Evolution of hydrogenosomes in anaerobic ciliates

William H. Lewis

Doctor of Philosophy

Institute for Cell and Molecular Biosciences

March 2017

Abstract

Within ciliates (protozoa of the phylum Ciliophora), anaerobic species are widespread and typically possess organelles which produce H_2 and ATP, called hydrogenosomes. Hydrogenosomes are mitochondrial homologues and are a product of evolutionary convergence, having been found in wide-ranging and diverse anaerobic eukaryotes. Ciliates seem to have evolved hydrogenosomes on multiple occasions from the mitochondria of their aerobic ancestors. The hydrogenosomes of the ciliate *Nyctotherus ovalis* were studied in detail previously but little is known about the hydrogenosomes from other ciliate species. In the present study seven species of ciliate, *Cyclidium porcatum, Metopus contortus, Metopus es, Metopus striatus, Nyctotherus ovalis, Plagiopyla frontata* and *Trimyema* sp. were cultured and their hydrogenosomes were investigated using genomic and transcriptomic sequencing from whole genome amplifications from single and small numbers of isolated cells. The data were then used to reconstruct putative hydrogenosome metabolic pathways. Components of these pathways are typically encoded by the ciliate nuclear genomes but *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus* and *Cyclidium porcatum* have also retained mitochondrial (now hydrogenosomal) genomes which were sequenced for the first time. The most complete of these genomes were from *Nyctotherus ovalis* and *Metopus contortus*. These have both retained genes for proton-pumping subunits of the electron transport chain Complex I and ribosomal subunits needed for their synthesis. The ciliates *Plagiopyla frontata* and *Trimyema* sp. appear to have completely lost the organelle genome during the conversion of mitochondria into hydrogenosomes.

The ciliate hydrogenosomes for which the most data was obtained appear to have retained some of the enzymes needed to produce energy by substrate-level phosphorylation but some species have also retained a partial electron transport chain and *Cyclidium porcatum* has retained nuclear encoded subunits of the mitochondrial F_1F_0 ATP synthase complex. Nuclear genes encoding enzymes that play a key role in H_2 production, FeFe-hydrogenase, pyruvate: ferredoxin oxidoreductase and pyruvate: $NADPH^+$ oxidoreductase, were also sequenced from the sampled ciliates and their evolutionary origins were investigated using phylogenies. These suggest that ciliate FeFe-hydrogenases are monophyletic and

have a separate bacterial origin from FeFe-hydrogenases in other eukaryotes. No evidence was found to support an alpha-proteobacteria or mitochondrial ancestry of these enzymes as predicted by the Hydrogen Hypothesis (Martin and Müller, 1998). Each of the ciliates investigated contained methanogenic Archaea endosymbionts, which can consume the H_2 produced by the hydrogenosomes. Some of these endosymbionts were identified to the species-level. The associations they have formed with their hosts appear to be stable over short time-scales but not over longer evolutionary periods, as closely related ciliates like *Nyctotherus ovalis* and *Metopus contortus* do not have closely related endosymbionts, providing no evidence for long-term co-speciation.

Acknowledgements

I would like to express my gratitude and appreciation to the following people. My supervisors, Professor Martin Embley, Professor Robert Hirt and Professor Genoveva Esteban, for their continued support and advice throughout this project. My colleagues Dr Kacper Sendra and Dr Tom Williams who have provided me with invaluable training and discussion. My collaborators Professor Thijs Ettema, Anders Lind and Henning Onsbring Gustafson, whose contributions to this project were indispensable. Ekaterina Kozhevnikova for her assistance in the laboratory. My colleagues Dr Cedric Bicep and Dr Andrew Watson for their discussions and help with bioinformatics. All my other colleagues from the Embley/Hirt lab: Dr Peter Major, Dr Paul Dean, Dr Heli Monttinen, Dr Alina Goldberg and Maxine Geggie.

Table of Contents

List of figures

List of tables

List of Abbreviations

ACO	Aconitase
ALDO	Fructose-bisphosphate aldolase
AOX	Alternative oxidase
ASCT	Acetate:succinate CoA transferase
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
CAI	Codon adaptation index
CS	Citrate synthase
DNA	Deoxyribonucleic acid
EM	Electron microscopy
ENO	Enolase
ETC	Electron transport chain
FH	Fumarate hydratase
FISH	Fluorescence *in situ* hybridisation
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GCK	Glucokinase
GPI	Glucose-6-phosphate isomerase
HMM	Hidden Markov model
IDH	Isocitrate dehydrogenase
IMP	Inner membrane peptidase
ISC	Iron-sulphur cluster
LSU	Large subunit
MCF	Mitochondrial carrier family
MDH	Malate dehydrogenase
MPP	Mitochondrial processing peptidase
OGDC	Oxoglutarate dehydrogenase complex
ORF	Open reading frame
OXA	Oxidase assembly
PFK	Phosphofructokinase

PFL	Pyruvate:formate lyase
PFO	Pyruvate:ferredoxin oxidoreductase
PGAM	Phosphoglycerate mutase
PGK	Phosphoglycerate kinase
PK	Phosphate kinase
PNO	Pyruvate:NADP+ oxidoreductase
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SDH	Succinate dehydrogenase
SCS	Succinate:CoA synthetase
SSU	Small subunit
STED	Stimulated emission depletion microscopy
TCA cycle	Tricarboxylic acid cycle
TEM	Transmission electron microscopy
TIM	Translocase of the inner membrane
TOM	Translocase of the outer membrane
TPI	Triose phosphate isomerase
tRNA	Transfer RNA

Chapter 1. General introduction

1.1 Eukaryote evolution and the significance of anaerobic metabolism

1.1.1 The evolution of eukaryotes

The chimeric nature of eukaryotes (Sagan, 1967) and their genomes (Golding and Gupta, 1995) suggests that they evolved as the result of a symbiosis between two prokaryotes. Since then eukaryotes have evolved to become more complex than their prokaryotic ancestors and are recognisable by their nucleus and the presence of double membrane-bound organelles (Margulis, 1970). Early studies attempting to understand how eukaryotes were related to other prokaryotic groups found that Archaea and eukaryotes had similar types of polymerase, distinct from the polymerases of Bacteria, indicating that they might be more closely related to each other (Huet et al., 1983). Further comparisons of ribosome structures suggested that eukaryotes were more closely related to a group of Archaea named eocytes, now named Crenarchaeota, than either were to other Archaea or Bacteria (Lake et al., 1984). From this finding a structure of the relationships between the main domains of life was inferred, which became known as the eocyte tree. This suggested that eukaryotes emerged from within the Archaea and that these two groups form a single clade to the exclusion of Bacteria. In contradiction to this view, phylogenetic analyses by Woese (1987), based on sequences of small subunit rRNA, recovered Bacteria, Archaea and eukaryotes as separate clades representing three primary domains of life, thereby placing eukaryotes as a sister clade to the Archaea. Despite some further support for the eocyte tree (Rivera and Lake, 1992), evidence from molecular phylogenetic analyses began to accumulate in favour of the three-domain structure of the tree of life, leading to this hypothesis becoming preferred by a consensus of researchers, which lasted for three decades (Gouy and Li, 1989; Woese et al., 1990; Ciccarelli et al., 2006). However this consensus in opinion has once more been overturned by a renewal of the eocyte tree hypothesis. This is due to recent studies, with more sophisticated methods and greater sampling, providing stronger evidence that eukaryotes emerged from within the Archaea (Cox et al., 2008; Foster et al., 2009; Williams et al., 2012; Williams et al., 2013), specifically within the Asgard superphylum (Zaremba-Niedzwiedzka et al., 2017), which is a recently discovered sister lineage to the TACK superphylum. Lokiarchaeota was the first phylum of

Asgard Archaea to be proposed (Guy and Ettema, 2011; Spang et al., 2015) but the closest related modern-day relatives to eukaryotes are now thought to be members of the Heimdallarchaeota (Zaremba-Niedzwiedzka et al., 2017) (Figure 1.1). Thus, current ideas have an archaeon related to members of the Asgard superphylum as the founders of the eukaryotic host lineage.

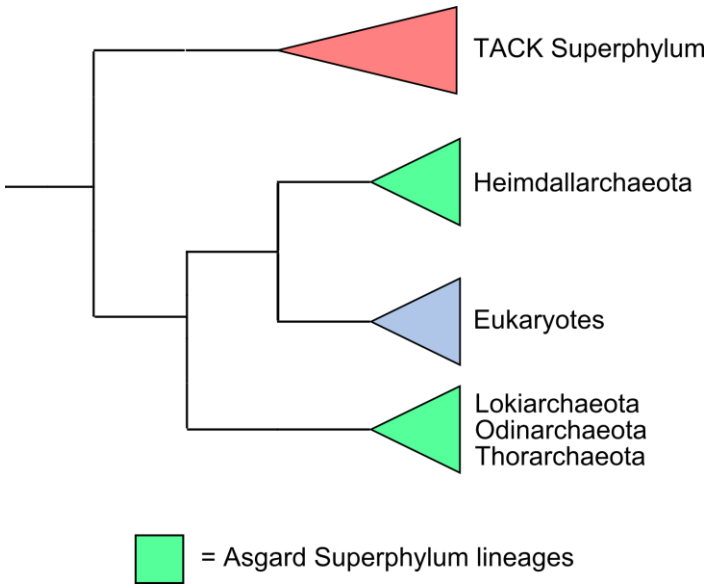


Figure 1.1 A cladogram representation of the relationships between eukaryotes, the Asgard Superphylum of Archaea and the TACK Superphylum of Archaea. Based on data from Zaremba-Niedzwiedzka et al. (2017).

The second prokaryotic contributor to the chimeric nature of eukaryotes is now thought to have been the ancestor of modern mitochondria (Embley and Martin, 2006). Evidence suggests that an archaeal ancestor of eukaryotes acquired a bacterium, related to contemporary alpha-proteobacteria, which became the mitochondrial endosymbiont and eventually evolved to become an organelle (Andersson et al., 1998). Mitochondria and their derivatives (collectively referred to as mitochondrial homologues) are an important feature of the eukaryotic cell and were retained by all but at least one species (Karnkowska et al., 2016) of extant eukaryotes studied to date, although they appear in many diverse functional and morphological guises (reviewed in Embley and Martin (2006) and Stairs et al. (2015)). The only features known to be conserved between all mitochondrial homologues are their double membranes and mechanisms of protein import (Dolezal et al., 2006; Embley and Martin, 2006). It is generally agreed that the initial selective success of mitochondria could not have been the provision of ATP to its host as no Bacteria are known to export ATP to their external environments (Martin and Müller,

1998) and the ATP/ADP translocases of mitochondria are not related to those found in alpha-proteobacteria (Andersson et al., 1998). Instead it has been suggested that the initial selective advantage of the mitochondrial endosymbiont was provision of a steady supply of H_2 needed as substrate for the metabolism of an autotrophic archaeon host (Martin and Müller, 1998). Mitochondria are now found in many metabolically diverse lineages of complex eukaryotes with both aerobic and anaerobic metabolisms. Mitochondrial ATP production is thought to have had an important role in the evolution of eukaryotic complexity, particularly driving selection for generally larger genomes in comparison to prokaryote genomes by providing energy for the expression of a larger repertoire of protein-coding genes (Lane and Martin, 2010). Many anaerobic eukaryotes do not have recognisable classical aerobic mitochondria and instead have other organelles called hydrogenosomes, some of which can produce ATP in the absence of O_2 (Muller et al., 2012).

1.1.2 The Archezoa Hypothesis

One of the earliest hypotheses for mitochondrial and early eukaryote evolution was based on the notion that modern-day anaerobic eukaryotes are descendants of deeply-branching eukaryotes that evolved prior to the mitochondrial endosymbiosis (Cavalier-Smith, 1987). These organisms were thought to have never contained mitochondria and were named Archezoa (Cavalier-Smith, 1987; Cavalier-Smith, 1989). The existence of what were thought to be contemporary Archezoa, such as *Giardia*, *Trichomonas*, *Entamoeba* and microsporidia, was the central piece of evidence supporting this hypothesis, which became known as the Archezoa Hypothesis. This provided a model for the early evolution of eukaryotes, as well as the existence of both aerobic and anaerobic metabolisms that are found in eukaryotes today. The Archezoa Hypothesis indicated that eukaryotes were already relatively complex and diverse, and exclusively inhabited anoxic environments, prior to the mitochondrial endosymbiosis (Cavalier-Smith, 1987). Following the acquisition of mitochondria, which therefore must have occurred relatively late in eukaryote evolution, following an initial period of diversification after the origin of the archezoans, eukaryotes then further diversified into aerobic lineages. This was also supported by phylogenies based on 18S rRNA gene sequences indicating that some

archezoan species were early branching eukaryotes (Vossbrinck et al., 1987; Sogin, 1989).

The Archezoa Hypothesis was challenged by improved phylogenies that no longer placed Archezoa species near the root of the eukaryotes (Cavalier-Smith, 1993; Hirt et al., 1999), as well as the discovery of mitochondrial genes in Archezoa (Clark and Roger, 1995; Bui et al., 1996; Horner et al., 1996; Roger et al., 1996; Germot et al., 1997; Hirt et al., 1997; Roger et al., 1998). Additionally, evidence from *Entamoeba histolytica* suggested that it possessed mitosomes, organelles of mitochondrial ancestry that have lost the ability to produce ATP (Mai et al., 1999; Tovar et al., 1999) and experimental data from the microsporidian *Trachipleistophora hominis* (Williams et al., 2002) provided more evidence to support the existence of mitosomes in an archezoan. Mitosomes have since been found in a variety of unrelated eukaryotes that have lost the ability to produce energy via mitochondria. Around the same period, hydrogenosomes (discussed in sections below), energy producing organelles found in the archezoan *Trichomonas vaginalis* (Lindmark and Müller, 1973), were discovered to be homologues of mitochondria (Hrdy et al., 2004; Sutak et al., 2004). The discovery of mitochondrial homologues in archezoan species suggested that mitochondria existed, in one form or another, in all eukaryotes. This not only provided strong evidence that the Archezoa Hypothesis should be rejected, since no organisms are now known to be true Archezoa, but also demonstrated previously unrecognised flexibility in mitochondrial form and function.

1.1.3 Evolutionary origins of hydrogenosomes in different eukaryotes

Unidentified microbodies found in the cytosol of trichomonads that produce molecular H_2 from protons were given the name ‘hydrogenosomes’ by their discoverers, Lindmark and Müller (1973). Although the true evolutionary nature and origin of these organelles was obscure, they were thought to have a role in the anaerobic metabolism and energy production of trichomonads, as they produced H_2 and metabolised pyruvate in similar ways to anaerobic *Clostridium* species (Bauchop, 1971; Lindmark and Müller, 1973). These similarities lead to the hypothesis that hydrogenosomes in trichomonads might have originated as a result of an endosymbiosis between a eukaryote and a *Clostridium*, similar to the endosymbioses known to have given rise to mitochondria and chloroplasts (Whatley et al., 1979).

Eventually hydrogenosomes were discovered in other anaerobic eukaryotes but initially the origins of these hydrogenosomes were thought to be different to those in *Trichomonas*. In the rumen fungus *Neocallimastix* sp. it was thought that hydrogenosomes were derived from peroxisomes since immunological evidence suggested that their hydrogenases had C-terminal peroxisomal-like targeting signals (Marvin-Sikkema et al., 1993). This was rejected however when later evidence suggested that their malic enzyme had an N-terminal mitochondrial-like targeting signal, and their hydrogenosomes had double membranes, mitochondrial-like ATP/ADP carriers and contained mitochondrial heat-shock proteins, Hsp60 and mtHsp70 (Van Der Giezen et al., 1997a; van der Giezen et al., 2002; van der Giezen et al., 2003). Finlay and Fenchel (1989) provided early evidence that hydrogenosomes in anaerobic ciliates had morphological features similar to mitochondria, suggesting that hydrogenosomes in other eukaryotic lineages had also evolved from mitochondria. Strong evidence for the mitochondrial origins of *Trichomonas* hydrogenosomes was provided by the discovery of mitochondrial heat-shock proteins and the 24 kDa and 51 kDa subunits of ETC Complex I in the hydrogenosomes of *Trichomonas vaginalis* (Bui et al., 1996; Germot et al., 1996; Horner et al., 1996; Roger et al., 1996; Hrdy et al., 2004) and perhaps the most direct evidence was obtained by the discovery of a hydrogenosome with a mitochondrial genome, in the ciliate *Nyctotherus ovalis* (Akhmanova et al., 1998)(further discussed in section 1.2.4).

The accumulation of evidence suggested that hydrogenosomes in all eukaryotic lineages had a mitochondrial ancestry (reviewed in Embley and Martin (2006); Hrdy et al. (2004); Gray (2005)). At the same time improved molecular phylogenies indicated that amitochondriate lineages did not branch near the base of eukaryotes but rather among mitochondria-containing eukaryotic lineages (Embley and Hirt, 1998; Hirt et al., 1999). This evidence allowed the Archezoa Hypothesis to be rejected for the groups for which it was originally proposed. Two main conclusions were drawn from this research: Firstly, all studied eukaryotes either contained mitochondria or organelles derived from mitochondria (mitosomes and hydrogenosomes), suggesting that the mitochondrial endosymbiosis occurred earlier in eukaryotic evolution than the archezoa hypothesis had predicted. Secondly, hydrogenosomes have convergently evolved from mitochondria in diverse, distantly

related eukaryotes, which suggests that the transformation of mitochondria to hydrogenosomes can occur with relative evolutionary ‘ease’.

1.1.4 The Hydrogen Hypothesis for the origin of eukaryotes

A number of hypotheses have been published to explain the origin of eukaryotes whilst accounting for the origins of both their aerobic and anaerobic metabolisms. Most also assume that eukaryotes are chimeric, based upon genomic analyses showing that they contain genes with archaeal and bacterial origins (Martin et al., 1993; Golding and Gupta, 1995). One such hypothesis is the Hydrogen Hypothesis (Martin and Müller, 1998), which provides a detailed metabolic argument for how the mitochondrial endosymbiosis occurred. Unlike some other hypotheses, such as the syntrophy hypothesis (Moreira and López-García, 1998), the Hydrogen Hypothesis is relatively parsimonious in that it only requires the occurrence of a single symbiosis event. The Hydrogen Hypothesis describes a scenario in which a H_2-producing facultatively anaerobic alpha-proteobacterium becomes the mitochondrial endosymbiont by forming a symbiosis with a H_2-dependant methanogenic archaeon host, which utilised the waste H_2, CO_2 and acetate produced by the bacterium as a substrate for its own metabolism and ATP production. The two organisms would have initially met in an anaerobic environment, where the archaeon could potentially benefit from a stable supply of substrates provided by H_2 producing Bacteria. In order to maximise transfer of substrates, the Hydrogen Hypothesis postulates that the archaeon would maximise its surface area contact with the symbiont, gradually encapsulating it. Being no longer exposed to the outer environment, in order for the symbiont to still acquire substrates for its own metabolism, it would be necessary for its genes encoding carbon substrate transporters to be transferred to the archaeon and for these transporters to be expressed and become functional in the archaeon outer membrane. This would enable the archaeon to import carbon-based substrates from the environment into its cytosol. In order to ensure the flow of these imported substrates from the host cytosol to the symbiont, the symbiont glycolytic pathway would also need to be transferred to the host cytosol, replacing the metabolic pathways of the host. Replacement of the host metabolic pathways would mean that the partnership had become irreversibly heterotrophic, both partners relying on the energy producing pathways that originated from the symbiont. The result of this

process would be a previously anaerobic archaeon with a hydrogen-producing organelle of bacterial ancestry which was also capable of aerobic respiration. This organism would have the metabolic pathways necessary to inhabit both oxic and anoxic environments, components of which are said to have been differentially lost or retained in various lineages as eukaryotes diversified (Martin and Müller, 1998).

The Hydrogen Hypothesis does not account for the origin of the eukaryotic nucleus or cytoskeleton (Martin and Müller, 1998) but these are argued to be the result of processes occurring independently after the mitochondrial acquisition and were supported by it through an enhanced capacity for energy generation (Lane and Martin, 2010). Recently it has been suggested from genomic analyses of Lokiarchaeota that these organisms are H_2-dependant (Sousa et al., 2016). If the last archaeal ancestor of eukaryotes was H_2-dependant too then this would be consistent with the Hydrogen Hypothesis (Martin and Müller, 1998). Others however, have raised the possibility that Lokiarchaeota and some TACK Archaea, could be phagotrophic, as genomic analyses have identified homologues of eukaryotic cytoskeleton components in these species (Ettema et al., 2011; Koonin, 2015; Spang et al., 2015). This indicates that the common ancestor of eukaryotes might have been phagotrophic and this could have facilitated the mitochondrial endosymbiosis (Ettema et al., 2011; Guy and Ettema, 2011). Therefore, despite the popularity of the Hydrogen Hypothesis, there has been some criticism of it as a strongly supported model for eukaryotic evolution (discussed in section 1.1.5).

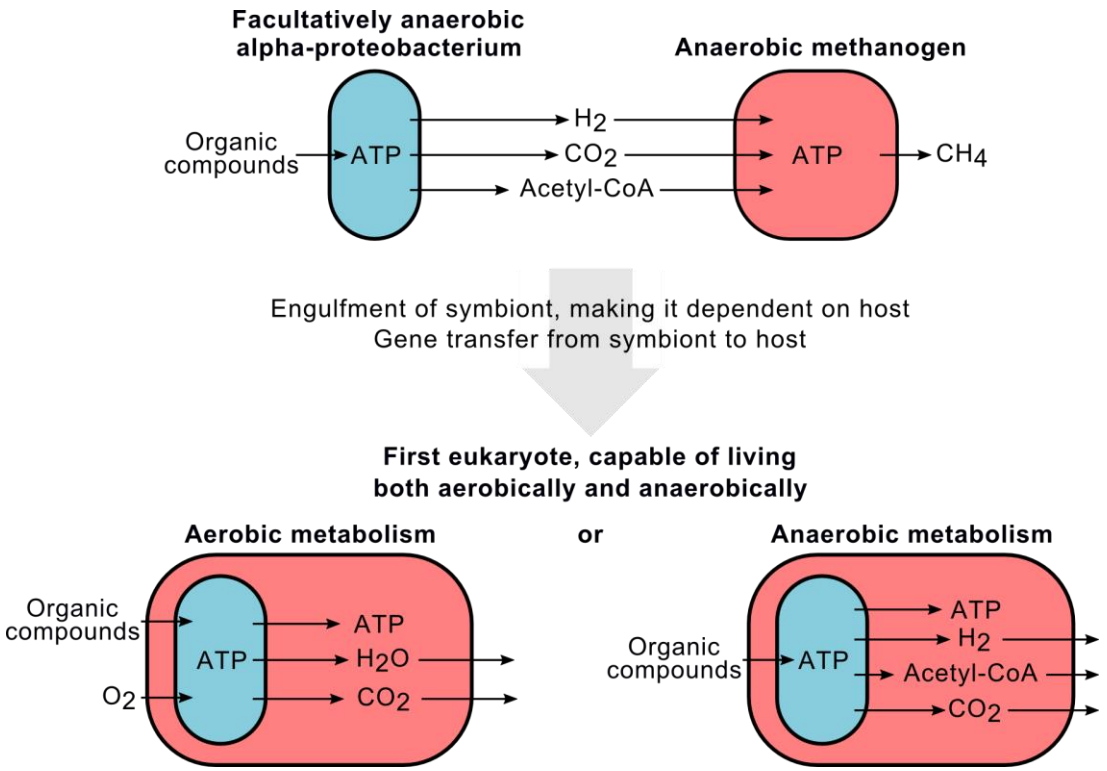


Figure 1.2 A schematic view of the hydrogen hypothesis for the first eukaryote (Martin and Müller, 1998).

1.1.5 Eukaryotic anaerobic metabolism enzymes: Ancestral acquisition with differential loss or repeated acquisition by lateral gene transfer?

