the LGT hypothesis, and so this aspect of my results does not support either hypothesis over the other. Furthermore, the phylogenetic affinities of eukaryotes remain unclear. An α-proteobacterial origin for the main clade of [FeFe]-hydrogenases and ASCT1B could not be ruled out by my analyses (Chapter 4). Thus, it cannot be excluded that the [FeFe]-hydrogenase in anaerobic eukaryotes derived from the mitochondrion in (or prior to) LECA. In contrast, the hypothesis of eukaryotic + α-proteobacterial monophyly was rejected for PFO and PNO. Meanwhile, hydrogenase maturases are extremely patchily distributed in α-proteobacteria, such that candidate 'ancestral' lineages would need to be reduced to a single taxon, suggesting that they likely originated from a lineage other than the α-proteobacterial symbiont that gave rise to mitochondria.

Proponents of the hydrogen hypothesis have suggested that this observation can be attributed to massive LGT and gene loss from α-proteobacteria following the endosymbiotic event (Müller et al. 2012). But it is not clear why this would disproportionately be the case in phylogenies of anaerobic ATP generation proteins. For instance, Min proteins (which are also sparsely distributed among eukaryotes and lack phylogenetic resolution) are retained in a much larger number of α-proteobacteria. The extremely sparse distribution of anaerobic ATP generation enzymes in α-proteobacteria in general reflects the pattern predicted by the LGT hypothesis and the associated predictions (see Chapter 1) that the anaerobic enzymes will be absent from, or sparsely and patchily distributed in, α-proteobacteria, and that homologs from α-proteobacteria are not expected to be over-represented relative to other bacterial groups as sister group taxa to eukaryote homologs.

Furthermore, subsets of some of these enzymes exhibit rare and unusual fusion patterns that are shared by distantly-related organisms. A fusion of HydE and HydF is found in green algae (Posewitz et al. 2004), but also in the alveolate *Vitrella brassicaformis* (Stairs, Leger, Roger 2015 (in press)). A fusion of SufC and SufB is found in a stramenopile (*Blastocystis*), in a breviate (*Pygsuia*), and in an excavate (*Andalucia incarcerata*). CysJ-fused [FeFe]-hydrogenases are found in the excavates *Trichomonas* and *Andalucia incarcerata*, but also in the breviate *Pygsuia* (Chapter 4). A scenario in which *all* of these enzymes were present in LECA, only to be lost in the vast majority of extant eukaryotes stretches credibility. In contrast, this type of pattern is perfectly consistent with predictions 1) and 3) of the LGT hypothesis: namely, the enzymes are patchily

distributed among eukaryotes, and heterogeneity is observed in the specific anaerobic ATP generation enzymes borne by individual eukaryotes.

Caveats

Overall, the distributions and phylogenies of anaerobic ATP enzymes appear more congruent with the LGT hypothesis than with the hydrogen hypothesis or similar alternatives (i.e., the syntrophic hypothesis). Similar patterns of genes found only in small numbers of mostly anaerobic and distantly-related eukaryote lineages are observed for flavodiiron protein, and for other enzymes described in the literature (Andersson et al. 2006; Takishita et al. 2012; Stairs et al. 2014), and robust examples of likely LGTs between eukaryotes have previously been reported (Simpson, Perley, Lara 2008), reviewed in (Andersson 2009b; Richards, Talbot 2013; Soanes, Richards 2014). This bolsters the idea that lateral gene transfer might be common adaptive mechanism for protists inhabiting similar low-oxygen environments. However, more genomic data is needed from multiple 'outgroup' taxa to the anaerobic taxa currently described. This will clarify whether the ATP generation enzymes found in these taxa were ancestrally present in the groups, as they should be if the hydrogen hypothesis were correct (they should have been vertically inherited from LECA under this hypothesis). It will also be useful to search newly available genomic data for evidence of more eukaryotes like *Acanthamoeba* and *Naegleria*: organisms that are often thought of as 'aerobic' eukaryotes that nevertheless experience transient hypoxia and possess anaerobic ATP generation enzymes.

Neither the LGT hypothesis for the origin of anaerobic enzymes in eukaryotes

nor the stepwise scenario for the emergence of MROs presented in Chapter 2 makes any claims as to the factors underpinning the original protomitochondrial endosymbiosis. Proponents of the hydrogen hypothesis argue that the initial endosymbiosis could not have arisen as a result of ATP exchange between the endosymbiont and the host, and that the protomitochondrial endosymbiont must therefore have been a facultative anaerobe. The first part of this argument is debatable: ATP could initially have been transferred to the host by digestion of a subset of endosymbionts, rather than their (admittedly unlikely) excretion of ATP. Even if it is true, the second part does not follow. As pointed out in the context of the pre-endosymbiont hypothesis (Gray 2014), modern endosymbioses involving *Acanthamoeba* and α-proteobacteria are not based on ATP exchange, yet the endosymbionts may still be aerobes (Yu et al. 2007).

First Steps in the Emergence of MROs

In Chapter 2, I presented a model for the sequence of events that might have led to the emergence of the spectrum of modern MROs. As discussed above, the evidence as it stands is in favour of a recent, secondary origin of anaerobic metabolism enzymes in modern MROs. The relatively close genomic proximity of [FeFe]-hydrogenase and its maturases in *Acanthamoeba castellanii* (Chapter 2) suggests that some of these enzymes might be transferred in tandem with others in the same pathway, providing a means for eukaryotes to rapidly adopt an anaerobic lifestyle. It seems likely that the recipient of such a transfer would already be able to transiently tolerate low-oxygen environments, where it would be likely to encounter potential donors - possibly using the type of metabolism currently found in animals that encounter hypoxia (Müller et al. 2012), or

possibly by beginning encystation. Upon acquiring an anaerobic energy-generating pathway, a previously obligate aerobe would be able to thrive in a more diverse range of habitats, as well as better surviving temporal fluctuations in oxygen levels: a modern analog of such an organism is *Acanthamoeba castellanii*. Subsequently, descendents of such a cell inhabiting exclusively low-oxygen environments, with a reduced need to perform oxidative phosphorylation, might lose components of the electron transport chain, as well as other mitochondrial functions (Stairs, Leger, Roger 2015 (in press)).

Final Conclusions

In this thesis, I have shed light on how anaerobic ATP generation enzymes might have been acquired in MROs - first by describing an organism possessing a mitochondrion capable of hydrogenosomal-like ATP production, and then by constructing phylogenies of the enzymes involved from a wider range of MROs. This work also expands our knowledge of other MRO functions. I have reconstructed the pathways present in a free-living jakobid and found more amino acid catabolism enzymes than previously described. I have provided the first example of division proteins in an MRO; and in examining the taxonomic distribution of these proteins, I have revealed a mechanism of division control previously unsuspected in mitochondria, yet widespread among eukaryotes.

References

Aguilera, P, T Barry, J Tovar. 2008. *Entamoeba histolytica* mitosomes: organelles in search of a function. Exp Parasitol 118:10-16.

Akhmanova, A, F Voncken, T van Alen, A van Hoek, B Boxma, G Vogels, M Veenhuis, JH Hackstein. 1998. A hydrogenosome with a genome. Nature 396:527-528.

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

Altman, R. 1890. Die Elementarorganismen und ihre Beziehungen zu den Zellen. Leipzig, Germany: http://www.deutschestextarchiv.de/book/show/altmann_elementarorganismen_1890.

Altschul, SF, TL Madden, AA Schaffer, J Zhang, Z Zhang, W Miller, DJ Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Anderson, E, HW Beams. 1959. The cytology of *Tritrichomonas* as revealed by the electon microscope. J Morphol 104:205-235.

Andersson, J, R Hirt, P Foster, A Roger. 2006. Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. BMC Evol Biol 6.

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

Andersson, JO. 2009b. Horizontal gene transfer between microbial eukaryotes. Methods Mol Biol 532:473-487.

Andersson, JO, AM Sjogren, LA Davis, TM Embley, AJ Roger. 2003. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Curr Biol 13:94-104.

Andersson, JO, AM Sjogren, DS Horner, CA Murphy, PL Dyal, SG Svard, JM Logsdon, Jr., MA Ragan, RP Hirt, AJ Roger. 2007. A genomic survey of the fish parasite *Spironucleus salmonicida* indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. BMC Genomics 8:51.

Arimura, S, GP Aida, M Fujimoto, M Nakazono, N Tsutsumi. 2004. *Arabidopsis* dynamin-like protein 2a (ADL2a), like ADL2b, is involved in plant mitochondrial division. Plant Cell Physiol 45:236-242.

Arimura, S, N Tsutsumi. 2002. A dynamin-like protein (ADL2b), rather than FtsZ, is involved in *Arabidopsis* mitochondrial division. Proc Natl Acad Sci U S A 99:5727-5731.

Atteia, A, A Adrait, S Brugiere, et al. 2009. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. Mol Biol Evol 26:1533-1548.

Auer, J, K Lechner, A Bock. 1989. Gene organization and structure of two transcriptional units from *Methanococcus* coding for ribosomal proteins and elongation factors. Can J Microbiol 35:200-204.

Aurrecoechea, C, M Heiges, H Wang, et al. 2007. ApiDB: integrated resources for the apicomplexan bioinformatics resource center. Nucleic Acids Res 35:D427-430.

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

Barton, RM, HJ Worman. 1999. Prenylated prelamin A interacts with Narf, a novel nuclear protein. J Biol Chem 274:30008-30018.

Basu, S, P Fey, Y Pandit, R Dodson, WA Kibbe, RL Chisholm. 2013. DictyBase 2013: integrating multiple Dictyostelid species. Nucleic Acids Res 41:D676-683.

Beech, PL, T Nheu, T Schultz, S Herbert, T Lithgow, PR Gilson, GI McFadden. 2000. Mitochondrial FtsZ in a chromophyte alga. Science 287:1276-1279.

Bendich, AJ, LP Gauriloff. 1984. Morphometric analysis of cucurbit mitochondria: the relationship between chondriome volume and DNA content Protoplasma 119:1-7.

Benson, DA, K Clark, I Karsch-Mizrachi, DJ Lipman, J Ostell, EW Sayers. 2014. GenBank. Nucleic Acids Res 42:D32-37.

Bleazard, W, JM McCaffery, EJ King, S Bale, A Mozdy, Q Tieu, J Nunnari, JM Shaw. 1999. The dynamin-related GTPase Dnm1 regulates mitochondrial fission in yeast. Nat Cell Biol 1:298-304.

Bowman, BH, JW Taylor, AG Brownlee, J Lee, SD Lu, TJ White. 1992. Molecular evolution of the fungi: relationship of the Basidiomycetes, Ascomycetes, and Chytridiomycetes. Mol Biol Evol 9:285-296.

Boxma, B, RM de Graaf, GW van der Staay, et al. 2005. An anaerobic mitochondrion that produces hydrogen. Nature 434:74-79.

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

Bradley, RK, A Roberts, M Smoot, S Juvekar, J Do, C Dewey, I Holmes, L Pachter. 2009. Fast statistical alignment. PLoS Comput Biol 5:e1000392.

Brindefalk, B, TJ Ettema, J Viklund, M Thollesson, SG Andersson. 2011. A phylometagenomic exploration of oceanic alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade. PLoS One 6:e24457.

Broderick, JB, AS Byer, KS Duschene, BR Duffus, JN Betz, EM Shepard, JW Peters. 2014. H-cluster assembly during maturation of the [FeFe]-hydrogenase. J Biol Inorg Chem 19:747-757.

Brown, DM, JA Upcroft, P Upcroft. 1995. Free radical detoxification in Giardia duodenalis. Mol Biochem Parasitol 72:47-56.

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

Buetow, DE. 1989. The mitochondrion. In: DE Buetow, editor. The Biology of Euglena, Vol 4, Subcellular Biochemistry and Molecular-Biology. San Diego: Academic Press. p. 247-314.

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

Bui, ET, PJ Bradley, PJ Johnson. 1996. A common evolutionary origin for mitochondria and hydrogenosomes. Proc Natl Acad Sci U S A 93:9651-9656.

Bult, CJ, O White, GJ Olsen, et al. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273:1058-1073.

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

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

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

Carlton, JM, RP Hirt, JC Silva, et al. 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. Science 315:207-212.

Cavalier-Smith, T. 1983a. Endosymbiotic origin of the mitochondrial envelope. In: W Schwemmler, HEA Schenck, editors. Endocytobiology II: Intracellular Space As Oligogenetic Ecosystem: Proceedings. Tübingen: Walter De Gruyter, Inc. p. 265–279.

Cavalier-Smith, T. 1983b. A six-kingdom classification and a unified phylogeny. In: W Schwemmler, HEA Schenck, editors. Endocytobiology II: Intracellular Space As Oligogenetic Ecosystem: Proceedings. Tübingen: Walter De Gruyter, Inc. p. 1027-1034.

Cavalier-Smith, T. 1987a. Eukaryotes with no mitochondria. Nature 326:332-333.

Cavalier-Smith, T. 1987b. The origin of eukaryotic and archaebacterial cells. Ann N Y Acad Sci 503:17-54.

Chan, KW, DJ Slotboom, S Cox, et al. 2005. A novel ADP/ATP transporter in the mitosome of the microaerophilic human parasite *Entamoeba histolytica*. Curr Biol 15:737-742.

Chevreux, B, T Pfisterer, B Drescher, AJ Driesel, WEG Muller, T Wetter, S Suhai. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14:1147-1159.

Chevreux, B, T Wetter, S Suhai. 1999. Genome sequence assembly using trace signals and additional sequence information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics 99:45-56.

Claros, MG. 1995. MitoProt, a Macintosh application for studying mitochondrial proteins. Comput Appl Biosci 11:441-447.

Colletti, KS, EA Tattersall, KA Pyke, JE Froelich, KD Stokes, KW Osteryoung. 2000. A homologue of the bacterial cell division site-determining factor MinD mediates placement of the chloroplast division apparatus. Curr Biol 10:507-516.

Cometá, I, S Schatz, W Trzyna, A Rogerson. 2011. Tolerance of naked amoebae to low oxygen levels with an emphasis on the genus *Acanthamoeba*. Acta Protozoologica 50:33-40.

Criscuolo, A, S Gribaldo. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10:210.

Ctrnacta, V, JG Ault, F Stejskal, JS Keithly. 2006. Localization of pyruvate:NADP+ oxidoreductase in sporozoites of *Cryptosporidium parvum*. J Eukaryot Microbiol 53:225-231.

Dagley, MJ, P Dolezal, VA Likic, O Smid, AW Purcell, SK Buchanan, J Tachezy, T Lithgow. 2009. The protein import channel in the outer mitosomal membrane of *Giardia intestinalis*. Mol Biol Evol 26:1941-1947.

Daniel, WA, CF Mattern, BM Honigberg. 1971. Fine structure of the mastigont system in *Tritrichomonas muris* (Grassi). J Protozool 18:575-586.

Danovaro, R, A Dell'Anno, A Pusceddu, C Gambi, I Heiner, RM Kristensen. 2010. The first metazoa living in permanently anoxic conditions. BMC Biol 8:30.

Davidson, EA, M van der Giezen, DS Horner, TM Embley, CJ Howe. 2002. An [Fe] hydrogenase from the anaerobic hydrogenosome-containing fungus *Neocallimastix frontalis* L2. Gene 296:45-52.

de Boer, PA. 2010. Advances in understanding *E. coli* cell fission. Curr Opin Microbiol 13:730-737.

de Graaf, RM, I Duarte, TA van Alen, JW Kuiper, K Schotanus, J Rosenberg, MA Huynen, JH Hackstein. 2009. The hydrogenosomes of *Psalteriomonas lanterna*. BMC Evol Biol 9:287.

de Graaf, RM, G Ricard, TA van Alen, et al. 2011. The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. Mol Biol Evol 28:2379-2391.

den Blaauwen, T, JM Andreu, O Monasterio. 2014. Bacterial cell division proteins as antibiotic targets. Bioorg Chem 55:27-38.

Denoeud, F, M Roussel, B Noel, et al. 2011. Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. Genome Biol 12:R29.

Di Matteo, A, FM Scandurra, F Testa, E Forte, P Sarti, M Brunori, A Giuffre. 2008. The O2-scavenging flavodiiron protein in the human parasite *Giardia intestinalis*. J Biol Chem 283:4061-4068.

Dickerson, RE. 1980. Evolution and gene transfer in purple photosynthetic bacteria. Nature 283:210-212.

Do, CB, MS Mahabhashyam, M Brudno, S Batzoglou. 2005. ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome Res 15:330-340.

Dolezal, P, MJ Dagley, M Kono, et al. 2010. The essentials of protein import in the degenerate mitochondrion of Entamoeba histolytica. PLoS Pathog 6:e1000812.

Dolezal, P, V Likic, J Tachezy, T Lithgow. 2006. Evolution of the molecular machines for protein import into mitochondria. Science 313:314-318.

Dolezal, P, O Smid, P Rada, Z Zubacova, D Bursac, R Sutak, J Nebesarova, T Lithgow, J Tachezy. 2005. *Giardia* mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting. Proc Natl Acad Sci U S A 102:10924-10929.

Douglas, SE, SL Penny. 1999. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. J Mol Evol 48:236-244.

Dyall, SD, W Yan, MG Delgadillo-Correa, A Lunceford, JA Loo, CF Clarke, PJ Johnson. 2004. Non-mitochondrial complex I proteins in a hydrogenosomal oxidoreductase complex. Nature 431:1103-1107.

Eckers, E, M Cyrklaff, L Simpson, M Deponte. 2012. Mitochondrial protein import pathways are functionally conserved among eukaryotes despite compositional diversity of the import machineries. Biol Chem 393:513-524.

Edgar, RC. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.

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

Edgar, RC. 2010. Quality measures for protein alignment benchmarks. Nucleic Acids Res 38:2145-2153.

Edwards, SW, D Lloyd. 1978. Properties of mitochondria isolated from cyanide-sensitive and cyanide-stimulated cultures of *Acanthamoeba castellanii*. Biochem J 174:203-211.

Egan, AJ, W Vollmer. 2013. The physiology of bacterial cell division. Ann N Y Acad Sci 1277:8-28.

Ellis, JE, R Williams, D Cole, R Cammack, D Lloyd. 1993. Electron transport components of the parasitic protozoon *Giardia lamblia*. FEBS Lett 325:196-200.

Emanuelsson, O, S Brunak, G von Heijne, H Nielsen. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2:953-971.

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

Embley, TM. 2006. Multiple secondary origins of the anaerobic lifestyle in eukaryotes. Philos Trans R Soc Lond B Biol Sci 361:1055-1067.

Embley, TM, BJ Finlay, PL Dyal, RP Hirt, M Wilkinson, AG Williams. 1995. Multiple origins of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. Proc Biol Sci 262:87-93.

Embley, TM, M van der Giezen, DS Horner, PL Dyal, S Bell, PG Foster. 2003a. Hydrogenosomes, mitochondria and early eukaryotic evolution. IUBMB life 55:387-395.

Embley, TM, M van der Giezen, DS Horner, PL Dyal, P Foster. 2003b. Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. Philos Trans R Soc Lond B Biol Sci 358:191-201; discussion 201-192.

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

Ewing, B, P Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186-194.

Ewing, B, L Hillier, MC Wendl, P Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175-185.

Fenchel, T, BJ Finlay. 1991. The biology of free-living anaerobic ciliates. Eur J Protistol 26:201-215.

Fey, P, RJ Dodson, S Basu, RL Chisholm. 2013. One stop shop for everything *Dictyostelium*: dictyBase and the Dicty Stock Center in 2012. Methods Mol Biol 983:59-92.

Filadoro, F. 1970. Fine structure of chromatic granules in *Trichomonas vaginalis* Donne. Experientia 26:213-214.

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

Fitzpatrick, DA, CJ Creevey, JO McInerney. 2006. Genome phylogenies indicate a meaningful alpha-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. Mol Biol Evol 23:74-85.

Flicek, P, MR Amode, D Barrell, et al. 2014. Ensembl 2014. Nucleic Acids Res 42:D749-755.

Forestier, M, P King, L Zhang, M Posewitz, S Schwarzer, T Happe, ML Ghirardi, M Seibert. 2003. Expression of two [Fe]-hydrogenases in *Chlamydomonas reinhardtii* under anaerobic conditions. Eur J Biochem 270:2750-2758.

Foster, PG, CJ Cox, TM Embley. 2009. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. Philos Trans R Soc Lond B Biol Sci 364:2197-2207.

Fraunholz, MJ, E Moerschel, UG Maier. 1998. The chloroplast division protein FtsZ is encoded by a nucleomorph gene in cryptomonads. Mol Gen Genet 260:207-211.

Friedman, JR, LL Lackner, M West, JR DiBenedetto, J Nunnari, GK Voeltz. 2011. ER tubules mark sites of mitochondrial division. Science 334:358-362.

Fritz-Laylin, LK, ML Ginger, C Walsh, SC Dawson, C Fulton. 2011. The *Naegleria* genome: a free-living microbial eukaryote lends unique insights into core eukaryotic cell biology. Res Microbiol 162:607-618.

Fritz-Laylin, LK, SE Prochnik, ML Ginger, et al. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. Cell 140:631-642.

Gaffron, H, J Rubin. 1942. Fermentative and photochemical production of hydrogen in algae. J Gen Physiol 26:219-240.

Galli, E, K Gerdes. 2010. Spatial resolution of two bacterial cell division proteins: ZapA recruits ZapB to the inner face of the Z-ring. Mol Microbiol 76:1514-1526.

Gaston, D, AD Tsaousis, AJ Roger. 2009. Predicting proteomes of mitochondria and related organelles from genomic and expressed sequence tag data. Meth Enzymol 457:21-47.

Gaudet, P, P Fey, S Basu, YA Bushmanova, R Dodson, KA Sheppard, EM Just, WA Kibbe, RL Chisholm. 2011. dictyBase update 2011: web 2.0 functionality and the initial steps towards a genome portal for the Amoebozoa. Nucleic Acids Res 39:D620-624.

Gawryluk, RM, KA Chisholm, DM Pinto, MW Gray. 2012. Composition of the mitochondrial electron transport chain in *Acanthamoeba castellanii*: structural and evolutionary insights. Biochim Biophys Acta 1817:2027-2037.

Gawryluk, RM, KA Chisholm, DM Pinto, MW Gray. 2014. Compositional complexity of the mitochondrial proteome of a unicellular eukaryote (*Acanthamoeba castellanii*, supergroup Amoebozoa) rivals that of animals, fungi, and plants. J Proteomics 109:400-416.

Gawryluk, RM, MW Gray. 2010. Evidence for an early evolutionary emergence of gamma-type carbonic anhydrases as components of mitochondrial respiratory complex I. BMC Evol Biol 10:176.

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

Georgiades, K, D Raoult. 2011. The rhizome of *Reclinomonas americana*, *Homo sapiens*, *Pediculus humanus* and *Saccharomyces cerevisiae* mitochondria. Biol Direct 6:55.

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

Germot, A, H Philippe, H Le Guyader. 1997. Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. Mol Biochem Parasitol 87:159-168.

Ghosal, D, D Trambaiolo, LA Amos, J Lowe. 2014. MinCD cell division proteins form alternating copolymeric cytomotive filaments. Nat Commun 5:5341.

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

Gilson, PR, XC Yu, D Hereld, C Barth, A Savage, BR Kiefel, S Lay, PR Fisher, W Margolin, PL Beech. 2003. Two *Dictyostelium* orthologs of the prokaryotic cell division protein FtsZ localize to mitochondria and are required for the maintenance of normal mitochondrial morphology. Eukaryot Cell 2:1315-1326.

Gogarten, JP, H Kibak, P Dittrich, et al. 1989. Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. Proc Natl Acad Sci U S A 86:6661-6665.

Grabherr, MG, BJ Haas, M Yassour, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644-652.

Gray, MW. 2014. The pre-endosymbiont hypothesis: a new perspective on the origin and evolution of mitochondria. Cold Spring Harb Perspect Biol 6.

Gray, MW, WF Doolittle. 1982. Has the endosymbiont hypothesis been proven? Microbiol Rev 46:1-42.

Grigoriev, IV, H Nordberg, I Shabalov, et al. 2012. The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res 40:D26-32.

Guy, L, TJ Ettema. 2011. The archaeal 'TACK' superphylum and the origin of eukaryotes. Trends Microbiol 19:580-587.

Hackstein, J. 2005. Eukaryotic Fe-hydrogenases—old eukaryotic heritage or adaptive acquisitions. Biochem. Soc. Trans 33:47-50.

Hale, CA, PA de Boer. 1997. Direct binding of FtsZ to ZipA, an essential component of the septal ring structure that mediates cell division in *E. coli*. Cell 88:175-185.

Hale, CA, PA de Boer. 1999. Recruitment of ZipA to the septal ring of *Escherichia coli* is dependent on FtsZ and independent of FtsA. J Bacteriol 181:167-176.

Hale, CA, H Meinhardt, PA de Boer. 2001. Dynamic localization cycle of the cell division regulator MinE in *Escherichia coli*. EMBO J 20:1563-1572.

Hampl, V, JD Silberman, A Stechmann, S Diaz-Trivino, PJ Johnson, AJ Roger. 2008. Genetic evidence for a mitochondriate ancestry in the 'amitochondriate' flagellate Trimastix pyriformis. PLoS One 3:e1383.

Hanson, RS, TE Hanson. 1996. Methanotrophic bacteria. Microbiol Rev 60:439-471.

Happe, T, A Kaminski. 2002. Differential regulation of the Fe-hydrogenase during anaerobic adaptation in the green alga *Chlamydomonas reinhardtii*. Eur J Biochem 269:1022-1032.

Hardham, AR. 1987. Ultrastructure and serial section reconstruction of zoospores of the fungus *Phytophthora cinnamomi*. Exp Mycol 11:297-306.

Harding, T, MW Brown, A Plotnikov, E Selivanova, JS Park, JH Gunderson, M Baumgartner, JD Silberman, AJ Roger, AG Simpson. 2013. Amoeba stages in the deepest branching heteroloboseans, including *Pharyngomonas*: evolutionary and systematic implications. Protist 164:272-286.

Hasegawa, M, T Hashimoto. 1993. Ribosomal RNA trees misleading? Nature 361:23.

Hashimoto, T, M Hasegawa. 1996. Origin and early evolution of eukaryotes inferred from the amino acid sequences of translation elongation factors 1alpha/Tu and 2/G. Adv Biophys 32:73-120.

Heidel, AJ, HM Lawal, M Felder, et al. 2011. Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. Genome Res 21:1882-1891.

Heinz, E, T Lithgow. 2013. Back to basics: A revealing secondary reduction of the mitochondrial protein import pathway in diverse intracellular parasites. Biochim Biophys Acta 1833:295-303.

Heiss, AA, G Walker, AG Simpson. 2013. The microtubular cytoskeleton of the apusomonad *Thecamonas*, a sister lineage to the opisthokonts. Protist 164:598-621.

Hell, K, W Neupert, RA Stuart. 2001. Oxa1p acts as a general membrane insertion machinery for proteins encoded by mitochondrial DNA. EMBO J 20:1281-1288.

Henderson, E, M Oakes, MW Clark, JA Lake, AT Matheson, W Zillig. 1984. A new ribosome structure. Science 225:510-512.

Henriquez, FL, TA Richards, F Roberts, R McLeod, CW Roberts. 2005. The unusual mitochondrial compartment of *Cryptosporidium parvum*. Trends Parasitol 21:68-74.

Hirt, RP, B Healy, CR Vossbrinck, EU Canning, TM Embley. 1997. A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. Curr Biol 7:995-998.

Hogan, MJ, C Yoneda, L Feeney, P Zweigart, A Lewis. 1960. Morphology and culture of *Toxoplasma*. Arch Ophthalmol 64:655-667.

Honigberg, BM, CF Mattern, WA Daniel. 1971. Fine structure of the mastigont system in *Tritrichomonas foetus* (Riedmuller). J Protozool 18:183-198.

Hoogenraad, NJ, LA Ward, MT Ryan. 2002. Import and assembly of proteins into mitochondria of mammalian cells. Biochim Biophys Acta 1592:97-105.