Hydrogenosomes are double membrane-bound organelles with an ability to produce H_2 (Hrdy et al., 2004) that are found in anaerobic eukaryotes such as anaerobic ciliates (Yarlett et al., 1981), trichomonads (Lindmark and Müller, 1973) and chytrid fungi (Yarlett et al., 1986). H_2 production is achieved by a FeFe-hydrogenase enzyme (Horner et al., 2002), an Fe-S cluster protein, which transfers electrons from ferredoxin or NADH-dehydrogenase to protons, thereby producing molecular H_2 (Hrdy et al., 2004; van der Giezen et al., 2005). FeFe-hydrogenases are also found in various Bacteria (Shepard et al., 2014) and also some anaerobic eukaryotes that do not possess hydrogenosomes, in which their FeFe-hydrogenases are localised to the cytosol (Nixon et al., 2003). The successful maturation and integration of Fe-S clusters into the FeFe-hydrogenase active site typically requires the activity of 3 maturases, HydE HydF and HydG. However, these enzymes have not been found in all eukaryotes that possess an FeFe-hydrogenase, including *Giardia intestinalis* and *Entamoeba histolytica*, for which whole genome data is available (Loftus et al., 2005; Morrison et al., 2007). In these organisms, maturation of FeFe-hydrogenase has been suggested to either occur by an alternative, unknown means (Pütz et al., 2006)

or that they are only partially matured by Fe-S cluster assembly systems (Nicolet and Fontecilla-Camps, 2012).

In aerobic eukaryotes pyruvate is decarboxylated by pyruvate dehydrogenase (PDH), forming acetyl-CoA, which is then oxidised by the TCA cycle. However PDH can be inhibited by high concentrations of NADH (Bremmer, 1969), which can increase under anaerobic conditions due to a reduced and less efficient ETC. In the hydrogenosomes of some anaerobic eukaryotes, the function of pyruvate decarboxylation can instead be performed by pyruvate: ferredoxin oxidoreductase (PFO)/pyruvate: $NADP^+$ oxidoreductase (PNO), in organisms including *Trichomonas* and *Blastocystis* (Lindmark et al., 1975; Gill et al., 2007) or pyruvate: formate lyase (PFL), in organisms including *Neocallimastix* (Akhmanova et al., 1999). The aerobic mitochondria-containing relatives of these species appear to be unable to produce H_2 because they lack the enzymes to do this. Therefore determining the origins of FeFe-hydrogenases and their maturases, as well as pyruvate oxidising enzymes, is essential for understanding hydrogenosome evolution.

The Hydrogen Hypothesis suggests that the mitochondrial endosymbiont already contained a set of genes capable of supporting hydrogenosome metabolism (Martin and Müller, 1998). If this were true then it could be predicted that anaerobic metabolic enzymes like FeFe-hydrogenase, HydE, HydF, HydG, PFO/PNO and PFL, found in polyphyletic anaerobic eukaryotes, should share a common origin from a facultatively anaerobic alpha-proteobacterium ancestor that became the mitochondrial endosymbiont. This would require that anaerobic enzymes were retained in some aerobic lineages, or perhaps more likely, in organisms that are facultative anaerobes. The idea that genes encoding anaerobic enzymes were present in the genome of the mitochondrial endosymbiont is supported by the presence of proteins that are divergent homologues of anaerobic metabolism enzymes, but have acquired novel functions, that are found in diverse aerobic eukaryotes (Muller et al., 2012). These include yeast sulfite reductases that seem to be derived from PNO (Horner et al., 1999; Rotte et al., 2001), and nuclear prelamin-A recognition factor (Narf/Nar1), which are homologues of FeFe-hydrogenase and are present in most aerobic eukaryotes (Horner et al., 2002; Freibert et al., 2017). Narf proteins are structurally similar to FeFe-hydrogenases but function in the maturation of cytosolic and nuclear iron-sulphur proteins (Vignais et al., 2001; Balk et al., 2004). It has been argued that these enzymes evolved from gene duplications of anaerobic

enzymes in the eukaryote common ancestor (Muller et al., 2012). Therefore, under the Hydrogen Hypothesis, it would mean that anaerobic metabolism enzymes were lost from all of the contemporary aerobic eukaryotes studied to date and only retained in the anaerobic lineages.

Phylogenetic analyses have been used to investigate the origins of the enzymes that function in hydrogenosomes (Horner et al., 1999; Horner et al., 2000; Hug et al., 2010; Hampl et al., 2011). Although phylogenies of PFO/PNO sometimes recover eukaryote monophyly (Horner et al., 1999; Rotte et al., 2001), the relationships recovered within eukaryote clades do not typically match the species trees for the organisms from which they are found and their closest prokaryotic relatives do not appear to be alpha-proteobacteria (Horner et al., 1999; Hug et al., 2010). Phylogenies of HydE, HydF and HydG maturases also recover eukaryotic monophyly but again they do not group with alpha-proteobacteria (Hug et al., 2010), whereas FeFe-hydrogenase phylogenies typically show eukaryotes to be polyphyletic, emerging from various bacterial groups but not alpha-proteobacteria (Horner et al., 2000; Embley et al., 2003; Nixon et al., 2003; Leger et al., 2016). Taken at face value, these results indicate that the key enzymes required for anaerobic metabolism in eukaryotes were not acquired from the mitochondrial endosymbiont. However to account for these patterns in a manner that can be reconciled with the Hydrogen Hypothesis, it has been argued (Muller et al., 2012) that the reasons why they group with bacterial groups other than the alpha-proteobacteria is firstly, because the genome of the organism that became the mitochondrial endosymbiont probably already contained a mosaic of genes of different prokaryotic origins, and secondly the alpha-proteobacteria, and also other lineages, have continued to evolve by acquiring genes from and transferring genes to other prokaryotes, since the mitochondrial endosymbiosis. Both of these suggestions are supported by available genome data for prokaryotes. This would allow genes present in relatives of the mitochondrial endosymbiont to have been acquired laterally by other types of Bacteria, such as the lineages that eukaryotic genes now cluster with in phylogenies. This could also mean that the genomes of modern-day alpha-proteobacteria have a different gene composition compared to their ancestors had at the time of the mitochondrial endosymbiosis. Under these conditions it is argued that phylogenetic analyses are unable to falsify a mitochondrial ancestry of these genes (Martin, 1999; Rotte et al., 2001; Muller et al., 2012) and it would

therefore be difficult to disprove this hypothesis. It is true that lateral gene transfer appears to have occurred on a large scale in prokaryotes since the mitochondrial endosymbiosis (Doolittle, 1999; Ochman et al., 2000) but a key argument against the process having muddied our understanding of the history of anaerobic enzymes is that the same patterns are not usually observed for genes encoding enzymes for aerobic mitochondrial metabolism. For example, phylogenies of the 51 kDa subunit of ETC Complex I (Hrdy et al., 2004) and enzymes of the Fe-S cluster biogenesis pathway (Freibert et al., 2017) show a clear ancestry leading to the alpha-proteobacteria, which is what would be expected for enzymes acquired via the mitochondrial endosymbiont.

An alternative explanation for these observations is that anaerobic enzymes are not exclusively inherited vertically in eukaryotes but rather in some cases they are acquired by lateral gene transfer from prokaryotic donors. This seems to have happened on multiple occasions for FeFe-hydrogenases since eukaryotic sequences do not appear to be monophyletic (Horner et al., 2000; Embley et al., 2003). Additionally, it is possible that such enzymes could then be passed on to other unrelated eukaryotic lineages by eukaryote-to-eukaryote lateral gene transfer (Stairs et al., 2015). This is supported by the conflicting topologies between phylogenies of anaerobic metabolism enzymes and the species tree of the organisms they are found in (Hug et al., 2010). The difficulty with this is that unlike lateral gene transfer in prokaryotes, which is a well-documented and widely accepted phenomenon, there is disagreement in the research community of whether the same occurs in eukaryotes to such a degree that it is a significant evolutionary process (Ku and Martin, 2016). One issue is that sequenced eukaryotic genome assemblies can suffer from contamination of prokaryote DNA that is mistaken for a case of lateral gene transfer (Boothby et al., 2015; Koutsovoulos et al., 2016) and it has been argued that there is not enough evidence of recently occurring lateral gene transfer from prokaryotes to eukaryotes to indicate that the process is ongoing and important (Ku and Martin, 2016). However numerous publications take an opposing stance, finding cases of lateral gene transfer to eukaryotes (Eme et al., 2017), which have been reviewed in detail by various authors (Andersson, 2009; Hirt et al., 2015; Soucy et al., 2015). These data suggest that lateral gene transfer is an occurring process in eukaryotes, most commonly affecting metabolic pathways (Alsmark et al., 2013).

1.1.6 The metabolic functions of mitochondria

Mitochondria perform several functions for eukaryotic cells, including ATP synthesis, Fe-S cluster biogenesis, heme biosynthesis and amino acid metabolism, as well as playing a role in apoptosis. The only function that was previously thought to be conserved by all mitochondrial homologues known so far was the biogenesis of Fe-S clusters, cofactors that have structural and catalytic roles in various proteins and are essential for life (Lill et al., 1999). However in *Monocercomonoides* sp. (Karnkowska et al., 2016) this function has been replaced by a system of bacterial origin, located in the cytosol. Another important role of mitochondria and hydrogenosomes is the production of ATP. In eukaryotes both aerobes with classical mitochondria and anaerobes with hydrogenosomes conserve energy from the chemical breakdown of complex substrates, such as glucose, via series of redox reactions. In eukaryotes glucose is broken down via the glycolytic pathway, which produces a relatively small net amount of 2 moles of ATP per mole of glucose (Muller et al., 2012).

In aerobic mitochondria pyruvate, the product of glycolysis, is transported across the outer and inner mitochondrial membranes via porins and mitochondrial pyruvate carriers, respectively (Bender et al., 2015), into the mitochondrial matrix where it is oxidatively decarboxylated to acetyl-CoA by PDH (Patel et al., 2014). Acetyl-CoA then enters the TCA cycle, where it undergoes a cycle of enzymatic steps producing a number of products, including NADH (Fernie et al., 2004). NADH is a cofactor that can readily undergo redox reactions (Fenchel and Finlay, 1995), one role of which is to transfer electrons to the ETC (Tielens et al., 2002). The ETC powers ATP synthesis by oxidative phosphorylation in aerobic mitochondria and uses O_2 as a terminal electron acceptor, which is effective due to its high redox potential relative to NAD^+ (Fenchel and Finlay, 1995). The ETC is formed by four Fe-S cluster-containing multi-protein complexes embedded in the phospholipid-bilayer of the inner mitochondrial membrane, as well as ubiquinone and cytochrome c (Saraste, 1999). ETC Complex I oxidises NADH, enabling H^+ to be pumped across the inner mitochondrial membrane into the inter-membrane space (Brandt, 2006). The released electrons are then transferred to ubiquinone, reducing it to ubiquinol. ETC Complex II, which is also a component of the TCA cycle, oxidises succinate to fumarate and in doing so transfers electrons to ubiquinone, reducing it to ubiquinol (Sun et al., 2005). Electrons from ubiquinol are transferred to ETC Complex III and then ETC Complex IV via cytochrome *c* (Xia et al., 1997). ETC Complex IV is finally

oxidised by O_2 producing H_2O (Iwata et al., 1995). As ETC Complexes III and IV transport electrons they also pump H^+ across the inner mitochondrial membrane, similarly to ETC Complex I, generating a proton gradient across the inner mitochondrial membrane (Iwata et al., 1995; Xia et al., 1997). The F_1F_0-ATP synthase complex is also embedded in the inner mitochondrial membrane, creating a channel connecting the matrix to the inter-membrane space, facilitating the flow of H^+ pumped by ETC Complexes I, III and IV back into the mitochondrial matrix and in doing so generates energy in the form of ATP (Boyer, 1997). There is a large difference in redox potential between NADH and O_2, therefore the flow of electrons from one molecule to the other via the ETC generates a large amount of available free energy that can be conserved as molecules of ATP (approximately an additional 34 moles of ATP per mole of glucose), more than is available to anaerobes that lack this ability (Fenchel and Finlay, 1995).

Anaerobic eukaryotes are unable to utilize the high redox potential of O_2 and therefore must maximise energy production by an alternative means. Anaerobes obtain pyruvate as an end-product of glycolysis but species such as *Trichomonas vaginalis* (Hrdý and Müller, 1995b) and *Neocallimastix* sp. (Van Der Giezen et al., 1997a) are also thought to produce pyruvate from malate via malic enzyme in the hydrogenosome matrix. As described in section 1.1.5, instead of PDH, hydrogenosomes often utilise alternative enzymes for the decarboxylation of pyruvate, such as PFO, PNO and PFL, the origins of which in eukaryotes are unclear. Instead of the produced acetyl-CoA entering the TCA cycle and being converted to citrate by citrate synthase, like in aerobic mitochondria, its CoA moiety is transferred to succinate by acetate:succinate CoA transferase (ASCT) to form succinyl-CoA and acetate (Müller, 1993; Tielens et al., 2010). The acetate is excreted but succinyl-CoA, along with ADP and P_i, can be used to produce ATP and succinate by substrate-level phosphorylation, via succinyl-CoA synthetase, a component of the TCA cycle. An alternative to ATP production by ASCT and SCS is found in the hydrogenosomes of the anaerobic amoeba *Mastigamoeba balamuthi*, which instead uses an acetyl-CoA synthase enzyme to produce ATP directly from acetyl-CoA, again by substrate-level phosphorylation (Gill et al., 2007).

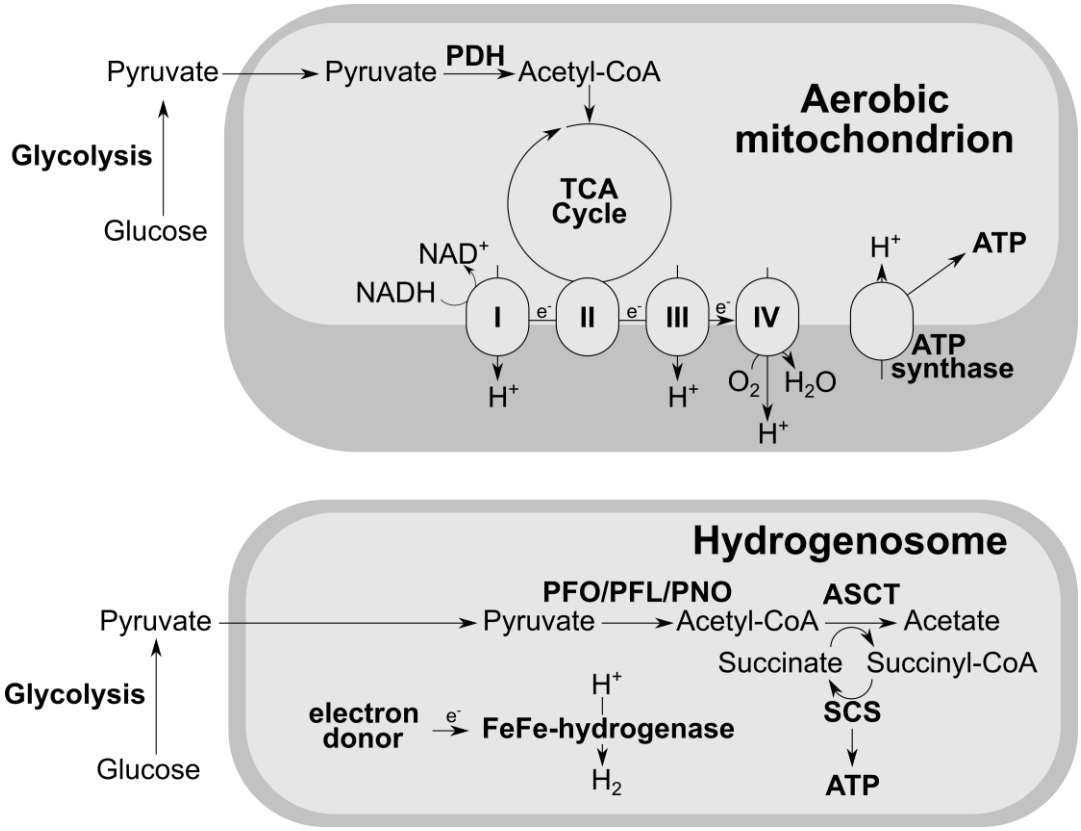


Figure 1.3 A generalised comparison of ATP production in aerobic mitochondria by oxidative phosphorylation and ATP production by substrate-level phosphorylation in a hydrogenosome that has lost the electron transport chain.

1.1.7 Hydrogenosome metabolism and morphology

Hydrogenosomes have been studied in a number of different anaerobic eukaryotes and although they share the phenotype of H_2 and ATP production, their metabolisms can differ. Recent reviews of the metabolic diversity and evolution of hydrogenosomes were published by Muller et al. (2012) and Stairs et al. (2015). Typically the evolution of hydrogenosomes from mitochondria involves a reduction in metabolic capacity by the loss of components from the TCA cycle and ETC. The most reduced hydrogenosomes have lost all ETC components and have been described in species such as *Spironucleus salmonicida*, which therefore produce energy by substrate-level phosphorylation (Jerlström-Hultqvist et al., 2013). Other hydrogenosomes, like those found in *Trichomonas vaginalis* (Hrdy et al., 2004) and *Sawyeria marylandensis* (Barberà et al., 2010) have retained the 24kDa and 51kDa NADH-dehydrogenase components of ETC Complex I, which they use to oxidise NADH, regenerating NAD^+, an essential electron carrier used in many metabolic pathways (Hrdy et al., 2004). Species like the anaerobic ciliate *Nyctotherus ovalis*

have more complete ETC Complexes I and II (de Graaf et al., 2011), although they still lack some subunits found in aerobic eukaryotes. Despite this *Nyctotherus ovalis* and other anaerobic eukaryotes have retained the main subunits of ETC Complex I involved in proton translocation (de Graaf et al., 2011), suggested to be homologous to bacterial antiporters (Mathiesen and Hägerhäll, 2002). Since these hydrogenosomes lack F_1F_0 ATP synthase, there is no need to translocate protons for the sake of ATP production by the F_1F_0 ATP synthase complex, like in aerobic eukaryotes. Since a membrane potential is however usually necessary for protein import across the inner membrane in mitochondria, this may be the reason proton translocation is preserved in these organisms (Chacinska et al., 2009). Hydrogenosomes and mitochondria that possess ETC Complex I oxidise NADH and produce ubiquinol. However, unlike aerobic mitochondria, many hydrogenosomes lack ETC Complex III, which oxidises ubiquinol in aerobic mitochondria. This would therefore lead to an accumulation of ubiquinol as it cannot be further oxidised by the ETC. In such cases it is thought that succinate is used as a sink to relinquish electrons from rhodoquinone (a quinone with a lower redox potential often found in anaerobes, replacing the role of ubiquinone in aerobes (Van Hellemond et al., 1995)). Electrons are transferred to succinate via ETC Complex II, which functions as succinate dehydrogenase in the TCA cycle of aerobes but in some anaerobes functions in reverse as fumarate reductase. Alternatively ubiquinol could potentially be reoxidised by alternative oxidase (AOX) if O_2 is present.

Cristae are invaginations of the mitochondrial inner membrane that form a complex mitochondrial architecture that is thought to be necessary for metabolic function, since certain regions of cristae membranes are more enriched in some ETC complexes than others (Davies et al., 2011; Milenkovic and Larsson, 2015). It has been demonstrated in *Saccharomyces cerevisiae* that subunits e and g of F_1F_0 ATP synthase complex are required for its dimerisation (Arnold et al., 1998; Velours and Arselin, 2000) and mitochondria of *Saccharomyces cerevisiae* mutants devoid of these F_1F_0 ATP synthase subunits exhibit atypical cristae structures (Arselin et al., 2004). This is because F_1F_0 ATP synthase dimers are directly involved in the formation in cristae tips, where they can be observed on the EM level organised into distinctive rows (Paumard et al., 2002; Minauro-Sanmiguel et al., 2005; Davies et al., 2011). The loss of these proteins therefore would be expected to result in either reduced or complete loss of cristae and this is observed in most hydrogenosomes.

Consistent with this, treatment of mammalian cells with ethidium bromide is thought to disrupt translation of the mitochondrial genome, which resulted in a reduction of the number of cristae observed at the EM level (Soslau and Nass, 1971). Interestingly, the hydrogenosomes of *Brevimastigamonas vehiculus*, which have retained an F_1F_0 ATP synthase complex do appear to have cristae (Gawryluk et al., 2016). This link between function and structure of mitochondrial homologues suggests that the presence of cristae in particular hydrogenosomes could provide an indication that a mitochondrial-type organelle genome is also present. The reasoning for this is that genes found on the mitochondrial genomes are required for the ETC, as well as components required for their translation. Therefore if the ETC is lost, not only would cristae become reduced, there would also no longer be any positive selection to retain an organelle genome as the genes it encodes would be no longer required.

1.1.8 The mitochondrial proteome: Encoded by nuclear and mitochondrial genomes

Some of the genes encoding proteins functioning in mitochondria are thought to have been acquired from the mitochondrial endosymbiont (Gabaldón and Huynen, 2003; Timmis et al., 2004). Modern mitochondria contain ancestral relics of this genome, located within the mitochondrial matrix (Taanman, 1999), which typically encodes ETC components and proteins involved in their transcription and translation. Many mitochondrial homologues, such as some hydrogenosomes and mitosomes, have however lost their organelle genomes by reductive evolution (Müller, 1993; Tovar et al., 1999) presumably due to a lack of positive selection ensuring their preservation. The majority of genes encoding other proteins that function in the mitochondria and originated from the mitochondrial endosymbiont, such as some of the mitochondrial ribosomal proteins and some of the core components of the ETC complexes, have been transferred to the nuclear genome (Karlberg et al., 2000). These proteins are synthesised by ribosomes in the cytosol of the cell and transported into mitochondria by specific targeting signals and import mechanisms. The selective forces driving this relocation are thought to occur due to the differences in the way these two genomes are reproduced (Allen and Martin, 2016). Mitochondrial genomes reproduce asexually, whereas nuclear genomes can reproduce sexually. In asexual populations deleterious mutations accumulate irreversibly in a process known as Muller's Ratchet

(Muller, 1964). Conversely, in sexual populations the accumulation of deleterious mutations can be reversed by recombining with another genome copy (homologus recombination) that does not share them. Not only that but mutations have also been suggested to occur more frequently in mitochondrial genomes due to the high levels of mutagenic reactive oxygen species (Allen and Martin, 2016). Therefore from a selection perspective genes are better off, with regards to their fitness, being in the nuclear genome. This also raises the question of why mitochondria retain any genome at all if there is so much selection against it. Numerous hypotheses have been proposed to explain this. One such explanation is the need for co-location of gene and gene product for redox regulation, known as the CoRR hypothesis (Allen, 1993, 2003), which suggests that an organelle genome is retained in order for the expression of the genes it encodes to be regulated by the current redox state of the organelle. This is because important energy generating ETC components, encoded by the organelle genome, are affected by changes in local redox conditions within the organelle in which they function and therefore their expression must be altered accordingly to be able to respond rapidly to compensate for these changes (Allen, 2015). Another key hypothesis is that certain protein subunits encoded by mitochondrial genomes are too hydrophobic to import across the mitochondrial membranes and it is therefore more efficient to synthesise these within the organelle (Adams and Palmer, 2003).

1.2 Ciliates

1.2.1 General description of ciliates

The SAR super-group (Burki et al., 2008) includes Stramenopiles, Alveolates and Rhizaria. Along with dinoflagelates, apicomplexa and protalveolata, ciliates (phylum Ciliophora) group within the Alveolates (Adl et al., 2012) and evolved approximately 1400 Ma, during the Mesoproterozoic (Parfrey et al., 2011). A distinctive characteristic of ciliates is their nuclear dimorphism: they have both a germline micronucleus (MIC), which meiotically divides to form haploids before they are exchanged during sexual conjugation (Prescott, 2000; Stover and Rice, 2011), and a somatic macronucleus (MAC), which is transcriptionally active and derived from mitotic division and nuclear rearrangement of the MIC (Katz, 2001). Except for within

the class Karyorelictea, the MAC can also divide by a unique form of amitosis during asexual reproduction (Zufall et al., 2006) and the number of MIC and MAC per cell varies between species (Prescott, 2000). A characteristic shared by all ciliates, which gives the group its name, is the presence of multiple cilia during at least one stage of their life cycles. Cilia are membrane-bound protrusions from a cell surface that contain an axoneme: A cytoskeleton consisting of microtubules in organised arrangements. The axoneme originates from modified centrosomes called the basal body (Pedersen et al., 2012). Cilia can be used for a range of behaviours and functions, including motility, sensory reception and feeding (Verni and Gualtieri, 1997). The cilia ultrastructure has been conserved across many eukaryotic lineages, both single- and multi-cellular, including humans and other mammals where they are found in cells of the kidneys, pancreas, respiratory tract, female reproductive tract and regions of the brain, amongst other organs (Satir and Christensen, 2007). This high level of conservation underlines the importance of cilia in eukaryotes and the study of cilia in ciliates has facilitated our understanding of the cilia found in other organisms (Pazour et al., 2005; Satir and Christensen, 2007).

Since their discovery in the 17th century by Anthony van Leeuwenhoek, ciliates have traditionally been identified and grouped based on microscopic observations of their morphology (Lynn, 2003). Since the advent of molecular technology, like most organisms ciliates are now more commonly classified based on comparisons of their DNA by phylogenetic reconstruction, which often provide support for the original morphology-based groupings (Lynn, 2003). Increasingly multiple locus-based phylogenetic reconstructions are improving the resolution of deeply diverging ciliate groups (Gao and Katz, 2014). Based on such analyses, ciliates can be divided into two sub-phyla: Intramacronucleata, which divide their MAC by microtubules within the MAC, and the Postciliodesmatophora, which divide their MAC with microtubules external to the MAC or are unable to divide them at all (Gao and Katz, 2014). Ciliates inhabit a range of environments, including freshwater and marine environments, metazoan digestive tracts, as well as soils that are at least periodically aquatic: some ciliates can survive drought, as well as other environmental stresses such as starvation, by forming temporary cysts (Gutiérrez et al., 2001; Lynn, 2008), whereby cells decrease their volume and metabolic activities, dehydrating their cytosol and forming a protective outer barrier (Verni and Rosati, 2011). Most free-living ciliates are heterotrophs, and their primary method of nutrient acquisition is by engulfing and

digesting Bacteria (Verni and Gualtieri, 1997) and other small eukaryotes (Siqueira-Castro et al., 2016) by phagocytosis. However, some species are mixotrophic as they have acquired means of phototrophic energy production either via photosynthesising endosymbionts or by sequestering chloroplasts from other organisms (Esteban et al., 2009; Esteban et al., 2010; Johnson, 2011). Some ciliates are also parasites of aquatic animals, acquiring nutrients directly from their hosts by feeding on their tissues (Nigrelli et al., 1976; Morado and Small, 1995). Some of these pathogenic species are well studied as they are pathogens that have significant negative impacts on human-associated activities, such as commercial fishery and aquaculture industries. The best studied species is *Ichthyophthirius multifiliis*, a common fish parasite that causes white-spot disease. The MAC and mitochondrial genomes of this species have been sequenced in order to aid the prevention of infection and transmission by identification of metabolic targets for therapies and vaccines (Coyne et al., 2011).

The relative ease at which some ciliates, such as species of the genera *Tetrahymena* and *Paramecium*, can be cultured and grown in the lab and their short generation times have facilitated their in-depth study and have therefore long been regarded as important and evidently useful eukaryotic model organisms. Cells of *Tetrahymena* can be stored long term in liquid nitrogen and there are numerous mutant and inbred strains available as well as strains that have been genetically engineered by a variety of methods (Cassidy-Hanley, 2012). In previous decades numerous important discoveries regarding general biological processes have been made from studying ciliates, such as the discovery of self-splicing RNA (Zaug and Cech, 1980), the discovery of ribozymes (Kruger et al., 1982) and the function of telomeres (Szostak and Blackburn, 1982). More recently MAC genomes from several model and economically relevant ciliate species have been sequenced (Aury et al., 2006; Eisen et al., 2006; Coyne et al., 2011; Swart et al., 2013; Chen et al., 2015) which help further facilitate many genomics and genetics-based studies. The mitochondrial proteome of the ciliate *Tetrahymena thermophila* has also been well studied by a combination of comparative analysis of its genome sequence and tandem mass spectrometry of isolated mitochondria (Smith et al., 2007). What is known about these mitochondria will facilitate our investigations into hydrogenosomes of anaerobic ciliate species and allow us to understand how they

have evolved to function in the absence of O_2 by identifying differences in their mitochondrial proteomes compared to aerobic relatives.

1.2.2 Mitochondrial genomes of ciliates and other eukaryotes

Large differences are observed in the architecture, gene content, size and overall general complexity of mitochondrial genomes in different eukaryotes (Lynch et al., 2006). All known mitochondrial genomes consisted of a single circular DNA chromosome until a linear mitochondrial genome consisting of a single chromosome, was discovered in the ciliate *Tetrahymena pyriformis* (Suyama and Miura, 1968). Further linear genomes were subsequently discovered in other ciliate species (Goddard and Cummings, 1975) and eukaryotes including green algae and yeast (Ryan et al., 1978; Wesolowski and Fukuhara, 1981). Linear mitochondrial genomes thus appear to have evolved repeatedly during eukaryotic evolution and genomes of even closely related organisms can differ in structure. For example, a number of yeasts exhibit intra-species variation in mitochondrial genome structure, suggesting the presence of a mechanism for conversion between a linear or circular state (Rycovska et al., 2004). As well as single chromosomal mitochondrial genomes, multi-chromosomal mitochondrial genomes have been discovered in a range of organisms, which can consist of multiple linear or circular genomes. The mitochondrial genome of the ichthyosporean *Amoebidium parasiticum* consists of hundreds of relatively short linear chromosomes each encoding a single gene (Burger et al., 2003a) and a number of unrelated animal mitochondrial genomes consist of multiple circular chromosomes. The mitochondrial genome of the mesozoan *Dicyema* features a number of DNA ‘minicircles’, each encoding a single gene (Watanabe et al., 1999), as does, similarly, that of the human body louse *Pediculus humanus* (Shao et al., 2009). The largest mitochondrial genomes are found in angiosperm plants, which can be as large as 2400 kb, yet they encode 50-60 genes, which is a fairly typical gene content compared to smaller mitochondrial genomes (Burger et al., 2003b; Kubo and Newton, 2008). The least reduced mitochondrial genomes with the greatest number of genes are believed to be those found in jakobid protists but even these contain only a small percentage of the genes that were likely present on the genome of the mitochondrial endosymbiont (Burger et al., 2013).

The genomes of mitochondria found in ciliate species studied to date consist of a primary linear chromosome which ranges in size from approximately 40 to 70kb. These genomes encode protein components of the ETC, as well as translational components, including ribosomal proteins, ribosomal RNAs and transfer RNAs (Swart et al., 2012). This linear chromosome is typically what is referred to when discussing the organelle genome in ciliates, although it is worth noting that the mitochondrial genome of the ciliate *Sterkiella histriomuscorum* (Swart et al., 2012) and some other eukaryotes (Meinhardt et al., 1990), such as the amoebezoan *Physarum polycephalum* (Kawano et al., 1991) and some plants (Handa, 2008), are accompanied by shorter secondary linear chromosomes that contain regions that have sequence similarity to the primary mitochondrial genome chromosome. Described as linear mitochondrial plasmids, in most cases they appear to have ORFs encoding unknown proteins and both RNA and DNA polymerases (Handa, 2008).

1.2.3 Anaerobic ciliates and anoxic habitats

Anaerobic ciliates are found inhabiting freshwater and marine sediments, as well as the digestive tracts of insects and mammals. These anoxic environments occur initially due to a redox imbalance resulting in a local reductive environment. This is typically caused as a result of aerobic heterotrophs, degrading organic material produced by phototrophic organisms, involving the conversion of O_2 to CO_2. Although these are usually balanced by the O_2 -producing metabolism of phototrophs, in environments where heterotrophs are abundant and are metabolising at high rates, they can cause a depletion of O_2. If the diffusion rate of O_2 in this environment is also limited, the principle example being aquatic environments as O_2 has a low diffusion rate through water, then this leads to anoxia and anaerobic microbial communities can begin to thrive. The major pathways of anaerobic communities are shown in Figure 1. Anoxia is further maintained due to the chemically reducing metabolic end-products of anaerobic microbial consortia, such as sulphides and H_2 (Fenchel and Finlay, 1995).

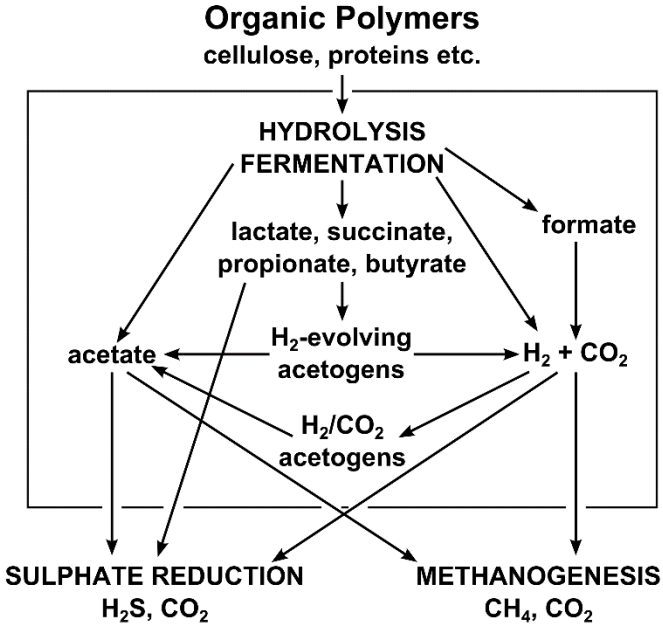


Figure 1.2 The structure of anaerobic communities with major pathways of anaerobic degradation of organic matter. Different metabolic processes within the pathways are fulfilled by diverse organisms and those within the box are fermentations. Adapted from Fenchel and Finlay (1995).

Anaerobic ciliates can be difficult to culture as they are sensitive to O_2 and therefore must be maintained in anoxic conditions. Anaerobic ciliates have only been successfully cultured xenically, with at least one species of food Bacteria present, which can cause a level of contamination in sequencing studies. Modern technologies however can help circumvent some of these issues since genomes and transcriptomes can now be isolated and sequenced from nucleic acids extracted from single purified cells (Nawy, 2014).

The only ciliate hydrogenosomes that have been studied in any detail are those of *Nyctotherus ovalis*. All known species of the genus *Nyctotherus* are commensal organisms of animal digestive tracts, including both vertebrate and invertebrate hosts, with *Nyctotherus ovalis* inhabiting the digestive tracts of cockroaches. *Nyctotherus* forms a clade with *Metopus* and *Clevelandella* which together comprise the class Armophorea, all known members of which are thought to be anaerobic. *Nyctotherus ovalis* provided the first direct evidence for the mitochondrial ancestry of hydrogenosomes, as it contains an organelle genome, homologous to mitochondrial genomes found in aerobic ciliates and encoding some of the same genes (Akhmanova et al., 1998; Boxma et al., 2005). Initially a mitochondrial-type rRNA gene was sequenced by Akhmanova et al. (1998) who also

reported a nuclear encoded FeFe-hydrogenase enzyme. Furthermore they demonstrated that antisera against a *Trichomonas vaginalis* FeFe-hydrogenase localises to the hydrogenosomes in immuno-gold labelling experiments. Two studies published partially sequenced hydrogenosome genomes from *Nyctotherus ovalis* (Boxma et al., 2005; de Graaf et al., 2011), as well as a limited transcriptome analysis investigating the coding potential of *Nyctotherus ovalis* MAC genome and an inference of its hydrogenosome metabolism (de Graaf et al., 2011). Interestingly, the two organelle genome sequences reported from *Nyctotherus ovalis* in these papers are not identical, indicating either that they were not from the same species, that organelle genomes in different *Nyctotherus ovalis* can be highly divergent, or the presence of sequencing and assembly errors.

Anaerobic species of ciliate with hydrogenosomes have been described in several classes of ciliates including the Oligohymenophorea, Plagiopylea, Armophorea and Litostomatea (Fenchel et al., 1977; Embley et al., 1995). More recently an anaerobic Karyorelictid ciliate has been described that appears to contain endosymbiotic methanogens and organelles that are purported to be hydrogenosomes, although this was not demonstrated directly (Edgcomb et al., 2011). This suggests that genes needed to evolve hydrogenosomes from aerobic mitochondria were present in the earliest ciliates, or at least the last common ancestor of the known hydrogenosomes containing ciliates. However the molecular basis for hydrogenosome metabolism is unknown for the majority of anaerobic ciliate species and so it is unclear if they are making H_2 in the same way. Given that anaerobic species with hydrogenosomes have evolved independently numerous times from ancestors with aerobic mitochondria in ciliates, this group of organisms provide an opportunity for studying how hydrogenosomes have evolved repeatedly in a relatively closely related group of organisms.

1.3 Methanogens

1.3.1 General description of methanogen Archaea

Until recently methanogenesis was only known to occur within one group of Archaea known as methanogens, found within the clade Euryarchaeaota. It has now been inferred from the genome sequences of species from another archaeal clade, the

Bathyarchaeaota, that these Archaea also have methanogenic metabolisms (Evans et al., 2015). Based on homology of key enzyme-encoding genes, it is thought that methanogenic metabolism may have evolved in an ancestor of these two clades and has been retained or lost in different lineages, meaning this trait is more widespread in Archaea than first thought (Lever, 2016). Methanogens use a restricted set of substrates including CO_2 and H_2, formate, methanol, methylamines and acetate, which they can reduce to methane in order to make ATP (Thauer et al., 2008). Although some small amounts of ATP are produced by substrate-level phosphorylation occurring in the methanogenesis pathway, the majority of methanogens are thought to make most of their energy by creating a transmembrane gradient from the translocation of protons and Na^+, that can be utilised by A_1A_0 ATP synthase complexes (distant homologues of the bacterial and mitochondrial F_1F_0ATP synthase complexes) to generate ATP (Deppenmeier et al., 1996). Methanogens are environmentally relevant, especially from an anthropogenic perspective since they contribute 69% of global CH_4 production, significantly higher in comparison to the 25% of CH_4 production associated with anthropogenic activities such as burning of fossil fuels (Conrad, 2009). CH_4 can escape into the atmosphere, where it is photochemically oxidised to CO_2, the major greenhouse gas responsible for global warming (Thauer, 2011). Another significant source of methane emissions comes from the enteric methanogen communities inhabiting ruminant livestock. In ruminants methanogens produce a range of cellulose-digesting enzymes, which benefit the ruminant by aiding in food digestion (Hill et al., 2016). As the human demand for livestock increases, so does the production of methane from these animals.

1.3.2 Detection of methanogens and their symbiosis with anaerobic ciliates

Methanogens can be putatively identified based on their auto-fluorescence when excited with 420nm wavelength light (Edwards and McBride, 1975; Mink and Dugan, 1977; Doddema and Vogels, 1978). The fluorescence is emitted by the oxidised form of the cofactor coenzyme F_{420} (Cheeseman et al., 1972). In methanogenic Archaea, oxidised coenzyme F_{420} is produced from the reduced form which acts as an electron donor to several enzymatic reactions in the methanogenesis pathway (DiMarco et al., 1990). Coenzyme F_{420} is also found in some species of Bacteria (Selengut and Haft, 2010) and therefore detection of auto-fluorescence at 420nm wavelength light does

not definitively identify methanogens but can be a useful preliminary screening method.

Auto-fluorescence detection was initially used to determine the endosymbionts some anaerobic ciliates as methanogens (Fenchel et al., 1977). The excess of available H_2 produced as a waste-product of hydrogenosome metabolism in anaerobic ciliates creates a niche which is occupied by hydrogenotrophic methanogens. The nature of these associations between ciliate host and methanogenic endosymbiont and whether they are transient or permanent is not fully understood. Likewise, whilst benefits of the symbiosis to the methanogen are likely to be the provision of a stable H_2 source and protection from extracellular stresses, the benefits to the ciliate is less clear. Previously it was suggested that the consumption of H_2 makes hydrogenosome energy production more thermodynamically favourable by lowering the pH_2 within the cell (van Bruggen et al., 1983). It is debatable however whether this is significant enough to benefit the host since anaerobic ciliates live in environments with sufficiently low pH_2 that H_2 production should not be inhibited (Fenchel and Finlay, 1991a). Despite this, the growth rate of *Plagiopyla frontata* and *Metopus contortus* was shown to decrease when their methanogens were inhibited, using 2-bromoethanesulfonic acid, although there appeared to be no significant effect on growth rate when they were eliminated from *Metopus palaeformis* (Fenchel and Finlay, 1991a).

The species of these methanogenic endosymbionts have been identified for a handful of anaerobic ciliates. All those identified to date come from the orders Methanobacteriales and Methanomicrobiales and each species is closely related to but distinct from free-living methanogen species (Embley and Finlay, 1994). Identification is commonly achieved by fluorescence in situ hybridisation (FISH) experiments, with fluorescently labelled DNA oligonucleotide probes designed specifically to target the 16S rRNA of ribosomes from specific endosymbiont species (Embley et al., 1992a; Embley et al., 1992b; Finlay et al., 1993b; Embley and Finlay, 1994; Shinzato et al., 2007).

It has been demonstrated experimentally that a strain of *Trimyema compressum* devoid of methanogen endosymbionts, acquired free-living *Methanobacterium formicicum* cells by phagocytosis when incubated together in the same medium and that these methanogens were transferred from food vacuoles to the ciliate cytosol, instating them as endosymbionts (Wagener et al., 1990). This

would suggest that these endosymbioses are transient rather than permanent associations since the strain of *Methanobacterium formicicum* can immediately adapt from a free-living lifestyle to surviving in the cytosol of an organism that can provide it with a steady supply of H_2. Furthermore, closely related ciliate species can have methanogenic endosymbionts from two different lineages, which would also indicate that these endosymbionts do not evolve linearly with their hosts, based on comparisons of phylogenies for hosts and endosymbionts (Embley and Finlay, 1994) and since they are so similar to free-living species, both morphologically and by comparison of their 16S rRNA gene sequences, they must have evolved symbiotic lifestyles relatively recently (Fenchel and Finlay, 1995). Conversely however, there are some of these endosymbionts that seem to have certain phenotypes that would indicate a degree of stability or adaptation involved in these relationships. For example the endosymbionts in *Metopus contortus* and *Trimyema* sp., both closely related to free-living species of the genus *Methanocorpusculum*, show evidence from EM that they undergo a polymorphic transformation by degrading their cell walls and attaching to hydrogenosomes, forming close interactions to presumably maximise H_2 transfer (Embley et al., 1992a; Finlay et al., 1993b). Also the endosymbionts of *Plagiopyla frontata* and *Cyclidium porcatum* form methanogen-hydrogenosome complexes, the organisation of which in the cytosol appears to be distinctive in each organism (Esteban et al., 1993; Embley and Finlay, 1994). These fairly complex interactions and symbiont-hydrogenosome arrangements are perhaps more indicative of a stable relationship between the two and that the endosymbionts have evolved the ability to form the interactions due to long-term co-evolution in the presence of hydrogenosomes, although of course free-living methanogens may also form these types of interactions with H_2 producing cells in extra-cellular environments. This study increases the sampling for species of anaerobic ciliates to identify what species of methanogenic endosymbionts they contain in order to infer their evolutionary basis for these relationships. Sequencing and analysis of genome sequences from some of these endosymbionts and comparison with the genomes of their closest free-living relatives should reveal much more regarding the nature of these symbiont-host interactions. The genome sequence should also provide insight into metabolic interdependencies between both of these partners.

1.4 General aims

The current study had the following main aims:

1. To investigate ciliate hydrogenosome evolution by isolating and culturing a taxonomically diverse sample of anaerobic ciliates with hydrogenosomes and using modern genomic sequencing methods to identify and characterise their hydrogenosome genomes. This is to test the hypothesis that ciliate hydrogenosomes are mitochondrial homologues and their hydrogenosome genomes encode components of mitochondrial ribosomes and the electron transport chain, like aerobic mitochondria.

2. To provide evidence of core mitochondrial processes within the hydrogenosomes of anaerobic ciliates. To test this, the metabolisms of ciliate hydrogenosomes will be inferred by generating transcriptomic data for individual ciliates using single cell methods and in combination with nuclear genomic data, in order to reconstruct their metabolic pathways and proteomes *in silico*.

3. To understand whether anaerobic ciliates and their endosymbiotic methanogens evolve in parallel, forming relatively stable associations, or whether they are acquired independently in different lineages. To test this, I will identify endosymbiont species living in anaerobic ciliates and to investigate the evolutionary history of both host and symbiont using a combination of molecular gene sequencing, phylogenetics and *in situ* fluorescent probing.

4. To identify the key hydrogenosomal proteins involved in hydrogen production and to investigate their origin(s) and evolutionary histories using phylogenetic methods, in order to understand whether they were acquired ancestrally from the mitochondrial endosymbiont, and inherited vertically in ciliates, or whether they were acquired more recently by lateral gene transfer.

Chapter 2. Methods

2.1 Field site sampling and culturing methods

To isolate free-living anaerobic ciliates, samples containing both water and sediment were collected, using a 500 ml beaker attached to the end of a 2 m pole, from field sites in Dorset, United Kingdom: Freshwater samples were taken from a pond known as East Stoke Fen (GPS 50.679064, -2.191587) adjacent to and occasionally flooded by the River Frome. Marine samples were taken from Poole Park Lake (GPS 50.715541, -1.971177), a brackish lake partially connected to Poole Harbour, which was sufficiently saline that only marine species of anaerobic ciliates were found in samples from this site. These samples were transferred into screw-top bottles for transport to the laboratory. Approximately 50 ml of liquid and suspended sediment from samples were transferred to 125ml glass vials to which another 50ml of medium was added, as well as a wheat grain and some dried cereal leaves, which were added in order to stimulate growth of natural food Bacteria present in the samples which ciliates can feed on. Ciliates were grown in SES (soil extract with added salts) medium and N75S (new cereal leaf 75% seawater) medium (recipes available from Culture Collection of Algae and Protozoa: https://www.ccap.ac.uk/), for freshwater and marine samples respectively. The pH of media were always adjusted to pH7 by addition of HCl or NaOH before sterilisation by autoclave. Soil for the production of SES medium was collected from a deciduous woodland close to East Stoke Fen. Seawater for the production of N75S:NSW medium was taken from Poole Park Lake. All culture vials were sealed with rubber stoppers and the gaseous headspace of vials were flushed using compressed N_2 for 3 minutes to remove O_2, creating anoxic conditions. Anaerobic ciliate species were identified in enriched samples by Genoveva Esteban (Bournemouth University) and were cultured monoxenically by transferring them to pre-incubated vials containing ciliate-free medium using glass micropipettes and were cultured as described above. All cultures were continually incubated at 18°C in the absence of light and subcultured every 4-6 weeks by dividing the culture from one vial into two, such that there were approximately 50 ml in each, and added another 50 ml of sterile medium to each, as well as a wheat grain and cereal leaves.

The isolation of *Nyctotherus ovalis* and continued growth of cockroach colonies were performed by Anders Lind (Ettema-Lab, Uppsala University), as described below.