Horner, DS, PG Foster, TM Embley. 2000. Iron hydrogenases and the evolution of anaerobic eukaryotes. Mol Biol Evol 17:1695-1709.

Horner, DS, B Heil, T Happe, TM Embley. 2002. Iron hydrogenases--ancient enzymes in modern eukaryotes. Trends Biochem Sci 27:148-153.

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

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

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

Hrdý, I, M Müller. 1995. Primary structure and eubacterial relationships of the pyruvate:ferredoxin oxidoreductase of the amitochondriate eukaryote *Trichomonas vaginalis*. J Mol Evol 41:388-396.

Hrdý, I, J Tachezy, M Müller. 2008. Metabolism of Trichomonad Hydrogenosomes. In: J Tachezy, editor. Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes: Springer. p. 113-146.

Hu, Z, J Lutkenhaus. 2001. Topological regulation of cell division in *E. coli*: spatiotemporal oscillation of MinD requires stimulation of its ATPase by MinE and phospholipid. Mol Cell 7:1337-1343.

Huang, X, A Madan. 1999. CAP3: A DNA sequence assembly program. Genome Res 9:868-877.

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

Itoh, R, M Fujiwara, N Nagata, S Yoshida. 2001. A chloroplast protein homologous to the eubacterial topological specificity factor MinE plays a role in chloroplast division. Plant Physiol 127:1644-1655.

Iwabe, N, K Kuma, M Hasegawa, S Osawa, T Miyata. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci U S A 86:9355-9359.

Jedelsky, PL, P Dolezal, P Rada, et al. 2011. The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. PLoS One 6:e17285.

Jerlstrom-Hultqvist, J, E Einarsson, SG Svard. 2012. Stable transfection of the diplomonad parasite *Spironucleus salmonicida*. Eukaryot Cell 11:1353-1361.

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

Kamikawa, R, M Kolisko, Y Nishimura, A Yabuki, MW Brown, SA Ishikawa, K Ishida, AJ Roger, T Hashimoto, Y Inagaki. 2014. Gene content evolution in Discobid mitochondria deduced from the phylogenetic position and complete mitochondrial genome of *Tsukubamonas globosa*. Genome Biol Evol 6:306-315.

Katoh, K, K Kuma, H Toh, T Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33:511-518.

Katoh, K, K Misawa, K Kuma, T Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066.

Katoh, K, DM Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780.

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

Keeling, PJ, F Burki, HM Wilcox, et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol 12:e1001889.

Keeling, PJ, WF Doolittle. 1997. Evidence that eukaryotic triosephosphate isomerase is of alpha-proteobacterial origin. Proc Natl Acad Sci U S A 94:1270-1275.

Keithly, JS, SG Langreth, KF Buttle, CA Mannella. 2005. Electron tomographic and ultrastructural analysis of the *Cryptosporidium parvum* relict mitochondrion, its associated membranes, and organelles. J Eukaryot Microbiol 52:132-140.

Kiefel, BR, PR Gilson, PL Beech. 2004. Diverse eukaryotes have retained mitochondrial homologues of the bacterial division protein FtsZ. Protist 155:105-115.

Kimura, M, E Arndt, T Hatakeyama, T Hatakeyama, J Kimura. 1989. Ribosomal proteins in halobacteria. Can J Microbiol 35:195-199.

Knoll, AH. 1992. The early evolution of eukaryotes: a geological perspective. Science 256:622-627.

Kreppel, L, P Fey, P Gaudet, E Just, WA Kibbe, RL Chisholm, AR Kimmel. 2004. dictyBase: a new *Dictyostelium discoideum* genome database. Nucleic Acids Res 32:D332-333.

Kurland, CG, SG Andersson. 2000. Origin and evolution of the mitochondrial proteome. Microbiol Mol Biol Rev 64:786-820.

Labrousse, AM, MD Zappaterra, DA Rube, AM van der Bliek. 1999. *C. elegans* dynamin-related protein DRP-1 controls severing of the mitochondrial outer membrane. Mol Cell 4:815-826.

LaGier, MJ, J Tachezy, F Stejskal, K Kutisova, JS Keithly. 2003. Mitochondrial-type iron-sulfur cluster biosynthesis genes (IscS and IscU) in the apicomplexan *Cryptosporidium parvum*. Microbiology 149:3519-3530.

Lake, JA, E Henderson, M Oakes, MW Clark. 1984. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc Natl Acad Sci U S A 81:3786-3790.

Lang, BF, G Burger, CJ O'Kelly, R Cedergren, GB Golding, C Lemieux, D Sankoff, M Turmel, MW Gray. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387:493-497.

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

Lara, E, A Chatzinotas, AGB Simpson. 2006. *Andalucia* (n. gen.) - the deepest branch within jakobids (Jakobida; Excavata), based on morphological and molecular study of a new flagellate from soil. J Eukaryot Microbiol 53:112-120.

Larkin, JM, MC Henk. 1996. Filamentous sulfide-oxidizing bacteria at hydrocarbon seeps of the Gulf of Mexico. Microsc Res Tech 33:23-31.

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

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

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

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

Leger, MM, RM Gawryluk, MW Gray, AJ Roger. 2013. Evidence for a hydrogenosomal-type anaerobic ATP generation pathway in *Acanthamoeba castellanii*. PLoS One 8:e69532.

Leger, MM, M Petru, V Zarsky, L Eme, C Vlcek, T Harding, BF Lang, M Elias, P Dolezal, AJ Roger. 2015. An ancestral bacterial division system is widespread in eukaryotic mitochondria. Proc Natl Acad Sci U S A.

Leipe, DD, JH Gunderson, TA Nerad, ML Sogin. 1993. Small subunit ribosomal RNA+ of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. Mol Biochem Parasitol 59:41-48.

Li, J, IB Heath, T Bauchop. 1990. *Piromyces mae* and *Piromyces dumbonica* , two new species of uniflagellate anaerobic chytridiomycete fungi from the hindgut of the horse and elephant. Can J Bot 68:1021-1033.

Li, J, IB Heath, K-J Cheng. 1991. The development and zoospore ultrastructure of a polycentric chytridiomycete gut fungus, *Orpinomyces joyonii* comb.nov. Can J Bot 69:580-589.

Lill, R, K Diekert, A Kaut, H Lange, W Pelzer, C Prohl, G Kispal. 1999. The essential role of mitochondria in the biogenesis of cellular iron-sulfur proteins. Biol Chem 380:1157-1166.

Lill, R, B Hoffmann, S Molik, AJ Pierik, N Rietzschel, O Stehling, MA Uzarska, H Webert, C Wilbrecht, U Muhlenhoff. 2012. The role of mitochondria in cellular iron-sulfur protein biogenesis and iron metabolism. Biochim Biophys Acta 1823:1491-1508.

Lindmark, DG. 1980. Energy metabolism of the anaerobic protozoon *Giardia lamblia*. Mol Biochem Parasitol 1:1-12.

Lindmark, DG, M Muller. 1973. Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. J Biol Chem 248:7724-7728.

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

Loftus, B, I Anderson, R Davies, et al. 2005. The genome of the protist parasite *Entamoeba histolytica*. Nature 433:865-868.

Lohan, AJ, MW Gray. 2007. Analysis of 5′– or 3′–Terminal tRNA Editing: Mitochondrial 5′ tRNA Editing in *Acanthamoeba castellanii* as the Exemplar. Meth Enzymol 424:221-242.

Loose, M, E Fischer-Friedrich, J Ries, K Kruse, P Schwille. 2008. Spatial regulators for bacterial cell division self-organize into surface waves in vitro. Science 320:789-792.

Lutkenhaus, J. 2007. Assembly dynamics of the bacterial MinCDE system and spatial regulation of the Z ring. Annu Rev Biochem 76:539-562.

Lutkenhaus, J, S Pichoff, SS Du. 2012. Bacterial cytokinesis: From Z ring to divisome. Cytoskeleton 69:778-790.

Maguire, F, TA Richards. 2014. Organelle evolution: a mosaic of 'mitochondrial' functions. Curr Biol 24:R518-520.

Mai, Z, S Ghosh, M Frisardi, B Rosenthal, R Rogers, J Samuelson. 1999. Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("crypton") in the microaerophilic protozoan parasite *Entamoeba histolytica*. Mol Cell Biol 19:2198-2205.

Makiuchi, T, T Nozaki. 2014. Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa. Biochimie 100:3-17.

Mano, S, C Nakamori, M Kondo, M Hayashi, M Nishimura. 2004. An *Arabidopsis* dynamin-related protein, DRP3A, controls both peroxisomal and mitochondrial division. Plant J 38:487-498.

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

Marciano-Cabral, F. 2004. Introductory remarks: bacterial endosymbionts or pathogens of free-living amebae1. J Eukaryot Microbiol 51:497-501.

Marciano-Cabral, F, G Cabral. 2003. *Acanthamoeba* spp. as agents of disease in humans. Clinical Microbiology Reviews 16:273-307.

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

Marvin-Sikkema, FD, MN Kraak, M Veenhuis, JC Gottschal, RA Prins. 1993a. The hydrogenosomal enzyme hydrogenase from the anaerobic fungus *Neocallimastix* sp. L2 is recognized by antibodies, directed against the C-terminal microbody protein targeting signal SKL. Eur J Cell Biol 61:86-91.

Marvin-Sikkema, FD, GA Lahpor, MN Kraak, JC Gottschal, RA Prins. 1992. Characterization of an anaerobic fungus from llama faeces. J Gen Microbiol 138:2235-2241.

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

McInerney, JO, MJ O'Connell, D Pisani. 2014. The hybrid nature of the Eukaryota and a consilient view of life on Earth. Nat Rev Microbiol 12:449-455.

Meier, EL, ED Goley. 2014. Form and function of the bacterial cytokinetic ring. Curr Opin Cell Biol 26:19-27.

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

Mereschkowsky, C. 1905. Über natur und ursprung der chromatophoren im pflanzenreiche. Biologisches Centralblatt 18:593-604.

Meyer, J. 2007. [FeFe] hydrogenases and their evolution: a genomic perspective. Cell Mol Life Sci 64:1063-1084.

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

Miyagishima, SY, Y Kabeya, C Sugita, M Sugita, T Fujiwara. 2014. DipM is required for peptidoglycan hydrolysis during chloroplast division. BMC Plant Biol 14:57.

Miyagishima, SY, K Nishida, T Kuroiwa. 2003. An evolutionary puzzle: chloroplast and mitochondrial division rings. Trends Plant Sci 8:432-438.

Miyagishima, SY, H Nozaki, K Nishida, K Nishida, M Matsuzaki, T Kuroiwa. 2004. Two types of FtsZ proteins in mitochondria and red-lineage chloroplasts: the duplication of FtsZ is implicated in endosymbiosis. J Mol Evol 58:291-303.

Moreira, D, P Lopez-Garcia. 1998. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J Mol Evol 47:517-530.

Müller, M. 1973. Biochemical cytology of trichomonad flagellates. I. Subcellular localization of hydrolases, dehydrogenases, and catalase in *Tritrichomonas foetus*. J Cell Biol 57:453-474.

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

Munoz-Gomez, SA, CH Slamovits, JB Dacks, KA Baier, KD Spencer, JG Wideman. 2015. Ancient homology of the mitochondrial contact site and cristae organizing system points to an endosymbiotic origin of mitochondrial cristae. Curr Biol 25:1489-1495.

Murcha, MW, B Kmiec, S Kubiszewski-Jakubiak, PF Teixeira, E Glaser, J Whelan. 2014a. Protein import into plant mitochondria: signals, machinery, processing, and regulation. J Exp Bot 65:6301-6335.

Murcha, MW, R Narsai, J Devenish, S Kubiszewski-Jakubiak, J Whelan. 2015. MPIC: a mitochondrial protein import components database for plant and non-plant species. Plant Cell Physiol 56:e10.

Murcha, MW, Y Wang, R Narsai, J Whelan. 2014b. The plant mitochondrial protein import apparatus - the differences make it interesting. Biochim Biophys Acta 1840:1233-1245.

Nasirudeen, AM, KS Tan. 2004. Isolation and characterization of the mitochondrion-like organelle from *Blastocystis hominis*. J Microbiol Methods 58:101-109.

Natale, P, M Pazos, M Vicente. 2013. The *Escherichia coli* divisome: born to divide. Environ Microbiol 15:3169-3182.

Neupert, W. 1997. Protein import into mitochondria. Annu Rev Biochem 66:863-917.

Nielsen, H, J Engelbrecht, S Brunak, G von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10:1-6.

Nishida, K, M Takahara, S Miyagishima, H Kuroiwa, M Matsuzaki, T Kuroiwa. 2003. Dynamic recruitment of dynamin for final mitochondrial severance in a primitive red alga. Proc Natl Acad Sci U S A 100:2146-2151.

Nixon, JEJ, J Field, AG McArthur, ML Sogin, N Yarlett, BJ Loftus, J Samuelson. 2003. Iron-dependent hydrogenases of *Entamoeba histolytica* and *Giardia lamblia*: activity of the recombinant entamoebic enzyme and evidence for lateral gene transfer. The Biological Bulletin 204:1-9.

Nordberg, H, M Cantor, S Dusheyko, S Hua, A Poliakov, I Shabalov, T Smirnova, IV Grigoriev, I Dubchak. 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res 42:D26-31.

Nyvltova, E, CW Stairs, I Hrdy, J Ridl, J Mach, J Paces, AJ Roger, J Tachezy. 2015. Lateral gene transfer and gene duplication played a key role in the evolution of *Mastigamoeba balamuthi* hydrogenosomes. Mol Biol Evol 32:1039-1055.

Nyvltova, E, R Sutak, K Harant, M Sedinova, I Hrdy, J Paces, C Vlcek, J Tachezy. 2013. NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. Proc Natl Acad Sci U S A 110:7371-7376.

O'Brien, EA, LB Koski, Y Zhang, L Yang, E Wang, MW Gray, G Burger, BF Lang. 2007. TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). Nucleic Acids Res 35:D445-451.

Osteryoung, KW, E Vierling. 1995. Conserved cell and organelle division. Nature 376:473-474.

Oulton, MM, R Amons, P Liang, TH MacRae. 2003. A 49 kDa microtubule cross-linking protein from *Artemia franciscana* is a coenzyme A-transferase. Eur J Biochem 270:4962-4972.

Park, KT, W Wu, S Lovell, J Lutkenhaus. 2012. Mechanism of the asymmetric activation of the MinD ATPase by MinE. Mol Microbiol 85:271-281.

Perkins, DN, DJ Pappin, DM Creasy, JS Cottrell. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551-3567.

Petitjean, C, P Deschamps, P Lopez-Garcia, D Moreira, C Brochier-Armanet. 2015. Extending the conserved phylogenetic core of archaea disentangles the evolution of the third domain of life. Mol Biol Evol 32:1242-1254.

Philippe, H, A Germot. 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. Mol Biol Evol 17:830-834.

Pichoff, S, J Lutkenhaus. 2002. Unique and overlapping roles for ZipA and FtsA in septal ring assembly in *Escherichia coli*. EMBO J 21:685-693.

Posewitz, MC, P King, SL Smolinski, L Zhang, M Seibert, M Ghirardi. 2004. Discovery of two novel radical S-adenosylmethionine proteins required for the assembly of an active [Fe] hydrogenase. J Biol Chem 279:25711-25720.

Price, DC, CX Chan, HS Yoon, et al. 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science 335:843-847.

Price, MN, PS Dehal, AP Arkin. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26:1641-1650.

Puhler, G, H Leffers, F Gropp, P Palm, HP Klenk, F Lottspeich, RA Garrett, W Zillig. 1989. Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc Natl Acad Sci U S A 86:4569-4573.

Purkanti, R, M Thattai. 2015. Ancient dynamin segments capture early stages of host-mitochondrial integration. Proc Natl Acad Sci U S A.

Pütz, S, P Dolezal, G Gelius-Dietrich, L Bohacova. 2006. Fe-Hydrogenase Maturases in the Hydrogenosomes of Trichomonas vaginalis†. Eukaryot Cell.

Regoes, A, D Zourmpanou, G Leon-Avila, M van der Giezen, J Tovar, AB Hehl. 2005. Protein import, replication, and inheritance of a vestigial mitochondrion. J Biol Chem 280:30557-30563.

Richards, TA, NJ Talbot. 2013. Horizontal gene transfer in osmotrophs: playing with public goods. Nat Rev Microbiol 11:720-727.

Richards, TA, M van der Giezen. 2006. Evolution of the Isd11-IscS complex reveals a single alpha-proteobacterial endosymbiosis for all eukaryotes. Mol Biol Evol 23:1341-1344.

Rivera, MC, R Jain, JE Moore, JA Lake. 1998. Genomic evidence for two functionally distinct gene classes. Proc Natl Acad Sci U S A 95:6239-6244.

Rivera, MC, JA Lake. 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. Nature 431:152-155.

Riviere, L, SW van Weelden, P Glass, P Vegh, V Coustou, M Biran, JJ van Hellemond, F Bringaud, AG Tielens, M Boshart. 2004. Acetyl:succinate CoA-transferase in procyclic *Trypanosoma brucei*. Gene identification and role in carbohydrate metabolism. J Biol Chem 279:45337-45346.

Roberts, CW, F Roberts, FL Henriquez, et al. 2004. Evidence for mitochondrial-derived alternative oxidase in the apicomplexan parasite *Cryptosporidium parvum*: a potential anti-microbial agent target. Int J Parasitol 34:297-308.

Rochette, NC, C Brochier-Armanet, M Gouy. 2014. Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. Mol Biol Evol 31:832-845.

Rodriguez, MA, ME Hidalgo, T Sanchez, E Orozco. 1996. Cloning and characterization of the *Entamoeba histolytica* pyruvate: ferredoxin oxidoreductase gene. Mol Biochem Parasitol 78:273-277.

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

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

Roger, AJ, Svärd, S. G., Tovar, J., Clark, C. G., Smith, M. W., Gillin, F. D. and Sogin, M. L. 1998. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: Evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria Proc Natl Acad Sci U S A 95:229-234.

Romeralo, M, JC Cavender, JC Landolt, SL Stephenson, SL Baldauf. 2011. An expanded phylogeny of social amoebas (Dictyostelia) shows increasing diversity and new morphological patterns. BMC Evol Biol 11.

Rotte, C, F Stejskal, G Zhu, J Keithly, W Martin. 2001. Pyruvate: NADP oxidoreductase from the mitochondrion of *Euglena gracilis* and from the apicomplexan *Cryptosporidium parvum*: a biochemical relic linking pyruvate metabolism in mitochondriate and amitochondriate protists. Mol Biol Evol 18:710.

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

Sato, M, T Nishikawa, T Yamazaki, S Kawano. 2005. Isolation of the plastid ftsZ gene from *Cyanophora paradoxa* (Glaucocystophyceae, Glaucocystophyta). Phycol Res 53:93-96.

Scheffler, I. 2007. Mitochondria. Hoboken, NJ, USA: John Wiley & Sons.

Schnarrenberger, C, W Martin. 2002. Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. A case study of endosymbiotic gene transfer. Eur J Biochem 269:868-883.

Schneider, G, S Sjoling, E Wallin, P Wrede, E Glaser, G von Heijne. 1998. Feature-extraction from endopeptidase cleavage sites in mitochondrial targeting peptides. Proteins 30:49-60.

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

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

Segui-Simarro, JM, MJ Coronado, LA Staehelin. 2008. The mitochondrial cycle of *Arabidopsis* shoot apical meristem and leaf primordium meristematic cells is defined by a perinuclear tentaculate/cage-like mitochondrion. Plant Physiol 148:1380-1393.

Sharma, NN, GH Bourne. 1963. Studies on the histochemical distribution of oxidative enzymes in *Trichomonas vaginalis*. J Histochem Cytochem 11:628-634.

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

Siddall, ME, H Hong, SS Desser. 1992. Phylogenetic analysis of the Diplomonadida (Wenyon, 1926) Brugerolle, 1975: evidence for heterochrony in protozoa and against *Giardia lamblia* as a "missing link". J Protozool 39:361-367.

Simpson, AG, DJ Patterson. 2001. On core jakobids and excavate taxa: the ultrastructure of *Jakoba incarcerata*. J Eukaryot Microbiol 48:480-492.

Simpson, AG, TA Perley, E Lara. 2008. Lateral transfer of the gene for a widely used marker, alpha-tubulin, indicated by a multi-protein study of the phylogenetic position of *Andalucia* (Excavata). Mol Phylogenet Evol 47:366-377.

Šmíd, O, A Matušková, SR Harris, et al. 2008. Reductive evolution of the mitochondrial processing peptidases of the unicellular parasites *Trichomonas vaginalis* and *Giardia intestinalis*. PLoS Pathog 4:e1000243.

Smirnova, E, L Griparic, DL Shurland, AM van der Bliek. 2001. Dynamin-related protein Drp1 is required for mitochondrial division in mammalian cells. Mol Biol Cell 12:2245-2256.

Smutna, T, VL Goncalves, LM Saraiva, J Tachezy, M Teixeira, I Hrdy. 2009. Flavodiiron protein from *Trichomonas vaginalis* hydrogenosomes: the terminal oxygen reductase. Eukaryot Cell 8:47-55.

Soanes, D, TA Richards. 2014. Horizontal gene transfer in eukaryotic plant pathogens. Annu Rev Phytopathol 52:583-614.

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

Spang, A, JH Saw, SL Jorgensen, K Zaremba-Niedzwiedzka, J Martijn, AE Lind, R van Eijk, C Schleper, L Guy, TJ Ettema. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521:173-179.

Spencer, DF, MN Schnare, MW Gray. 1984. Pronounced structural similarities between the small subunit ribosomal RNA genes of wheat mitochondria and *Escherichia coli*. Proc Natl Acad Sci U S A 81:493-497.

Stairs, CW, L Eme, MW Brown, C Mutsaers, E Susko, G Dellaire, DM Soanes, M van der Giezen, AJ Roger. 2014. A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist *Pygsuia*. Curr Biol 24:1176-1186.

Stairs, CW, MM Leger, AJ Roger. 2015 (in press). Diversity and origins of anaerobic metabolism in mitochondria and related organelles. Philos Trans R Soc Lond B Biol Sci.

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

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

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

Stechmann, A, K Hamblin, V Perez-Brocal, D Gaston, GS Richmond, M van der Giezen, CG Clark, AJ Roger. 2008. Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. Curr Biol 18:580-585.

Steinbuchel, A, M Muller. 1986. Anaerobic pyruvate metabolism of *Tritrichomonas foetus* and *Trichomonas vaginalis* hydrogenosomes. Mol Biochem Parasitol 20:57-65.

Stiller, JW, BD Hall. 1997. The origin of red algae: implications for plastid evolution. Proc Natl Acad Sci U S A 94:4520-4525.

Stiller, JW, BD Hall. 1999. Long-branch attraction and the rDNA model of early eukaryotic evolution. Mol Biol Evol 16:1270-1279.

Sun, S, J Chen, W Li, et al. 2011. Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. Nucleic Acids Res 39:D546-551.

Szklarczyk, R, MA Huynen. 2010. Mosaic origin of the mitochondrial proteome. Proteomics 10:4012-4024.

Takahara, M, H Kuroiwa, S Miyagishima, T Mori, T Kuroiwa. 2001. Localization of the mitochondrial FtsZ protein in a dividing mitochondrion. Cytologia 66:421-425.

Takahara, M, H Takahashi, S Matsunaga, S Miyagishima, H Takano, A Sakai, S Kawano, T Kuroiwa. 2000. A putative mitochondrial ftsZ gene is present in the unicellular primitive red alga *Cyanidioschyzon merolae*. Mol Gen Genet 264:452-460.

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

Tamura, Y, Y Harada, S Nishikawa, et al. 2013. Tam41 Is a CDP-diacylglycerol synthase required for cardiolipin biosynthesis in mitochondria. Cell Metabolism 17:709-718.

TerBush, AD, Y Yoshida, KW Osteryoung. 2013. FtsZ in chloroplast division: structure, function and evolution. Curr Opin Cell Biol 25:461-470.

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

Tian, HF, JM Feng, JF Wen. 2012. The evolution of cardiolipin biosynthesis and maturation pathways and its implications for the evolution of eukaryotes. BMC Evol Biol 12:32.

Tielens, AG, KW van Grinsven, K Henze, JJ van Hellemond, W Martin. 2010. Acetate formation in the energy metabolism of parasitic helminths and protists. Int J Parasitol 40:387-397.

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

Tong, J, P Dolezal, J Selkrig, S Crawford, AG Simpson, N Noinaj, SK Buchanan, K Gabriel, T Lithgow. 2011. Ancestral and derived protein import pathways in the mitochondrion of *Reclinomonas americana*. Mol Biol Evol 28:1581-1591.

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

Tovar, J, G Leon-Avila, LB Sanchez, R Sutak, J Tachezy, M van der Giezen, M Hernandez, M Muller, JM Lucocq. 2003. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. Nature 426:172-176.

Tsaousis, AD, D Gaston, A Stechmann, PB Walker, T Lithgow, AJ Roger. 2011. A functional Tom70 in the human parasite *Blastocystis* sp.: implications for the evolution of the mitochondrial import apparatus. Mol Biol Evol 28:781-791.

Tsaousis, AD, MM Leger, CAW Stairs, AJ Roger. 2012a. The biochemical adaptations of mitochondrion-related organelles of parasitic and free-living microbial eukaryotes to low oxygen environments. In: AV Altenbach, JM Bernhard, J Seckbach, editors. Anoxia: Evidence for Eukaryote Survival and Paleontological Strategies. Dordrecht: Springer. p. 51-81.

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

Tsaousis, AD, S Ollagnier de Choudens, E Gentekaki, et al. 2012b. Evolution of Fe/S cluster biogenesis in the anaerobic parasite *Blastocystis*. Proc Natl Acad Sci U S A 109:10426-10431.

Turner, NA, GA Biagini, D Lloyd. 1997. Anaerobiosis-induced differentiation of *Acanthamoeba castellanii*. FEMS Microbiology Letters 157:149-153.

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

van der Giezen, M, G Leon-Avila, J Tovar. 2005. Characterization of chaperonin 10 (Cpn10) from the intestinal human pathogen Entamoeba histolytica. Microbiology 151:3107-3115.

van der Giezen, M, J Tovar. 2005. Degenerate mitochondria. EMBO Rep 6:525-530.

van Dooren, GG, M Marti, CJ Tonkin, LM Stimmler, AF Cowman, GI McFadden. 2005. Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum*. Mol Microbiol 57:405-419.

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

van Grinsven, KW, JJ van Hellemond, AG Tielens. 2009. Acetate:succinate CoA-transferase in the anaerobic mitochondria of *Fasciola hepatica*. Mol Biochem Parasitol 164:74-79.

Van Hellemond, JJ, M Klockiewicz, CP Gaasenbeek, MH Roos, AG Tielens. 1995. Rhodoquinone and complex II of the electron transport chain in anaerobically functioning eukaryotes. J Biol Chem 270:31065-31070.

van Hoek, AH, AS Akhmanova, MA Huynen, JH Hackstein. 2000. A mitochondrial ancestry of the hydrogenosomes of Nyctotherus ovalis. Mol Biol Evol 17:202-206.

Vicente, JB, F Testa, D Mastronicola, E Forte, P Sarti, M Teixeira, A Giuffre. 2009. Redox properties of the oxygen-detoxifying flavodiiron protein from the human parasite *Giardia intestinalis*. Arch Biochem Biophys 488:9-13.

Vicente, JB, V Tran, L Pinto, M Teixeira, U Singh. 2012. A detoxifying oxygen reductase in the anaerobic protozoan *Entamoeba histolytica*. Eukaryot Cell 11:1112-1118.

Vieler, A, G Wu, CH Tsai, et al. 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. PLoS Genet 8:e1003064.

Vignais, PM, B Billoud. 2007. Occurrence, classification, and biological function of hydrogenases: an overview. Chem Rev 107:4206-4272.

Vignais, PM, B Billoud, J Meyer. 2001. Classification and phylogeny of hydrogenases. FEMS Microbiol Rev 25:455-501.

Vincent, KA, A Parkin, O Lenz, SP Albracht, JC Fontecilla-Camps, R Cammack, B Friedrich, FA Armstrong. 2005. Electrochemical definitions of O2 sensitivity and oxidative inactivation in hydrogenases. J Am Chem Soc 127:18179-18189.

Voncken, FG, B Boxma, AH van Hoek, AS Akhmanova, GD Vogels, M Huynen, M Veenhuis, JH Hackstein. 2002. A hydrogenosomal [Fe]-hydrogenase from the anaerobic chytrid *Neocallimastix* sp. L2. Gene 284:103-112.

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

Wakasugi, T, T Nagai, M Kapoor, et al. 1997. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. Proc Natl Acad Sci U S A 94:5967-5972.

Waller, RF, C Jabbour, NC Chan, N Celik, VA Likic, TD Mulhern, T Lithgow. 2009. Evidence of a reduced and modified mitochondrial protein import apparatus in microsporidian mitosomes. Eukaryot Cell 8:19-26.

Wallin, IE. 1926. Bacteria and the Origin of Species. Science 64:173-175.

Wang, X, J Huang, A Mukherjee, C Cao, J Lutkenhaus. 1997. Analysis of the interaction of FtsZ with itself, GTP, and FtsA. J Bacteriol 179:5551-5559.

Wawrzyniak, I, M Roussel, M Diogon, A Couloux, C Texier, KS Tan, CP Vivares, F Delbac, P Wincker, H El Alaoui. 2008. Complete circular DNA in the mitochondria-like organelles of *Blastocystis hominis*. Int J Parasitol 38:1377-1382.

Wexler-Cohen, Y, GC Stevens, E Barnoy, AM van der Bliek, PJ Johnson. 2014. A dynamin-related protein contributes to *Trichomonas vaginalis* hydrogenosomal fission. Faseb J 28:1113-1121.

Wienke, DC, ML Knetsch, EM Neuhaus, MC Reedy, DJ Manstein. 1999. Disruption of a dynamin homologue affects endocytosis, organelle morphology, and cytokinesis in *Dictyostelium discoideum*. Mol Biol Cell 10:225-243.

Wilding, M, A Fiorentino, ML De Simone, V Infante, L De Matteo, M Marino, B Dale. 2002. Energy substrates, mitochondrial membrane potential and human preimplantation embryo division. Reprod Biomed Online 5:39-42.

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

Williams, BA, PJ Keeling. 2003. Cryptic organelles in parasitic protists and fungi. Adv Parasitol 54:9-68.

Williams, BA, PJ Keeling. 2005. Microsporidian mitochondrial proteins: expression in *Antonospora locustae* spores and identification of genes coding for two further proteins. J Eukaryot Microbiol 52:271-276.

Williams, K, PN Lowe, PF Leadlay. 1987. Purification and characterization of pyruvate: ferredoxin oxidoreductase from the anaerobic protozoon *Trichomonas vaginalis*. Biochem J 246:529-536.

Williams, TA, PG Foster, TM Nye, CJ Cox, TM Embley. 2012. A congruent phylogenomic signal places eukaryotes within the Archaea. Proc Biol Sci 279:4870-4879.

Woese, CR, O Kandler, ML Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87:4576-4579.

Wojtkowska, M, D Buczek, O Stobienia, A Karachitos, M Antoniewicz, M Slocinska, W Makalowski, H Kmita. 2015. The TOM complex of Amoebozoans: the cases of the amoeba *Acanthamoeba castellanii* and the slime mold *Dictyostelium discoideum*. Protist 166:349-362.

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

Yarlett, N, AC Hann, D Lloyd, A Williams. 1981. Hydrogenosomes in the rumen protozoon *Dasytricha ruminantium* Schuberg. Biochem J 200:365-372.

Yarlett, N, CG Orpin, EA Munn, NC Yarlett, CA Greenwood. 1986. Hydrogenosomes in the rumen fungus *Neocallimastix patriciarum*. Biochem J 236:729-739.

Yoshida, Y, H Kuroiwa, S Hirooka, T Fujiwara, M Ohnuma, M Yoshida, O Misumi, S Kawano, T Kuroiwa. 2009. The bacterial ZapA-like protein ZED is required for mitochondrial division. Curr Biol 19:1491-1497.

Yu, HS, HJ Jeong, YC Hong, SY Seol, DI Chung, HH Kong. 2007. Natural occurrence of *Mycobacterium* as an endosymbiont of *Acanthamoeba* isolated from a contact lens storage case. Korean J Parasitol 45:11-18.

Zimorski, V, P Major, K Hoffmann, XP Bras, WF Martin, SB Gould. 2013. The N-terminal sequences of four major hydrogenosomal proteins are not essential for import into hydrogenosomes of *Trichomonas vaginalis*. J Eukaryot Microbiol 60:89-97.

Zinder, SH, M Koch. 1984. Non-aceticlastic methanogenesis from acetate: acetate oxidation by a thermophilic syntrophic coculture. Arch Microbiol 138:263-272.

Zubacova, Z, L Novak, J Bublikova, V Vacek, J Fousek, J Ridl, J Tachezy, P Dolezal, C Vlcek, V Hampl. 2013. The mitochondrion-like organelle of *Trimastix pyriformis* contains the complete glycine cleavage system. PLoS One 8:e55417.

Appendix A. Supplementary Material for Chapter 2

Supporting Methods

Anaerobic induction

Prior to anaerobic induction, Neff base medium, glucose and ferric citrate were placed in an anaerobic chamber (Forma Scientific Anaerobic System model 1024) containing 79.8% N_2, 10.4% H_2, 9.8% CO_2, and allowed to de-gas for 24 hr. High-density *A. castellanii* cells were subjected to centrifugation at 200*g,* and the supernatant was replaced with de-gassed medium, supplemented with glucose, ferric citrate, vitamins and $CaCl_2$, inside the anaerobic chamber.

Antibody production

[FeFe]-hydrogenase was amplified from cDNA and cloned into pET-16b (Novagen) downstream of a 6xHis tag for recombinant expression in OverExpress™ C41(DE) cells, kindly provided by Prof. John Walker (Medical Research Council Mitochondrial Biology Unit, Cambridge, UK). The recombinant protein was expressed in inclusion bodies, which were purified using BugBuster reagent (Novagen). An antibody against the purified inclusion bodies was raised in rats by GenScript Corporation (Piscataway, NJ, USA).

Western blotting

The anti-[FeFe]-hydrogenase antibody was tested against inclusion bodies from C41(DE) cells expressing recombinant *A. castellanii* [FeFe]-hydrogenase from the pET-16b vector, or the empty pET-16b vector. Proteins were blotted onto PVDF membranes and blocked overnight at 4°C in 5% milk. To remove cross-reaction of the primary antibody, and to confirm the identity of the bound

protein, antibody competition assays were performed. The anti-*A. castellanii* [FeFe]-hydrogenase antibody was incubated overnight at 4°C and for a further 90 min at room temp the following day with either BugBuster reagent, inclusion bodies from cells expressing empty pET-16b vector in BugBuster reagent, or inclusion bodies from cells expressing recombinant *A. castellanii* [FeFe]-hydrogenase from pET-16b. Following competition, blots were incubated with primary antibody at a ratio of 1:50000 in 1% milk for 1 hr at room temp, washed, and incubated with horseradish peroxidase-conjugated secondary anti-rat IgG antibody (Sigma) 1:2000 for 1 hr at room temp, washed again, and incubated with ECL western blotting reagents (Amersham).

Immunolocalization

Anaerobically induced cells were subjected to centrifugation at 100*g* for 2 min, and fixed for 1 hr with 4% paraformaldehyde/0.5% glutaraldehyde diluted with 0.1 M sodium cacodylate buffer. In order to prevent the living cells from being exposed to oxygen, all steps up to and including fixation were performed either inside the anaerobic chamber itself, or in containers that had been sealed inside the anaerobic chamber. Fixed cells were rinsed three times for a minimum of 10 min each with 0.1 M sodium cacodylate buffer, and dehydrated with a graduated ethanol series. The dehydrated samples were then embedded in 100% LR White resin and cured for 48 hr in a 60°C oven. Thin sections were cut using an LKB Huxley ultramicrotome with a diamond knife, and placed onto 300 mesh nickel grids.

Sections were blocked overnight at 4°C by incubating the grids on droplets of blocking agent (phosphate-buffered saline pH 7.4 containing 0.8% bovine serum

albumin and 0.01% Tween 20), and incubated for 3 hr at room temp on droplets of anti-[FeFe]-hydrogenase primary antibody diluted 1:10 in blocking agent. The grids were washed four times for 10 min each on droplets of blocking agent, then incubated for 1 hr on droplets of gold-conjugated goat anti-rat IgG secondary antibody conjugated to 10 nm gold particles (Sigma; Electron Microscopy Services) diluted 1:20 in blocking agent. Following incubation with the secondary antibody, grids were washed three times for 10 min each in blocking agent, then rinsed three times for 30 sec each in sterile ddH_2O. Antibody-stained grids were stained for 10 min with 2% aqueous uranyl acetate, rinsed twice with distilled water for 5 min each, stained for 4 min with lead citrate, rinsed, and air-dried. The sections were viewed using a JEOL JEM 1230 transmission electron microscope at 80 kV, and images were captured using a Hamamatsu ORCA-HR digital camera.

The areas of the nucleus, mitochondria and cytosol of each cell cross-section were measured using ImageJ, and the gold particles in each part of the cell were counted by eye.

Supporting Results

[FeFe]-hydrogenase localizes to the mitochondria

To experimentally examine the localization of [FeFe]-hydrogenase, the characteristic hydrogenosomal metabolism enzyme, we raised an antibody against recombinant *A. castellanii* [FeFe]-hydrogenase expressed in *E. coli*. The resulting antibody recognized the expressed recombinant enzyme on western blots (Figure S2). Attempts to detect bands in whole cell lysate or crude mitochondrial preparations from *A. castellanii* were unsuccessful (data not

shown). This failure is likely attributable to exceptionally low expression levels of this enzyme; the very high concentration of antibody required for localization in immunoelectron microscopy experiments would seem to support this inference. Upon exposure to oxygen, [FeFe]-hydrogenases typically become rapidly and irreversibly inactivated, and are degraded (Vincent et al. 2005). For this reason, we performed immunogold labeling experiments on *A. castellanii* cells that had been exposed to anaerobic conditions for 6 or 24 hr. Antibody staining in these cells was higher in mitochondria than in the cytosol (approx. 2.9-fold) or the nucleus (approx. 1.7-fold), consistent with the presence of a predicted mitochondrial targeting peptide for [FeFe]-hydrogenase (Figures S3, S4). Antibody staining was also enriched in the nucleus compared with the cytosol (approx. 1.8-fold), although to a much lesser extent than in mitochondria. A distant homolog of [FeFe]-hydrogenase, nuclear prelamin A recognition factor (NARF), is localized to the nucleus in some organisms (Barton, Worman 1999; Horner, Foster, Embley 2000); cross-reaction with a nuclear NARF homolog may therefore explain this elevated staining pattern.

Table A.S1 Accession numbers of genes and ESTs encoding anaerobic energy generation enzymes.

	Genomic DNA NCBI accession numbers	**EST NCBI accession numbers**
[FeFe]-hydrogenase	AEYA01000219, (75972 – 75673, 75581 – 75413, 75351 – 75282, 75212 – 75041, 74960 – 74885, 74787 – 74630, 74547 – 74221, 74156 – 73995, 73905 – 73487, 73421 – 73279, 73195 – 73115)	EC110970, EC109037, GANQ00000000
PFO	AEYA01001097 (15473 – 15542, 15611 – 15727, 15806 – 15924, 16047 – 16397, 16494 – 17192, 17297 – 17400, 17510 – 18330, 18435 – 18558, 19407 – 19554, 19630 – 20010)	EC107872, GANQ00000000
HydE	AEYA01000219 (120492 – 120286, 120217 – 120163, 120101 – 119992, 119912 – 119849, 119787 – 119712, 119627 – 119387, 119326 – 119073, 118996 – 118803, 118730 – 118489)	EC100760, EC110418, EC106931, GANQ00000000
HydF	AEYA01000219 (120678 – 120792, 120860 – 120966, 120137 – 121153, 121228 – 121326, 121395 – 121532, 121601 – 121719, 121795 – 121907, 121970 – 122342, 122410 – 122701)	GANQ00000000
HydG	AEYA01000219 (103720 – 103594, 103484 – 103336, 103238 – 102957, 102857 – 102737, 102658 – 101825, 101734 – 101583)	EC108023, EC108506, GANQ00000000
ASCT 1A	AEYA01002158 (112023 – 111879, 111691 – 111528, 111428 – 111284. 111178 – 111086, 110997 – 110914, 110803 – 110631, 110497 – 110329, 110257 – 110153, 110030 – 109923, 109815 – 109735, 109566 – 109491, 109418 – 109332. 109220 – 109140, 109056 – 108922)	EC108981, EC106432, GANQ00000000
ASCT 1B	AEYA01000777 (11122 – 11029, 10854 – 10782, 10599 – 10509, 10225 – 10043, 9865 – 9752, 9628 – 9510, 9356 – 9203, 9112 – 8934, 8822 – 8710, 8609 – 8497, 8335 – 8150)	EC101929, GANQ00000000

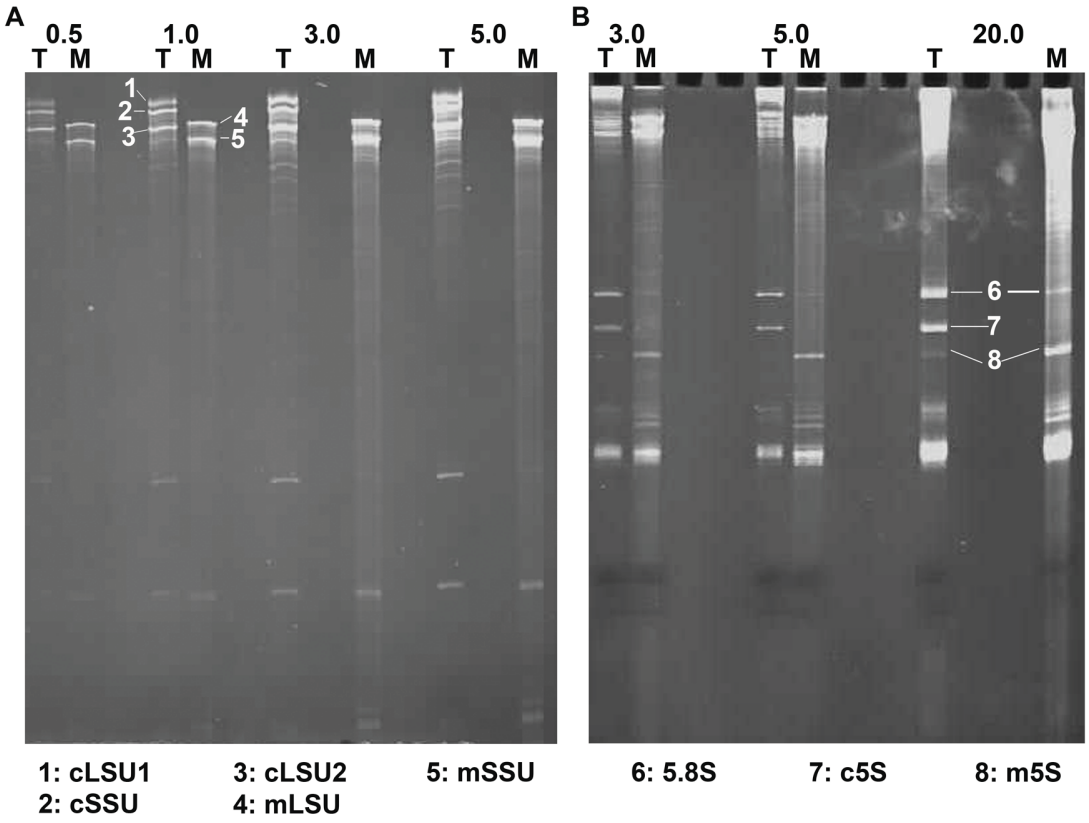


Figure A.S1 Assessment of mitochondrial purity by comparing rRNA profiles of total cellular RNA and total mitochondrial RNA from *A. castellanii*. A) An EtBr-stained 6% (w:v) acrylamide, 7 M urea gel loaded with varying quantities of total cellular (T) and total mitochondrial (M) RNA. Numbers are quantities of RNA (in **µ**g) loaded in each lane. The profiles of large rRNA species in T and M are distinct, suggesting that the mitochondrial fraction is relatively free of contaminating cytosolic rRNA (and presumably cytosolic ribosomal proteins). *A. castellanii* cytosolic (c) and mitochondrial (m) LSU and SSU species are identified (note that the cLSU is split). B) 10% acrylamide gel as in A). Note that a mitochondrial 5S rRNA (m5S) is visualized in the M lane and, to a much lower extent, in the T lane. A band corresponding in size to cytosolic 5.8S rRNA is visible in an overloaded M lane; however, no cytosolic 5S (c5S) is detectable.

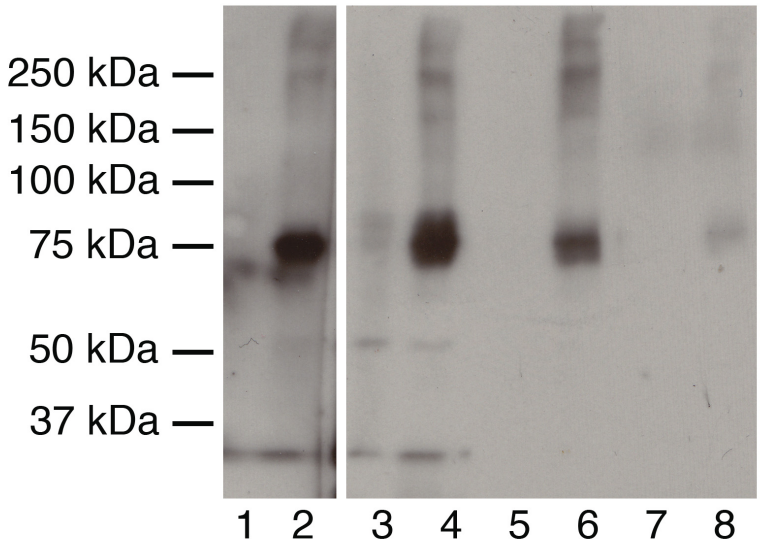


Figure A.S2 Western blot showing recognition of the recombinant [FeFe]-hydrogenase by the homologous anti-[FeFe]-hydrogenase antibody. Lanes 1, 3, 5 and 7: inclusion bodies from C41(DE) cells expressing empty pET-16b vector. Lanes 2, 4, 6 and 8: inclusion bodies from C41(DE) cells expressing recombinant *A. castellanii* [FeFe]-hydrogenase from pET-16b. Lanes 1 and 2: anti-His-tag antibody, exposed for a shorter period of time than lanes 3-8 in order to avoid overexposure. Lanes 3 and 4: anti [FeFe]-hydrogenase antibody only. Lanes 5 and 6: anti-[FeFe]-hydrogenase antibody incubated with inclusion bodies from C41(DE) cells expression empty pET-16b. Lanes 7 and 8: anti-[FeFe]-hydrogenase antibody incubated with inclusion bodies from C41(DE) cells expressing recombinant *A. castellanii* [FeFe]-hydrogenase from pET-16b.

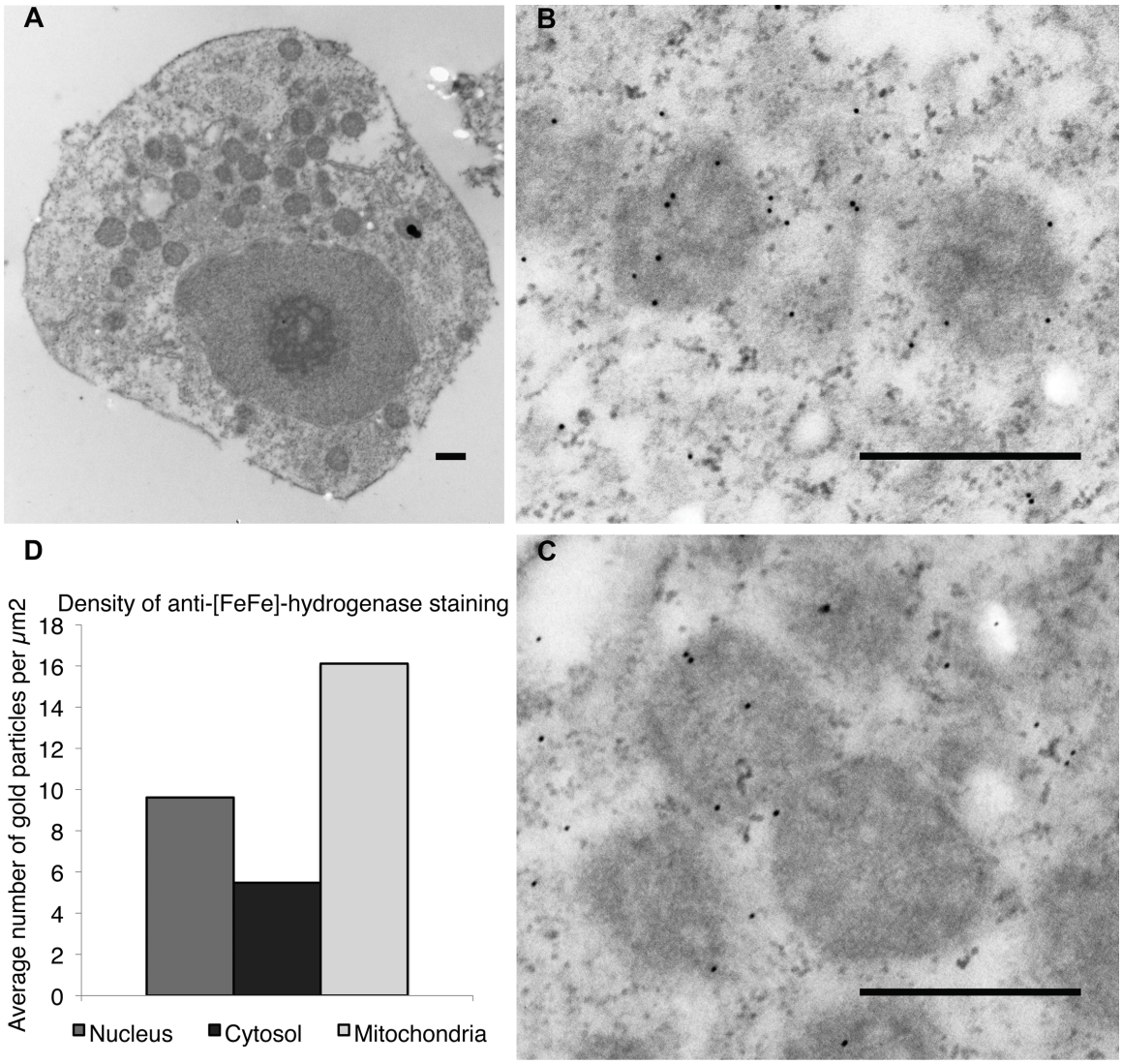


Figure A.S3 Immunogold localization of [FeFe]-hydrogenase in *A. castellanii* trophozoites exposed to anaerobic conditions for 24 hr. A. Whole cell fixed for immunogold staining; scale bar, 500 nm; this image has been cropped in order to show only the whole cell from which the insets shown were derived. B. and C. Magnified sections from the cell depicted in (A), showing gold particles corresponding to [FeFe]-hydrogenase localization; scale bar, 500 nm. D. Mean density of immunogold labeling in the cytosol, nucleus and mitochondria (8 cells). Brightness and contrast have been adjusted in each image to enhance visibility of the mitochondria and gold particles.

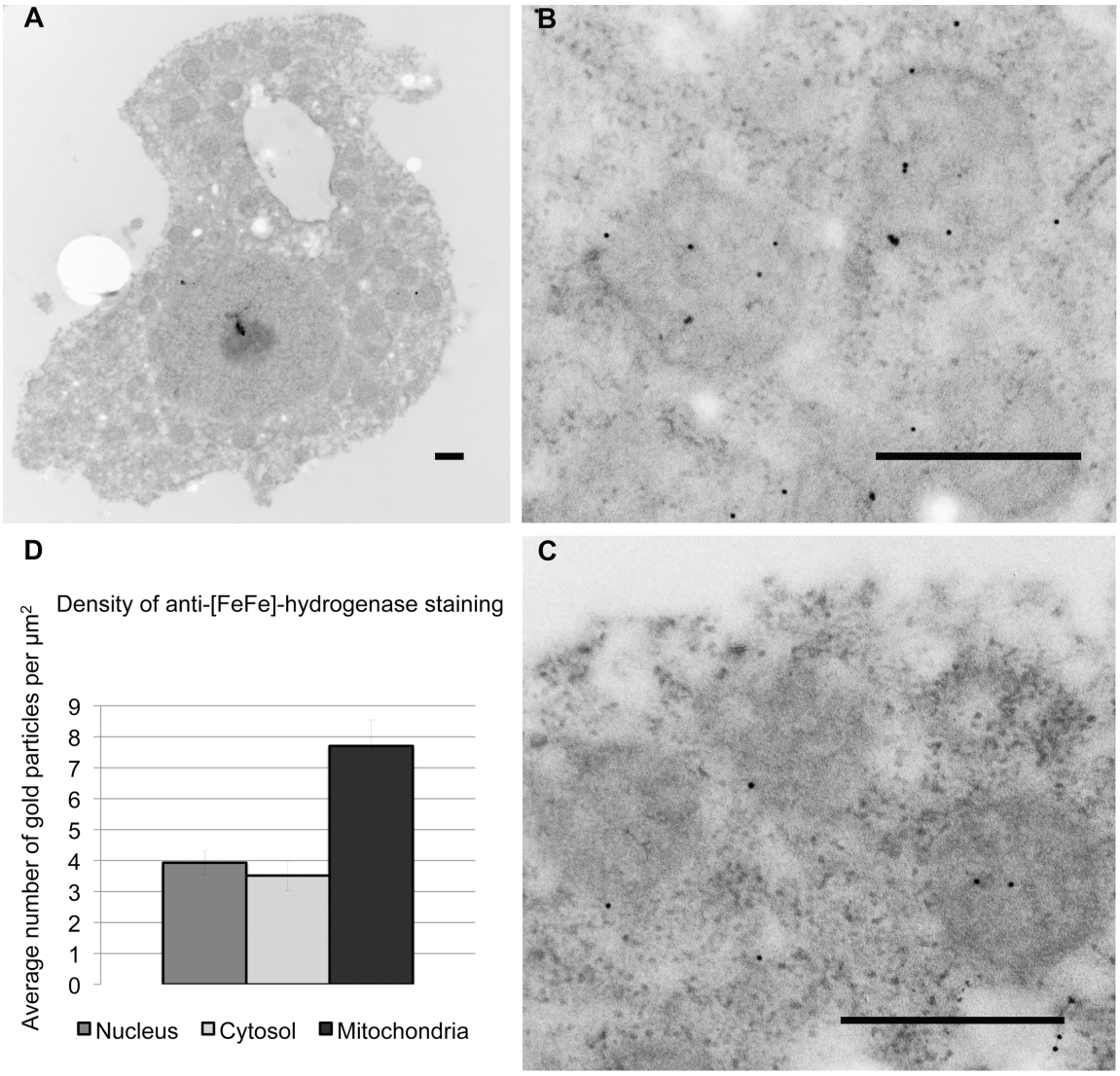
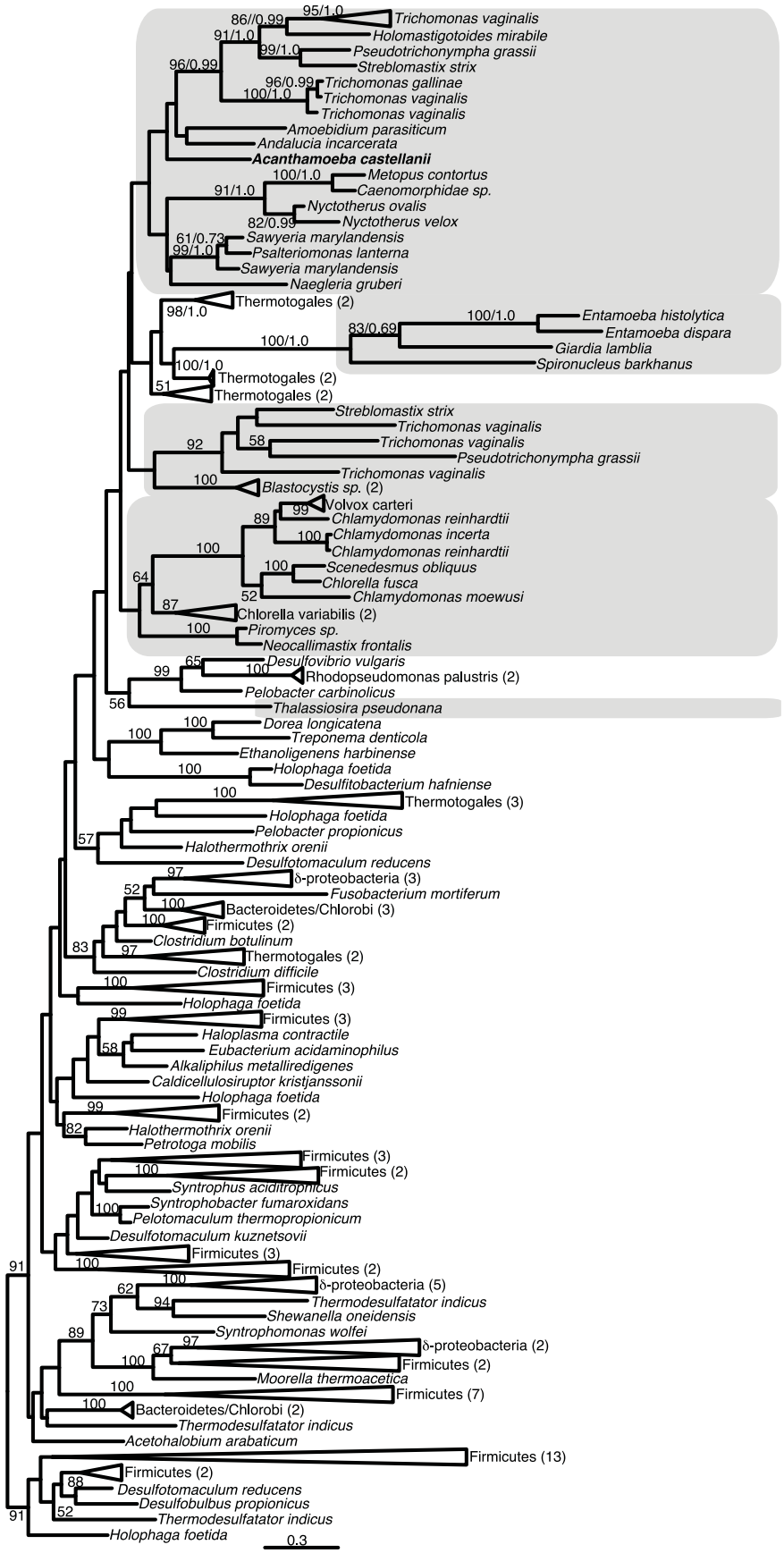


Figure A.S4 Immunogold localization of [FeFe]-hydrogenase in *A. castellanii* trophozoites exposed to anaerobic conditions for 6 hr. A. Whole cell fixed for immunogold staining; scale bar, 500 nm; this image has been cropped in order to show only the whole cell from which the insets shown were derived. B. and C. Magnified sections from the cell depicted in (A), showing gold particles corresponding to [FeFe]-hydrogenase localization; scale bar, 500 nm. D. Mean density of immunogold labeling in the cytosol, nucleus and mitochondria (7 cells). Brightness and contrast have been adjusted in each image to enhance visibility of the mitochondria and gold particles.