Cells of *Nyctotherus ovalis* were obtained from cockroaches of the species *Blaptica dubia*, acquired commercially from Cricket Express (cricketexpress.se). Cockroaches were reared in plastic boxes and fed dried dog food and fresh fruit *ad libitum*. Two male and two female cockroaches were dissected under electromigration buffer (2.7 mM KH_2PO_4, 1.8 mM KH_2PO_4, 21.5 mM KCl, 20 mM NaCl, 6.1 mM $MgSO_4.7H_2O$, 0.5 mM L-cysteine, 0.5mM $CaCl_2.2H_20$, 0.5mM titanium citrate, and 1 mM $NaHCO_3$, pH adjusted to 7.5 using 1M NaOH) and their hindguts extracted. The ciliates were then extracted by electromigration, in electromigration buffer as described by Hoek et al. (1999).

2.2 PCR, cloning and sequencing

2.2.1 DNA template preparation for PCR

PCR was performed using either isolated cells or purified DNA as template. Typically up to 10 cells were isolated using a glass micropipette whilst being observed via a stereoscopic microscope and washed in sterile PBS or sterile medium. Cells were then transferred to microcentrifuge tubes and heated in a heat block at 90°C until all liquid had evaporated. These dried samples were then used directly as templates in PCR reactions. Purified DNA was obtained by centrifuging 200ml of ciliate cultures at 1500 x g for 45 minutes. Supernatant was carefully removed to leave pellets intact, which were transferred to microcentrifuge tubes. Cells were lysed by a combination of vortex mixing and free-thaw cycles between -80°C or -20°C and 90°C. DNA was purified from these samples using a QIAamp DNA Mini Kit (QIAGEN) following the protocol for crude cell lysates in the kit handbook and eluted in sterile 10mM Tris buffer adjusted to pH8. Purified DNA not being used immediately was kept at -20°C.

2.2.2 PCR reagents

PCRs were performed using KOD Hot Start DNA polymerase (Merck Millipore) and reagents were combined using concentrations and volume ratios recommended by the manufacturer's protocol. Total reaction volumes ranged from 10 – 50µl. For reactions using dried cells as DNA template, the volume of buffer containing DNA in the reaction mixture was replaced with sterile H_2O. All primers used in PCR experiments are listed in Table 2.1 and the corresponding thermal cycler conditions used with each primer are listed in Table 2.2. The annealing temperatures in Table 2.2 were suboptimal for amplification in the case of some primer combinations. In these cases PCR replicates were performed across a gradient of annealing temperatures, typically ranging between 50°C and 70°C, in order to identify more optimal annealing temperatures that would result in increased product and clearer gel bands.

Table 2.1 Primers used in PCR experiments

Primer specificity	Forward/Reverse Primer	Primer name	Primer sequence (5'-3')	Reference	Thermal cycler conditions (Table 2.1)
Eukaryote 18S rDNA	Forward	EMBF	AYCTGGTTGATYYTGCCAG	Embley et al. 1992b	A
	Reverse	EMBR	TGATCCATCTGCAGGTTCACCT	Embley et al. 1992b	A
	Forward	EK-555F	TCYGKTTGATCCYGSCRGAG	López-García et al. 2001	A
	Reverse	EK-1269R	TGGGTCTCGCTCGTTG	López-García et al. 2001	A
Methanogen endosymbiont 16S rDNA	Forward	1A	TCYGKTTGATCCYGSCRGAG	Embley et al. 1992b	A
	Reverse	1100A	TGGGTCTCGCTCGTTG	Embley et al. 1992b	A
	Forward	340F	CCCTAYGGGGYGCASCAG	Gantner et al. 2011	A
	Reverse	1000R	GGCCATGCACYWCYTCTC	Gantner et al. 2011	A
Prokaryote 16S rDNA	Forward	A519F	CAGCMGCCGCGGTAA	Wang and Qian 2009	A
	Reverse	U1391R	ACGGGCGGTGWGTRC	Lanzen et al. 2011	A
Ciliate mitochondria/hydrogenosome 12S rDNA	Forward	VH59	TGTGCCAGCAGCCGCGGTAA	Van Hoek et al. 2000	B
	Reverse	VH60	CCCMTACCRGTACCTTGTGT	Van Hoek et al. 2000	B
Armophorea hydrogenosome 12S rDNA	Forward	TE59	ATTGGCGGGAGCYTGTATAAA	Anders Lind (Uppsala University)	C
	Reverse	TE60	CGTGATGGGCGGTGTGTG	Anders Lind (Uppsala University)	C
Metopus contortus hydrogenosome 12S rDNA	Reverse	RevA	GTGCAAAGTTAGAAAGAAACA	This study	C
Cyclidium porcatum hydrogenosome 12S rDNA	Forward	cpTE59	TTTGGCGGGAAAAATG	This study	C
	Reverse	cpTE60	GGGCAGTGTGTAGAAT	This study	C
	Forward	cpTE59.1	CTCTAATTAACTAACG	This study	C
	Reverse	cpTE60.1	AATTCCACGTAATTACC	This study	C

Table 2.2 Thermal cycler conditions used in PCR experiments

A. Embley et al. 1992b			B. Dunthorne et al. 2011			C. Anders Lind (Uppsala University)		
1.	95°C	2:00	1.	95°C	3:00	1.	95°C	3:00
2.	94°C	0:40	2.	95°C	0:15	2.	95°C	0:30
3.	55°C	0:40	3.	67°C	0:15	3.	52.7°C	0:15
4.	72°C	2:00	4.	72°C	1:15	4.	72°C	1:15
5.	Go to Step 2	10x	5.	Go to Step 2	40x	5.	Go to Step 2	32x
6.	92°C	0:30	6.	72°C	10:00	6.	72°C	10:00
7.	55°C	0:40	7.	4°C	Infinite hold	7.	4°C	Infinite hold
8.	72°C	2:30						
9.	Go to Step 6	20x						
10.	72°C	6:00						
11.	4°C	Infinite hold						

2.2.3 Agarose gel electrophoresis

PCR products were analysed and visualised by agarose gel electrophoresis. Gels were prepared by adding agarose, with a final concentration of 1 – 2%, according to the expected size of the DNA bands in the PCR product, to 1x TAE buffer (40mM Tris acetate, 1 mM EDTA, pH8). The agarose was dissolved by heating the suspension and then cooling before the addition of ethidium bromide to a final concentration of 0.5 µg/ml. Agarose solutions were poured into gel moulds, containing combs to make wells when removed, and allowed to cool at room temperature until the gels solidified. Solidified gels were removed from moulds, transferred into electrophoresis chambers and submerged in 1x TAE buffer. 5-20µl aliquots of PCR product, depending on expected band intensities, were mixed with 6x DNA Gel Loading Dye (Thermo Scientific) in a ratio of 5:1, and loaded into wells of the gels. MassRuler DNA Ladder Mix (Thermo Scientific) was also loaded in wells adjacent to PCR products, for DNA band size comparison and molecular weight/DNA length estimation of PCR products. Electrophoresis was carried out at 60 – 90V until loading dye was observed to have migrated into the bottom quarter of the gel. Migrated DNA in PCR products and markers in gels were visualised using a UV transilluminator.

2.2.4 Nested and semi-nested PCR

In experiments that provided limited quantities of PCR product that were insufficient for downstream applications, nested and semi-nested PCRs were performed using purified DNA product from initial PCRs as template for second PCRs. DNA from the initial reaction was purified using a QIAquick PCR Purification Kit (QIAGEN) following the manufacturer's recommended protocol. The second PCRs used a forward primer designed to bind to the target DNA template downstream from the forward primer used in the first reaction and/or a reverse primer designed to bind upstream from the reverse primer used in the first reaction. PCRs were performed using the same methods described in sections 2.2.2 – 2.2.5.

2.2.5 PCR product purification, plasmid ligation, cloning and sequencing

Following agarose gel electrophoresis of PCR products, DNA bands of expected size for cloning were excised from gel using a sterile scalpel and purified using the QIAquick Gel Extraction Kit (QIAGEN), following the manufacturer's recommended protocol to dissolve the agarose, bind DNA to the purification column, wash the DNA and then elute in 10mM Tris-Cl, pH 8.5. Purified DNA fragments were ligated into linearised pJET1.2 blunt ended plasmids using the CloneJET PCR Cloning Kit (Thermo Scientific) following the manufacturer's recommended protocol to perform the ligation reaction. 50µl aliquots of Subcloning Efficiency DH5α Competent Cells (Invitrogen) were thawed on ice. 5µl ligation reactions were added to cell suspensions and incubated on ice for 10 minutes. Cell suspensions were then heat-shocked by transferring to a 42°C water bath for 40 seconds, before returning them to ice and incubating for a further 10 minutes. 1ml LB medium was then added to each suspension and incubated at 37°C with continual shaking at 200rpm for a 1 hour recovery period. 100µl of these suspensions were then spread on selective LB agar plates containing carbenicillin (100µg/ml) and incubated at 37°C overnight. Colonies grown on selective plates were cultured overnight in 5ml LB medium containing carbenicillin (100 µg/ml). Plasmids were purified from overnight cultures using the QIAprep Miniprep Kit (QIAGEN) following the manufacturer's recommended protocol. To assess

whether plasmids in purified samples contained inserts of the expected size, aliquots of these samples were digested with the restriction enzyme BglII (Thermo Scientific) following the manufacturer's recommended protocol. Digested samples were visualised by agarose gel as described in section 2.2.5 and samples observed to contain plasmids with expected inserts were selected for sequencing. Sanger sequencing was performed by GATC Biotech, whom were provided with aliquots of plasmid samples with concentrations adjusted according to GATC Biotech requirement guidelines. All linear DNA and plasmid concentrations were estimated using a NanoDrop 2000 Spectrophotometer (Thermo Scientific) and samples were sequenced in both directions using forward and reverse sequencing primers provided in the CloneJET PCR Cloning Kit. Nucleotide sequences were inferred from chromatogram files, which were assessed manually using the program FinchTV 1.4.0 (Geospiza). Reverse and forward read nucleotide sequences were then aligned and assembled using the program Sequencher 4.2.2 (Gene Codes Corporation), which was also used to remove primer and plasmid sequences.

2.3 Fluorescence *in situ* hybridisation

2.3.1 Probe design

Where available probe sequences used in other publications or from probeBase (Loy et al., 2003) were used to target specific organisms. In cases where no probe sequences were available that were specific to a particular taxonomic group, 16S rDNA sequences of target organisms were aligned with the most similar 16S rDNA sequences of non-target organisms obtained by blastn searches. Probes were manually designed to gene regions conserved between all target organisms and differed in the non-target organisms by at least 3 nucleotides.

2.3.2 Sample fixation, hybridisation and imaging

Ciliate endosymbiont species were identified by fluorescence *in situ* hybridisation (FISH) using the oligonucleotide probes in Table 2.3. All probes were synthesised by biomers.net GmbH and double-labelled at the 5' and 3'- ends, in order to

increase the fluorescent signal, with either 6-Fam, Cy3 or Cy5 fluorescent dyes. Due to auto-fluorescence emitted from the sample at similar emission spectra to 6-Fam and Cy3, only Cy5 could be used to visualise endosymbionts of *Nyctotherus ovalis*. Ciliate cells were fixed in 4% paraformaldehyde at 4°C and washed in PBS. All cells were attached to poly-L-lysine coated slides, except cells of *Nyctotherus ovalis*, which were attached to gelatine coated slides. Sample dehydration, probe hybridisation and washing were the same as in Daims et al. (2005), except when hybridising species of *Metopus* and *Nyctotherus*. In these experiments formamide was removed from the hybridisation buffer as its presence caused non-specific binding of probes to samples. In these cases the stringency of the hybridisation reactions was ensured by increasing the hybridisation temperature according to the estimated probe dissociation temperatures (T_d), which were estimated as described by Stahl and Amann (1991). Probes were hybridised for 2 hours at 2°C lower than their respective T_d. After washing, dried samples were mounted with ProLong Diamond antifade mountant. Z-sections were imaged using a confocal microscope (A1R, Nikon) with a 63x/1.4 objective lens and a Leica SP8 confocal gSTED microscope was used for super-resolution imaging. Vertical z-stacks were deconvolved using Huygens deconvolution software (Scientific Volume Imaging B.V.) with empirically measured point spread functions and image projections were reconstructed using the program Fiji (ImageJ) (Schindelin et al., 2012).

Table 2.3 Fluorescent probes used in FISH experiments

Probe name	Probe Sequence (5'-3')	Target Organisms	Reference
ARCH915	GTGCTCCCCCGCCAATTCCT	Archaea	Stahl and Amann, 1991
EUB338	GCTGCCTCCCGTAGGAGT	Most bacteria	Amann et al., 1990
NONEUB	ACTCCTACGGGAGGCAGC	None (Negative control)	Wallner et al., 1993
SYM5	CTGCATCGACAGGCACT	*Metopus contortus* and *Trimyema* sp. endosymbionts, and other species of the genus *Methanocorpusculum*	Embley et al., 1992a
MB	CCGTTAAGGATGGCACT	*Nyctotherus ovalis* endosymbionts and other species of the genus *Methanobrevibacter*	This study

2.4 Organelle genome sequencing

2.4.1 Organelle enrichment and DNA sequencing

The workflow for sample preparation and organelle genome sequencing methods is described in Figure 2.1. In order to optimise the enrichment process PCR was used to test each fraction (pellets and supernatants) for the presence of

macronuclei, using primers targeting 18S rRNA; Bacteria, using primers targeting 16S small-subunit rRNA; and hydrogenosomes, using primers targeting mitochondrial 12S small-subunit rRNA.

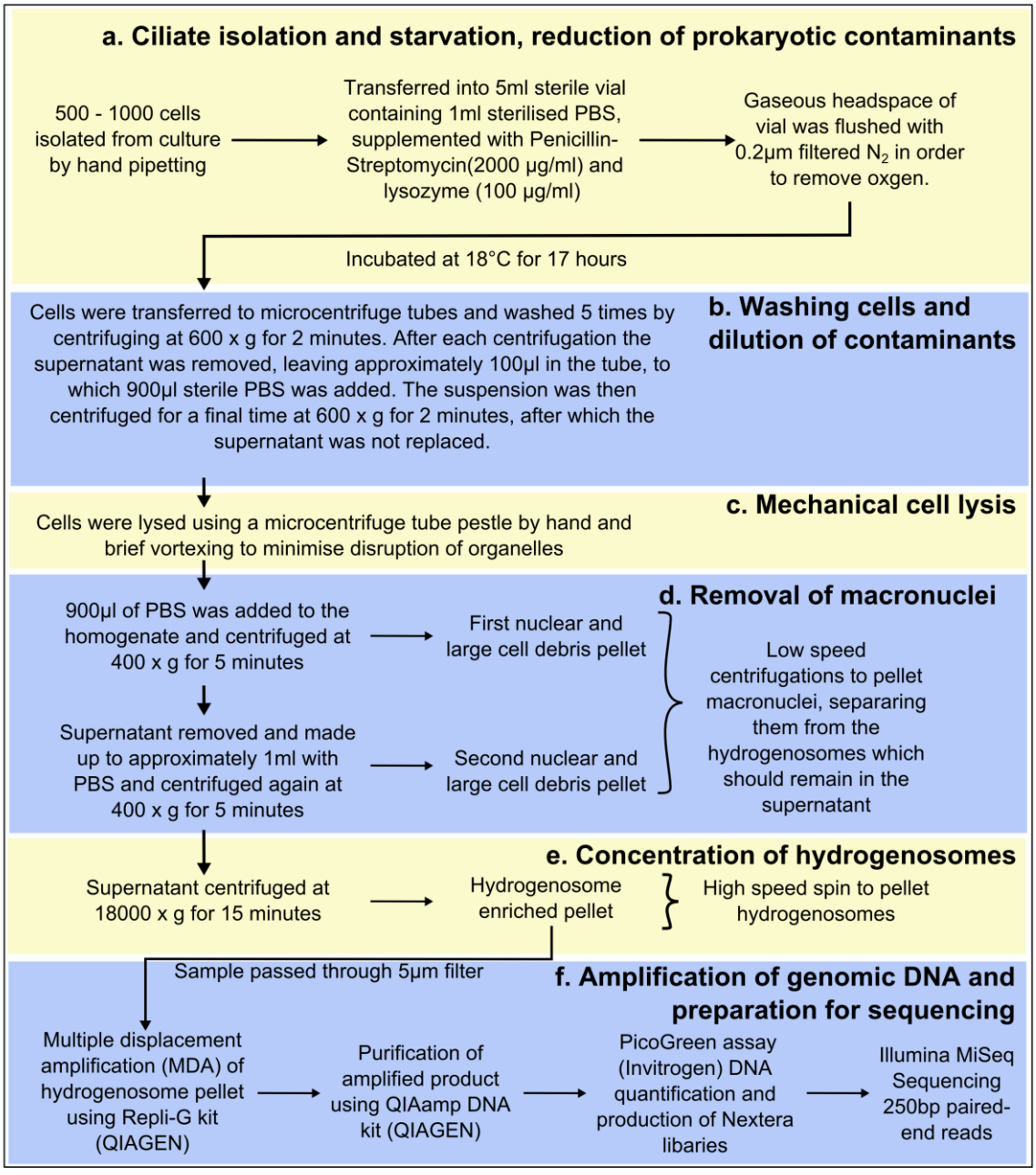


Figure 2.1 Workflow of methods used for preparation of hydrogenosome DNA for sequencing a. & b. Reduction of prokaryotes in sample; c. Lysis of ciliates in order to release organelles; d. Reduction of macronuclei; e. Concentrated hydrogenosomes in preparation for whole-genome amplification; f. Amplification of enriched hydrogenosome DNA and prepare libraries for sequencing.

2.4.2 Organelle genome assembly and annotation

All assembly of genomic datasets, from raw reads to assembled contigs, was performed by Anders Lind (Ettema-lab, Uppsala University), as he had more experience doing this for other datasets, using the following methods: Read quality was assessed using FastQC (Andrews, 2010) and SeqPrep (https://github.com/jstjohn/SeqPrep) was used to remove short reads, remove Illumina adapters from reads and merge overlapping paired-end reads. Low quality bases were removed using Trimmomatic (Bolger et al., 2014) with the following parameters: TRAILING:20, MINLEN:150. Paired-end reads were assembled into contigs using the SPAdes Genome Assembler (Bankevich et al., 2012) with the following parameters to reduce mismatches and indels: --sc --careful.

The following analysis was carried out by myself, William Lewis: A dataset of the available ciliate mitochondria and hydrogenosome genomes was used as a query in blastn searches against assembled datasets, in order to identify a set of potential ciliate hydrogenosome genome contigs. In order to increase search sensitivity, the following parameters values were used in these blastn searches instead of their default settings: -word_size 11, -reward 2, -penalty 3, -gapopen 5, -gapextend 2. Potential hydrogenosome genome contigs were either identified or dismissed as being true ciliate hydrogenosome genome sequences based on further blastn searches against the NCBI nucleotide collection and blastx searches against the Nr database (NCBI) and their similarity to other ciliate mitochondria and hydrogenosome genome sequences. Hydrogenosome genome contigs were annotated using several methods. Firstly, FACIL (Dutilh et al., 2011) predicted the genetic code used by each hydrogenosome genome, using sequenced contigs as input, as being most similar to the protozoan mitochondrial code (NCBI genetic code 4). Genes were predicted from contigs and translate *in silico* using Prodigal (Hyatt et al., 2010) with either -meta or -single settings and translation table 4 specified. rRNA genes were predicted by searching the Rfam database (Gardner et al., 2009) using the program cmscan, included in the Infernal software package (Nawrocki and Eddy, 2013). tRNA genes were predicted using tRNAscan-SE (Lowe and Chan, 2016).

2.5 RNA-Seq

2.5.1 Transcriptome sequencing and assembly

All cDNA libraries were produced and transcriptomic datasets assembled by Henning Onsbring Gustafson (Ettema-lab, Uppsala University), as he had more experience doing this for other datasets, using the following methods: Transcriptome datasets were produced from five micro-manipulated ciliate cells per sample as it was easier to wash this number of cells effectively and minimize contamination. The methods followed the Smart-Seq2 protocol (Picelli et al., 2014), which includes all necessary steps for successful mRNA isolation, reverse transcription and library preparation. Nextera libraries were sequenced with an Illumina MiSeq using 250bp paired-end reads. Quality control and trimming of sequencing reads was performed as described in section 2.4.2 and assembled with Trinity (Grabherr et al., 2011). Reads were mapped to transcripts to quantify coverage for the purposes of assessing dataset quality using Rsem (Li and Dewey, 2011) with Bowtie2 (Langmead and Salzberg, 2012). Taxonomic diversity of transcript dataset was estimated using MEGAN5 (Nawrocki and Eddy, 2013) to assess contamination by transcripts from non-target organisms, using alignments from blastx searches against the Nr database as described in Section 2.5.2.

2.5.2 Gene identification from transcript datasets

Putative transcripts encoding genes involved in ciliate hydrogenosome metabolisms were identified using a combination of methods. Initial dataset refinement was performed using CBOrg (Gaston et al., 2009), in order to identify transcripts that likely function in ciliate hydrogenosomes based on sequence similarity to compiled hydrogenosome and mitochondria datasets from other organisms. Automatic gene identification was performed using the EggNOG-mapper (Huerta-Cepas et al., 2016) with HMMER (Eddy, 1998) mapping. Additional gene identification was performed by searching assembled transcript datasets against the Nr database (NCBI) using blastx. Transcripts were translated

to protein sequences *in silico* using TransDecoder (https://transdecoder.github.io/). Ciliates have evolved a variety of genetic codes in different lineages (Lozupone et al., 2001) and in the present study, based on which code did not introduce stop codons within protein sequences, it was determined that *Nyctotherus ovalis, Metopus contortus, Metopus es* and *Metopus striatus* all use the standard genetic code (NCBI genetic code 1), whereas *Trimyema* sp.*, Plagiopyla frontata* and *Cyclidium porcatum* use the ciliate nuclear genetic code (NCBI genetic code 6). These codes were selected in TransDecoder to produce protein translations.

2.5.3 Prediction of N-terminal targeting signals

The programs TargetP (Emanuelsson et al., 2007), MitoProt II (Claros and Vincens, 1996), Predotar (Small et al., 2004) and MitoFates (Fukasawa et al., 2015), each utilise a combination of different methods to predict the presence of mitochondrial N-terminal targeting signals from the amino acid sequences of proteins. These programs were used to predict whether ciliate proteins, translated from transcripts, had mitochondria-like N-terminal targeting sequences.

2.5.4 Codon usage analysis for transcript verification

Cultures used for isolating cells for transcriptome sequencing each contained a single ciliate species and a diverse consortia of different prokaryotic organisms. Despite isolating ciliate cells by pipetting and washing in sterile buffer, prokaryotes still remained in the sample as undigested prey in the ciliate food vacuoles, as endosymbionts in the ciliate cytoplasm, as well as extracellular contaminating organisms carried over during washing steps. Eukaryotes, including ciliates, have poly-adenylated mRNA transcripts that were enriched in the Smart-seq2 (Picelli et al., 2014) sequencing process, using a primer that binds to the poly-adenylated sequence of the mRNA. This enriched the number of ciliate transcripts in the sequenced dataset but the samples still contained mRNA from prokaryotic organisms, which were also sequenced and assembled in the final dataset. Due to low sequence conservation of some genes and the possibility of lateral acquisitions of genes from other organisms, the taxonomic identification of genes

can be ambiguous from blast-based sequence comparisons alone. Therefore when identifying genes in the transcriptomic datasets, it was necessary to determine a level of confidence that transcripts from other organisms are not being mistaken for ciliate genes. A number of methods were used to provide evidence that particular transcripts are true ciliate genes. For some transcripts, the poly-adenylated tail was partially sequenced and therefore the sequence was terminated by a series of repetitive ‘A’s. This signature provides some evidence that such transcripts originate from the ciliate nuclear genome, although this is not completely definitive as some gene sequences may happen to contain repetitive ‘A’-sequences.

The codon usage of transcripts was analysed as a quantitative method to provide evidence of transcripts being encoded by ciliate genomes. Since the nucleotide composition of a gene is affected by mutational biases that apply to the whole genome in which it is resident, genes generally conform to a codon usage that is typical of that genome. This means that laterally acquired genes over time acquire a codon usage pattern typical of the genome of which it is resident in a process known as amelioration (Lawrence and Ochman, 1997). A dataset was assembled of all transcripts that had best hits in blastx searches with known ciliate genes in the Nr database. This dataset contained the transcripts that were considered most likely to be encoded by the macronuclear genome of each ciliate. Possible ORFs for these transcripts were predicted using TransDecoder (https://transdecoder.github.io/), which also translates ORFs to protein sequences. All ORFs were retained that had positive hits in a HMMER search (Eddy, 1998) to protein domains in the Pfam database (Finn et al., 2016), indicating that they encode functional proteins. From the nucleotide sequences of these ORFs a codon usage table was generated for the datasets from each ciliate species, using the program cusp (Rice et al., 2000). Using the codon usage tables, CAI (Sharp and Li, 1987) scores were calculated for each of the ORFs in the dataset, using the program cai (Rice et al., 2000), as a measure of the degree of deviation in codon usage of each ORF compared to what is typical for the total dataset. These values were then plotted as a distribution. CAI scores were also calculated from the most common bacterial and methanogen species present in each dataset using ORFs from their published complete genome sequences. The CAI scores

for each of these species were plotted as distributions on the same axes as the ciliate scores for each dataset.

2.6 Phylogenetics

2.6.1 Multiple sequence alignment and phylogenetic inference

Nucleotide and amino-acid sequences were automatically aligned using MUSCLE (Edgar, 2004). Multiple sequence alignments were checked for misaligned sites manually, which were realigned using the alignment viewers SEAVIEW (Galtier et al., 1996) and Jalview (Waterhouse et al., 2009). For phylogenetic analysis, alignments were trimmed using trimAl (Capella-Gutiérrez et al., 2009). As an initial screen of the datasets, maximum likelihood trees were inferred using RAxML (Stamatakis, 2014) with 100 rapid bootstrap replicates, in order to help refine the datasets. More in-depth analyses were then carried out using IQ-TREE (Nguyen et al., 2015) with 1000 ultrafast bootstrap replicates (Minh et al., 2013). Bayesian trees were inferred using Phylobayes MPI (Lartillot et al., 2013) running three independent MCMC chains until two had converged. Convergence was assessed using the bpcomp and tracecomp programs, which are part of the Phylobayes MPI package (Lartillot et al., 2013).

2.6.2 Database sampling for phylogenies

To assemble a taxonomically representative dataset of sequences for phylogenetic analyses aimed at investigating the origins of ciliate proteins, these ciliate protein sequences were used as queries for searches against the Nr database (NCBI), using blastp. These searches were restricted to the groups listed in Table 2.4 to ensure sampling from diverse taxonomic groups. The top 5 best hits were saved from each blastp search and were aligned with any additional key proteins of interest that were manually downloaded. Redundancy was reduced in sequence datasets (aligned as described in section 2.6.1) by removing the most highly similar sequences using the program Decrease Redundancy (available at: http://web.expasy.org/decrease_redundancy/).

Table 2.4. Secondary level taxonomic groups according to the NCBI taxonomy browser. These groups were used to restrict blast searches in order to ensure diverse sampling for phylogenetic analyses.

Bacteria	Archaea	Eukaryota
Acidobacteria	Aenigmarchaeota	Alveolata
Aquificae	Diapherotrites	Amoebozoa
Caldiserica	Micrarchaeota	Apusozoa
Calditrichaeota	Nanohaloarchaeota	Breviatea
Chrysiogenetes	Pacearchaeota	Centroheliozoa
Deferribacteres	Parvarchaeota	Cryptophyta
Dictyoglomi	Woesearchaeota	Euglenozoa
Elusimicrobia	Nanoarchaeota	Fornicata
FCB group	Archaeoglobi	Glaucocystophyceae
Fusobacteria	Hadesarchaea	Haptophyceae
Nitrospinae	Halobacteria	Heterolobosea
Nitrospirae	Methanobacteria	Jakobida
Proteobacteria	Methanococci	Katablepharidophyta
PVC group	Methanomicrobia	Malawimonadidae
Rhodothermaeota	Methanopyri	Opisthokonta
Spirochaetes	Thermococci	Oxymonadida
Synergistetes	Thermoplasmata	Parabasalia
Terrabacteria group	Korarchaeota	Rhizaria
Thermodesulfobacteria	Verstraetearchaeota	Rhodophyta
Thermotogae	Crenarchaeota (eocytes)	Stramenopiles
unclassified Bacteria	Thaumarchaeota	Viridiplantae
	unclassified Archaea	unclassified eukaryotes

2.7 Light microscopy

Living ciliate cells were imaged using an Olympus BH-2 light microscope mounted with a Micropublisher 3.3 RTV camera (QImaging). Methanogenic endosymbionts were visualised by fixing ciliates cells in 4% paraformaldehyde and transferring them to a black polycarbonate membrane filter, mounted on a microscope slide using Type FF Immersion Oil (Cargille). F420 fluorescence emission from methanogens was imaged using the microscope and camera described above, under 420nm wavelength epifluorescent illumination. Confocal microscopy was performed as described in section 2.3.2.

2.8 Transmission electron microscopy

Samples were prepared for transmission electron microscopy (TEM) by centrifuging 200ml of ciliate cultures at 1500 x g for 45 minutes. Supernatant was carefully removed to leave pellets intact, which were transferred to microcentrifuge

tubes. Cells were fixed in 2.5% glutaraldehyde in 0.15M HEPES-buffer at 4°C. A commercial service provided by Benoît Zuber and Beat Haenni (Microscopy Imaging Centre, Institute of Anatomy, University of Bern, Switzerland) was used for sample preparation and imaging of fixed samples.

Chapter 3. The metabolism of ciliate hydrogenosomes

3.1 Introduction

3.1.1 Hydrogenosomes in anaerobic ciliates

Hydrogenosomes have evolved independently in diverse and often distantly related eukaryote lineages (Embley and Martin, 2006). On a smaller evolutionary scale, ciliates are a single clade of eukaryotes that appear to have evolved hydrogenosomes independently on multiple occasions (Embley et al., 1995). This provides a unique opportunity to study hydrogenosomes from a group of closely related organisms in order to better understand their similarities and differences, and from this to infer the processes that have shaped their evolution. The repeated evolution of hydrogenosomes in ciliates (Embley et al., 1995), suggests that they are in some way predisposed to evolving hydrogenosomes from aerobic mitochondria relative to other groups. Exactly how they have achieved this however, and what evolutionary processes were involved, are unknown since they have not been studied in sufficient detail.

The first hydrogenosome found to contain an organelle genome, homologous to the genomes of mitochondria, was discovered in the anaerobic ciliate *Nyctotherus ovalis* (de Graaf et al., 2011). Organelle genomes have since been found in the mitochondrial homologues of other anaerobic eukaryotes, including the stramenopile *Blastocystis* sp. (Stechmann et al., 2008) and the rhizarian *Brevimastigamonas vehiculus* (Gawryluk et al., 2016). The anaerobic ciliate *Nyctotherus ovalis* belongs to the class Armophorea and lives commensally in the hindgut of cockroaches (Gijzen et al., 1991). The hydrogenosome metabolism of *Nyctotherus ovalis* was previously partially reconstructed by analysis of its hydrogenosome genome (Boxma et al., 2005; de Graaf et al., 2011) and by transcriptome sequencing of its macronuclear genome, from which genes were identified that were thought to encode hydrogenosomal proteins based on the presence of N-terminal targeting signals that were similar to mitochondrial targeting signals (de Graaf et al., 2011). These studies suggest that the hydrogenosomes of *Nyctotherus ovalis* have a partial ETC and TCA cycle, similar to aerobic mitochondria, and can produce energy exclusively by substrate-level

phosphorylation as they do not have an F_1F_0 ATP synthase (de Graaf et al., 2011). It is not known however, if the hydrogenosomes of other anaerobic ciliates also have organelle genomes and whether their metabolisms are similar to *Nyctotherus ovalis*.

In the present study, genomic and transcriptomic sequencing methods were used to investigate the hydrogenosomes of seven ciliate species, six of which have not been investigated with such methods previously. By inferring the phylogeny of ciliates it was possible to strategically sample species from diverse lineages, these were *Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum*, *Trimyema* sp. and *Plagiopyla frontata*. *Metopus contortus, Metopus es* and *Metopus striatus* were sampled as they are closely related to *Nyctotherus ovalis* and sampling over short genetic distances allows the fine scale evolution of hydrogenosomes to be investigated. *Cyclidium porcatum* was sampled as it has a complex symbiosis based upon its hydrogenosomes (Esteban et al., 1993) and is distinct from the other ciliate groups that were sampled. *Plagiopyla frontata* and *Trimyema* sp. were sampled based on them being closely related members of another anaerobic group of ciliates and their hydrogenosomes and endosymbionts show unusual morphologies. *Nyctotherus ovalis* was also reinvestigated with improved methods in order to expand what is already known about the hydrogenosomes of this species from previous studies (Akhmanova et al., 1998; Boxma et al., 2005; de Graaf et al., 2011). Little is currently known about many of these anaerobic ciliates and the present study has provided the first molecular and morphological data obtained from most of them. Some key questions include how do the hydrogenosomes of these anaerobic ciliates produce energy, what enzymes do they use to make H_2 and do they all contain organelle genomes like *Nyctotherus ovalis* (Akhmanova et al., 1998)? Studying these species makes it possible to compare the hydrogenosomes of anaerobic ciliate species within a single lineage and also between separate ciliate lineages that have evolved hydrogenosomes independently, whereas previously there has not been enough data available to facilitate this. Comparing the hydrogenosomes of these ciliates to the mitochondria of their aerobic relatives will also be crucial to understanding what changes occur when evolving a hydrogenosome from an aerobic mitochondrion.

3.1.2 Evolution of the electron transport chain (ETC) in hydrogenosomes

In aerobic eukaryotes, ATP production via oxidative phosphorylation occurs in mitochondria (Kennedy and Lehninger, 1949). The transfer of electrons from NADH and $FADH_2$ to O_2, via the ETC, pumps protons across the inner mitochondrial membrane generating a membrane potential that drives ATP production by F_1F_0 ATP synthase (Hatefi, 1985). Most studied hydrogenosomes do not have F_1F_0 ATP synthase and so do not use oxidative phosphorylation to produce ATP (reviewed in Muller et al. (2012) and Stairs et al. (2015)). It is unclear therefore why hydrogenosomes belonging to species such as *Nyctotherus ovalis* (de Graaf et al., 2011) and *Blastocystis* (Stechmann et al., 2008) appear to have retained ETC Complex I, a proton-pumping component of the ETC. One possible explanation suggested previously is that ETC Complex I is retained as it maintains a membrane potential across the inner membrane which might be required for protein import (Stairs et al., 2015), since a membrane potential has been shown to be required for protein import into yeast mitochondria (Gasser et al., 1982). If this is the case however, it does not explain why, based on localisation data and the presence of mitochondria-like protein targeting signals, some hydrogenosomes, including those of *Spironucleus salmonicida* (Jerlström-Hultqvist et al., 2013), can import proteins yet seem to have lost ETC Complex I and other proton-pumping ETC Complexes III and IV. Since some mitochondrial carrier proteins co-transport H^+, such proteins might contribute to the formation of a transmembrane H^+ gradient. In aerobic mitochondria, the ETC typically consists of the multi-protein ETC Complexes I, II, III and IV, as well as ubiquinone and cytochrome *c*.

The ETC Complex I has been well studied from some Bacteria and from the mitochondria of some aerobic eukaryotes (reviewed in Brandt (2006)). Based on the homology of different subunits to various hydrogenases (Tran-Betcke et al., 1990; Pilkington et al., 1991; Albracht, 1993), ETC Complex I has been functionally divided into three modules that have distinct roles (Brandt, 2006). The N-module and the Q-module together form the hydrophilic peripheral arm, which extends into the mitochondrial matrix (Brandt, 2006). Electrons are transferred to ETC Complex I via the 51 kDa subunit of the N-module as it oxidises NADH (Yagi

and Dinh, 1990). Fe-S cluster protein subunits transfer electrons (Ohnishi, 1998; Hinchliffe and Sazanov, 2005) through the N-module to the Q-module, where the electrons are used to reduce ubiquinone (Magnitsky et al., 2002). The third module of ETC Complex I is the hydrophobic P-module that spans the inner-mitochondrial membrane (Brandt, 2006). This module uses redox energy transferred from the Q-module when it reduces ubiquinone, to pump protons across the inner mitochondrial membrane via protein subunits that are distant homologues of bacterial antiporters (Mathiesen and Hägerhäll, 2002). A total of fourteen core protein subunits make up these three ETC Complex I modules and these subunits are found in both eukaryotes and prokaryotes (Brandt, 2006). In eukaryotes the fourteen core subunits of ETC Complex I can be encoded either by the nuclear or the mitochondrial genome and exactly which genes are encoded by each genome varies between species (Brandt, 2006). Numerous additional ETC Complex I accessory subunits have also been identified in different eukaryote lineages (Carroll et al., 2003; Marques et al., 2005). The functions of accessory subunits are mostly unknown but they are thought to have roles in assembly, stabilisation and regulation of the complex (Brandt, 2006; Hunte et al., 2010). In the ciliates studied so far, the majority of the ETC Complex I core subunits are typically encoded by the mitochondrial genome and only nad8, nad11 and the 24 kDa and 51 kDa subunits are encoded by the nuclear genome (Smith et al., 2007; Swart et al., 2012).

The catalytic site of ETC Complex II typically consists of two hydrophilic subunits that form a heterodimer in the mitochondrial matrix, anchored to the inner mitochondrial membrane by up to two other peptides (Sun et al., 2005). In aerobic mitochondria ETC Complex II functions as succinate dehydrogenase and has a dual role in the TCA cycle and the ETC, oxidising succinate to fumarate whilst transferring electrons to ubiquinone (Yankovskaya et al., 2003). In some anaerobic Bacteria and the hydrogenosomes of some anaerobic eukaryotes that have retained ETC Complex I, however, ETC Complex II can function as a fumarate reductase in the malate dismutation pathway (Massey and Singer, 1957; Tielens, 1994; Van Hellemond and Tielens, 1994; Tielens and Van Hellemond, 1998). This pathway transforms malate to succinate, utilising some enzymes of the TCA cycle but catalysing their reactions in reverse (Tielens, 1994; Tielens et al., 2002). In anaerobic eukaryotes, the fumarate reduction step of malate

dismutation requires the replacement of ubiquinone in the ETC with rhodoquinone (Van Hellemond and Tielens, 1994). Like ubiquinone, rhodoquinone accepts electrons from ETC Complex I but because it has a lower redox potential, it is capable of donating electrons to ETC Complex II acting as fumarate reductase, thereby using fumarate as a terminal electron acceptor and transforming it to succinate (Tielens and Van Hellemond, 1998; Lonjers et al., 2012). This process is thought to be required in the hydrogenosomes of anaerobic eukaryotes that use ETC Complex I as they are unable to use O_2 as an electron sink (Van Hellemond and Tielens, 1994).

3.1.3 ATP production by substrate-level phosphorylation in the hydrogenosomes of anaerobic eukaryotes

How the hydrogenosomes of most anaerobic ciliates make energy is unknown. Generally hydrogenosomes are inferred to make energy using a form of substrate-level phosphorylation, as was shown for *Trichomonas* (Steinbüchel and Müller, 1986). For this to occur pyruvate is first transformed to acetyl-CoA and the CoA moiety of this molecule is then transferred by ASCT to succinate, producing succinyl-CoA (Müller and Lindmark, 1978). Succinyl-CoA can then be re-oxidised to succinate by the TCA cycle enzyme SCS and in doing so, ATP is produced from ADP and inorganic phosphate (Jenkins et al., 1991). Several different enzymes were identified in hydrogenosomes from different species that are potentially capable of fulfilling the steps of this pathway. In *Nyctotherus ovalis* pyruvate is thought to be oxidised to acetyl-CoA by the pyruvate dehydrogenase complex (de Graaf et al., 2011), which is used for the same role in aerobic mitochondria. In the hydrogenosomes and mitochondrial homologues of other anaerobic eukaryotes however, including those of *Trichomonas, Blastocystis* sp. and *Mastigamoeba balamuthi*, it is thought that this role can be performed by PFO/PNO (Lindmark and Müller, 1973; Lindmark et al., 1975; Gill et al., 2007; Lantsman et al., 2008), or PFL in the hydrogenosomes of some other eukaryotes, including the chytrid fungi *Piromyces* and *Neocallimastix* (Akhmanova et al., 1999). Hydrogenosomes also possess FeFe-hydrogenases which enable them to use protons as electron acceptors (Bui and Johnson, 1996) and a variety of structural types have been found in anaerobic eukaryotes and species with hydrogenosomes (Horner et al., 2000). Three different sub-families of ASCT (sub-families 1A, 1B and 1C) have been identified in anaerobic eukaryotes, of which

Nyctotherus ovalis uses an ASCT of the sub-family 1A (de Graaf et al., 2011). Some ASCT enzymes of the subfamilies 1B and 1C appear to have been acquired by lateral gene transfer in some anaerobic eukaryotes (Stairs et al., 2014), it would therefore be of interest to know whether these enzymes exist in anaerobic ciliates and if so what their origins are. Knowing which enzymes and pathways are used in the hydrogenosomes of different anaerobic ciliate species would provide a better understanding of how hydrogenosomes have evolved on numerous occasions in this clade.

3.1.4 Organelle genomes from the mitochondria and hydrogenosomes of ciliates

Mitochondrial genomes have been previously studied from several aerobic ciliates (Pritchard et al., 1990; Burger et al., 2000; Brunk et al., 2003; Moradian et al., 2007; de Graaf et al., 2009; Barth and Berendonk, 2011; Coyne et al., 2011). Each of these genomes encode a mostly common set of genes encoding subunits of ETC Complexes I, II and III; a cytochrome *c* maturation protein, ccmF; the atp9 subunit of F_1F_0 ATP synthase; LSU and SSU ribosomal proteins; tRNAs; and rRNAs. The hydrogenosome genome of the ciliate *Nyctotherus ovalis* has been sequenced twice previously and appears to lack genes encoding subunits of ETC Complexes II and III, F_1F_0 ATP synthase and ccmF (Boxma et al., 2005; de Graaf et al., 2011). Sequencing the hydrogenosome genomes of anaerobic ciliates and identifying what genes they encode will complement analyses of their macronuclear genomes. Identifying proteins encoded by each of these genomes and that potentially function in hydrogenosomes will enable the reconstruction of their metabolisms. Existing data about ciliate mitochondrial genomes will facilitate detailed comparisons with the hydrogenosome genomes from anaerobic ciliates and improve understanding of their evolution.

3.1.5 The connection between cristae, the ETC and the presence of an organelle genome

Electron microscopy data has indicated that the hydrogenosomes of some anaerobic eukaryotes have invaginations in their inner membranes that are likely

to be homologues of the cristae observed in aerobic mitochondria (Zierdt et al., 1988; Finlay and Fenchel, 1989). Some of these species, including *Nyctotherus ovalis* (Gijzen et al., 1991; Akhmanova et al., 1998), *Blastocystis hominis* (Zierdt et al., 1988; Wawrzyniak et al., 2008) and *Brevimastigamoeba motovehiculus* (Gawryluk et al., 2016), have organelle genomes too. Interestingly the hydrogenosomes of *Neocallimastix frontalis* were reported to contain cristae-like structures but in this case the organelle genome appears to have been lost (van der Giezen et al., 1997b). In aerobic mitochondria, cristae morphology is directly linked to metabolic function, as cristae have been shown to be enriched in transmembrane protein complexes of the ETC and F_1F_0 ATP synthase in model organisms (Gilkerson et al., 2003; Wurm and Jakobs, 2006). F_1F_0 ATP synthase complexes are also thought to have a structural role in mitochondrial cristae biogenesis as they are located and organised in rows of dimers, along cristae edges in various species (Allen et al., 1989; Davies et al., 2011). Mitochondrial genomes encode several subunits of the ETC and F_1F_0 ATP synthase complexes and components involved in their translation (Anderson et al., 1981; Bibb et al., 1981). In organisms that no longer require the ETC, including some anaerobic eukaryotes with hydrogenosomes, the organelle genome would serve no clear purpose and hence its retention would potentially have no selective benefit. The absence of cristae in hydrogenosomes could thus provide visual indication that both the ETC and organelle genome have been lost. For this reason, the morphology of hydrogenosomes from anaerobic ciliates were investigated, to identify those that contain evidence of cristae, as the most likely species to have retained an organelle genome.

3.2 Aims

1. To culture strategically chosen free-living anaerobic ciliates using morphology and 18S rRNA gene data for identification, to infer their phylogeny and to use TEM to obtain data regarding their hydrogenosome ultrastructure, and to use these data to choose species for further in-depth analysis of their hydrogenosomes.
2. To use RNA sequencing and bioinformatic analysis to identify transcripts for key proteins that are likely to function in the hydrogenosomes of the studied anaerobic ciliate species and to use these data to reconstruct their metabolic pathways, and investigate how they make energy and H_2.
3. To establish and use small-scale genomic sequencing methods to sequence and assemble hydrogenosome genomes in anaerobic ciliates, and to bioinformatically analyse their sequence features.

3.3 Results

3.3.1 Isolation and morphological identification of anaerobic ciliates

Six species of free-living ciliates were successfully isolated and cultured in the present study using the methods described in Section 2.1. *Metopus contortus* and *Plagiopyla frontata* are marine species and were isolated from Poole Park Lake brackish samples. *Metopus es, Metopus striatus, Cyclidium porcatum* and *Trimyema* sp. are freshwater species and were isolated from East Stoke Fen freshwater samples. *Nyctotherus ovalis* is not free-living and the cells used in the present study were isolated from cockroaches by Anders Lind (Ettema-Lab, Uppsala University). The free-living ciliate species were initially identified visually by Professor Genoveva Esteban (Bournemouth University), based on their morphological features (Figure 3.1.). The images in Figure 3.1 were generated in the present study.