Figure A.S5 Phylogeny of [FeFe]-hydrogenase in eukaryotes and bacteria excluding long-branching taxa. **(Following page)** Taxa forming a long-branching clade, corresponding to Clade B identified by Hug et al. (2010), have been excluded from these analyses. The topology shown is the ML tree generated by RAxML analyses; 328 sites were examined across 151 taxa. Bootstrap support values ≥50% and posterior probabilities ≥0.5 are shown. Eukaryotes are shaded gray.


Trichomonas vaginalis
Holomastigotoides mirabile
Pseudotrichonympha grassii
Streblomastix strix
Trichomonas gallinae
Trichomonas vaginalis
Trichomonas vaginalis
Amoebidium parasiticum
Andalucia incarcerata
Acanthamoeba castellanii
Metopus contortus
Caenomorphidae sp.
Nyctotherus ovalis
Nyctotherus velox
Sawyeria marylandensis
Psalteriomonas lanterna
Sawyeria marylandensis
Naegleria gruberi
Thermotogales (2)
Entamoeba histolytica
Entamoeba dispara
Giardia lamblia
Spironucleus barkhanus
Thermotogales (2)
Thermotogales (2)
Streblomastix strix
Trichomonas vaginalis
Trichomonas vaginalis
Pseudotrichonympha grassii
Trichomonas vaginalis
Blastocystis sp. (2)
Volvox carteri
Chlamydomonas reinhardtii
Chlamydomonas incerta
Chlamydomonas reinhardtii
Scenedesmus obliquus
Chlorella fusca
Chlamydomonas moewusi
Chlorella variabilis (2)
Piromyces sp.
Neocallimastix frontalis
Desulfovibrio vulgaris
Rhodopseudomonas palustris (2)
Pelobacter carbinolicus
Thalassiosira pseudonana
Dorea longicatena
Treponema denticola
Ethanoligenens harbinense
Holophaga foetida
Desulfitobacterium hafniense
Thermotogales (3)
Holophaga foetida
Pelobacter propionicus
Halothermothrix orenii
Desulfotomaculum reducens
δ-proteobacteria (3)
Fusobacterium mortiferum
Bacteroidetes/Chlorobi (3)
Firmicutes (2)
Clostridium botulinum
Thermotogales (2)
Clostridium difficile
Firmicutes (3)
Holophaga foetida
Firmicutes (3)
Haloplasma contractile
Eubacterium acidaminophilus
Alkaliphilus metalliredigenes
Caldicellulosiruptor kristjanssonii
Holophaga foetida
Firmicutes (2)
Halothermothrix orenii
Petrotoga mobilis
Firmicutes (3)
Firmicutes (2)
Syntrophus aciditrophicus
Syntrophobacter fumaroxidans
Pelotomaculum thermopropionicum
Desulfotomaculum kuznetsovii
Firmicutes (3)
Firmicutes (2)
δ-proteobacteria (5)
Thermodesulfatator indicus
Shewanella oneidensis
Syntrophomonas wolfei
δ-proteobacteria (2)
Firmicutes (2)
Moorella thermoacetica
Firmicutes (7)
Bacteroidetes/Chlorobi (2)
Thermodesulfatator indicus
Acetohalobium arabaticum
Firmicutes (13)
Firmicutes (2)
Desulfotomaculum reducens
Desulfobulbus propionicus
Thermodesulfatator indicus
Holophaga foetida
0.3

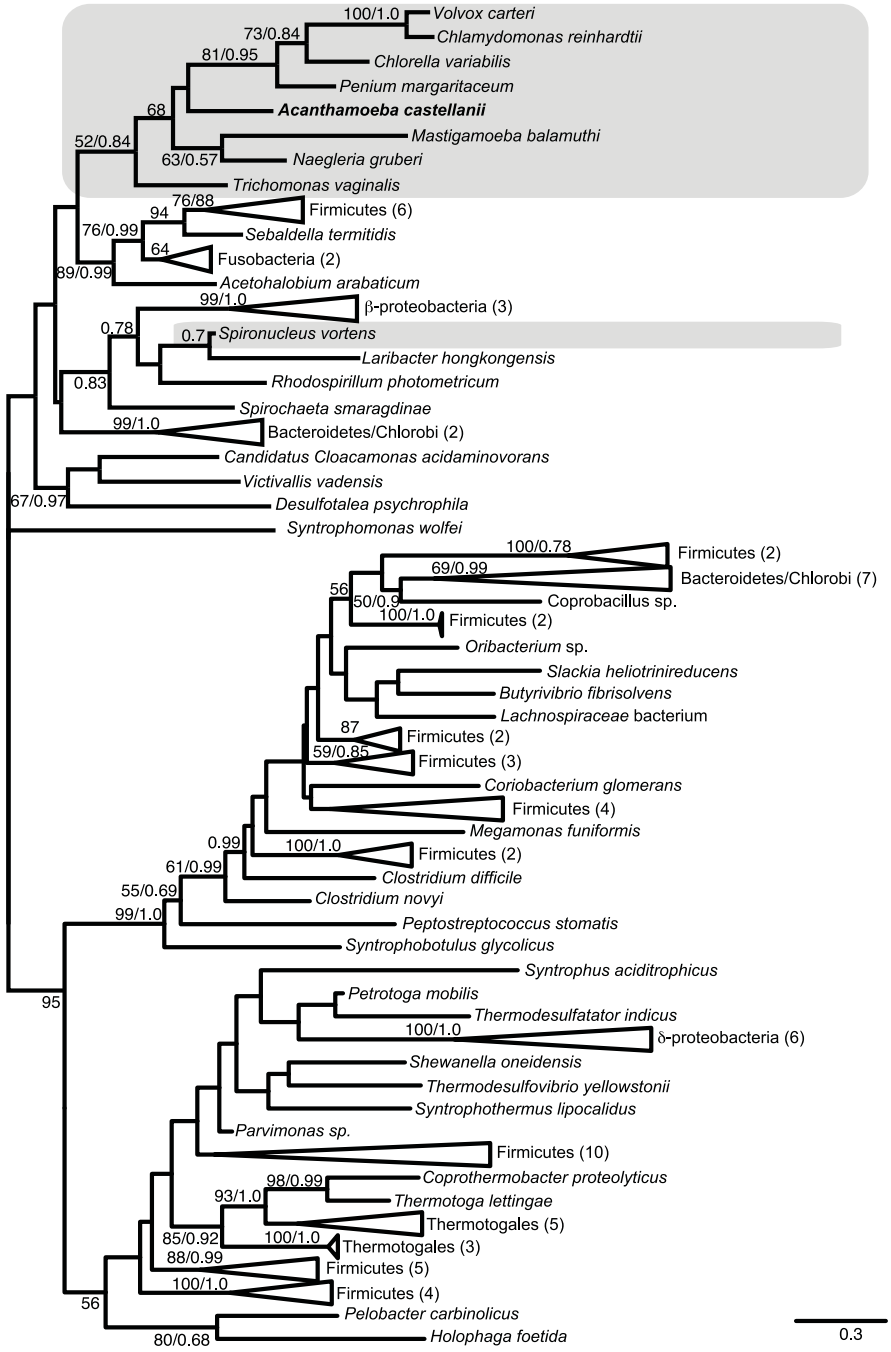


Figure A.S6 Phylogeny of HydE in eukaryotes and bacteria. The topology shown is the ML tree generated by RAxML analyses; 264 sites were examined across 109 taxa. Bootstrap support values ≥50% and posterior probabilities ≥0.5 are shown. Eukaryotes are shaded gray.

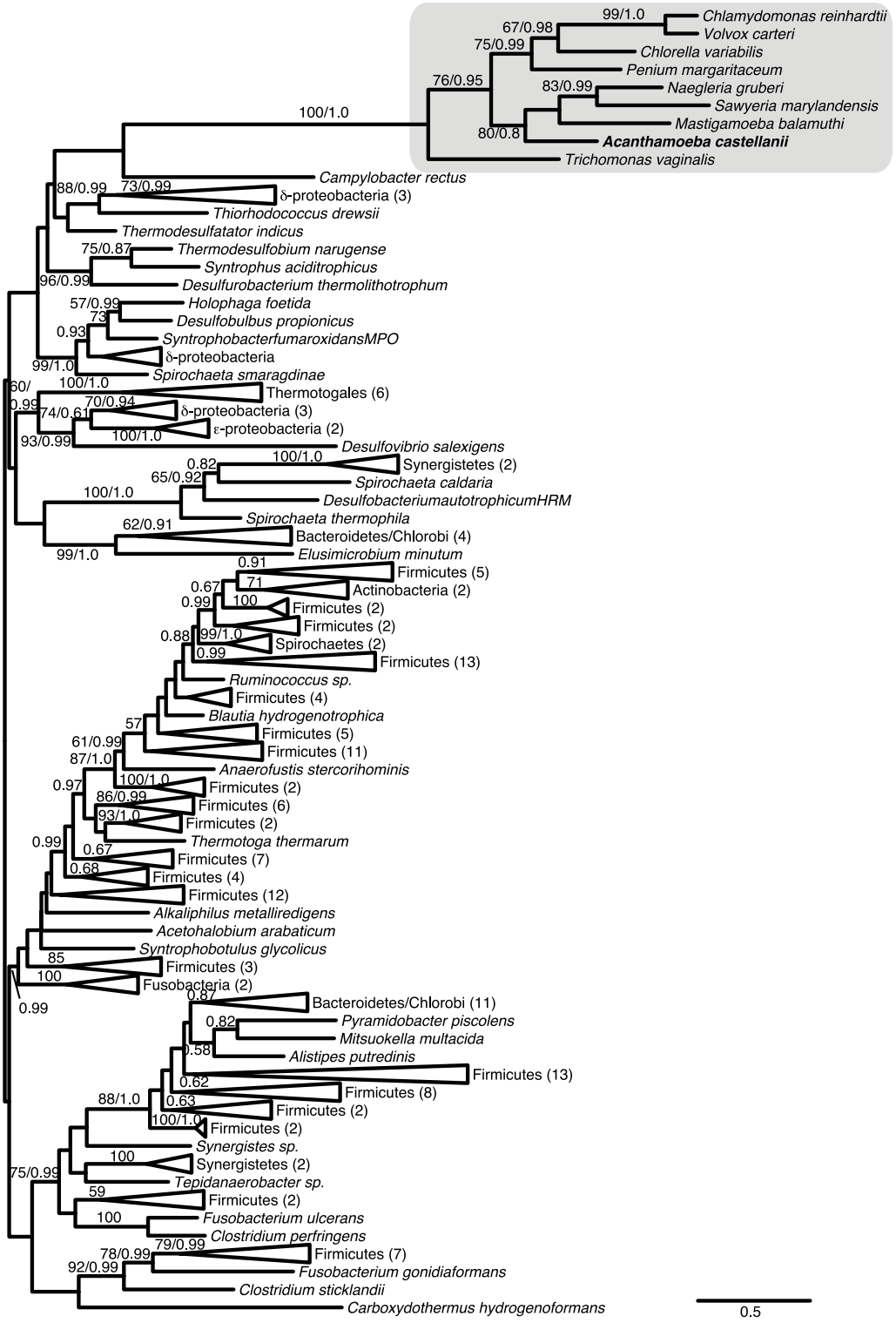


Figure A.S7 Phylogeny of HydF in eukaryotes and bacteria. The topology shown is the ML tree generated by RAxML analyses; 307 sites were examined across 196 taxa. Bootstrap support values ≥50% and posterior probabilities ≥0.5 are shown. Eukaryotes are shaded gray.

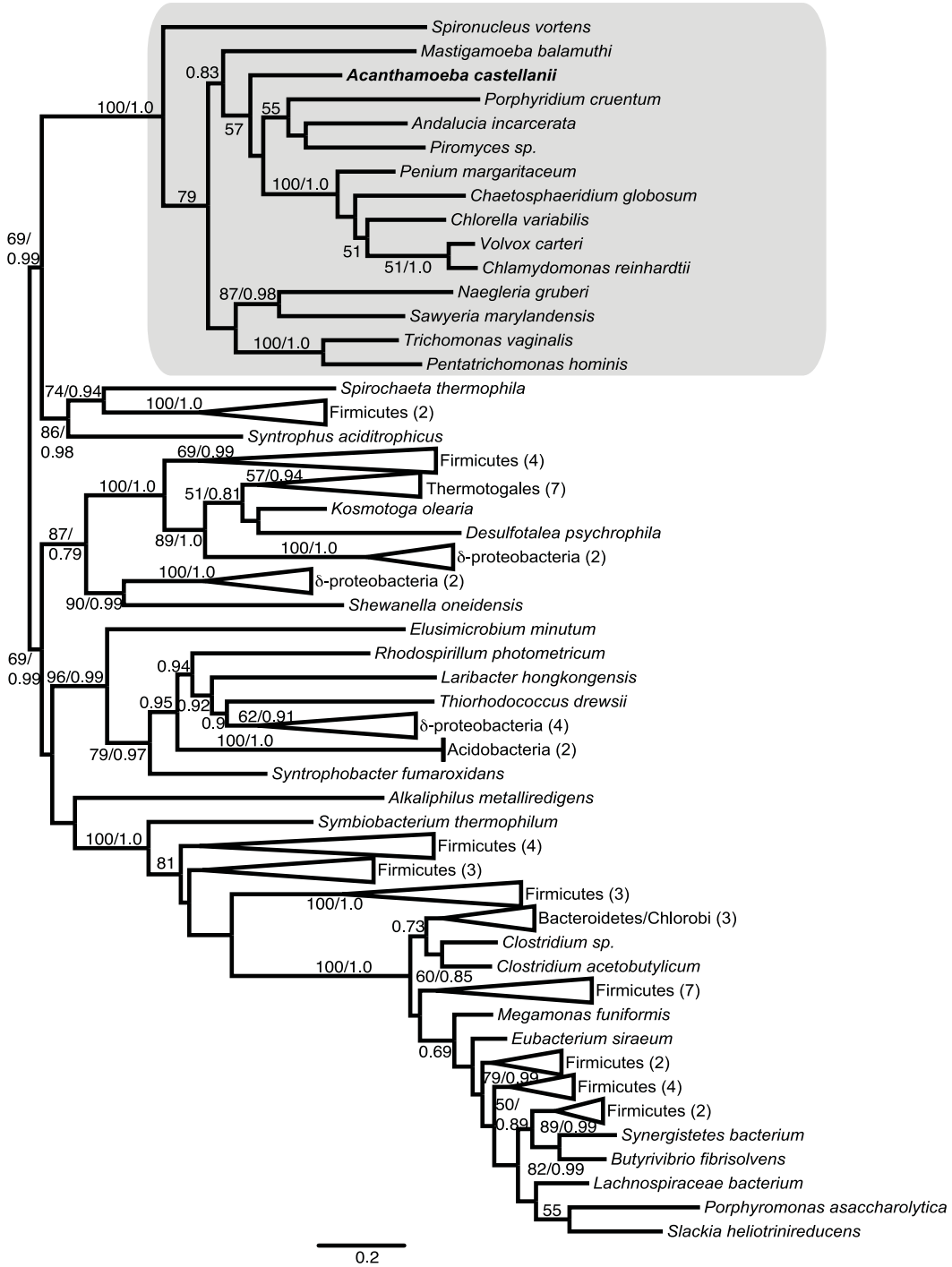
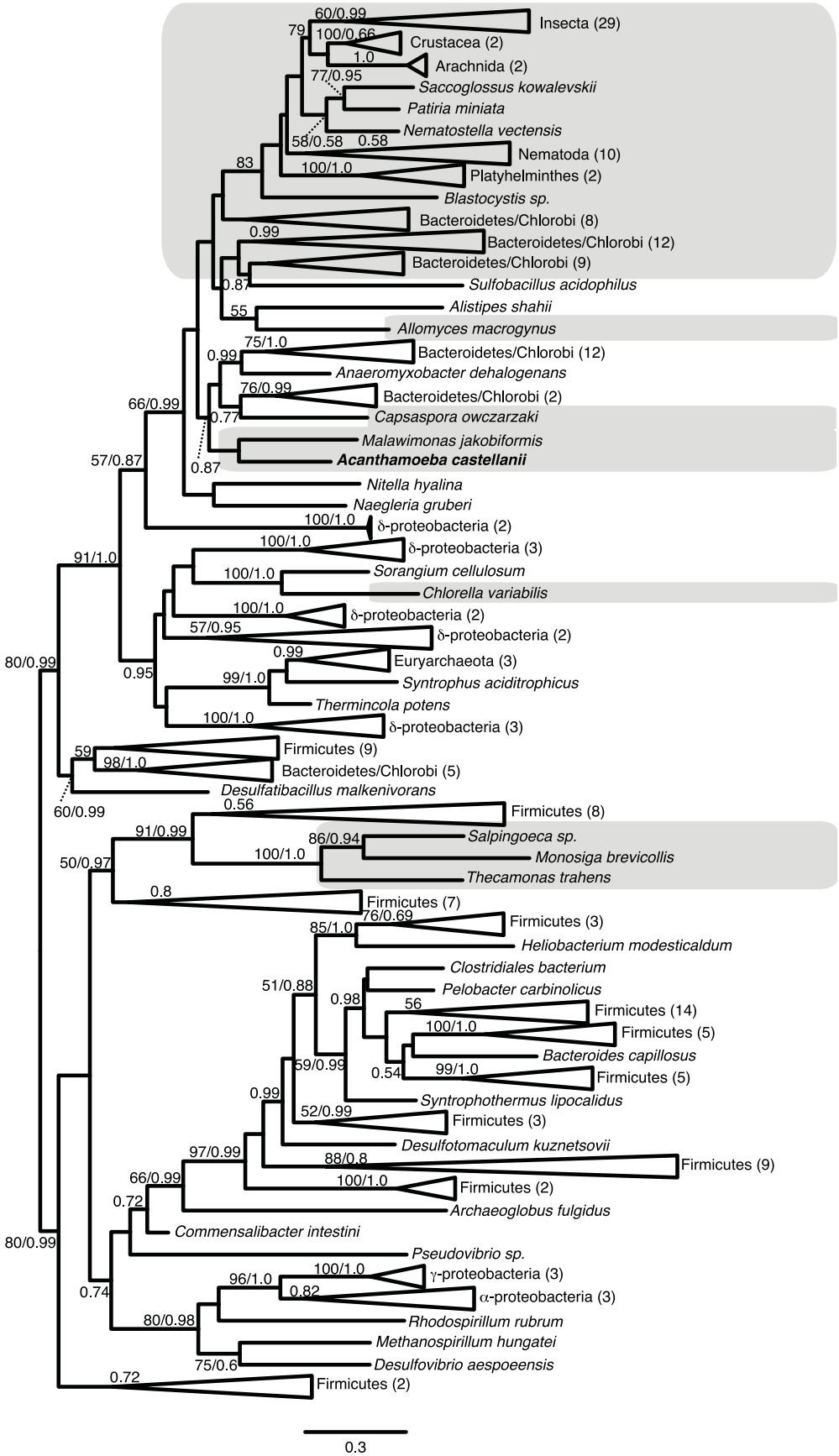


Figure A.S8 Phylogeny of HydG in eukaryotes and bacteria. The topology shown is the ML tree generated by RAxML analyses. 391 sites were examined across 87 taxa; bootstrap support values greater than or equal to 50%, and posterior probabilities greater than or equal to 0.5, are shown. Eukaryotes are shaded gray.

Figure A.S9 Phylogeny of ASCT1B in eukaryotes and bacteria. **(Following page)** The topology shown is the ML tree generated by RAxML analyses; 315 sites were examined across 221 taxa. Bootstrap support values ≥50% and posterior probabilities ≥0.5 are shown. Eukaryotes are shaded gray.


Insecta (29)
Crustacea (2)
Arachnida (2)
Saccoglossus kowalevskii
Patiria miniata
Nematostella vectensis
Nematoda (10)
Platyhelminthes (2)
Blastocystis sp.
Bacteroidetes/Chlorobi (8)
Bacteroidetes/Chlorobi (12)
Bacteroidetes/Chlorobi (9)
Sulfobacillus acidophilus
Alistipes shahii
Allomyces macrogynus
Bacteroidetes/Chlorobi (12)
Anaeromyxobacter dehalogenans
Bacteroidetes/Chlorobi (2)
Capsaspora owczarzaki
Malawimonas jakobiformis
Acanthamoeba castellanii
Nitella hyalina
Naegleria gruberi
δ-proteobacteria (2)
δ-proteobacteria (3)
Sorangium cellulosum
Chlorella variabilis
δ-proteobacteria (2)
δ-proteobacteria (2)
Euryarchaeota (3)
Syntrophus aciditrophicus
Thermincola potens
δ-proteobacteria (3)
Firmicutes (9)
Bacteroidetes/Chlorobi (5)
Desulfatibacillus malkenivorans
Firmicutes (8)
Salpingoeca sp.
Monosiga brevicollis
Thecamonas trahens
Firmicutes (7)
Firmicutes (3)
Heliobacterium modesticaldum
Clostridiales bacterium
Pelobacter carbinolicus
Firmicutes (14)
Firmicutes (5)
Bacteroides capillosus
Firmicutes (5)
Syntrophothermus lipocalidus
Firmicutes (3)
Desulfotomaculum kuznetsovii
Firmicutes (9)
Firmicutes (2)
Archaeoglobus fulgidus
Commensalibacter intestini
Pseudovibrio sp.
γ-proteobacteria (3)
α-proteobacteria (3)
Rhodospirillum rubrum
Methanospirillum hungatei
Desulfovibrio aespoeensis
Firmicutes (2)
0.3

Appendix B. Supplementary Material for Chapter 3

Table B.S1 Presence and absence of bacterial Min proteins and FtsZ in eukaryotes. Blue, proteobacteria-derived proteins; green, cyanobacteria-derived proteins; - and grey, no protein found encoded in complete genome data; ?, no protein found encoded in transcriptome or incomplete genome data. P, reported as plastid-encoded in the literature (4-16); P*, reported as chromatophore-encoded in the literature (17); N, reported as nucleomorph-encoded in the literature (18-20); M, TargetP-predicted mitochondrial location and targeting peptide (with mitochondrial targeting probability ≥ 0.5); C, TargetP-predicted plastid location and targeting peptide (with plastid targeting probability ≥ 0.5); +, no predicted subcellular localization, or incomplete sequence; †, predicted pseudogene. TargetP analyses were carried out using 'plant' parameters for plastid-bearing taxa, and 'non-plant' parameters for taxa lacking plastids. It should be noted that these parameters could give rise to artefacts in taxa with secondary plastids, as nuclear-encoded plastid proteins in these taxa have bipartite targeting peptides that TargetP is not designed to take into account. Mitochondrial or plastid predictions are based on single-protein phylogenies, predicted subcellular localization, and localization in yeast (*Andalucia incarcerata, Dictyostelium discoideum*). *chromatophore-encoded †predicted pseudogene

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Obazoa									
Breviatea									
Pygsuia biforma	?	?	M	?	?	?	M	?	GenBank
Apusomonadida									
Thecamonas trahens	M	-	M	-	-	-	+	-	Broad Institute
Opisthokonta									
Allomyces macrogynus	-	-	-	-	-	-	-	-	Broad Institute
Amoebidium parasiticum	?	?	?	?	?	?	?	?	GenBank
Capsaspora owczarzaki	-	-	-	-	-	-	-	-	Broad Institute
Fonticula alba	-	-	-	-	-	-	-	-	Broad Institute
Monosiga brevicolis	-	-	-	-	-	-	-	-	Broad Institute
Mortierella verticillata	-	-	-	-	-	-	-	-	Broad Institute
Salpingoeca rosetta	-	-	-	-	-	-	-	-	Broad Institute
Sphaeroforma arctica	-	-	-	-	-	-	-	-	Broad Institute
Spizellomyces punctatus	-	-	-	-	-	-	-	-	Broad Institute
Fungi, Metazoa, other Opisthokonta	-	-	-	-	-	-	-	-	GenBank

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Amoebozoa									
Dictyostelia									
Acytostelium subglobosum	M	-	M	-	M	-	+	-	GenBank
Dictyostelium citrinum	-	-	-	-	-	-	+	-	GenBank
Dictyostelium discoideum	-	-	-	-	-	-	M	-	DictyBase, GenBank
Dictyostelium fasciculatum	-	-	-	-	-	-	M	-	DictyBase, GenBank
Dictyostelium firmibasis	-	-	-	-	-	-	M	-	GenBank
Dictyostelium intermedium	-	-	-	-	-	-	M	-	GenBank
Dictyostelium purpureum	+	-	+	-	+	-	+	-	DictyBase, JGI, GenBank
Physarum polycephalum	+	?	?	?	?	?	+	?	GenBank
Polysphondylium pallidum	+	-	+	-	+	-	+	-	DictyBase, GenBank
Polysphondylium violaceum	+	-	+	-	+	-	+	-	GenBank
Archamoebae									
Entamoeba dispar SAW760	-	-	-	-	-	-	-	-	EuPathDB
Entamoeba histolytica	-	-	-	-	-	-	-	-	EuPathDB
Entamoeba invadens IP1	-	-	-	-	-	-	-	-	EuPathDB
Entamoeba moshkovskii Laredo	-	-	-	-	-	-	-	-	EuPathDB
Entamoeba nuttalli	-	-	-	-	-	-	-	-	EuPathDB
Mastigamoeba balamuthi	-	-	-	-	-	-	-	-	GenBank
Gracilipodida									
Filamoeba nolandi NC-AS-23-1	?	?	?	?	?	?	?	?	MMETSP
Discosea									
Acanthamoeba castellanii	-	-	-	-	-	-	-	-	GenBank
Mayorella sp. BSH-02190019	?	?	?	?	?	?	?	?	MMETSP
Paramoeba aestuarina (listed in MMETSP records by an older name , *Neoparamoeba aestuarina* SoJaBio B1-5/56/2 (1))	?	?	?	?	?	?	?	?	MMETSP
Paramoeba atlantica 621/1 / CCAP 1560/9	?	?	?	?	?	?	?	?	MMETSP
Pessonella sp. PRA-29	?	?	?	?	?	?	?	?	MMETSP
Sapocribrum chincoteaguense (misidentified in MMETSP records as *Sexangularia* sp. ATCC50979 (2))	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Stygamoeba regulata BSH-02190019	?	?	?	?	?	?	?	?	MMETSP
Trichosphaerium sp Am-I-7 wt	?	?	?	?	?	?	?	?	MMETSP
Vannella robusta DIVA3 518/3/11/1/6	?	?	?	?	?	?	?	?	MMETSP
Vannella sp. DIVA3 517/6/12	+	?	M	?	?	?	+	?	MMETSP
Vexillifera sp. DIVA3 564/2	?	?	?	?	?	?	?	?	MMETSP
Stereomyxa									
Stereomyxa ramosa Chinc5	?	?	?	?	?	?	?	?	MMETSP
Excavata									
Malawimonads									
Malawimonas californiana	M	-	M	-	M	-	M	-	Ongoing genome project, GenBank
Malawimonas jakobiformis	?	?	+	?	?	?	+	?	TBestDB, GenBank
Metamonada									
Giardia assemblage A	-	-	-	-	-	-	-	-	EuPathDB
Giardia assemblage B	-	-	-	-	-	-	-	-	EuPathDB
Giardia assemblage E	-	-	-	-	-	-	-	-	EuPathDB
Spironucleus salmonicida	-	-	-	-	-	-	-	-	EuPathDB
Trichomonas vaginalis G3	-	-	-	-	-	-	-	-	EuPathDB
Discoba									
Andalucia godoyi	+	-	M	-	M	-	M	-	Ongoing genome project, GenBank
Andalucia incarcerata	M	?	M	?	M	?	M	?	Ongoing RNASeq project, GenBank
Bodo saltans	-	-	-	-	-	-	-	-	Wellcome Trust Sanger Institute
Crithidia fasciculata strain Cf-Cl	-	-	-	-	-	-	-	-	EuPathDB
Euglena spp.	?	?	?	?	?	?	?	?	GenBank
Eutreptiella gymnastica NIES-381	?	?	?	?	?	?	?	?	MMETSP
Eutreptiella gymnastica-like CCMP1594	?	?	?	?	?	?	?	?	MMETSP
Leishmania braziliensis	-	-	-	-	-	-	-	-	EuPathDB
Leishmania donovani BPK282A1	-	-	-	-	-	-	-	-	EuPathDB
Leishmania infantum JPCM5	-	-	-	-	-	-	-	-	EuPathDB
Leishmania major strain Friedlin	-	-	-	-	-	-	-	-	EuPathDB

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Leishmania mexicana MHOM/GT/2001/U1103	-	-	-	-	-	-	-	-	EuPathDB
Leishmania tarentolae Parrot-TarII	-	-	-	-	-	-	-	-	EuPathDB
Leptomonas seymouri	-	-	-	-	-	-	-	-	Wellcome Trust Sanger Institute
Naegleria gruberi	-	-	-	-	-	-	-	-	GenBank, JGI
Neobodo designis CCAP 1951/1	?	?	?	?	?	?	?	?	MMETSP
Percolomonas cosmopolitus	?	?	?	?	?	?	?	?	MMETSP
Pharyngomonas kirbyi	+	?	+	?	?	?	M	?	Ongoing RNASeq project, GenBank
Trypanosoma brucei	-	-	-	-	-	-	-	-	EuPathDB
Trypanosoma congolense IL3000	-	-	-	-	-	-	-	-	EuPathDB
Trypanosoma cruzi	-	-	-	-	-	-	-	-	EuPathDB
Trypanosoma vivax Y486	-	-	-	-	-	-	-	-	EuPathDB
Tsukubamonas globosa	?	?	?	?	?	?	?	?	GenBank
Stramenopiles									
Blastocystis									
Blastocystis sp.	-	-	-	-	-	-	-	-	GenBank
Bicosoecida									
Bicosoecid sp. ms1	?	?	?	?	?	?	?	?	MMETSP
Cafeteria roenbergensis E4-10	?	?	?	?	?	?	?	?	MMETSP
Cafeteria sp. Caron Lab Isolate	?	?	?	?	?	?	?	?	MMETSP
Halocafeteria seosinensis	?	?	?	?	?	?	?	?	Ongoing RNASeq project
Peronosporomycota									
Albugo candida	M	-	M	-	-	-	M	-	GenBank
Albugo laibachii	+	?	M	?	+	?	M	?	GenBank
Aphanomyces astaci	+	-	M	-	M	-	M	-	GenBank
Aphanomyces invadans	M	-	M	-	M	-	M	-	GenBank
Hyaloperonospora arabidopsis	M	-	M	-	M	-	M	-	EnsemblProtists
Phytophthora capsici	M	-	M	-	M	-	M	-	JGI
Phytophthora cinnamomi	M	-	M	-	M	-	M	-	JGI
Phytophthora infestans	M	-	M	-	M	-	M	-	GenBank

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Phytophthora kernoviae	M	-	M	-	M	-	M	-	EnsemblProtists
Phytophthora lateralis	M	-	M	-	M	-	M	-	EnsemblProtists, GenBank
Phytophthora parasitica	M	-	M	-	M	-	M	-	GenBank
Phytophthora ramorum	M	-	M	-	M	-	+	-	EnsemblProtists
Phytophthora sojae	M	-	+	-	M	-	M	-	GenBank, JGI
Pythium aphanidermatum	M	-	M	-	M	-	M	-	EnsemblProtists
Pythium arrhenomanes	M	-	M	-	+	-	M	-	EnsemblProtists
Pythium irregulare	+	-	M	-	M	-	M	-	EnsemblProtists
Pythium iwayamai	M	-	M	-	M	-	M	-	EnsemblProtists
Pythium ultimum	M	-	M	-	M	-	M	-	EnsemblProtists
Pythium vexans	+	-	M	-	M	-	M	-	EnsemblProtists
Saprolegnia diclina	M	-	M	-	M	-	M	-	GenBank
Saprolegnia parasitica	M	-	M	-	M	-	M	-	Broad Institute
Eumastigatales									
Nannochloropsis gaditana	M	-	M	-	+	-	+	+	GenBank
Nannochloropsis oceanica	M	-	M	-	+	-	M	C	*Nannochloropsis oceanica* genome project, Michigan State University, https://bmb.natsci.msu.edu/about/directory/faculty/christoph-benning/nannochloropsis-oceanica-ccmp1779/
Chrysophyceae									
Chromulina nebulosa UTEXLB2642	?	?	?	?	?	?	?	?	MMETSP
Dinobryon sp. UTEXLB2267	?	?	?	?	?	?	?	?	MMETSP
Mallomonas splendens	?	?	?	?	?	?	M	+	GenBank
Ochromonas sp. BG-1	?	?	?	?	?	?	?	?	MMETSP
Ochromonas sp. CCMP1393	?	?	?	?	?	?	?	?	MMETSP
Ochromonas sp. CCMP1899	?	?	?	?	?	?	?	?	MMETSP
Paraphysomonas bandaiensis Caron Lab Isolate	?	?	?	?	?	?	?	?	MMETSP
Paraphysomonas imperforata PA2	+	?	M	?	?	?	+	?	MMETSP
Paraphysomonas vestita GFlagA	?	?	?	?	?	?	?	?	MMETSP
Spumella elongata CCAP 955/1	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Raphidophyceae									
Chattonella subsalsa CCMP2191	C	?	+	?	C	?	+	+	MMETSP
Fibrocapsa japonica CCMP1661	+	?	+	?	M	?	+	M	MMETSP
Heterosigma akashiwo	?	?	+	?	+	?	+	+	MMETSP
Phaeophyceae									
Ectocarpus siliculosus	+	-	M	-	+	-	+	M	GenBank
Dictyochophyceae									
Dictyocha speculum CCMP1381	?	?	?	?	?	?	?	?	MMETSP
Florenciella parvula	?	?	?	?	?	?	?	?	MMETSP
Florenciella sp. RCC1007	?	?	?	?	?	?	?	?	MMETSP
Florenciella sp. RCC1587	?	?	?	?	?	?	?	?	MMETSP
Pseudopedinella elastica CCMP716	?	?	?	?	?	?	?	?	MMETSP
Pteridomonas danica PT	+	?	+	?	?	?	+	?	MMETSP
Rhizochromulina marina cf CCMP1243	?	?	?	?	?	?	?	?	MMETSP
Pelagophyceae									
Aureococcus anophagefferens CCMP1850	-	-	-	-	-	-	+	-	JGI, MMETSP
Aureoumbra lagunensis CCMP1510	?	?	?	?	?	?	?	?	MMETSP
Chrysocystis fragilis CCMP3189	?	?	?	?	?	?	?	?	MMETSP
Chrysoreinhardia sp. CCMP2950	?	?	?	?	?	?	?	?	MMETSP
Chrysoreinhardia sp. CCMP3193	?	?	?	?	?	?	?	?	MMETSP
Genus nov. species nov. RCC1024	?	?	?	?	?	?	?	?	MMETSP
Pelagococcus subviridis CCMP1429	?	?	?	?	?	?	?	?	MMETSP
Pelagomonas calceolata	?	?	?	?	?	?	?	?	MMETSP
Sarcinochrysis sp. CCMP770	?	?	?	?	?	?	?	?	MMETSP
Bolidophyceae									
Bolidomonas pacifica	+	?	+	?	C	?	M	+	MMETSP
Bolidomonas sp. RCC1657	?	?	?	?	?	?	?	?	MMETSP
Bolidomonas sp. RCC2347	?	?	?	?	?	?	?	?	MMETSP
Pinguiochrysidales									
Phaeomonas parva CCMP2877	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Pinguiococcus pyrenoidosus CCMP2078	?	?	?	?	?	?	?	?	MMETSP
Synchromophyceae									
Synchroma pusillum CCMP3072	?	?	?	?	?	?	?	?	MMETSP
Xanthophyceae									
Vaucheria litorea CCMP2940	?	?	?	?	?	?	?	?	MMETSP
Labyrinthulomycetes									
Aplanochytrium sp. PBS07	?	?	?	?	?	?	?	?	MMETSP
Aplanochytrium kerguelense	-	-	-	-	-	-	-	-	JGI
Aplanochytrium stocchinoi GSBS06	?	?	?	?	?	?	?	?	MMETSP
Aurantiochytrium limacinum ATCCMYA-1381	-	-	-	-	-	-	-	-	JGI, MMETSP
Schizochytrium aggregatum ATCC28209	-	-	-	-	-	-	-	-	JGI, MMETSP
Thraustochytrium sp. LLF1b	?	?	?	?	?	?	?	?	MMETSP
Diatomea									
Amphiprora paludosa CCMP125	?	?	?	?	?	?	?	?	MMETSP
Amphiprora sp. CCMP467	?	?	?	?	?	?	?	?	MMETSP
Amphora coffeaeformis CCMP127	?	?	?	?	?	?	?	?	MMETSP
Asterionellopsis glacialis	?	?	?	?	?	?	?	?	MMETSP
Astrosyne radiata 13vi08-1A	?	?	?	?	?	?	?	?	MMETSP
Attheya septentrionalis CCMP2084	?	?	?	?	?	?	?	?	MMETSP
Aulacoseira subarctica CCAP 1002/5	?	?	?	?	?	?	?	?	MMETSP
Chaetoceros affinis CCMP159	?	?	?	?	?	?	?	?	MMETSP
Chaetoceros brevis CCMP164	?	?	?	?	?	?	?	?	MMETSP
Chaetoceros cf. neogracile RCC1993	?	?	?	?	?	?	?	?	MMETSP
Chaetoceros curvisetus Unknown	?	?	?	?	?	?	?	?	MMETSP
Chaetoceros debilis MM31A-1	?	?	?	?	?	?	?	?	MMETSP
Chaetoceros dichaeta CCMP1751	?	?	?	?	?	?	?	?	MMETSP
Chaetoceros neogracile	?	?	?	?	?	?	+	C	GenBank
Chaetoceros sp. GSL56	?	?	?	?	?	?	?	?	MMETSP
Chaetoceros sp. UNC1202	?	?	?	?	?	?	?	?	MMETSP
Corethron hystrix 308	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Corethron pennatum L29A3	?	?	?	?	?	?	?	?	MMETSP
Coscinodiscus wailesii CCMP2513	?	?	?	?	?	?	?	?	MMETSP
Craspedostauros australis CCMP3328	?	?	?	?	?	?	?	?	MMETSP
Cyclophora radiata	+	?	+	?	?	?	+	+	MMETSP
Cyclophora tenuis ECT3854	+	?	+	?	+	?	+	+	MMETSP
Cyclotella meneghiniana CCMP 338	?	?	?	?	?	?	?	?	MMETSP
Cylindrotheca closterium KMMCC:B-181	?	?	?	?	?	?	?	?	MMETSP
Dactyliosolen fragilissimus Unknown	?	?	?	?	?	?	?	?	MMETSP
Detonula confervacea CCMP 353	?	?	?	?	?	?	?	?	MMETSP
Ditylum brightwellii	?	?	?	?	?	?	?	?	MMETSP
Entomoneis sp. CCMP2396	?	?	?	?	?	?	?	?	MMETSP
Eucampia antarctica CCMP1452	?	?	?	?	?	?	?	?	MMETSP
Extubocellulus spinifer CCMP396	?	?	?	?	?	?	?	?	MMETSP
Fragilariopsis cylindricus CCMP 1102	-	-	-	-	-	-	-	-	JGI
Fragilariopsis kerguelensis	?	?	?	?	?	?	?	?	MMETSP
Grammatophora oceanica CCMP 410	?	?	?	?	?	?	?	?	MMETSP
Helicotheca tamensis CCMP826	?	?	?	?	?	?	?	?	MMETSP
Leptocylindrus danicus CCMP1856	+	?	+	?	+	?	+	+	MMETSP
Licmophora paradoxa CCMP2313	?	?	?	?	?	?	?	?	MMETSP
Minutocellus polymorphus	?	?	?	?	?	?	?	?	MMETSP
Nitzschia punctata CCMP561	?	?	?	?	?	?	?	?	MMETSP
Nitzschia sp. RCC80	?	?	?	?	?	?	?	?	MMETSP
Odontella aurita isolate 1302-5	+	?	M	?	C	?	+	+	MMETSP
Odontella sinensis Grunow 1884	?	?	?	?	?	?	?	?	MMETSP
Phaeodactylum tricornutum	-	-	-	-	-	-	-	+	JGI
Proboscia alata PI-D3	?	?	?	?	?	?	?	?	MMETSP
Proboscia inermis CCAP1064/1	?	?	?	?	?	?	?	?	MMETSP
Pseudo-nitzschia arenysensis B593	?	?	?	?	?	?	?	?	MMETSP
Pseudo-nitzschia australis 10249 10 AB	?	?	?	?	?	?	?	?	MMETSP
Pseudo-nitzschia delicatissima	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Pseudo-nitzschia fraudulenta WWA7	?	?	?	?	?	?	?	?	MMETSP
Pseudo-nitzschia heimii UNC1101	?	?	?	?	?	?	?	?	MMETSP
Pseudo-nitzschia multiseries CLN-47	-	-	-	-	-	-	-	-	JGI
Pseudo-nitzschia pungens	?	?	?	?	?	?	?	?	MMETSP
Rhizosolenia setigera CCMP 1694	?	?	?	?	?	?	?	?	MMETSP
Skeletonema costatum 1716	?	?	?	?	?	?	?	+	MMETSP
Skeletonema dohrnii SkelB	?	?	?	?	?	?	?	?	MMETSP
Skeletonema grethea CCMP 1804	?	?	?	?	?	?	?	?	MMETSP
Skeletonema japonicum CCMP2506	?	?	?	?	?	?	?	?	MMETSP
Skeletonema marinoi	?	?	?	?	?	?	?	+	MMETSP
Skeletonema menzelii CCMP793	?	?	?	?	?	?	?	?	MMETSP
Stauroneis constricta CCMP1120	?	?	?	?	?	?	?	?	MMETSP
Staurosira complex sp. CCMP2646	?	?	?	?	?	?	?	?	MMETSP
Stephanopyxis turris CCMP 815	?	?	?	?	?	?	?	?	MMETSP
Striatella unipunctata CCMP2910	?	?	?	?	+	?	?	+	MMETSP
Synedropsis recta cf CCMP1620	?	?	?	?	?	?	?	?	MMETSP
Thalassionema frauenfeldii CCMP 1798	?	?	?	?	?	?	?	?	MMETSP
Thalassionema nitzschioides	?	?	?	?	?	?	?	?	MMETSP
Thalassiosira antarctica CCMP982	?	?	?	?	?	?	?	?	MMETSP
Thalassiosira gravida GMp14c1	?	?	?	?	?	?	?	?	MMETSP
Thalassiosira miniscula CCMP1093	?	?	?	?	?	?	?	?	MMETSP
Thalassiosira oceanica CCMP1005	?	?	?	?	?	?	?	M	MMETSP
Thalassiosira pseudonana CCMP1335	+	-	C	-	+	-	+	+	JGI
Thalassiosira punctigera Tpunct2005C2	?	?	?	?	?	?	?	?	MMETSP
Thalassiosira rotula	?	?	?	?	?	?	?	?	MMETSP
Thalassiosira sp. FW	?	?	?	?	?	?	?	?	MMETSP
Thalassiosira sp. NH16	?	?	?	?	?	?	?	?	MMETSP
Thalassiosira weissflogii	?	?	?	?	?	?	?	?	MMETSP
Thalassiothrix antarctica L6-D1	?	?	?	?	?	?	?	?	MMETSP
Triceratium dubium CCMP147	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Alveolata									
Protalveolata									
Amoebophrya sp. Ameob2	?	?	?	?	?	?	?	?	MMETSP
Chromera velia CCMP2878	?	?	?	?	?	?	?	?	MMETSP
Chromera velia J2	-	-	-	-	-	-	-	-	GenBank
Perkinsus chesapeaki ATCC PRA-65	?	?	?	?	?	?	?	?	MMETSP
Perkinsus marinus ATCC 50439	?	?	?	?	?	?	?	?	MMETSP
Perkinsus marinus ATCC 50983	-	-	-	-	-	-	-	-	GenBank
Vitrella brassicaformis CCMP3346	?	?	?	?	?	?	?	?	MMETSP
Dinoflagellata									
Akashiwo sanguinea CCCM 885	?	?	?	?	?	?	?	?	MMETSP
Alexandrium andersonii CCMP2222	?	?	?	?	?	?	?	?	MMETSP
Alexandrium catenella OF101	?	?	?	?	?	?	?	?	MMETSP
Alexandrium fundyense CCMP1719	?	?	?	?	?	?	?	?	MMETSP
Alexandrium margalefi AMGDE01CS-322	?	?	?	?	?	?	?	?	MMETSP
Alexandrium minutum CCMP113	?	?	?	?	?	?	?	?	MMETSP
Alexandrium monilatum CCMP3105	?	?	?	?	?	?	?	?	MMETSP
Alexandrium tamarense CCMP1771	?	?	?	?	?	?	?	?	MMETSP
Alveolata sp. CCMP3155	?	?	?	?	?	?	?	?	MMETSP
Amphidinium carterae CCMP1314	?	?	?	?	?	?	?	?	MMETSP
Amphidinium massartii CS-259	?	?	?	?	?	?	?	?	MMETSP
Azadinium spinosum 3D9	?	?	?	?	?	?	?	?	MMETSP
Brandtodinium nutriculum RCC3387	?	?	?	?	?	?	?	?	MMETSP
Ceratium fusus PA161109	?	?	?	?	?	?	?	?	MMETSP
Crypthecodinium cohnii Seligo	?	?	?	?	?	?	?	?	MMETSP
Dinophysis acuminata DAEP01	?	?	?	?	?	?	?	?	MMETSP
Durinskia baltica CSIRO CS-38	?	?	?	?	?	?	?	?	MMETSP
Gambierdiscus australes CAWD 149	?	?	?	?	?	?	?	?	MMETSP
Glenodinium foliaceum CCAP 1116/3	?	?	?	?	?	?	?	?	MMETSP
Gonyaulax spinifera CCMP409	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Gymnodinium catenatum GC744	?	?	?	?	?	?	?	?	MMETSP
Gyrodinium dominans SPMC 103	?	?	?	?	?	?	?	?	MMETSP
Heterocapsa arctica CCMP445	?	?	?	?	?	?	?	?	MMETSP
Heterocapsa rotundata SCCAP K-0483	?	?	?	?	?	?	?	?	MMETSP
Heterocapsa triquestra CCMP 448	?	?	?	?	?	?	?	?	MMETSP
Karenia brevis	?	?	?	?	?	?	?	?	MMETSP
Karlodinium micrum CCMP2283	?	?	?	?	?	?	?	?	MMETSP
Kryptoperidinium foliaceum CCMP 1326	?	?	?	?	?	?	?	?	MMETSP
Lessardia elongata SPMC 104	?	?	?	?	?	?	?	?	MMETSP
Lingulodinium polyedra CCMP 1738	?	?	?	?	?	?	?	?	MMETSP
Noctiluca scintillans Unknown	?	?	?	?	?	?	?	?	MMETSP
Oxyrrhis marina	?	?	?	?	?	?	?	?	MMETSP
Pelagodinium beii RCC1491	?	?	?	?	?	?	?	?	MMETSP
Peridinium aciculiferum PAER-2	?	?	?	?	?	?	?	?	MMETSP
Polarella glacialis	?	?	?	?	?	?	?	?	MMETSP
Prorocentrum lima CCMP 684	?	?	?	?	?	?	?	?	MMETSP
Prorocentrum micans CCCM 845	?	?	?	?	?	?	?	?	MMETSP
Prorocentrum minimum	?	?	?	?	?	?	?	?	MMETSP
Protoceratium reticulatum CCCM 535 (=CCMP 1889)	?	?	?	?	?	?	?	?	MMETSP
Pyrocystis lunula CCCM 517	?	?	?	?	?	?	?	?	MMETSP
Pyrodinium bahamense pbaha01	?	?	?	?	?	?	?	?	MMETSP
Scrippsiella Hangoei SHTV-5	?	?	?	?	?	?	?	?	MMETSP
Scrippsiella hangoei-like SHHI-4	?	?	?	?	?	?	?	?	MMETSP
Scrippsiella trochoidea CCMP3099	?	?	?	?	?	?	?	?	MMETSP
Symbiodinium kawagutii CCMP2468	?	?	?	?	?	?	?	?	MMETSP
Symbiodinium minutum	-	-	-	-	-	-	-	-	GenBank
Symbiodinium sp. C1	?	?	?	?	?	?	?	?	MMETSP
Symbiodinium sp. C15	?	?	?	?	?	?	?	?	MMETSP
Symbiodinium sp. CCMP2430	?	?	?	?	?	?	?	?	MMETSP
Symbiodinium sp. CCMP421	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Symbiodinium sp. cladeA	?	?	?	?	?	?	?	?	MMETSP
Symbiodinium sp. D1a	?	?	?	?	?	?	?	?	MMETSP
Symbiodinium sp. Mp	?	?	?	?	?	?	?	?	MMETSP
Thoracosphaera heimii CCCM 670 (=CCMP 1069)	?	?	?	?	?	?	?	?	MMETSP
Togula jolla CCCM 725	?	?	?	?	?	?	?	?	MMETSP
Apicomplexa									
Babesia bigemina	-	-	-	-	-	-	-	-	Wellcome Trust Sanger Institute
Babesia bovis T2Bo	-	-	-	-	-	-	-	-	EuPathDB
Babesia microti RI	-	-	-	-	-	-	-	-	EuPathDB
Cryptosporidium hominis TU502	-	-	-	-	-	-	-	-	EuPathDB
Cryptosporidium muris RN66	-	-	-	-	-	-	-	-	EuPathDB
Cryptosporidium parvum Iowa II	-	-	-	-	-	-	-	-	EuPathDB
Eimeria acervulina Houghton	-	-	-	-	-	-	-	-	EuPathDB
Eimeria brunetti Houghton	-	-	-	-	-	-	-	-	EuPathDB
Eimeria falciformis Bayer Haberkorn 1970	-	-	-	-	-	-	-	-	EuPathDB
Eimeria maxima Weybridge	-	-	-	-	-	-	-	-	EuPathDB
Eimeria mitis Houghton	-	-	-	-	-	-	-	-	EuPathDB
Eimeria necatrix Houghton	-	-	-	-	-	-	-	-	EuPathDB
Eimeria praecox Houghton	-	-	-	-	-	-	-	-	EuPathDB
Eimeria tenella Houghton	-	-	-	-	-	-	-	-	EuPathDB
Gregarina niphrandrodes Unknown strain	-	-	-	-	-	-	-	-	EuPathDB
Lankesteria abbottii Grappler Inlet BC	?	?	?	?	?	?	?	?	MMETSP
Neospora caninum Liverpool	-	-	-	-	-	-	-	-	EuPathDB
Plasmodium berghei ANKA	-	-	-	-	-	-	-	-	EuPathDB
Plasmodium chabaudi chabaudi	-	-	-	-	-	-	-	-	EuPathDB
Plasmodium cynomolgi	-	-	-	-	-	-	-	-	Wellcome Trust Sanger Institute
Plasmodium falciparum 3D7	-	-	-	-	-	-	-	-	EuPathDB
Plasmodium falciparum IT	-	-	-	-	-	-	-	-	EuPathDB
Plasmodium gallinaceum 8A	-	-	-	-	-	-	-	-	EuPathDB
Plasmodium knowlesi strain H	-	-	-	-	-	-	-	-	EuPathDB