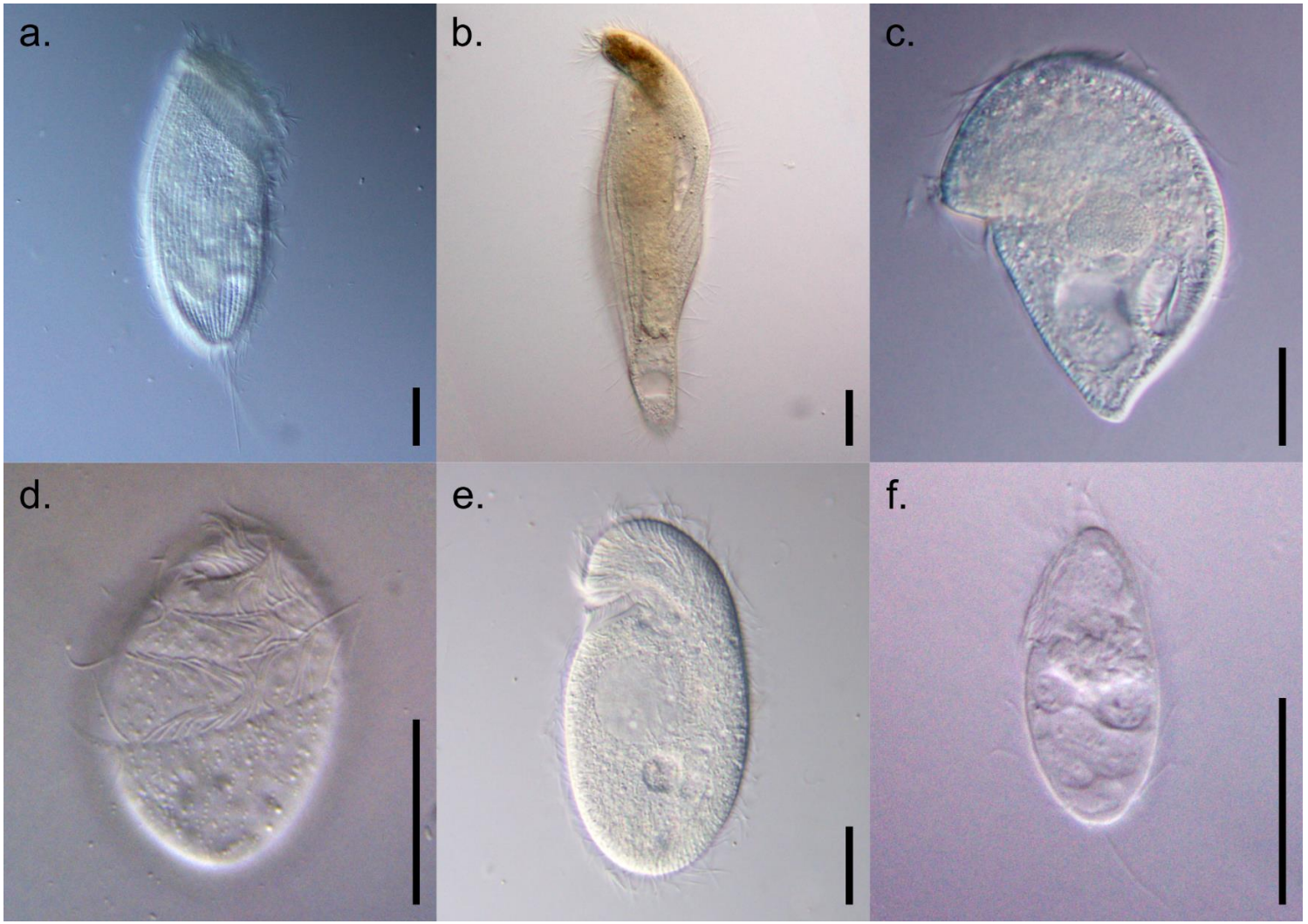


Figure 3.1. The morphology of cultured free-living anaerobic ciliates, visualised whilst living, using light microscopy. *Metopus contortus* (a), *Metopus es* (b), *Metopus striatus* (c), *Trimyema* sp. (d), *Plagiopyla frontata* (e), *Cyclidium porcatum* (f). Scale bars represent 20µm.

3.3.2 PCR and sequencing of 18S rRNA genes

In order to confirm the identity of the isolated free-living species, attempts were made to sequence their 18S rRNA genes using a variety of PCR and cloning strategies (Figures 3.2 and 3.3). In addition to the bands corresponding to amplified region of the intended target genes, many PCR experiments also produced additional bands, corresponding to various sizes, which were probably due to unspecific binding of primers to alternative templates. These sequences obtained from these experiments were compared to reference ciliate sequences in the NCBI nucleotide collection. Partial 18S rRNA sequences were obtained for each of the species *Metopus es, Metopus striatus, Metopus contortus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. The 18S rRNA genes sequenced from these ciliates were similar to those that had been sequenced from these species previously and were found in the NCBI nucleotide collection. This supported the species identities that were assigned to them based upon morphology (section 3.3.1). Later in the present study, full-length 18S rRNA gene sequences were also sequenced and retrieved for each of these species from transcriptomic datasets.

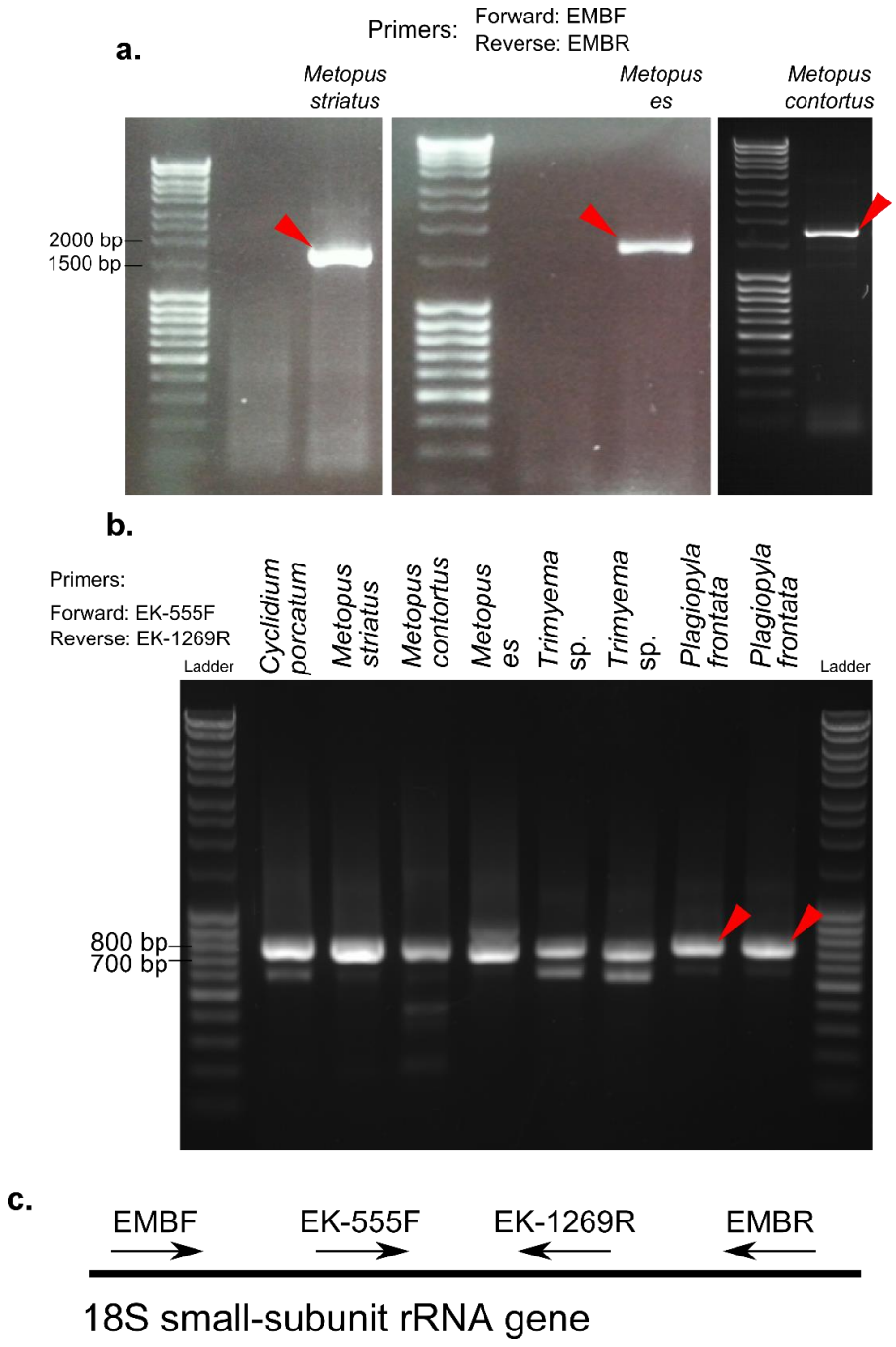


Figure 3.2. Results of PCR targeting 18S rRNA genes

a. PCR products from reactions using primers EMBF and EMBR, targeting 18S rRNA genes from dried cell samples were run on a 1% agarose gel. Red arrowheads indicate products that were cloned and Sanger sequenced. The species used to provide the DNA templates are indicated within the figure.

b. PCR products from reactions using primers EK-555F and EK-1269R, targeting 18S rRNA genes from dried cell samples were run on a 1% agarose gel. Red arrowheads indicate products that were cloned and Sanger sequenced. The remaining bands were not sequenced as larger portions of these genes were amplified and sequenced using other primers (Figure 3.2.a. and Figure 3.3). The species used to provide the DNA templates are indicated within the figure.

c. Diagram indicating positions of primer binding sites relative to each other on the 18S rRNA gene, demonstrates that the binding sites of primer pair EK-555F and EK-1269F are nested within the binding sites of EMBF and EMBR (not drawn to scale).

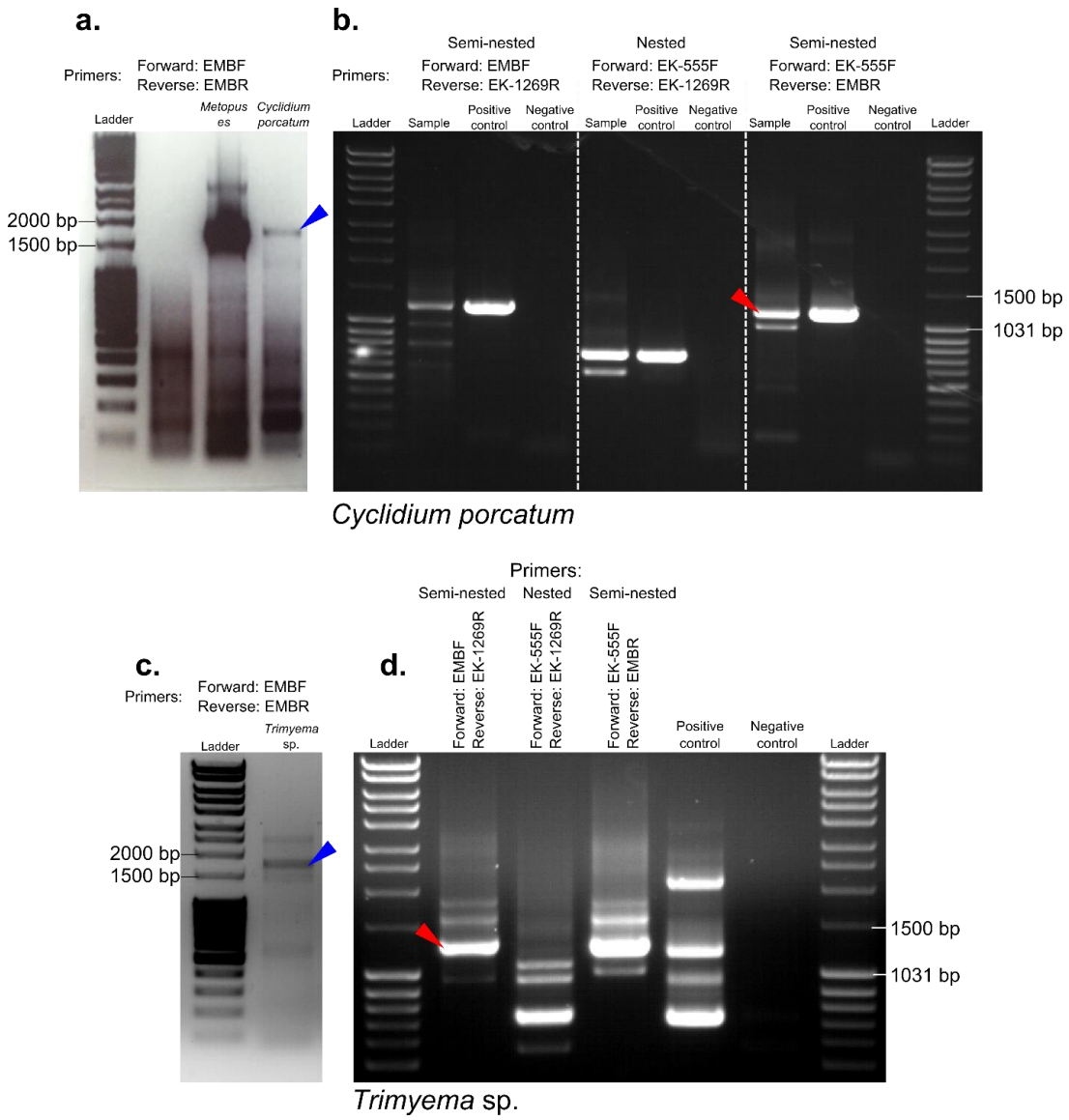


Figure 3.3. Results of nested PCR targeting 18S rRNA genes

a. PCR products from reactions using primers EMBF and EMBR, targeting 18S rRNA genes from dried cell samples were run on a 1% agarose gel. Initial PCR amplified lower quantities of DNA from *Cyclidium porcatum* compared to *Metopus es*. In order to amplify sufficient quantities of DNA for cloning and sequencing the PCR product from the *Cyclidium porcatum* sample (blue arrowhead) was purified and used to provide the DNA template for nested and semi-nested reactions (b.).

b. Semi-nested and nested PCR products using DNA from (a.) to provide the template. Products were run on a 1% agarose gel. Red arrowheads indicate products that were cloned and Sanger sequenced.

c. PCR products from reactions using primers EMBF and EMBR, targeting 18S rRNA genes from dried cell samples were run on a 1% agarose gel. Initial PCR amplified low quantities of DNA from *Trimyema* sp. and in order to amplify sufficient quantities of DNA for cloning and sequencing, the PCR product from the *Cyclidium porcatum* sample (blue arrowhead) was purified and used to provide the DNA template for nested and semi-nested reactions (d.).

d. Semi-nested and nested PCR products using DNA from (c.) to provide the template. Products run on a 1% agarose gel. Red arrowheads indicate products that were cloned and Sanger sequenced.

3.3.3 Phylogenetic inference of relationships between anaerobic ciliates

In order to understand the relationships between the anaerobic ciliate species, their phylogeny was inferred by analysing their full-length 18S rRNA gene sequences using the CAT+GTR model (Lartillot and Philippe, 2004) (Figure 3.4). The analysis included 18S rRNA gene sequences from *Metopus contortus, Metopus es, Metopus striatus, Nyctotherus ovalis, Cyclidium porcatum, Plagiopyla frontata,* and *Trimyema* sp. and two additional anaerobic species recovered from the NCBI nucleotide collection: *Entodinium caudatum* and *Spathidium foissneri*. 18S rRNA gene sequences were also sampled from at least one species belonging to each of the eleven ciliate classes, as recognised by Adl et al. (2012). The analysis recovered four groups of anaerobic ciliates with moderate to high support. These groups were *Spathidium foissneri* and *Entodinium caudatum* (posterior probability of 0.88); *Nyctotherus ovalis, Metopus striatus*, *Metopus es* and *Metopus contortus* (posterior probability of 1); *Cyclidium porcatum*, forming an anaerobic clade on its own (posterior probability of 0.99); and *Trimyema* sp. and *Plagiopyla frontata* (posterior probability of 1). These groups are consistent with other analyses (Embley et al., 1995), suggesting that hydrogenosomes have evolved from aerobic mitochondria in ciliates on at least four occasions. The classes which are known to contain anaerobes are Litostomatea, Armophorea, Oligohymenophorea and Plagiopylea, and species of Armophorea, Oligohymenophorea and Plagiopylea that were investigated in the present study are highlighted in Figure 3.4.

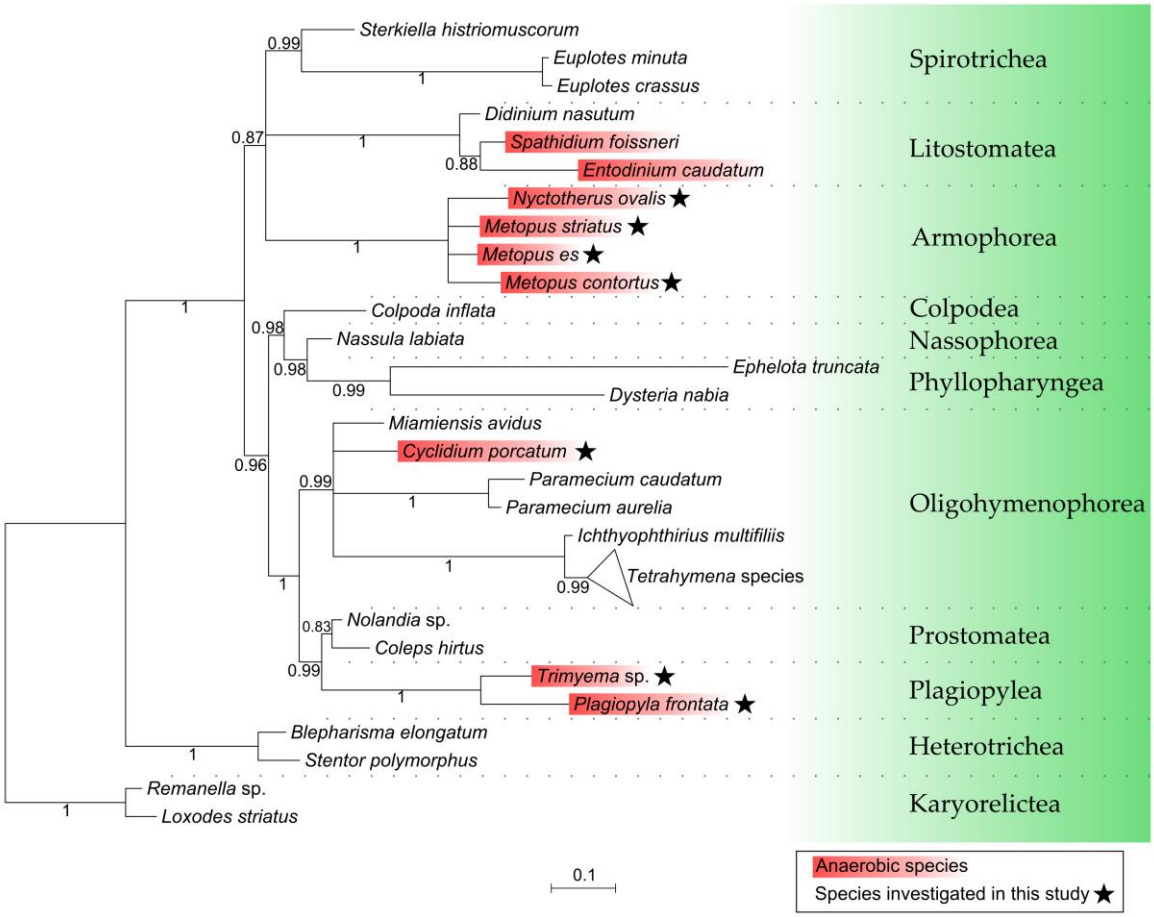
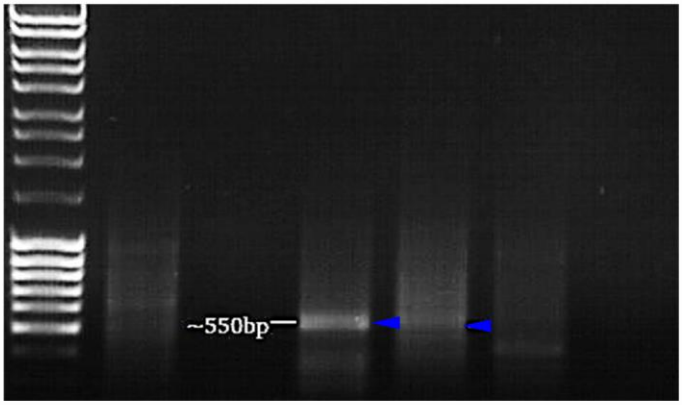


Figure 3.4. Phylogeny of ciliates inferred from 18S rRNA gene sequences by Bayesian analysis using CAT+GTR model (Lartillot and Philippe, 2004) from an alignment of 1754 nucleotide sites. Sequences were sampled from each major class of ciliates (green box). Support values indicate posterior probabilities. Scale bar represents estimated number of substitutions per site.

3.3.4 Identifying ciliate species with hydrogenosome genomes using PCR

Attempts were made to identify which anaerobic ciliate species had hydrogenosome genomes by using PCR to amplify a fragment of the hydrogenosome genome 12S rRNA genes with primers TE59 and TE60. The sequences of these primers were provided by Anders Lind (Ettema-Lab, Uppsala University) who had previously used them to amplify a fragment of the 12S rRNA gene from the hydrogenosome genome of *Nyctotherus ovalis*. The only species for which a putative organelle genome PCR product was obtained was *Metopus contortus* (Figure 3.5. a.). The primers did not amplify any DNA product of expected size or sequence from *Metopus striatus, Metopus es, Cyclidium porcatum, Trimyema* sp. or *Plagiopyla frontata* despite repeated attempts to optimise the reaction. Likewise, published primers targeting 12S rRNA genes (VH59 and VH60) (van Hoek et al., 2000a), and others designed *de novo* based on alignments of ciliate mitochondrial 12S rRNA gene sequences, did not amplify DNA from these species. DNA bands of approximately 500-600bp in length (Figure 3.5. a.) were amplified from two samples of five *Metopus contortus* cells, after they were washed and dried. This DNA was cloned and sequenced using the methods described in Section 2.2. This produced a 534bp sequence (Box 3.1), which blastn searches indicated was most similar (e-value: 3e-44) to the 12S rRNA gene of the *Nyctotherus ovalis* hydrogenosome genome (accession: Y16670.1) available in the NCBI nucleotide collection.

a. PCR with the primers TE59 and TE60, targeting 12S small-subunit rRNA from the hydrogenosome genome of *Metopus contortus*



b. Plasmids purified from *E. coli* DH5α clones, digested with BglII restriction enzyme and ran on a 1% agarose gel. Lanes 1, 5, 6 and 7 appear to have an insert of expected size and were Sanger sequenced by GATC Biotech.

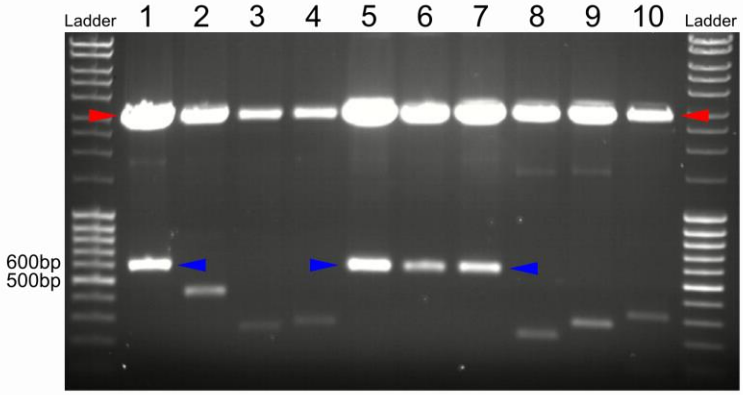


Figure 3.5. a. Products of PCR reactions with TE59 and TE60 primers targeting 12S rRNA genes were run on a 1% agarose gel. 5 dried *Metopus contortus* cells were used as the template in each sample. Blue arrowheads indicate bands of expected size that were gel purified. b. Products of restriction digests using BglII restriction enzyme were run on a 1% agarose gel. Red arrowheads indicate pJET 1.2 plasmid cut at two restriction sites flanking insert. Blue arrowheads indicate DNA inserts similar in size to band amplified in (a.).

3.3.5 Fractionation of cell samples by differential centrifugation

Fractionation of cell samples was performed as described in section 2.4.1, using cell lysis and differential centrifugation methods. This was based on a protocol devised and provided by Anders Lind (Ettema-lab, Uppsala University), which was used previously to successfully sequence a hydrogenosome genome from *Nyctotherus ovalis*. The fractionation procedure attempted to enrich for the hydrogenosomes in samples and decrease the number of macronuclei (Figure 3.6). This procedure was optimised initially using *Metopus contortus* samples as it was possible to amplify part of the hydrogenosome genome from this species by using PCR with primers TE59 and TE60 (section 3.3.4). The same procedures were then used for *Metopus* es, *Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. PCR was used to test which fractions were the most enriched in macronuclei, prokaryotes and hydrogenosomes, using primers targeting ciliate 18S rRNA, prokaryote 16S rRNA and hydrogenosome 12S rRNA genes (Figure 3.6), as markers for each these entities in the sample. The process was not completely successful, as some signal from 18S rRNA and 16S rRNA genes often remained in the final pellets that were used for downstream amplification and sequencing. However signal for macronuclei and prokaryote marker genes was reduced in the final pellet and were most enriched in earlier fractions. Also the final pellet did appear to be the fraction most enriched in hydrogenosomes as was concluded from increased PCR product for primers targeting 12S rRNA genes.