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Plasmodium malariae	-	-	-	-	-	-	-	-	Wellcome Trust Sanger Institute
Plasmodium ovale	-	-	-	-	-	-	-	-	Wellcome Trust Sanger Institute
Plasmodium reichenowi Dennis	-	-	-	-	-	-	-	-	EuPathDB
Plasmodium vivax Sal-1	-	-	-	-	-	-	-	-	EuPathDB
Plasmodium yoelli yoelli	-	-	-	-	-	-	-	-	EuPathDB
Theilera equi WA	-	-	-	-	-	-	-	-	EuPathDB
Theilera orientalis Shintoku	-	-	-	-	-	-	-	-	EuPathDB
Theilera parva Muguga	-	-	-	-	-	-	-	-	EuPathDB
Theileria annulata Ankara	-	-	-	-	-	-	-	-	EuPathDB
Toxoplasma gondii	-	-	-	-	-	-	-	-	EuPathDB
Ciliophora									
Anophryoides haemophila AH6	?	?	?	?	?	?	?	?	MMETSP
Aristerostoma sp. ATCC 50986	?	?	?	?	?	?	?	?	MMETSP
Blepharisma japonicum Stock R1072	?	?	?	?	?	?	?	?	MMETSP
Climacostomum virens Stock W-24	?	?	?	?	?	?	?	?	MMETSP
Condylostoma magnum COL2	?	?	?	?	?	?	?	?	MMETSP
Euplotes crassus CT5	?	?	?	?	?	?	?	?	MMETSP
Euplotes focardii TN1	?	?	?	?	?	?	?	?	MMETSP
Euplotes harpa FSP1.4	?	?	?	?	?	?	?	?	MMETSP
Fabrea salina	?	?	?	?	?	?	?	?	MMETSP
Favella ehrenbergii Fehren 1	?	?	?	?	?	?	?	?	MMETSP
Favella taraikaensis Fe Narragansett Bay	?	?	?	?	?	?	?	?	MMETSP
Ichthyophthirius multifiliis	?	?	?	?	?	?	?	?	GenBank
Litonotus pictus P1	?	?	?	?	?	?	?	?	MMETSP
Mesodinium pulex SPMC105	?	?	?	?	?	?	?	?	MMETSP
Myrionecta rubra CCMP2563	?	?	?	?	?	?	?	?	MMETSP
Oxytricha trifallax strain SB310	-	-	-	-	-	-	-	-	GenBank
Paramecium tetraurelia	-	-	-	-	-	-	-	-	EnsemblProtists
Platyophrya macrostoma WH	?	?	?	?	?	?	?	?	MMETSP
Protocruzia adherens Boccale	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Pseudokeronopsis sp. Brazil	?	?	?	?	?	?	?	?	MMETSP
Pseudokeronopsis sp. OXSARD2	?	?	?	?	?	?	?	?	MMETSP
Strombidinopsis acuminatum SPMC142	?	?	?	?	?	?	?	?	MMETSP
Strombidinopsis sp. SopsisLIS2011	?	?	?	?	?	?	?	?	MMETSP
Strombidium inclinatum S3	?	?	?	?	?	?	?	?	MMETSP
Strombidium rassoulzadegani ras09	?	?	?	?	?	?	?	?	MMETSP
Tetrahymena thermophila	-	-	-	-	-	-	-	-	EnsemblProtists
Tiarina fusus LIS	?	?	?	?	?	?	?	?	MMETSP
Uronema sp. Bbcil	?	?	?	?	?	?	?	?	MMETSP
Rhizaria									
Cercozoa									
Bigelowiella longifila CCMP242	?	?	?	?	?	?	?	?	MMETSP
Bigelowiella natans	-	-	-	-	-	-	-	C	JGI
Chlorarachnion reptans CCCM449	?	?	?	?	?	?	?	?	MMETSP
Gymnochlora sp. CCMP2014	?	?	?	?	?	?	?	?	MMETSP
Lotharella amoebiformis CCMP2058	?	?	?	?	?	?	?	?	MMETSP
Lotharella globosa	?	?	?	?	?	?	?	?	MMETSP
Lotharella oceanica CCMP622	?	?	?	?	?	?	?	?	MMETSP
Minchinia chitonis Unknown	?	?	?	?	?	?	?	?	MMETSP
Norrisiella sphaerica BC52	?	?	?	?	?	?	?	?	MMETSP
Partenskyella glossopodia RCC365	?	?	?	?	?	?	?	?	MMETSP
Paulinella chromatophora	?	P*	?	P*	?	P*	?	P*	GenBank
Retaria									
Ammonia sp. Unknown	?	?	?	?	?	?	?	?	MMETSP
Elphidium margaritaceum Unknown	?	?	?	?	?	?	?	?	MMETSP
Reticulomyxa filosa	-	-	-	-	-	-	-	-	GenBank
Rosalina sp. Unknown	?	?	?	?	?	?	?	?	MMETSP
Sorites sp. Unknown	?	?	?	?	?	?	?	?	MMETSP
Cryptophyceae									
Palpitomonas bilix NIES-2562	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Cryptomonadales									
Chroomonas mesostigmatica	?	?	?	?	?	?	?	?	GenBank
Cryptomonas curvata CCAP979/52	?	?	?	?	?	?	?	?	MMETSP
Cryptomonas paramecium CCAP977/2a	?	?	?	?	?	?	?	+	MMETSP
Geminigera cryophila CCMP2564	?	?	?	?	?	?	?	?	MMETSP
Geminigera sp. Caron Lab Isolate	?	?	?	?	?	?	?	?	MMETSP
Guillardia theta	-	-	-	P	-	P	-	N	JGI, MMETSP
Hanusia phi CCMP325	?	?	?	?	?	?	?	?	MMETSP
Hemiselmis andersenii	?	?	?	?	?	?	?	N	GenBank, MMETSP
Hemiselmis rufescens PCC563	?	?	?	?	?	?	?	?	MMETSP
Hemiselmis tepida CCMP443	?	?	?	?	?	?	?	?	MMETSP
Hemiselmis viresens PCC157	?	?	?	?	?	?	?	?	MMETSP
Proteomonas sulcata CCMP704	?	?	?	?	?	?	?	?	MMETSP
Rhodomonas abbreviata Caron Lab Isolate	?	?	?	?	?	?	?	?	MMETSP
Rhodomonas lens RHODO	?	?	?	?	?	?	?	?	MMETSP
Rhodomonas salina	?	?	?	P	?	P	?	N	GenBank, MMETSP
Rhodomonas sp. CCMP768	?	?	?	?	?	?	?	?	MMETSP
Goniomonas									
Goniomonas pacifica CCMP1869	?	?	?	?	?	?	?	?	MMETSP
Goniomonas sp. m	?	?	?	?	?	?	?	?	MMETSP
Haptophyta									
Prymnesio-phyceae									
Emiliania huxleyi	-	-	-	P	-	+	-	+	JGI, MMETSP
Calcidiscus leptoporus RCC1130	?	?	?	?	?	?	?	?	MMETSP
Chrysochromulina brevifilum UTEX LB 985	?	?	?	?	?	?	?	?	MMETSP
Chrysochromulina ericina CCMP281	?	?	?	?	?	?	?	?	MMETSP
Chrysochromulina polylepis	?	?	?	+	?	+	?	+	MMETSP
Chrysochromulina rotalis UIO044	?	?	?	?	?	?	?	+	MMETSP
Chrysoculter rhomboideus RCC1486	?	?	?	?	?	?	?	?	MMETSP

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Coccolithus pelagicus ssp braarudi PLY182g	?	?	?	?	?	?	?	?	MMETSP
Gephyrocapsa oceanica RCC1303	?	?	?	?	?	?	?	?	MMETSP
Imantonia sp. RCC918	?	?	?	?	?	?	?	?	MMETSP
Isochrysis galbana CCMP1323	?	?	?	?	?	?	?	?	MMETSP
Isochrysis sp. CCMP1244	?	?	?	?	?	?	?	?	MMETSP
Isochrysis sp. CCMP1324	?	?	?	?	?	?	?	?	MMETSP
Phaeocystis antarctica	?	?	?	P	?	?	?	?	GenBank, MMETSP
Phaeocystis cordata RCC1383	?	?	?	?	?	?	?	?	MMETSP
Phaeocystis globosa	?	?	?	P	?	?	?	?	GenBank
Phaeocystis sp. CCMP2710	?	?	?	?	?	?	?	?	MMETSP
Pleurochrysis carterae CCMP645	?	?	?	?	?	?	?	?	MMETSP
Prymnesium parvum Texoma1	?	?	?	+	?	?	?	M	MMETSP
Scyphosphaera apsteinii RCC1455	?	?	?	?	?	?	?	?	MMETSP
Pavlovophyceae									
Diacronema lutheri (listed in GenBank and MMETSP records by an older name, *Pavlova lutheri* (3))	?	?	?	P	?	?	?	?	GenBank, MMETSP
Exanthemachrysis gayraliae RCC1523	?	?	?	?	?	?	?	?	MMETSP
Pavlova gyrans CCMP608	?	?	?	?	?	?	?	?	MMETSP
Pavolva pinguis	?	?	?	?	?	?	?	+	GenBank
Pavlova sp. CCMP459	?	?	?	?	?	?	?	?	MMETSP
Rhodophyta									
Chondrus crispus	-	-	-	-	-	-	-	-	GenBank
Compsopogon coeruleus SAG 36.94	?	?	?	?	?	?	?	?	MMETSP
Cyanidioschyzon merolae	-	-	-	-	-	-	M	M	GenBank
Cyanidium caldarium	?	?	?	?	?	?	?	C	GenBank
Erythrolobus australicus CCMP3124	?	?	?	?	?	?	?	?	MMETSP
Erythrolobus madagascarensis CCMP3276	?	?	?	?	?	?	?	?	MMETSP
Galdieria sulphuraria	?	?	?	+ †	?	?	?	C	Michigan State University *Galdieria* Database http://genomics.msu.edu/galdieria

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Madagascaria erythrocladiodes CCMP3234	?	?	?	?	?	?	?	?	MMETSP
Porphyridium aerugineum SAG 1380-2	?	?	?	?	?	?	?	?	MMETSP
Porphyridium purpureum	-	+	-	-	-	-	-	-	GenBank
Pyropia yezoensis	-	-	-	-	-	-	-	-	Susabi-nori (*Pyropia yezoensis*) genome project, *National Research Institute of Fisheries Science,* http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/nori/
Rhodella maculata CCMP736	?	?	?	?	?	?	?	?	MMETSP
Rhodosorus marinus	?	?	?	?	?	?	?	+	MMETSP
Timspurckia oligopyrenoides CCMP3278	?	?	?	?	?	?	?	?	MMETSP
Glaucophyta									
Cyanophora paradoxa	-	-	-	+	-	C	-	+	GenBank, *Cyanophora* Genome Project
Cyanoptyche gloeocystis	?	?	?	?	?	?	?	?	MMETSP
Gloeochaete wittrockiana	?	?	?	?	?	?	?	?	MMETSP
Chlorophyta									
Prasinophytes									
Nephroselmis olivacea	?	?	?	P	?	?	?	+	GenBank
Pedinomonas minor	?	?	?	P	?	?	?	?	GenBank
Picocystis salinarum	?	M	?	P	?	?	?	+	MMETSP
Pyramimonas parkeae	?	?	?	?	?	?	?	C	GenBank
Mamiellophyceae									
Bathycoccus prasinos	-	C	-	C	-	-	-	C	GenBank
Micromonas pusilla	-	C	-	C	-	M	-	+	GenBank, JGI
Micromonas sp. RCC299	-	C	-	+	-	M	-	+	GenBank
Ostreococcus lucimarinus	-	M	-	+	-	M	-	+	GenBank
Ostreococcus tauri	-	M	-	+	-	+	-	M	GenBank, JGI
Ulvophyceae									
Oltmannsiellopsis viridis	?	?	?	P	?	?	?	?	GenBank
Pseudendoclonium akinetum	?	?	?	P	?	?	?	?	GenBank

Taxon name	MinC		MinD		MinE		FtsZ		Data source
	M	C	M	C	M	C	M	C	
Trebouxiophyceae									
Auxenochlorella protothecoides	-	-	-	P	-	C	-	C	GenBank
Chlorella sorokiniana	?	?	?	P	?	?	?	?	GenBank
Chlorella sp. ArM0029B	?	?	?	P	?	?	?	?	GenBank
Chlorella variabilis	-	-	-	P	-	-	-	+	GenBank
Chlorella vulgaris	?	?	?	P	?	+	?	+	GenBank
Coccomyxa subellipsoidea	-	-	-	P	-	-	-	+	GenBank
Leptosira terrestris	?	?	?	P	?	?	?	?	GenBank
Nannochloris bacillaris	?	?	?	?	?	?	?	M	GenBank
Oocystis solitaria	?	?	?	P	?	?	?	?	GenBank
Parachlorella kessleri	?	?	?	P	?	?	?	?	GenBank
Prototheca wickerhamii	?	?	?	P	?	?	?	?	GenBank
Trebouxia aggregata	?	?	?	P	?	?	?	?	GenBank
Trebouxiophyceae sp. MX-AZ01	?	?	?	P	?	?	?	?	GenBank
Chlorophyceae									
Chlamydomonas reinhardtii	-	+	-	C	-	M	-	C	GenBank
Volvox carteri	-	+	-	+	-	C	-	C	GenBank
Streptophyta									
Physcomitrella patens	-	C	-	C	-	C	-	C	GenBank
Other streptophytes	-	-	-	C	-	C	-	C	GenBank

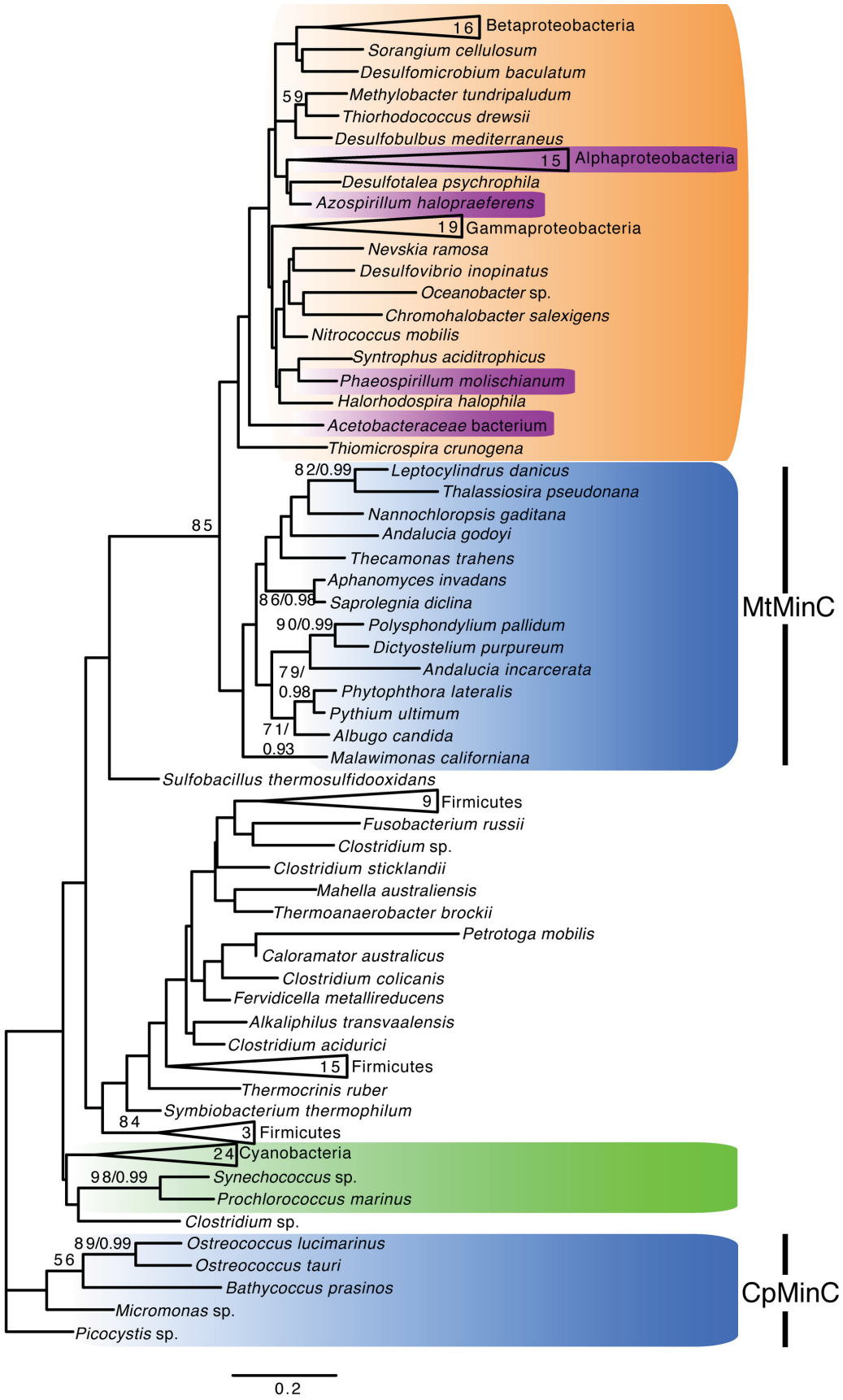
.

Table B.S2 Results of Approximately Unbiased topology tests. *Rejected with P<0.05 **Rejected with P<0.01

Constrained tree	**Topology**	**AU p-value**
MinC Topology 1	MtMinC + α-proteobacteria constrained, CpMinC and cyanobacteria unconstrained	0.760
MinC Topology 2	Eukaryotes monophyletic	0.143
MinD Topology 1	Mitoeuks + α-proteobacteria, other eukaryotic sequences unconstrained	0.424
MinD Topology 2	MtMinD + CpMinD constrained, *Paulinella* unconstrained	0.030*
MinE Topology 1	MtMinE + α-proteobacteria constrained, other eukaryotic sequences unconstrained	0.961
MinE Topology 2	CpMinE1 + CpMinE2 + CpMinE3 constrained, remaining eukaryotes and cyanobacteria unconstrained	0.955
MinE Topology 3	MtMinE + CpMinE1 + CpMinE2 constrained, cyanobacteria and CpMinE3 unconstrained	0.786
MinE Topology 4	MtMinE + CpMinE3 constrained, cyanobacteria and CpMin3 unconstrained	0.986
MinE Topology 5	MtMinE + CpMinE1 + CpMinE2 + CpMinE3 constrained, *Paulinella* and cyanobacteria unconstrained	0.679
FtsZ Topology 1	MtFtsZ1 + MtFtsZ2 constrained	0.617
FtsZ Topology 2	MtFtsZ1 + MtFtsZ2 + α-proteobacteria constrained, CpFtsZ and *Paulinella* unconstrained	0.449
FtsZ Topology 3	MtFtsZ1 + CpFtsZ + α-proteobacteria constrained, MtFtsZ2 and *Paulinella* unconstrained	0.006**
FtsZ Topology 4	MtFtsZ2 + CpFtsZ + α-proteobacteria constrained, MtFtsZ1 and *Paulinella*	0.018*

Constrained tree	**Topology**	**AU p-value**
	unconstrained	
FtsZ Topology 5	MtFtsZ1 + MtFtsZ2 + CpFtsZ constrained, *Paulinella* unconstrained	4e-04**

Figure B.S1 Unrooted Maximum Likelihood (ML) tree of MinC sequences. **(Following page)** Phylogenetic analyses were performed on 157 sequences and 82 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Eukaryotes are shaded blue, cyanobacteria green, proteobacteria orange, and alphaproteobacteria magenta. *Chlamydomonas* and *Volvox* sequences mentioned in Fig. 2 and Table S1 were excluded from the analysis because of a high degree of divergence from other MinC sequences.


16 Betaproteobacteria
Sorangium cellulosum
Desulfomicrobium baculatum
59 Methylobacter tundripaludum
Thiorhodococcus drewsii
Desulfobulbus mediterraneus
15 Alphaproteobacteria
Desulfotalea psychrophila
Azospirillum halopraeferens
19 Gammaproteobacteria
Nevskia ramosa
Desulfovibrio inopinatus
Oceanobacter sp.
Chromohalobacter salexigens
Nitrococcus mobilis
Syntrophus aciditrophicus
Phaeospirillum molischianum
Halorhodospira halophila
Acetobacteraceae bacterium
Thiomicrospira crunogena
82/0.99 Leptocylindrus danicus
Thalassiosira pseudonana
Nannochloropsis gaditana
85 Andalucia godoyi
Thecamonas trahens
Aphanomyces invadans
86/0.98 Saprolegnia diclina
90/0.99 Polysphondylium pallidum
Dictyostelium purpureum
79/ Andalucia incarcerata
0.98 Phytophthora lateralis
Pythium ultimum
71/ Albugo candida
0.93 Malawimonas californiana
MtMinC
Sulfobacillus thermosulfidooxidans
9 Firmicutes
Fusobacterium russii
Clostridium sp.
Clostridium sticklandii
Mahella australiensis
Thermoanaerobacter brockii
Petrotoga mobilis
Caloramator australicus
Clostridium colicanis
Fervidicella metallireducens
Alkaliphilus transvaalensis
Clostridium acidurici
15 Firmicutes
Thermocrinis ruber
Symbiobacterium thermophilum
84 3 Firmicutes
24 Cyanobacteria
98/0.99 Synechococcus sp.
Prochlorococcus marinus
Clostridium sp.
89/0.99 Ostreococcus lucimarinus
56 Ostreococcus tauri
Bathycoccus prasinos
Micromonas sp.
Picocystis sp.
CpMinC
0.2

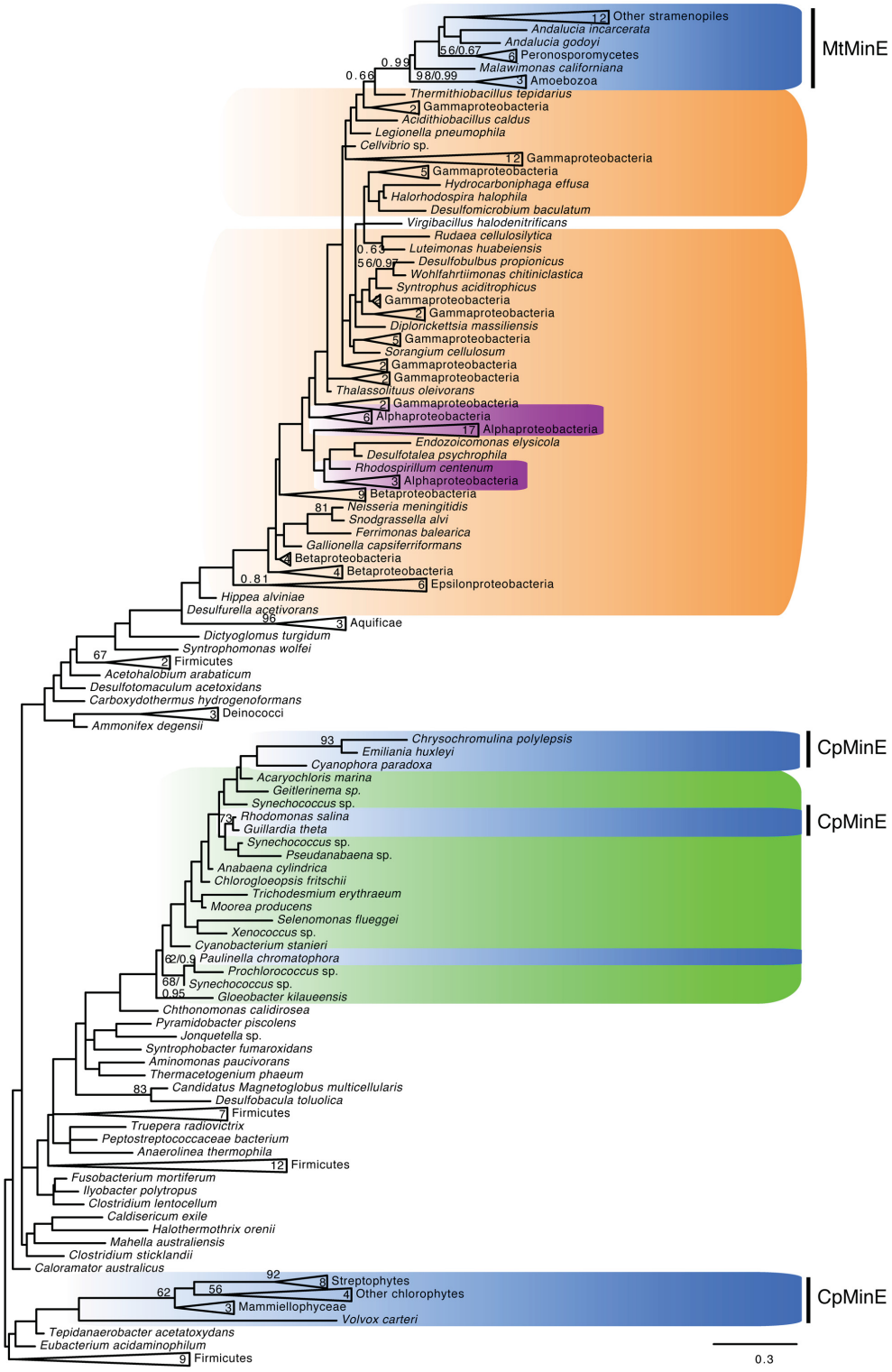
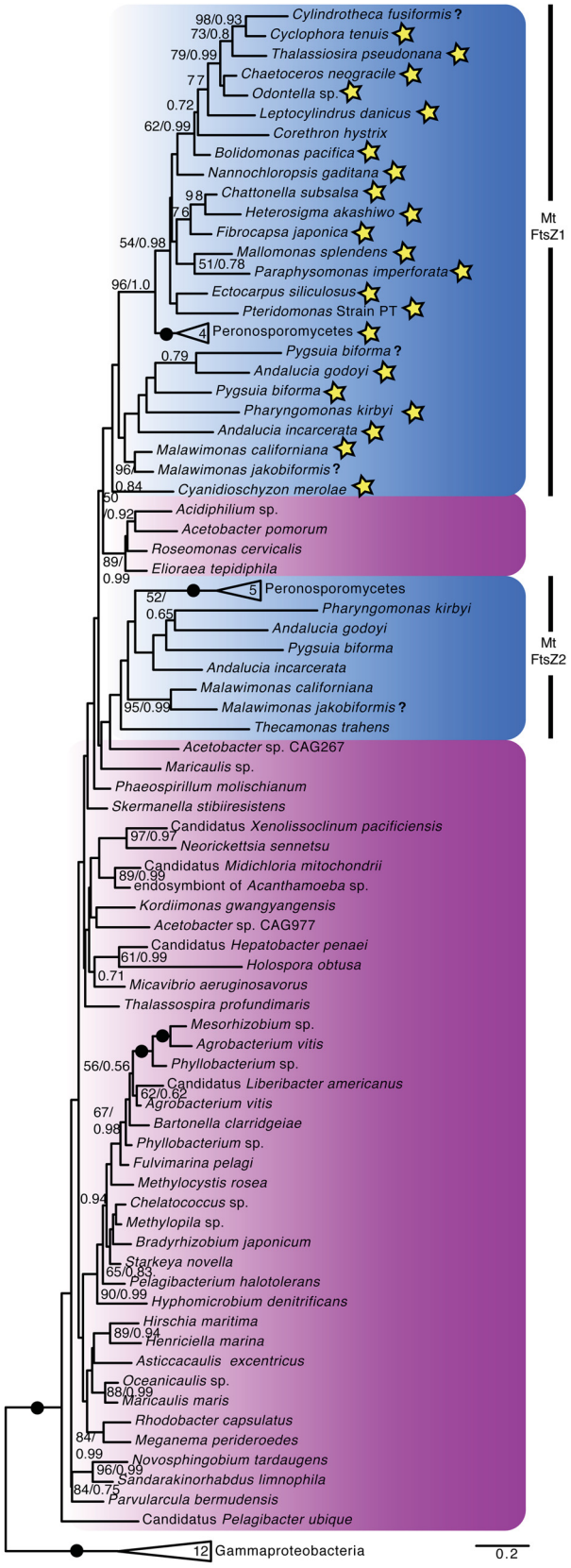


Figure B.S2 Unrooted Maximum Likelihood (ML) tree of MinE sequences. Phylogenetic analyses were performed on 235 sequences and 40 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Eukaryotes are shaded blue, cyanobacteria green, proteobacteria orange, and alphaproteobacteria magenta.

Figure B.S3 Unrooted Maximum Likelihood (ML) tree of subsampled FtsZ sequences from eukaryotes, alphaproteobacteria, and gammaproteobacteria. **(Following page)** Amoebozoan sequences were excluded due to their high AT content and long branches. Phylogenetic analyses were performed on 96 sequences and 292 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and alphaproteobacteria magenta. Eukaryotic paralogs lacking the variable C-terminal spacer region are indicated by stars, while those with incomplete sequence at the C-terminus are indicated by question marks.