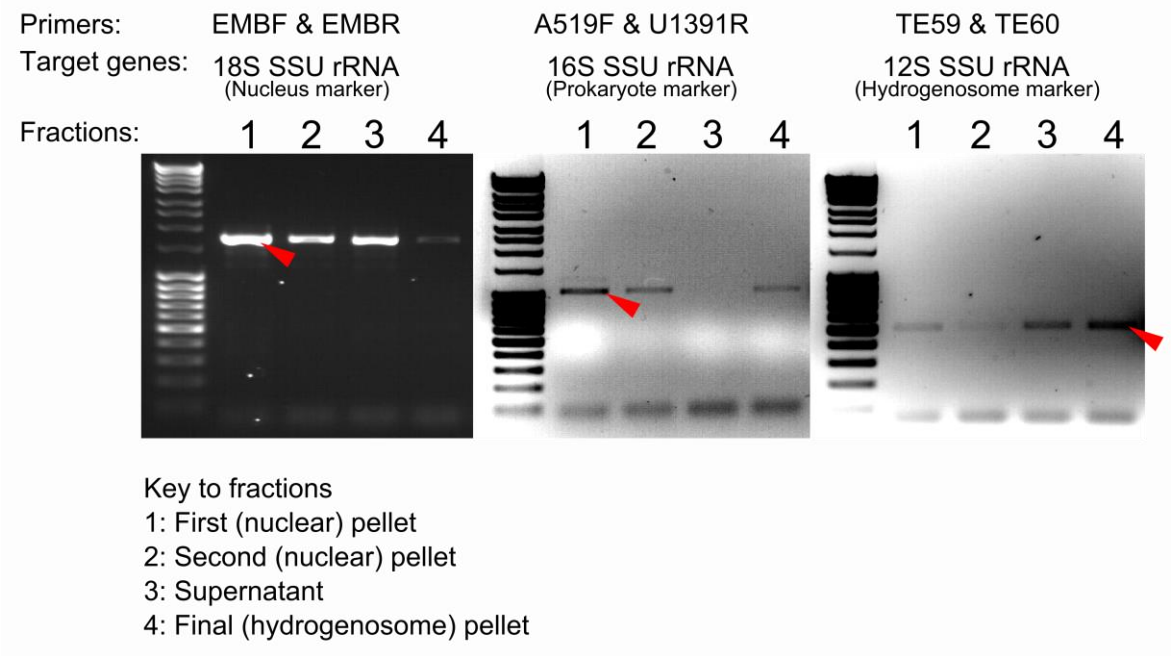


Figure 3.6. Assessment of relative abundance of macronuclei, prokaryotic cells and hydrogenosomes in fractionated *Metopus contortus* cell sample using PCR of individual marker RNA genes. Primers targeted 18S rRNA genes (ciliate nuclear genome), 16S rRNA genes (prokaryotes) and 12S rRNA genes (hydrogenosomes) provide an indication of material being enriched in each fraction. Red arrowheads indicate the most enriched fraction in each PCR reaction. The final pellet had the strongest 12S rRNA gene signal, and the least enriched in 18S rRNA genes was used as template for MDA reactions.

3.3.6 Multiple displacement amplification (MDA) of DNA from fractions

DNA from fraction 4 (Figure 3.6) was amplified using MDA to provide sufficient DNA (50ng) for the production of Nextera sequencing libraries. The results of this process for five biological replicate samples are shown in Figure 3.7. All of the MDA reactions appeared to produce large quantities of high-molecular weight DNA (Figure 3.7. a.), which was purified using a QIAamp purification kit in order to remove primers and reaction buffers (Figure 3.7. b.). The purified samples were reassessed and shown to contain 12S rRNA genes using PCR (Figure 3.7. c.) before being used for library production. Nextera library preparation was performed by Lina Juzokaite (Ettema-lab, Uppsala University) and Illumina sequencing was performed by the Uppsala Genome Centre (Uppsala University).

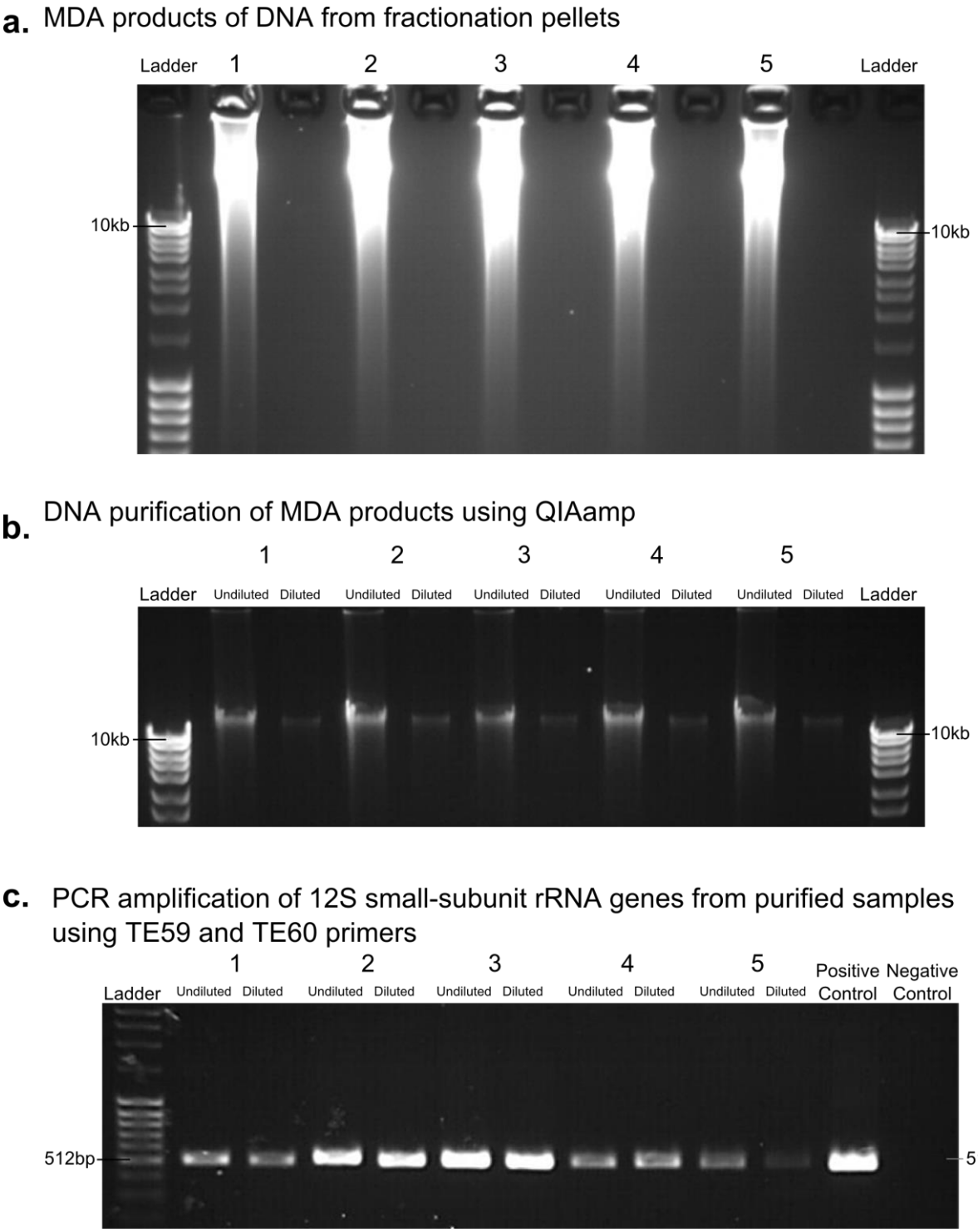


Figure 3.7. MDA products of 5 (labelled 1 - 5) *Metopus contortus* hydrogenosome biological replicate enrichment samples. 3µl of product loaded in each labelled lane were run on a 1% agarose gel. The size of relevant bands and band markers are shown.

a. Unpurified MDA products;

b. Undiluted and diluted QIAquick purified MDA products run on a 1% agarose gel, samples were diluted 1:10 with H_2O;

c. 5µl of PCR products from PCR reactions with primers targeting 12S rRNA genes, run on a 1% agarose gel, using 2µl of purified product from the samples in (b.) as template, except Positive Control in which the template was purified *Metopus contortus* 12S rDNA, amplified using primers TE59 and TE60, and Negative Control in which the template was replaced with water. Relevant sizes of DNA bands and marker bands in ladders are shown.

3.3.7 Genome assembly and identification of hydrogenosome genome contigs

Assembled genomic datasets were produced by Anders Lind (Ettema-lab, Uppsala University) and contigs that contained putative ciliate hydrogenosome genome sequences were identified using blast searches (described in Section 2.4.2). Contigs that appeared to represent hydrogenosome genome sequences were identified from the species *Nyctotherus ovalis, Metopus contortus, Metopus es* and *Metopus striatus*. The lengths and read coverage of these contigs are listed in Table 3.1. From this there generally appeared to be a trend that samples with higher coverage had a larger mean contig length. An exception to this however was that the mean length of contigs from *Metopus contortus* (9787.8 bp) was longer than the mean length of contigs from *Metopus es* (5537.2 bp) but the mean contig k-mer coverage from *Metopus contortus* (11.783) was lower than the mean contig k-mer coverage from *Metopus es* (15.145). The contigs recovered from *Metopus striatus* were on average the smallest (1063 bp) and also had the lowest coverage (4.723), whereas the single contig recovered from *Nyctotherus ovalis* was the longest (48118 bp) and had the highest coverage (649.397).

One contig, Contig ID 3623, that was recovered from the *Metopus es* dataset was unusual in that it had exceptionally high k-mer coverage (22708.3). Further inspection of this sequence using gene prediction software and blast searches (described in Section 2.4.2), suggested that this contig was a chimeric sequence as it appeared to encode a number of proteins with similarity to viral and bacterial proteins (Figure. 3.8). The formation of sequence chimeras is a well-known problem when using methods that involve MDA due to synthesised DNA strands being displaced and priming to a second DNA template during amplification (Lasken and Stockwell, 2007). Therefore the contig with Contig ID 3623 was ‘cut’ into two sections at the position at which similarity to other ciliate mitochondrial genomes could no longer be detected, which was immediately after the ORF encoding the gene *rpl2* (Figure 3.8). The 3040 bp section of Contig 3623 which had similarity to ciliate mitochondrial genome sequences was retained for downstream analyses, whilst the remaining 12636 bp section of Contig 3623 was discarded (Figure 3.8).

Table 3.1. Contigs assembled from genomic sequencing data that were identified as being hydrogenosome genome sequences. The total and mean lengths and mean k-mer coverage for contigs corresponding to hydrogenosome genome sequences from each species dataset is displayed. *Metopus es* contig 3623 was identified as a possible chimera and had exceptionally high coverage. *Metopus es* Contig 3623 was not included in the calculated mean totals.

Species	Contig ID	Length (bp)	K-mer Coverage
Nyctotherus ovalis	66	**48118**	**649.397**
Metopus contortus	273	21866	22.344
	400	12951	8.192
	589	8586	14.432
	1727	2771	9.405
	1865	2765	11.542
	Total:	**48939**	
	Mean:	**9787.800**	**11.783**
Metopus striatus	**21386**	2278	3.301
	35466	1546	4.297
	73798	787	1.954
	55930	1034	2.732
	66780	869	6.192
	52629	1096	4.756
	146059	389	4.084
	58689	988	7.245
	79035	733	3.726
	60293	963	8.939
	Total:	**10683**	
	Mean:	**1068.300**	**4.723**
Metopus es	7642	8779	19.771
	15872	4761	20.943
	17274	4431	4.968
	18487	4191	6.484
	5427	11635	20.480
	5862	10957	14.721
	105885	950	24.173
	30317	2796	18.379
	73053	1335	6.381
	Total:	**49835**	
	Mean:	**5537.222**	**15.145**
Metopus es	3623	15676	22708.3

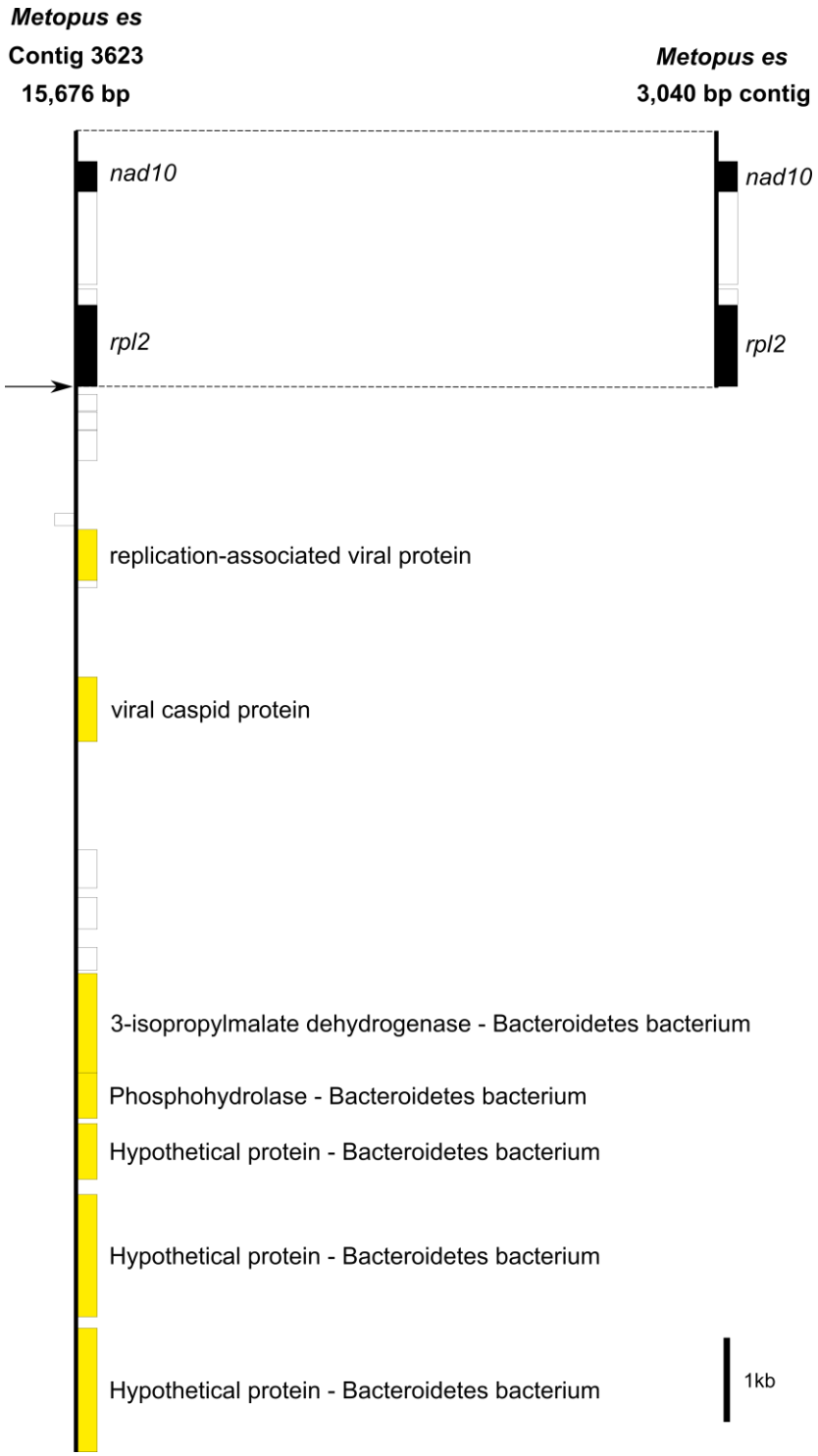


Figure 3.8. Map of *Metopus es* Contig 3623 that appears to be a chimeric sequence. The positions of ORFs are displayed. Black boxes indicate ORFs with similarity to genes from ciliate mitochondrial genomes. Yellow boxes indicate ORFs with no similarity to ciliate genes and similarity to genes from other sources, for each of these ORFs the best hits from blast searches against the Nr database (NCBI) are indicated within the figure. Arrow indicates the point at which contig was separated, leaving a 3,040 bp contig which is considered as part of the hydrogenosome genome of *Metopus es* based on similarity to other ciliate mitochondria and hydrogenosome genomes. The remainder of Contig 3623 was discarded and excluded from any downstream analyses.

3.3.8 Co-assembly of hydrogenosome genomic contigs and RNA-seq reads

Following the initial genomic assembly (section 3.3.7), the contigs listed in Table 3.1 were used for a second, co-assembly with RNA-seq reads, which were sequenced from ciliates isolated from the same culture stocks that were used for genomic sequencing with the help of Henning Onsbring Gustafson (Ettema-lab, Uppsala University). The co-assemblies were performed by Anders Lind (Ettema-lab, Uppsala University). For some contigs from *Metopus contortus*, *Metopus es*, and *Metopus striatus* the co-assembly of genomic contigs and RNA-seq reads extended the length of contigs from the initial genomic assembly and also joined some contigs together to form a single contig. In the case of *Nyctotherus ovalis* however, these procedures did not change the contig from the initial assembly. Additionally some transcripts were identified from assembled transcriptomic datasets, from *Metopus contortus, Metopus es, Metopus striatus* and *Cyclidium porcatum* (Section 3.3.14.), that were not detected in the original genomic data. These were also predicted to be encoded by hydrogenosome genomes based on their sequence similarity to other ciliate mitochondrial genome sequences. These transcripts that did not already map to contigs of hydrogenosome genomes from the genomic assemblies are displayed in Figure 3.10, Figure 3.11 and Figure 3.12. Attempts to join contigs using PCR with specific primers were unsuccessful. Genomic maps of the final co-assembled contigs from the hydrogenosome genomes of *Nyctotherus ovalis*, *Metopus contortus, Metopus es*, *Metopus striatus* and *Cyclidium porcatum* are displayed in Figures 3.9-3.13 (These contig sequences annotated in GenBank format can be found in Appendix A) and the products of the genes encoded by these genomes are listed in Table 3.2.

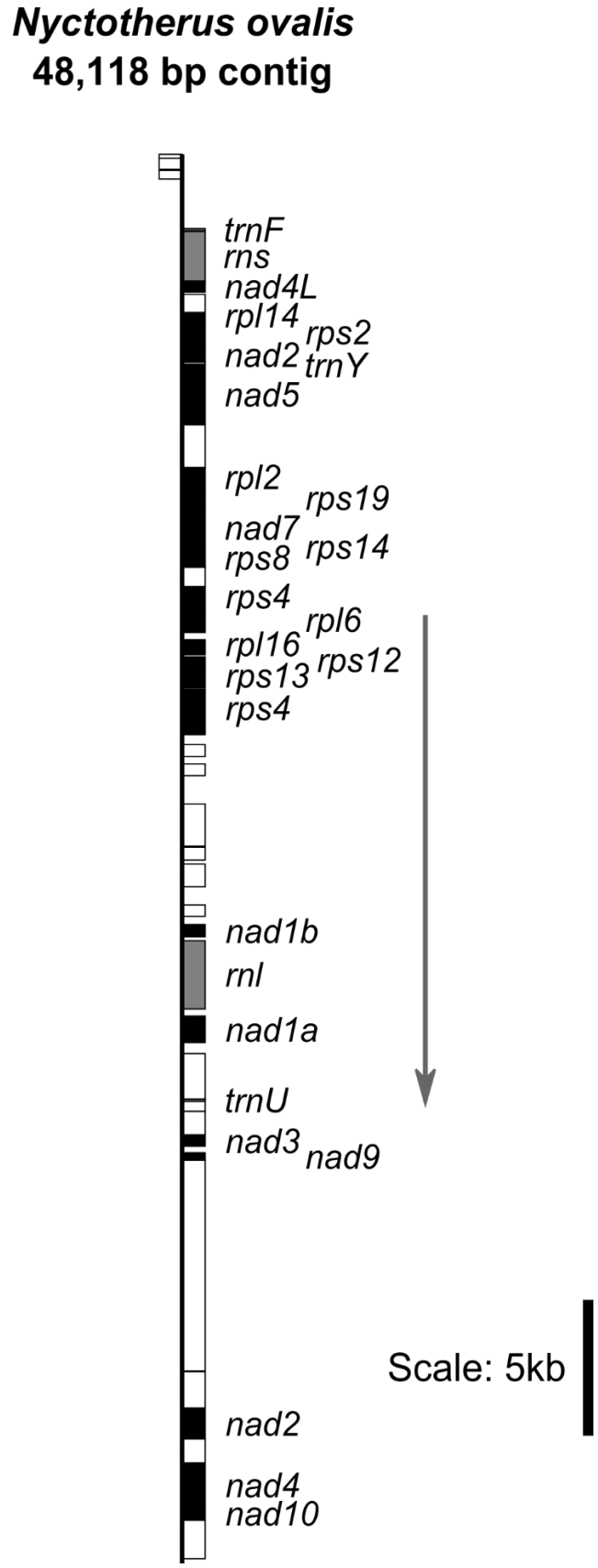


Figure 3.9 Map of the single contig identified as being a partial sequence of the hydrogenosome genome from *Nyctotherus ovalis*. Black boxes indicate ORFs of known function. White boxes indicate ORFs of unknown function and tRNA and rRNA genes are indicated by grey boxes. Arrows indicate the direction of transcription.

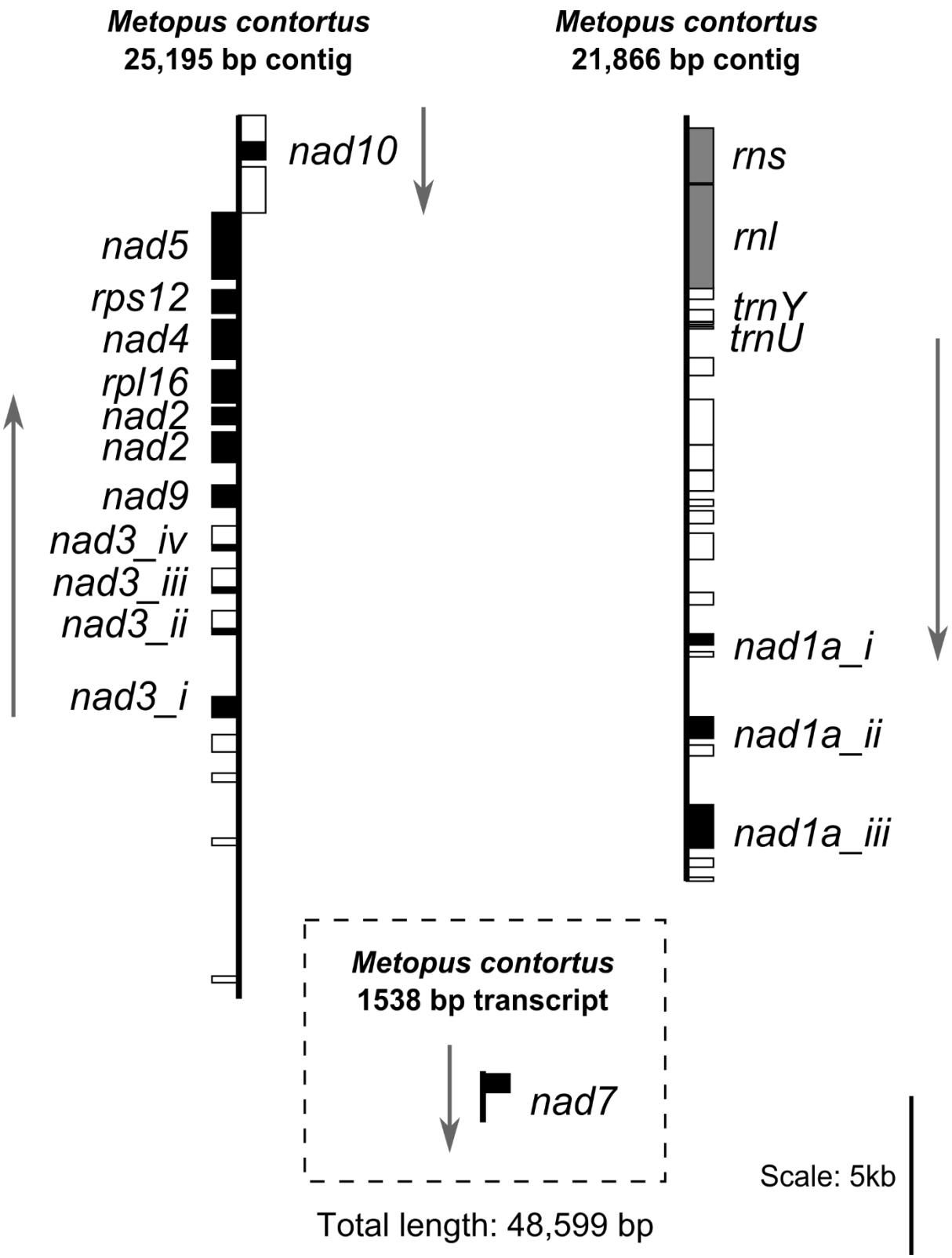


Figure 3.10. Map of the two contigs and one transcript identified as being partial sequences of the hydrogenosome genome from *Metopus contortus*. Black boxes indicate ORFs of known function, white boxes indicate ORFs of unknown function and tRNA and rRNA genes are indicated by grey boxes. Arrows indicate the direction of transcription.