Cylindrotheca fusiformis?
Cyclophora tenuis
Thalassiosira pseudonana
Chaetoceros neogracile
Odontella sp.
Leptocylindrus danicus
Corethron hystrix
Bolidomonas pacifica
Nannochloropsis gaditana
Chattonella subsalsa
Heterosigma akashiwo
Fibrocapsa japonica
Mallomonas splendens
Paraphysomonas imperforata
Ectocarpus siliculosus
Pteridomonas Strain PT
Peronosporomycetes
Pygsuia biforma?
Andalucia godoyi
Pygsuia biforma
Pharyngomonas kirbyi
Andalucia incarcerata
Malawimonas californiana
Malawimonas jakobiformis?
Cyanidioschyzon merolae
Acidiphilium sp.
Acetobacter pomorum
Roseomonas cervicalis
Elioraea tepidiphila
Peronosporomycetes
Pharyngomonas kirbyi
Andalucia godoyi
Pygsuia biforma
Andalucia incarcerata
Malawimonas californiana
Malawimonas jakobiformis?
Thecamonas trahens
Acetobacter sp. CAG267
Maricaulis sp.
Phaeospirillum molischianum
Skermanella stibiiresistens
Candidatus Xenolissoclinum pacificiensis
Neorickettsia sennetsu
Candidatus Midichloria mitochondrii
endosymbiont of Acanthamoeba sp.
Kordiimonas gwangyangensis
Acetobacter sp. CAG977
Candidatus Hepatobacter penaei
Holospora obtusa
Micavibrio aeruginosavorus
Thalassospira profundimaris
Mesorhizobium sp.
Agrobacterium vitis
Phyllobacterium sp.
Candidatus Liberibacter americanus
Agrobacterium vitis
Bartonella clarridgeiae
Phyllobacterium sp.
Fulvimarina pelagi
Methylocystis rosea
Chelatococcus sp.
Methylopila sp.
Bradyrhizobium japonicum
Starkeya novella
Pelagibacterium halotolerans
Hyphomicrobium denitrificans
Hirschia maritima
Henriciella marina
Asticcacaulis excentricus
Oceanicaulis sp.
Maricaulis maris
Rhodobacter capsulatus
Meganema perideroedes
Novosphingobium tardaugens
Sandarakinorhabdus limnophila
Parvularcula bermudensis
Candidatus Pelagibacter ubique
Gammaproteobacteria
Mt FtsZ1
Mt FtsZ2
0.2

Appendix C. Supplementary Material for Chapter 4

Table C.S1 Predicted MRO proteins found in the *Andalucia incarcerata* transcriptome, and TargetP predictions. P(M): probability of a mitochondrial location

Predicted MRO protein	Status	TargetP P(M)	TargetP-predicted targeting peptide
Energy generation			
Pyruvate:ferredoxin oxidoreductase (1)	Complete	0.857	MFSLSRRIPR
Pyruvate:ferredoxin oxidoreductase (2)	Complete	0.679	MALLQTFTGGKSFPAGLSKAIFSRFKR HIA
Pyruvate:ferredoxin oxidoreductase (3)	Complete	0.964	MVSILSVLRSQRA
Pyruvate:ferredoxin oxidoreductase (4)	Complete	0.828	MLRSTFTSVGASSSSRS
Pyruvate:ferredoxin oxidoreductase (5)	Missing 3' and 5'		
[Fe]-hydrogenase (1)	Complete	0.886	MSFLAKSFRVSSAASTFARFLSGKM
[Fe]-hydrogenase (4)	Complete	0.733	MLSSSFQLFDRIIRSKAACFSRPLA
HydE	Complete	M 0.401	
HydF	Complete	0.867	MLGLGTLSSVA
HydG	Complete	0.874	MLSRLSSMLSRGAHA
Acetate:succinate CoA-transferase (1)	Complete	0.939	MLKSFSRISSSASRFLFSSRL
Acetate:succinate CoA-transferase (2)	Complete	0.597	MLTLSVCAVSAPFFSTLRRTV
Electron transport ferredoxin	Complete	0.931	MLRSFRFGGVFGGLSGLSVRSLSKFVY VR
Complex I 51kDa	Complete	0.847	MLSSLRSLTASRSFP
Complex I 24kDa	Complete	0.81	MLSSLFRGFSSPVRS
Succinate thiokinase subunit a	Complete	0.812	MLSRFVSASRSIAGPVLLRF
Succinate thiokinase subunit b	Complete	0.907	80a.a.s
D-lactate dehydrogenase	Complete	M 0.414	
Ferredoxin reductase Arh1	Complete	0.903	115a.a.s

Predicted MRO protein	Status	TargetP P(M)	TargetP-predicted targeting peptide
Electron transfer			
Alternative oxidase	Complete	0.543	92a.a.s
ETF-alpha	Complete	0.899	MLSSFLATRF
ETF-beta	Complete	0.428	
ETF-ubiquinone oxidoreductase 1	Complete	0.701	MLSLLPPFVRRNIVSSVCTG
ETF-ubiquinone oxidoreductase 2	Complete	0.877	MLCRGVLGGRISATFVSFSRTL
glycerol-3-phosphate DH	Complete	0.191	
Fe-S cluster assembly			
IscS	Complete	0.896	MLSYLRGSLLPGSRRAIFASSMLR
IscU	Complete	0.872	MLGHVVSRFRGGRIPSSLPSIVGLFSR WKY
IscA1	Complete	0.951	MSLWGFMRVSQRLSLYPSLARSFGTSL WGS
IscA2	Complete	0.685	MLSRGVFGVFSVFSRS
Isd11	Complete	0.148	
IND1	Missing 5'		
Frataxin	Complete	0.727	MLSRFS
Ferredoxin Yah1	Complete	0.932	MLFRRFGVLTRGSNSS
Jac1	Missing 5'		
Mge1	Complete	0.289	
Grx5	Complete	0.946	MSSHGSLSFFRGTLSVFRSVPLVRSLSS LMTSAS
Nfu1	Complete	0.863	MLQRFSFASYGGSLFSALVRSLRVKIET T
Atm1	Complete	M 0.424	
Protein import and folding, MRO dynamics			
mitochondrial processing peptidase A subunit	Complete	0.854	MFFLSSLSKRNSRVSGLWT
mitochondrial processing peptidase B subunit	Complete	0.946	MLRLGHRF

Predicted MRO protein	Status	TargetP P(M)	TargetP-predicted targeting peptide
SAM50	Complete	0.376	
Metaxin	Complete	0.393	
UPS1	Complete	0.217	
UPS2	Complete	0.266	
Tim8	Complete	0.129	
Tim9	Complete	0.123	
Tim10	Complete	0.152	
Tim13	Complete	0.293	
Tim22	Complete	0.34	
Pam18/Tim14	Complete	0.378	
Pam16/Tim16	Complete	0.792	MSIVRSGLVVGWRS
Tim44	Complete	0.622	MSSGLQRSLVLCGGQSALSS
Mmp37/Tam41	Complete	0.216	
Tim17	Complete	0.083	
Tim23	Complete	0.132	
Tom40	Complete	0.039	
Mia40	Complete	0.023	
Erv-family protein 1	Complete	0.12	
Erv-family protein 2	Complete	0.16	
Cpn60	Complete	0.647	MLQSLSSSLQSGFFGATKVLGGRLF
Cpn10	Complete	0.894	MSLLRSF
mtHsp70	Complete	0.957	MLSSFARGAFQSTMKSIMATSPRFLSR SF
Mitochondrial chaperone BCS-1	Complete	0.018	
Prohibitin	Complete	0.346	
Nek1 (1)	Complete	0.097	
Nek1 (2)	Missing 3' and 5'		
Mitochondrial Rho GTPase	Complete	0.049	

Predicted MRO protein	Status	TargetP P(M)	TargetP-predicted targeting peptide
Mitochondrial division protein 1	Complete	0.2	
MinD	Complete	0.864	114a.a.s
MinC	Complete	0.772	MLSFFSFPDSWTKCLRKGLFSRGLAISSFQFPSVSRKPELKFRTFFMPTISASQNL
MinE	Complete	0.663	MGFLSKLFGGGKSTAVAESVVKSASVAKDRLHIILASQRA
FtsZ (1)	Complete	0.887	MHLCRPF
FtsZ (2)	Complete	0.505	MLSGVSSVWAHGLKGVRFFSTLFSPLFPNCDSIAPLSHFPYPLRSFSTSASSS
dynamin-related protein/Drp1	Complete	0.02	
ERMES			
Mdm12	Complete	0.0355	
Mdm34	Missing 3'		
Mitochondrial outer membrane protein MMM1	Complete	0.367	
Amino acid and fatty acid metabolism			
Glycine cleavage system L protein	Complete	0.913	MLCRLVSGLGAAMSPPLRAMSSMPSMSLFQRMAQKGLRFFSSSAASS
Glycine cleavage system H protein	Complete	0.092	
Glycine cleavage system P protein	Complete	0.766	MLSTLFPNGGLRHVVAASVSKLRIPRWL
Glycine cleavage system T protein	Complete	0.732	MLSTLSRFCQPSKFTRSLSAAVGKK
mitochondrial serine hydroxymethyl transferase	Complete	0.717	107a.a.s
Brached-chain amino acid aminotransferase	Complete	0.741	MLAFSSLLRNAQKTSYAF
branched-chain alphaketoacid dehydrogenase E1a	Complete	0.744	MSFLVNHCVGSLGSMRPLQAPLRAFARF
branched-chain alphaketoacid dehydrogenase E1b	Complete	0.197	

Predicted MRO protein	Status	TargetP P(M)	TargetP-predicted targeting peptide
branched-chain alphaketoacid dehydrogenase E2	Complete	0.197	
Branched-chain acyl-coA dehydrogenase	Complete	0.591	MLSRFF
Enoyl-coA hydratase	Missing 5'		
acetyl-CoA C-acetyltransferase/acetyl-CoA thiolase	Complete	0.795	MFRRALSSSLREA
3-hydroxyisobutyrate dehydrogenase	Complete	0.542	112a.a.s
hydroxymethylglutaryl-CoA synthase	Missing 5'		
propionyl-coA carboxylase a-subunit	Complete	0.776	MAFSIVRGMFPRAGTYARFLAVKAKLF DKILIANRGEIACRV
propionyl-coA carboxylase b-subunit	Complete	0.63	MLSRFASVSKKIL
methylmalonyl semialdehyde DH	Complete	0.861	MLSRFFLSRVTKP
methylmalonyl coA epimerase	Complete	0.067	
methylmalonyl-CoA mutase	Complete	0.58	MLQRYGRPCALFSRSL
alanine-glyoxylate aminotransferase	Complete	0.667	MFSRIHHAQSKVFVRC
4-aminobutyrate transaminase	Complete	0.647	MLSRFGPLRSLPSFRFL
glutamate dehydrogenase (NADP)	Complete	0.725	MLSRAVSSSLASL
Succinate semialdehyde dehydrogenase (NADP)	Complete	0.866	110a.a.s
aspartate aminotransferase	Complete	0.64	MLSRFC
alanine aminotransferase	Missing 5'		
2-oxoglutarate dehydrogenase, E1 component	Complete	0.918	MYCLCGHPHFRNSGLRVFPVASFLRS EAQNVFRWTRTL
2-oxoglutarate dehydrogenase, E2 component	Complete	0.661	MMKPSKPRNPLALSSTPSCLLPPPSGW LRRFL

Predicted MRO protein	Status	TargetP P(M)	TargetP-predicted targeting peptide
2-oxoglutarate dehydrogenase, E2 component	Complete	0.703	65a.a.s
kynurenine/alpha-aminoadipate aminotransferase	Complete	0.64	MEVKTSDFVTRVSSRR
tyrosine aminotransferase	Complete	0.571	80a.a.s
3-mercaptopyruvate sulfurtransferase	Complete	0.796	70a.a.s
S-adenosylmethionine synthetase	Complete	0.935	MLSMSRLPALFRSSWKRGAAFLLRPF
Saccharopine dehydrogenase	Complete	0.643	MRRILVLGSGFVATALVKHFCRRPDQSLTIA
glycine-C-acetyl transferase	Missing 5'		
Acyl-CoA synthetase	Complete	0.887	MLSRFISRRVLPF
FAD-dependent oxidoreductase	Complete	0.618	MLRRFLHSTVPKV
Delta-1-pyrroline-5-carboxylate dehydrogenase	Complete	0.069	
Acetyl-CoA carboxylase	Complete	0.711	MSWKSVADYVRDRGGVRPIQRLLIANNGIAAVKAIRSIRKWAYQT
Succinate:3-oxoacid CoA-transferase	Complete	0.508	MFLSKCVYSI
long-chain-fatty-acid-coA-ligase (1)	Complete	0.07	
long-chain-fatty-acid-coA-ligase (2)	Complete	0.089	
ROS and oxygen detoxification			
flavoprotein A - ferredoxin fused to N-terminus	Complete	0.594	MLSNAFRKNLAPLSSPLFSSFLSQFTFVRS
peroxiredoxin V	Complete	0.503	MLSSLKFNSIFSSVCGVCVERALASRFL
thioredoxin (1)	Complete	0.958	113a.a.s
thioredoxin (2)	Complete	0.502	MLSRLFSHSLIGS
superoxide dismutase	Complete	0.512	MLRTFMQVKPFSGLLRN
Solute carriers			

Predicted MRO protein	Status	TargetP P(M)	TargetP-predicted targeting peptide
arginine/ornithine transport system ATPase	Complete	0.702	MFHMFKSGACLTRDVGVGWRS
MCF1 (Solute carrier family 25, member 43)	Complete	0.263	
MCF2 (Solute carrier family 25, member 38)	Complete	0.042	
MCF3	Complete	0.555	MLSRFAVGLGPLR
MCF4, possible ADP/ATP translocator	Missing 3' and 5'		
Aralar Asp/Glu carrier	Complete	0.191	
Glycerophospholipid metabolism			
Cardiolipin synthase	Complete	0.551	MPFSNIFTIPNVLSTARLVSAPFLAHRL PHCSPQKA
phosphoenolpyruvate phosphomutase	Complete	0.723	MLSGLSHIRSLLGKSF
glycerol-3-phosphate acyltransferase	Complete	0.152	
CDP-DAG synthase	Complete	0.095	
Miscellaneous			
Steroidogenic acute regulatory protein	Complete	0.885	MNQVFRFIRNQLTLVRKWRQNANY
cob(I)yrinic acid a,c-diamide adenosyltransferase	Complete	0.817	MLSSSRRSLVLPFLGR
molybdenum cofactor sulfurase	Complete	0.76	MWFKRFPSKKKRRDFLSMAASLWVR HRLLIAFLVFMGLGFFAIGAWIHMSTK GNGSKDEDAIGHAKKEFLSSSANER
lipoate-protein ligase A	Complete	0.856	MISSLLFLHRGFQSAANVI
Oligopeptidase A	Complete	0.739	MNKFFSFISSIPSRFR
Nuclease EXOG	Complete	0.095	
lon protease homolog	Complete	0.713	61a.a.s
DNA ligase I	Complete	0.595	75a.a.s
pyridine nucleotide transhydrogenase	Complete	0.045	
nitroreductase family protein	Complete	0.577	MSVSDIIHSRRSNRAFDSSRAVP

Predicted MRO protein	Status	TargetP P(M)	TargetP-predicted targeting peptide
FAD-dependent oxidoreductase/sulfurtransferase	Complete	0.776	MLARAFGRVSAFSSSFLAQ
FAD-dependent oxidoreductase/sulfurtransferase	Complete	0.729	81a.a.s
Predicted cytosolic proteins of interest			
Pyruvate:NADP oxidoreductase	Complete	0.073	-
[Fe]-hydrogenase (2) (flavodoxin-fused)	Complete	0.069	-
[Fe]-hydrogenase (3) (flavodoxin-fused)	Complete	0.077	-
Pyruvate:formate lyase	Complete	0.063	-
SufCB	Complete	0.073	-
flavoprotein A - non-fused	Complete	0.028	-

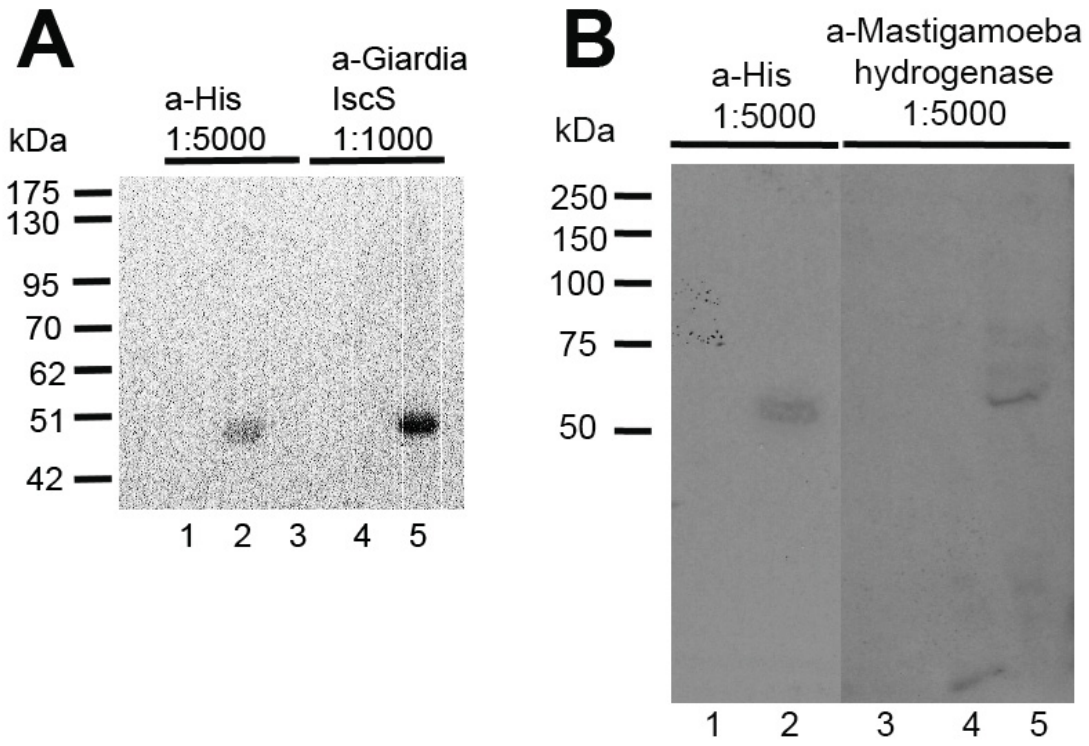


Figure C.S1 Western blot showing recognition of recombinant *Andalucia incarcerata* proteins by antibodies used in immunogold localization experiments. **A**, recognition of recombinant *A. incarcerata* IscS. Lane 1, 5µg inclusion bodies purified from *Escherichia coli* C41 cells expressing empty pET-16b plasmid (negative control). Lane 2, 5µg inclusion bodies purified from *E. coli* C41 cells expressing *A. incarcerata* IscS cloned into the pET-16b plasmid. Lane 3, Pink Plus prestained protein ladder. Lane 4, 1µg inclusion bodies purified from *E. coli* C41 cells expressing empty pET-16b plasmid (negative control). Lane 5, 1µg inclusion bodies purified from *E. coli* C41 cells expressing *A. incarcerata* IscS cloned into the pET-16b plasmid. Lanes 1 and 2 and half of Lane 3 were incubated with a monoclonal anti-His tag antibody (1:5000) as a positive control, then with a horseradish peroxidase-conjugated anti-mouse IgG secondary antibody. Lanes 4 and 5 and half of Lane 3 were incubated with the polyclonal anti-*Giardia* IscS primary antibody (1:1000), then with a horseradish peroxidase-conjugated anti-rabbit IgG secondary antibody. All lanes were subsequently incubated with chemiluminescent substrate, and visualized using a FluorChem E imaging system. Lanes 1, 3 and 4 were not recognized by the anti-His or anti-IscS antibodies, as expected; Lanes 2 and 5 each show one band corresponding to the expected molecular weight of *A. incarcerata* IscS, ~51.0 kDa (including His-tag). **B**, recognition of recombinant *A. incarcerata* [FeFe]-hydrogenase 1. Lane 1, 3µg inclusion bodies purified from *Escherichia coli* C41 cells expressing empty pET-16b plasmid (negative control). Lane 2, 3µg inclusion bodies purified from *E. coli* C41 cells expressing *A. incarcerata* [FeFe]-hydrogenase 1 cloned into the pET-16b plasmid. Lane 3, Precision Plus protein ladder. Lane 4, 3µg inclusion bodies purified from *E. coli* C41 cells expressing empty pET-16b plasmid (negative control). Lane 5, 3µg inclusion bodies purified from *E. coli* C41 cells expressing *A. incarcerata* [FeFe]-hydrogenase 1 cloned into the pET-16b plasmid. Lanes 1 and 2 were incubated with a monoclonal anti-His tag antibody (1:5000) as a positive control, then with a horseradish peroxidase-conjugated anti-mouse IgG secondary antibody. Lanes 3, 4 and 5 were incubated with the polyclonal anti-*Mastigamoeba* [FeFe]-hydrogenase

primary antibody (1:250), then with a horseradish peroxidase-conjugated anti-rabbit IgG secondary antibody. All lanes were subsequently incubated with chemiluminescent substrate, and photographed using film. Lanes 1, and 3 were not recognized by the anti-His or anti-[FeFe]-hydrogenase antibodies, as expected. A very low molecular weight band can be seen in Lane 4 – likely cross-reaction with an *E. coli* protein that is not obvious in Lane 5 because of the lower overall concentration of *E. coli* proteins in this lane relative to recombinant [FeFe]-hydrogenase. Lanes 2 and 5 each show one band corresponding to the expected molecular weight of *A. incarcerata* [FeFe]-hydrogenase 1, ~68.2 kDa (including His tag).

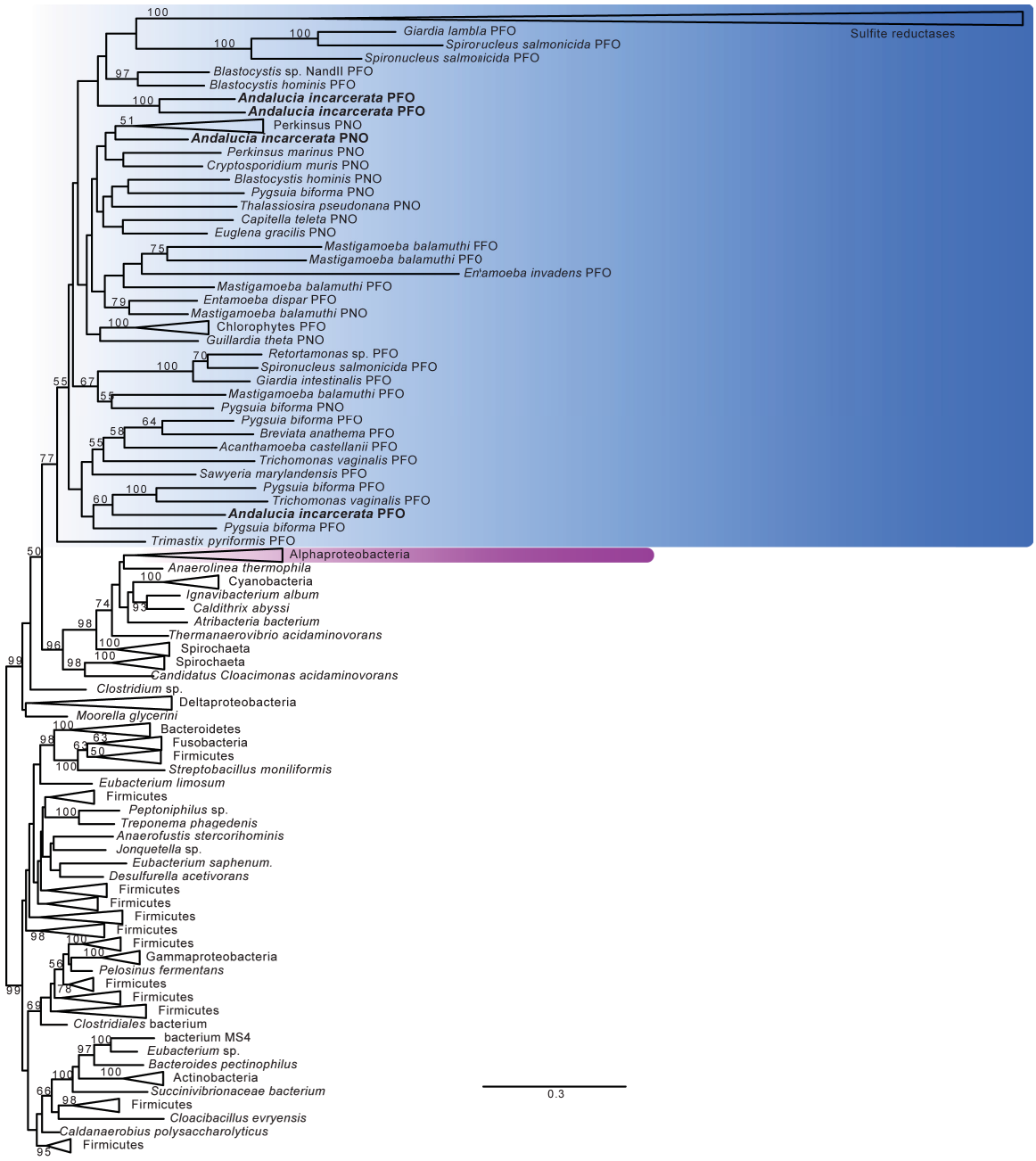


Figure C.S2 Unrooted Maximum Likelihood (ML) tree of PFO and sulfite reductase sequences. Phylogenetic analyses were performed on 286 sequences and 486 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and α-proteobacteria magenta.

Figure C.S3 Unrooted Maximum Likelihood (ML) tree of [FeFe]-hydrogenase sequences. **(Following page)** Phylogenetic analyses were performed on 283 sequences and 285 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and α-proteobacteria magenta. [FeFe]-hydrogenases with C-terminal CysJ or flavodoxin (FLD) domains are indicated.

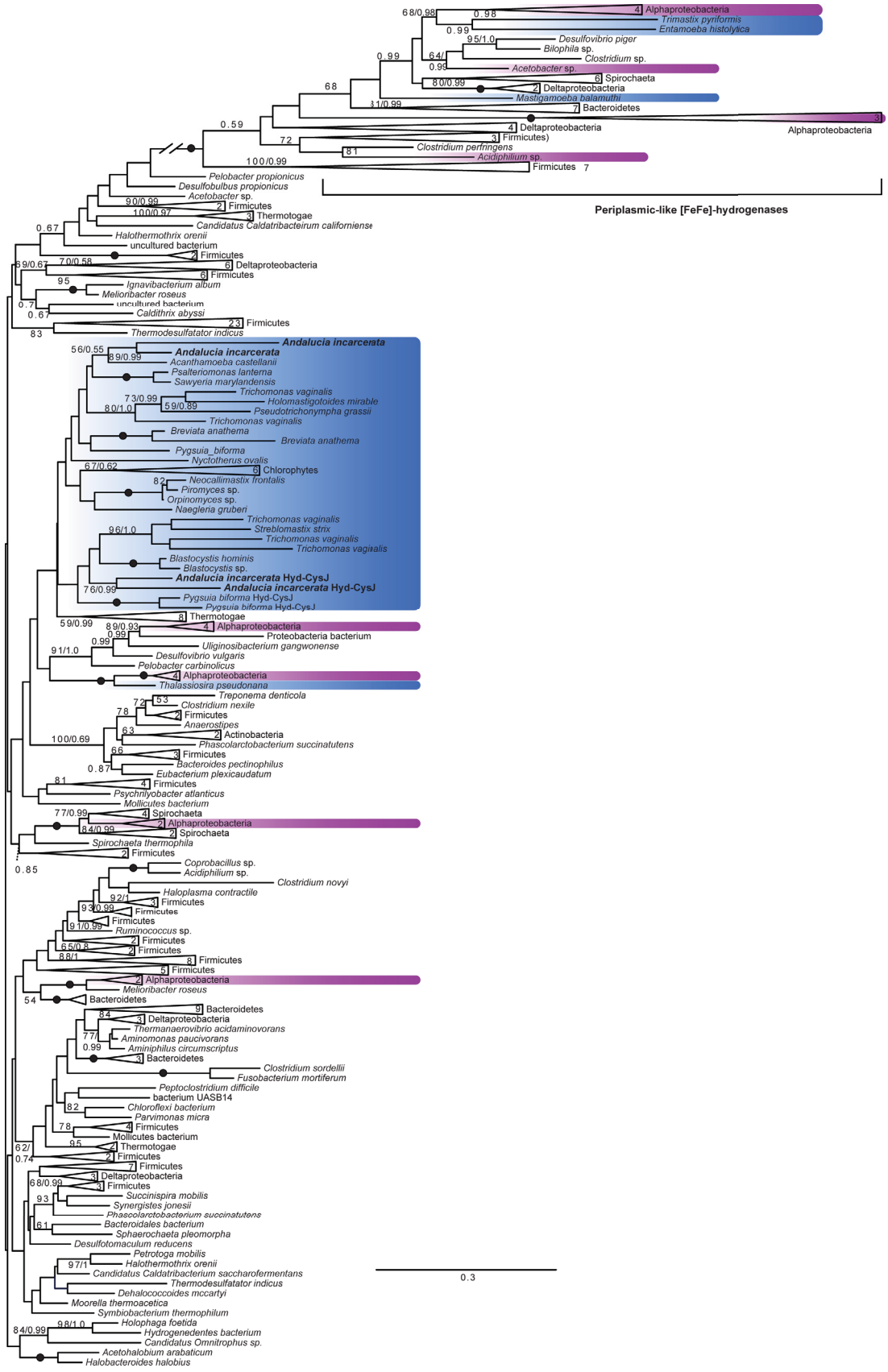

Periplasmic-like [FeFe]-hydrogenases
Alphaproteobacteria
Trimastix pyriformis
Entamoeba histolytica
Desulfovibrio piger
Bilophila sp.
Clostridium sp.
Acetobacter sp.
Spirochaeta
Deltaproteobacteria
Mastigamoeba balamuthi
Bacteroidetes
Alphaproteobacteria
Deltaproteobacteria
Firmicutes)
Clostridium perfringens
Acidiphilium sp.
Firmicutes
Pelobacter propionicus
Desulfobulbus propionicus
Acetobacter sp.
Firmicutes
Thermotogae
Candidatus Caldatribacteirum californiense
Halothermothrix orenii
uncultured bacterium
Firmicutes
Deltaproteobacteria
Firmicutes
Ignavibacterium album
Melioribacter roseus
uncultured bacterium
Caldithrix abyssi
Firmicutes
Thermodesulfatator indicus
Andalucia incarcerata
Andalucia incarcerata
Acanthamoeba castellanii
Psalteriomonas lanterna
Sawyeria marylandensis
Trichomonas vaginalis
Holomastigotoides mirabile
Pseudotrichonympha grassii
Trichomonas vaginalis
Breviata anathema
Breviata anathema
Pygsuia_biforma
Nyctotherus ovalis
Chlorophytes
Neocallimastix frontalis
Piromyces sp.
Orpinomyces sp.
Naegleria gruberi
Trichomonas vaginalis
Streblomastix strix
Trichomonas vaginalis
Trichomonas vaginalis
Blastocystis hominis
Blastocystis sp.
Andalucia incarcerata Hyd-CysJ
Andalucia incarcerata Hyd-CysJ
Pygsuia biforma Hyd-CysJ
Pygsuia biforma Hyd-CysJ
Thermotogae
Alphaproteobacteria
Proteobacteria bacterium
Uliginosibacterium gangwonense
Desulfovibrio vulgaris
Pelobacter carbinolicus
Alphaproteobacteria
Thalassiosira pseudonana
Treponema denticola
Clostridium nexile
Firmicutes
Anaerostipes
Actinobacteria
Phascolarctobacterium succinatutens
Firmicutes
Bacteroides pectinophilus
Eubacterium plexicaudatum
Firmicutes
Psychrilyobacter atlanticus
Mollicutes bacterium
Spirochaeta
Alphaproteobacteria
Spirochaeta
Spirochaeta thermophila
Firmicutes
Coprobacillus sp.
Acidiphilium sp.
Clostridium novyi
Haloplasma contractile
Firmicutes
Firmicutes
Firmicutes
Ruminococcus sp.
Firmicutes
Firmicutes
Firmicutes
Firmicutes
Alphaproteobacteria
Melioribacter roseus
Bacteroidetes
Bacteroidetes
Deltaproteobacteria
Thermanaerovibrio acidaminovorans
Aminomonas paucivorans
Aminiphilus circumscriptus
Bacteroidetes
Clostridium sordellii
Fusobacterium mortiferum
Peptoclostridium difficile
bacterium UASB14
Chloroflexi bacterium
Parvimonas micra
Firmicutes
Mollicutes bacterium
Thermotogae
Firmicutes
Firmicutes
Deltaproteobacteria
Firmicutes
Succinispira mobilis
Synergistes jonesii
Phascolarctobacterium succinatutens
Bacteroidales bacterium
Sphaerochaeta pleomorpha
Desulfotomaculum reducens
Petrotoga mobilis
Halothermothrix orenii
Candidatus Caldatribacterium saccharofermentans
Thermodesulfatator indicus
Dehalococcoides mccartyi
Moorella thermoacetica
Symbiobacterium thermophilum
Holophaga foetida
Hydrogenedentes bacterium
Candidatus Omnitrophus sp.
Acetohalobium arabaticum
Halobacteroides halobius
0.3

Figure C.S4 Unrooted Maximum Likelihood (ML) tree of HydE sequences. **(Following page)** Phylogenetic analyses were performed on 100 sequences and 262 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and α-proteobacteria magenta.


Spironucleus salmonicida
Trepomonas sp.
Trichomonas vaginalis
Orpinomyces sp.
Piromyces sp.
Pygsuia biforma
Chlorophytes
Penium margaritaceum
Chromera velia
Vitrella brassicaformis
Acanthamoeba castellanii
Naegleria gruberi
Andalucia incarcerata
Firmicutes
Sebaldella termitidis
Fusobacteria
Acetohalobium arabaticum
Betaproteobacteria
Phaeospirillum fulvum
Rhodospirillum photometricum
Laribacter hongkongensis
Spirochaeta smaragdinae
Bacteroidetes
Candidatus Cloacamonas acidaminovorans
Desulfotalea psychrophila
Victivallis vadensis
Syntrophomonas wolfei
Pleomorphomonas oryzae
Pelobacter carbinolicus
Holophaga foetida
Firmicutes
Shewanella oneidensis
Desulfotomaculum acetoxidans
Symbiobacterium thermophilum
Moorella thermoacetica
Dethiobacter alkaliphilus
Firmicutes
Parvimonas sp.
Deltaproteobacteria
Petrotoga mobilis
Thermodesulfatator indicus
Syntrophus aciditrophicus
Thermodesulfovibrio yellowstonii
Syntrophothermus lipocalidus
Firmicutes
Firmicutes
Alphaproteobacteria
Alphaproteobacteria
Thermotogae
Coprothermobacter proteolyticus
Thermotoga lettingae
Thermotoga petrophila
Bacteroidetes
Coprobacillus sp.
Firmicutes
Firmicutes
Slackia heliotrinireducens
Lachnospiraceae bacterium
Butyrivibrio fibrisolvens
Oribacterium sp.
Firmicutes
Firmicutes
Firmicutes
Coriobacterium glomerans
Megamonas funiformis
Anaerococcus hydrogenalis
Clostridium difficile
Clostridium novyi
Peptostreptococcus stomatis
Syntrophobotulus glycolicus
0.2

Figure C.S5 Unrooted Maximum Likelihood (ML) tree of HydF sequences. **(Following page)** Phylogenetic analyses were performed on 218 sequences and 284 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and α-proteobacteria magenta.

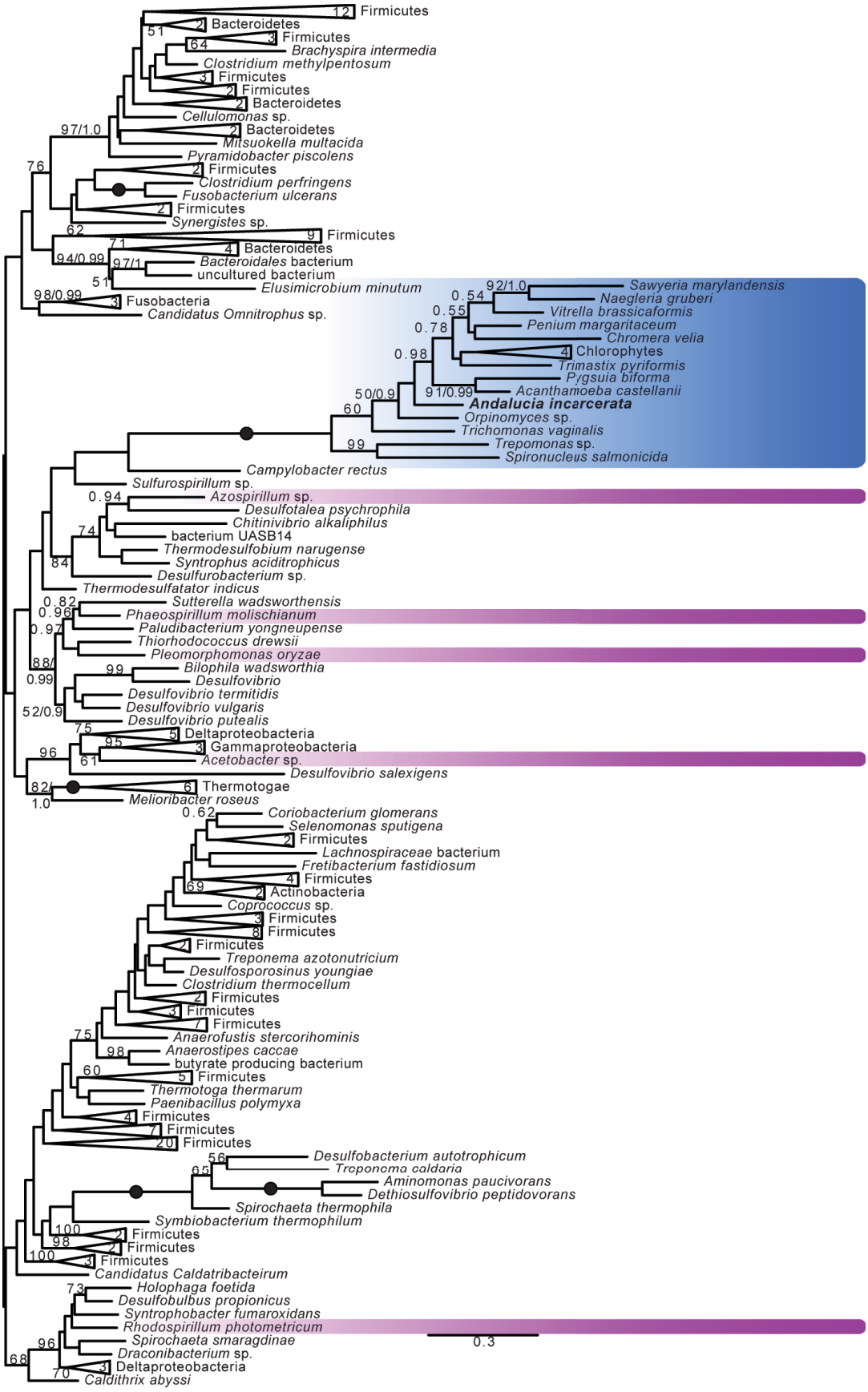

12 Firmicutes
51
2 Bacteroidetes
64
3 Firmicutes
Brachyspira intermedia
Clostridium methylpentosum
3 Firmicutes
2 Firmicutes
2 Bacteroidetes
Cellulomonas sp.
97/1.0
2 Bacteroidetes
Mitsuokella multacida
Pyramidobacter piscolens
76
2 Firmicutes
Clostridium perfringens
Fusobacterium ulcerans
2 Firmicutes
Synergistes sp.
62
9 Firmicutes
71
4 Bacteroidetes
94/0.99
97/1
Bacteroidales bacterium
uncultured bacterium
51
Elusimicrobium minutum
98/0.99
3 Fusobacteria
Candidatus Omnitrophus sp.
92/1.0
Sawyeria marylandensis
0.54
Naegleria gruberi
0.55
Vitrella brassicaformis
Penium margaritaceum
0.78
Chromera velia
0.98
4 Chlorophytes
Trimastix pyriformis
Pygsuia biforma
50/0.9
91/0.99
Acanthamoeba castellanii
Andalucia incarcerata
60
Orpinomyces sp.
Trichomonas vaginalis
99
Trepomonas sp.
Spironucleus salmonicida
Campylobacter rectus
Sulfurospirillum sp.
0.94
Azospirillum sp.
Desulfotalea psychrophila
Chitinivibrio alkaliphilus
74
bacterium UASB14
Thermodesulfobium narugense
84
Syntrophus aciditrophicus
Desulfurobacterium sp.
Thermodesulfatator indicus
0.82
Sutterella wadsworthensis
0.96
Phaeospirillum molischianum
0.97
Paludibacterium yongneupense
Thiorhodococcus drewsii
Pleomorphomonas oryzae
88/
0.99
99
Bilophila wadsworthia
Desulfovibrio
Desulfovibrio termitidis
Desulfovibrio vulgaris
52/0.9
Desulfovibrio putealis
75
5 Deltaproteobacteria
95
3 Gammaproteobacteria
96
61
Acetobacter sp.
Desulfovibrio salexigens
82/
1.0
6 Thermotogae
Melioribacter roseus
0.62
Coriobacterium glomerans
Selenomonas sputigena
2 Firmicutes
Lachnospiraceae bacterium
Fretibacterium fastidiosum
4 Firmicutes
69
2 Actinobacteria
Coprococcus sp.
3 Firmicutes
8 Firmicutes
2 Firmicutes
Treponema azotonutricium
Desulfosporosinus youngiae
Clostridium thermocellum
2 Firmicutes
3 Firmicutes
7 Firmicutes
75
Anaerofustis stercorihominis
98
Anaerostipes caccae
butyrate producing bacterium
60
5 Firmicutes
Thermotoga thermarum
Paenibacillus polymyxa
4 Firmicutes
7 Firmicutes
20 Firmicutes
56
Desulfobacterium autotrophicum
65
Treponema caldaria
Aminomonas paucivorans
Dethiosulfovibrio peptidovorans
Spirochaeta thermophila
Symbiobacterium thermophilum
100
2 Firmicutes
98
2 Firmicutes
100
3 Firmicutes
Candidatus Caldatribacteirum
73
Holophaga foetida
Desulfobulbus propionicus
Syntrophobacter fumaroxidans
Rhodospirillum photometricum
96
Spirochaeta smaragdinae
Draconibacterium sp.
68
3 Deltaproteobacteria
70
Calditrix abyssi
0.3

Figure C.S6 Unrooted Maximum Likelihood (ML) tree of ASCT1B sequences. **(Following page)** Phylogenetic analyses were performed on 171 sequences and 330 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and α-proteobacteria magenta.

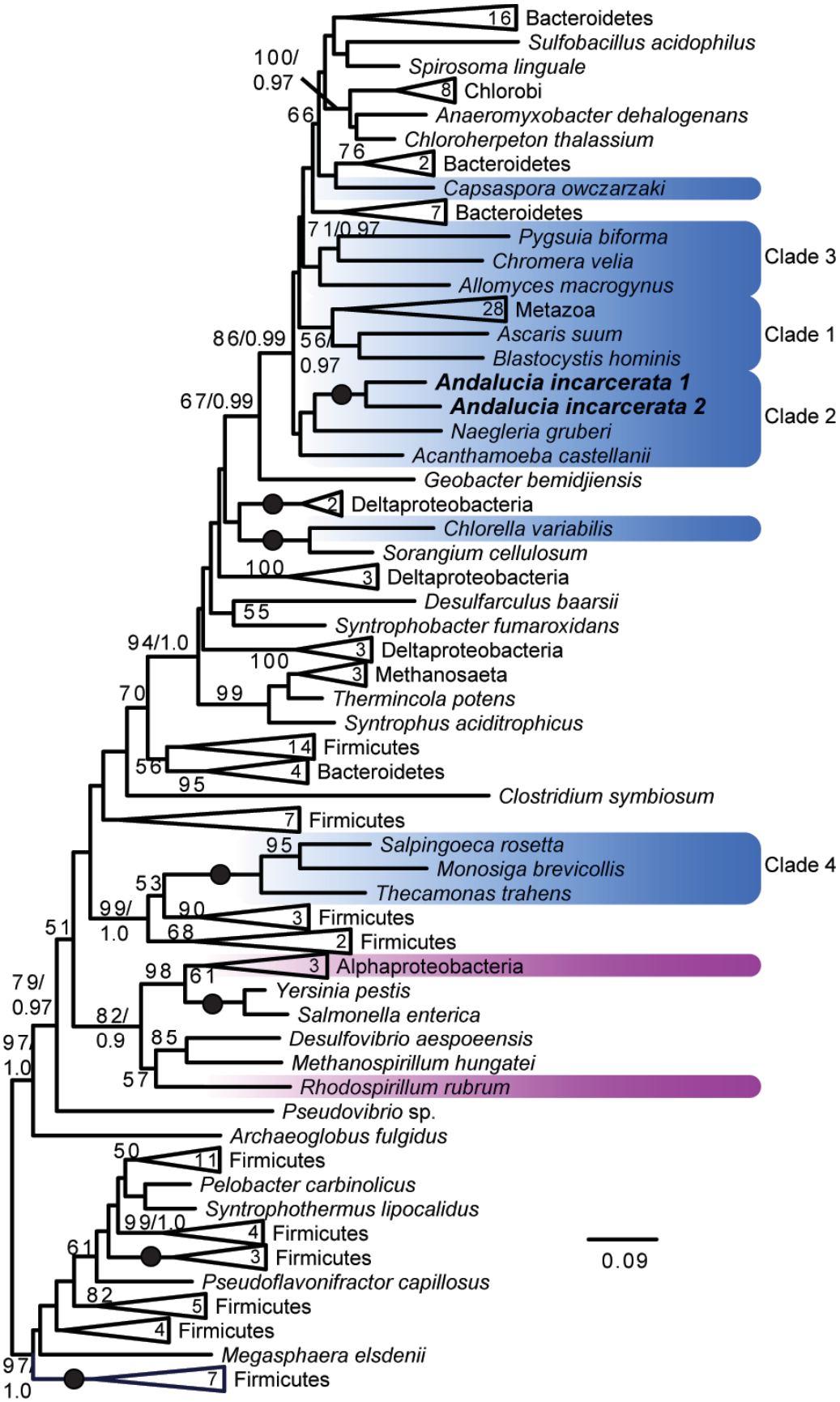
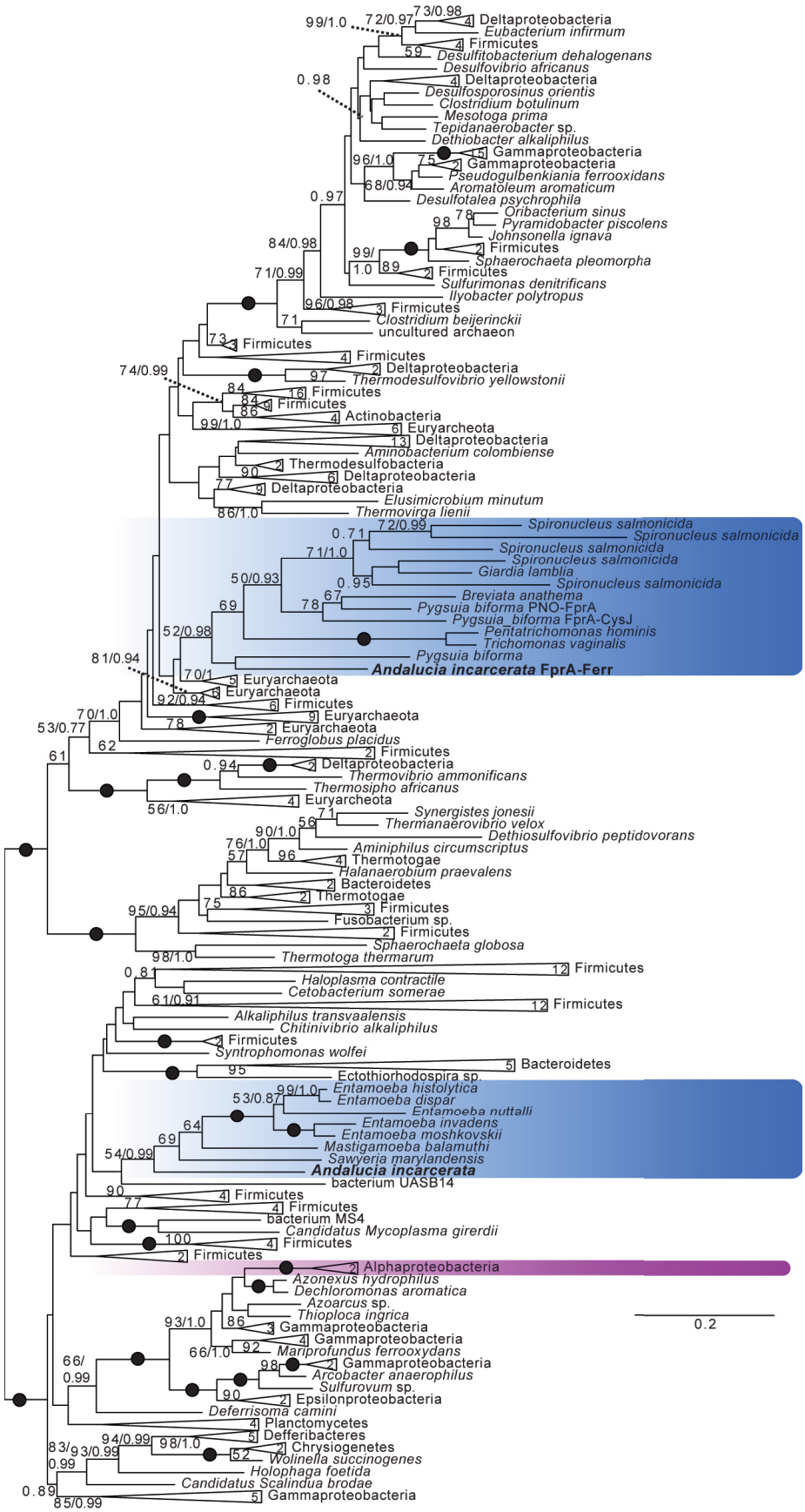

16 Bacteroidetes
Sulfobacillus acidophilus
100/ 0.97
Spirosoma linguale
8 Chlorobi
66
Anaeromyxobacter dehalogenans
Chloroherpeton thalassium
76
2 Bacteroidetes
Capsaspora owczarzaki
7 Bacteroidetes
71/0.97
Pygsuia biforma
Chromera velia
Clade 3
Allomyces macrogynus
28 Metazoa
86/0.99
56/ 0.97
Ascaris suum
Clade 1
Blastocystis hominis
67/0.99
Andalucia incarcerata 1
Andalucia incarcerata 2
Clade 2
Naegleria gruberi
Acanthamoeba castellanii
Geobacter bemidjiensis
2 Deltaproteobacteria
Chlorella variabilis
Sorangium cellulosum
100
3 Deltaproteobacteria
Desulfarculus baarsii
55
Syntrophobacter fumaroxidans
94/1.0
3 Deltaproteobacteria
100
3 Methanosaeta
70
99
Thermincola potens
Syntrophus aciditrophicus
14 Firmicutes
56
4 Bacteroidetes
95
Clostridium symbiosum
7 Firmicutes
95
Salpingoeca rosetta
Monosiga brevicollis
Clade 4
53
Thecamonas trahens
99/ 1.0
90
3 Firmicutes
51
68
2 Firmicutes
3 Alphaproteobacteria
98
61
Yersinia pestis
79/ 0.97
Salmonella enterica
82/ 0.9
Desulfovibrio aespoeensis
97/ 1.0
85
Methanospirillum hungatei
57
Rhodospirillum rubrum
Pseudovibrio sp.
Archaeoglobus fulgidus
50
11 Firmicutes
Pelobacter carbinolicus
Syntrophothermus lipocalidus
99/1.0
4 Firmicutes
61
3 Firmicutes
0.09
Pseudoflavonifractor capillosus
82
5 Firmicutes
4 Firmicutes
Megasphaera elsdenii
97/ 1.0
7 Firmicutes

Figure C.S7 Unrooted Maximum Likelihood (ML) tree of flavodiiron protein sequences. **(Following page)** Phylogenetic analyses were performed on 306 sequences and 273 sites, using RAxML and PhyloBayes. Bootstrap support values greater than 50%, and posterior probabilities greater than 0.5, are shown. Branches with 100% bootstrap support and posterior probability of 1.0 are indicated by black circles. Eukaryotes are shaded blue, and α-proteobacteria magenta. The ferredoxin-fused flavodiiron protein is indicated


Deltaproteobacteria
Eubacterium infirmum
Firmicutes
Desulfitobacterium dehalogenans
Desulfovibrio africanus
Deltaproteobacteria
Desulfosporosinus orientis
Clostridium botulinum
Mesotoga prima
Tepidanaerobacter sp.
Dethiobacter alkaliphilus
Gammaproteobacteria
Gammaproteobacteria
Pseudogulbenkiania ferrooxidans
Aromatoleum aromaticum
Desulfotalea psychrophila
Oribacterium sinus
Pyramidobacter piscolens
Johnsonella ignava
Firmicutes
Sphaerochaeta pleomorpha
Firmicutes
Sulfurimonas denitrificans
Ilyobacter polytropus
Firmicutes
Clostridium beijerinckii
uncultured archaeon
Firmicutes
Firmicutes
Deltaproteobacteria
Thermodesulfovibrio yellowstonii
Firmicutes
Firmicutes
Actinobacteria
Euryarcheota
Deltaproteobacteria
Aminobacterium colombiense
Thermodesulfobacteria
Deltaproteobacteria
Deltaproteobacteria
Elusimicrobium minutum
Thermovirga lienii
Spironucleus salmonicida
Spironucleus salmonicida
Spironucleus salmonicida
Spironucleus salmonicida
Giardia lamblia
Spironucleus salmonicida
Breviata anathema
Pygsuia biforma PNO-FprA
Pygsuia_biforma FprA-CysJ
Pentatrichomonas hominis
Trichomonas vaginalis
Pygsuia biforma
Andalucia incarcerata FprA-Ferr
Euryarchaeota
Euryarchaeota
Firmicutes
Euryarchaeota
Euryarchaeota
Ferroglobus placidus
Firmicutes
Deltaproteobacteria
Thermovibrio ammonificans
Thermosipho africanus
Euryarcheota
Synergistes jonesii
Thermanaerovibrio velox
Dethiosulfovibrio peptidovorans
Aminiphilus circumscriptus
Thermotogae
Halanaerobium praevalens
Bacteroidetes
Thermotogae
Firmicutes
Fusobacterium sp.
Firmicutes
Sphaerochaeta globosa
Thermotoga thermarum
Firmicutes
Haloplasma contractile
Cetobacterium somerae
Firmicutes
Alkaliphilus transvaalensis
Chitinivibrio alkaliphilus
Firmicutes
Syntrophomonas wolfei
Bacteroidetes
Ectothiorhodospira sp.
Entamoeba histolytica
Entamoeba dispar
Entamoeba nuttalli
Entamoeba invadens
Entamoeba moshkovskii
Mastigamoeba balamuthi
Sawyeria marylandensis
Andalucia incarcerata
bacterium UASB14
Firmicutes
Firmicutes
bacterium MS4
Candidatus Mycoplasma girerdii
Firmicutes
Firmicutes
Alphaproteobacteria
Azonexus hydrophilus
Dechloromonas aromatica
Azoarcus sp.
Thioploca ingrica
Gammaproteobacteria
Gammaproteobacteria
Mariprofundus ferrooxydans
Gammaproteobacteria
Arcobacter anaerophilus
Sulfurovum sp.
Epsilonproteobacteria
Deferrisoma camini
Planctomycetes
Defferibacteres
Chrysiogenetes
Wolinella succinogenes
Holophaga foetida
Candidatus Scalindua brodae
Gammaproteobacteria
0.2