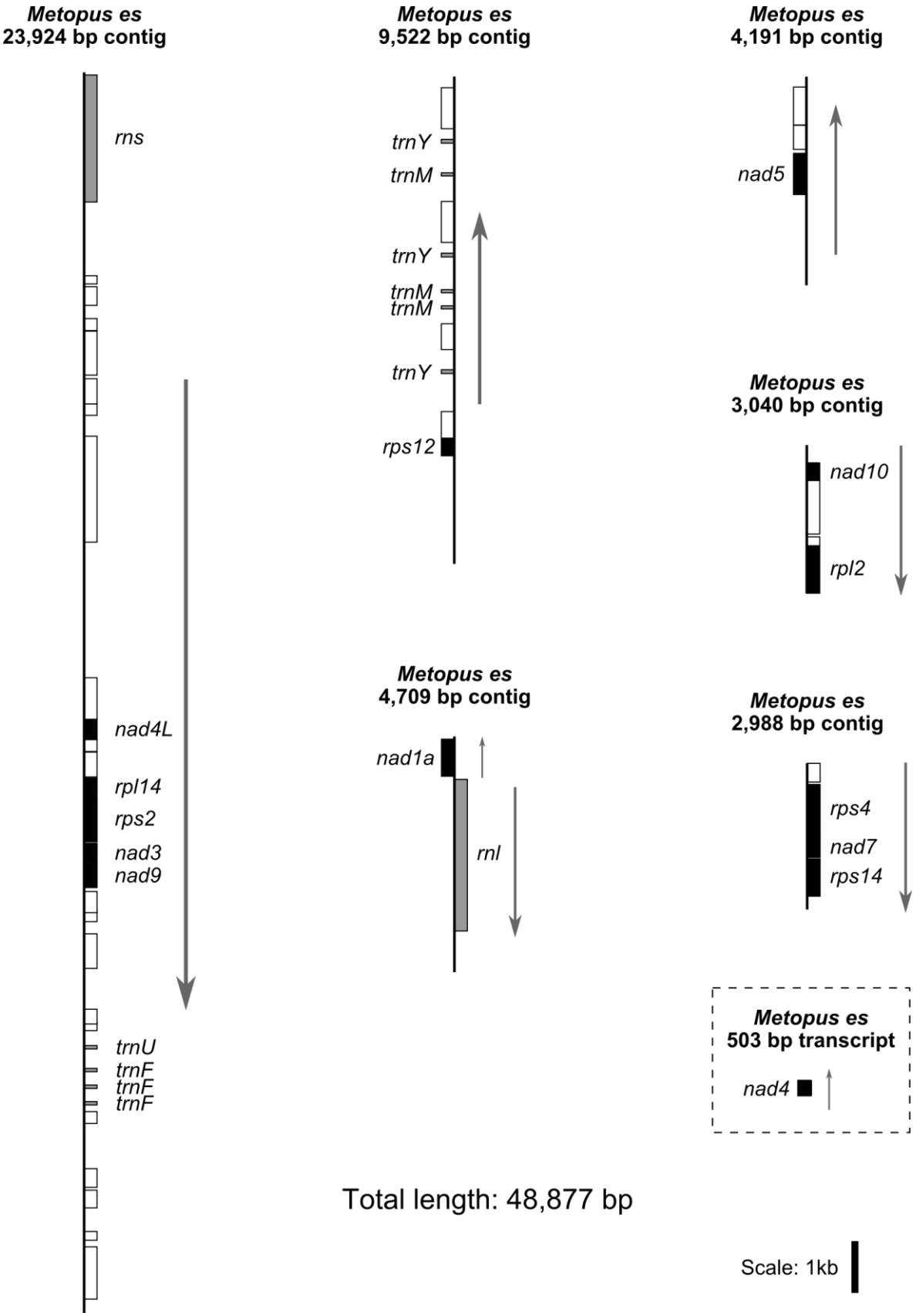


Figure 3.11. Map of the six contigs and one transcript identified as being partial sequences of the hydrogenosome genome from *Metopus es*. Black boxes indicate ORFs of known function, white boxes indicate ORFs of unknown function and tRNA and rRNA genes are indicated by grey boxes. Arrows indicate the direction of transcription.

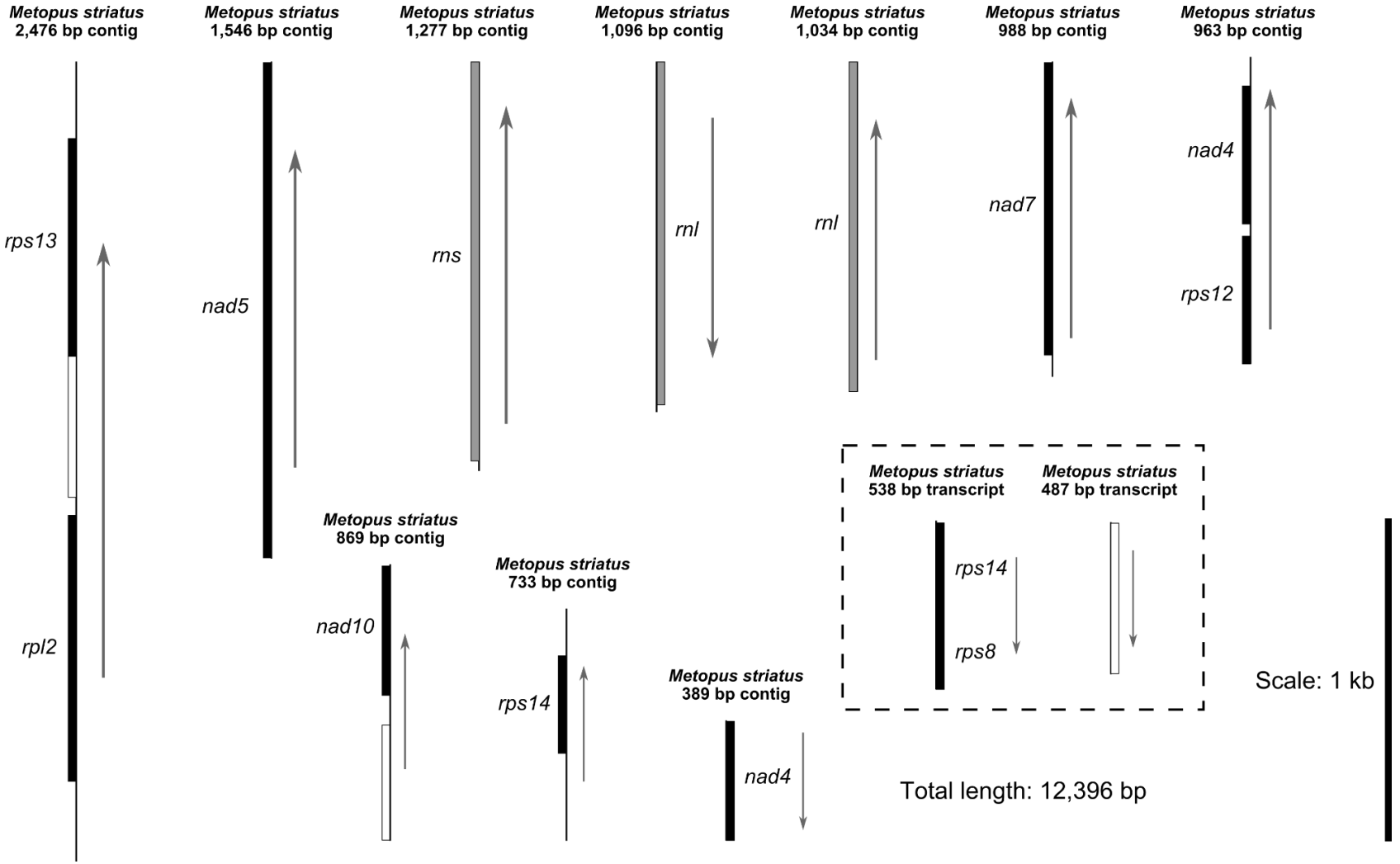


Figure 3.12. Map of the ten contigs and two transcripts identified as being partial sequences of the hydrogenosome genome from *Metopus striatus*. Black boxes indicate ORFs of known function, white boxes indicate ORFs of unknown function and tRNA and rRNA genes are indicated by grey boxes. Arrows indicate the direction of transcription.

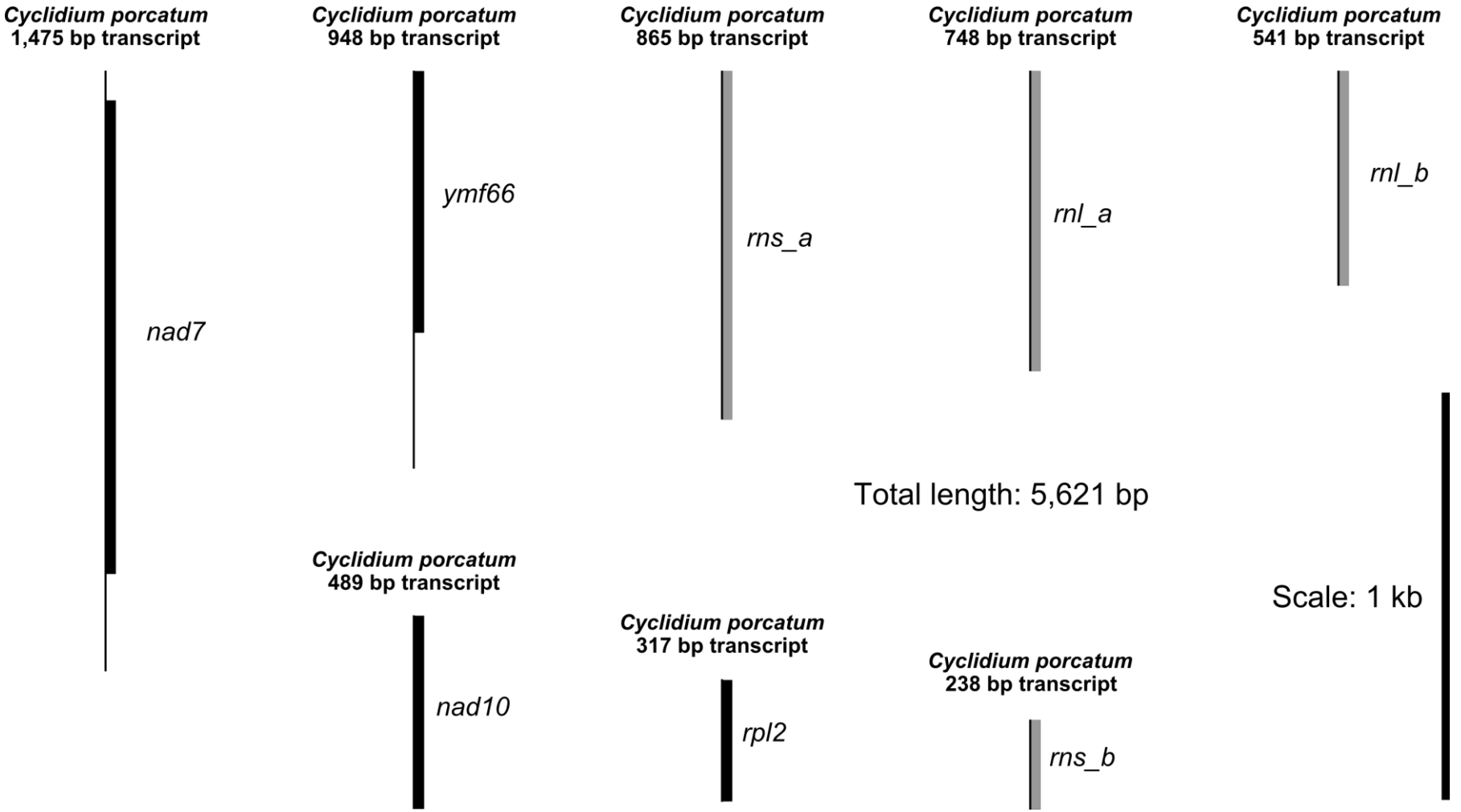


Figure 3.13. Map of the eight transcripts identified as being partial sequences of the hydrogenosome genome from *Cyclidium porcatum*. Black boxes indicate ORFs of known function and rRNA genes are indicated by grey boxes.

Gene	Gene product	Other aliases	Complex	
nad1	NADH-dehydrogenase subunit 1	ND1	P-module	ETC Complex I
nad2	NADH-dehydrogenase subunit 2	ND2		
nad3	NADH-dehydrogenase subunit 3	ND3		
nad4	NADH-dehydrogenase subunit 4	ND4		
nad4L	NADH-dehydrogenase subunit 4L	ND4L		
nad5	NADH-dehydrogenase subunit 5	ND5		
nad6	NADH-dehydrogenase subunit 6	ND6		
nad7	NADH-dehydrogenase subunit 7	49 kDa, NDUFS2	Q-module	
nad9	NADH-dehydrogenase subunit 9	30 kDa, NDUFS3		
nad10	NADH-dehydrogenase subunit 10	PSST, NDUFS7		
ymf66			F_1F_0 ATP synthase	
rps2	Ribosomal protein S2		Ribosome	
rps3	Ribosomal protein S3			
rps4	Ribosomal protein S4			
rps7	Ribosomal protein S7			
rps8	Ribosomal protein S8			
rps10	Ribosomal protein S10			
rps12	Ribosomal protein S12			
rps13	Ribosomal protein S13			
rps14	Ribosomal protein S14			
rps19	Ribosomal protein S19			
rpl2	Ribosomal protein L2	60S ribosomal protein		
rpl6	Ribosomal protein L6			
rpl14	Ribosomal protein L14			
rpl16	Ribosomal protein L19			
rns	Small subunit rRNA	12S	Ribosome	
rnl	Large subunit rRNA	16S		

Table 3.2. A key to the names of genes and their products that were identified from the ciliate hydrogenosome genomes, annotated in Figures 3.9-3.13.

3.3.9 Analysis of hydrogenosome genomes

Hydrogenosome genomes were partially sequenced from *Nyctotherus ovalis, Metopus contortus, Metopus es* and *Metopus striatus*. Indirect evidence of a hydrogenosome genome in *Cyclidium porcatum* was also obtained since eight transcripts were identified in transcriptome datasets that share similarity with genes found in other ciliate mitochondrial genomes. No evidence was found from analysis of either genomic or transcriptomic sequencing, to suggest that the hydrogenosomes of *Trimyema* sp. and *Plagiopyla frontata* have a genome. Since *Trimyema* sp. and *Plagiopyla frontata* are members of the same clade (Figure 3.4), it is possible that the hydrogenosome genome was lost in their common ancestor. Consistent with the lack of an organelle genome, no genes for components of the ETC were identified in these two species from transcriptomic datasets (section 3.3.14) and their hydrogenosomes do not appear to contain cristae (Figure 3.19).

Metopus contortus, Metopus es, Metopus striatus and *Nyctotherus ovalis* appear to have hydrogenosome genomes and contigs with sequence similarity to other ciliate mitochondrial genomes were identified for each of these four species from genomic assembly data using blast searches. This indicates that hydrogenosome genomes could be a conserved feature shared by all species of Armophorea. The 48 118 bp hydrogenosome genome contig from *Nyctotherus ovalis* sequenced in the present study is predicted to be almost complete and is consistent with the size of the size of the hydrogenosome genome from another isolate of *Nyctotherus ovalis*, predicted as ‘exceeding 48 kb’ by de Graaf et al. (2011) from the results of Southern blotting experiments. From analysis of transcript data several sequences were identified that either map to these contigs, or are likely to be encoded by the hydrogenosome genomes based on their sequence similarity to genes encoded by ciliate mitochondrial genomes. Evidence for a hydrogenosome genome in *Cyclidium porcatum* was found from analysis of transcriptomic data, suggesting that the fractionation enrichment methods that were used to sequence sections of *Metopus contortus*, *Metopus es* and *Metopus striatus* were inadequate for *Cyclidium porcatum*. This was possibly due to *Cyclidium porcatum* being smaller in size than those ciliates (Figure 3.1) with far fewer hydrogenosomes per cell. Estimations from these species isolated in other

studies suggested that *Metopus contortus* contains approximately 21000 hydrogenosomes per cell (Finlay and Fenchel, 1989), whereas *Cyclidium porcatum* was estimated to contain approximately 15 hydrogenosomes per cell (Esteban et al., 1993). Like other ciliate mitochondrial genomes the hydrogenosome genomes are expected to be linear-mapping but this was not clearly evident from the incomplete data generated. One end of the hydrogenosome genome contig sequenced from *Nyctotherus ovalis* in the present study however, has a region of three 38 bp (TATTGTAATACTAATAATATGT-GTGTTAATGCGCGTAC) tandem repeats, which is similar in structure to the known telomeric regions of mitochondrial genomes found in ciliates such as *Sterkiella histriomuscorum* and *Tetrahymena thermophila* (Morin and Cech, 1986; Swart et al., 2012). Ciliate mitochondrial telomeres are typically found to be constructed of tandemly repeated sequences, the length of which can vary and can have no significant sequence similarity, even between closely related species of *Tetrahymena* (Morin and Cech, 1988). The function of mitochondrial telomeres in ciliates is thought to be similar to that of nuclear telomeres, to prevent shortening of the genome during replication that would lead to degradation (Morin and Cech, 1988).

3.3.10 Protein-coding genes in ciliate hydrogenosome genomes

The predicted protein coding genes of known function identified from ciliate hydrogenosome genomes are summarised in Table 3.3, alongside two mitochondrial genomes from aerobic ciliates that were sequenced in other studies (Pritchard et al., 1990; Swart et al., 2012). Several ORFs from ciliate mitochondrial genomes appear to encode proteins that can be assigned functions based on their similarity to known genes. In addition to these, however, all ciliate mitochondrial genomes studied to date (Pritchard et al., 1990; Burger et al., 2000; Brunk et al., 2003; Moradian et al., 2007; de Graaf et al., 2009; Barth and Berendonk, 2011; Coyne et al., 2011; Swart et al., 2012) also encode numerous ORFs for which no function can be inferred, even when employing sensitive HMM search methods. This also appears to be the case for the ciliate hydrogenosome genomes

sequenced in the present study, which are also predicted to encode numerous ORFs of unknown function (Figures 3.9 – 3.13).

Table 3.3. Protein coding genes of known function encoded by the partial hydrogenosome genomes of *Cyclidium porcatum, Metopus striatus, Metopus es, Metopus contortus* and *Nyctotherus ovalis*, sequenced in the present study, the hydrogenosome genome of *Nyctotherus ovalis* sequenced in a previous study (de Graaf et al., 2011) and the mitochondrial genomes from two aerobic ciliates, *Paramecium aurelia* (Pritchard et al., 1990) and *Sterkiella histriomuscorum* (Swart et al., 2012). Genes encoded by a single ORF (*) and split genes (*a* + *b*) are shown.

		Hydrogenosomes						Mitochondria	
Length of genome sequence:		5.6 kb *Cyclidium porcatum*	12.4 kb *Metopus striatus*	48.9 kb *Metopus es*	48.6 kb *Metopus contortus*	48.1 kb *Nyctotherus ovalis*	47.1 kb *Nyctotherus ovalis*	40.5 kb *Paramecium aurelia*	69.8 kb *Sterkiella histriomuscorum*
ETC Complex I	*nad1*			*	*	*a* + *b*	*a* + *b*	*a* + *b*	*a* + *b*
	nad2				*a* + *b*	*a* + *b*	*a* + *b*	*a* + *b*	*a* + *b*
	nad3			*	*	*	*	*	*
	nad4		*	*	*	*	*	*	*
	nad4L			*		*	*	*	*
	nad5		*	*	*	*	*	*	*
	nad6							*	*
	nad7	*	*	*	*	*	*	*	*
	nad9			*	*	*	*	*	*
	nad10	*	*	*	*	*	*	*	*
ETC Complex III	*cob*							*	*
ETC Complex IV	*cox1*							*	*
	cox2							*	*
ATP-synthase	*atp9*							*	*
Cytochrome *c* maturation	*ccmF/yejR*							*	*
Ribosomal Proteins	*rps2*			*		*	*		*
	rps3							*	*
	rps4			*		*	*	*	*
	rps7							*	*
	rps8		*			*	*	*	*
	rps10							*	*
	rps12		*	*	*	*	*	*	*
	rps13		*					*	*
	rps14		*			*	*	*	*
	rps19							*	*
	rpl2	*	*	*		*	*	*	*
	rpl6					*	*		*
	rpl14			*		*	*	*	*
	rpl16				*	*	*	*	*

Since these genomes are only partially sequenced, it is difficult to form conclusions based on their patterns of gene loss but with this caveat in mind, some observations can nevertheless be made. Firstly, none of the components of ETC Complexes III and IV, or F_1F_0 ATP synthase, were detected in the data for *Nyctotherus ovalis, Metopus contortus, Metopus es* or *Metopus striatus*. By contrast, the genes for *cob, cox1, cox2, atp9* and the cytochrome *c* maturation protein-coding gene *ccmF*, are typically found on the mitochondrial genomes of aerobic ciliates. This is consistent with previous studies (Boxma et al., 2005; de Graaf et al., 2011) and suggests the loss of these complexes from the anaerobic species. The gene content of the *Nyctotherus ovalis* hydrogenosome genome sequenced in the present study and that which has been sequenced previously (de Graaf et al., 2011) appears to be the same, although the primary sequences of these genes and genomes show differences in their nucleotide composition. Although no genes for components of these complexes were identified in *Cyclidium porcatum*, one transcript, which is likely to be encoded by the hydrogenosome genome based on its predicted genetic code, was identified that has homology to a protein named Ymf66 encoded by the *Tetrahymena thermophila* mitochondrial genome (Brunk et al., 2003; Smith et al., 2007). This protein is thought to be a component of the divergent F_0 sub-complex of F_1F_0 ATP synthase (Nina et al., 2010). This suggests that the *Cyclidium porcatum* hydrogenosome genome may encode components of an F_1F_0 ATP synthase, as well as ETC Complex I and ribosomal proteins (Table 3.3). The only protein coding genes found in all hydrogenosome genomes in the present study are *nad7* and *nad10*, which encode two electron transferring Q-module subunits, and which have been identified from all other sequenced ciliate mitochondrial genomes. Previous studies suggest that these two genes are the most highly conserved protein-coding genes in ciliate mitochondrial genomes (Swart et al., 2012) and the retention of these genes in the hydrogenosome genomes sequenced in the present study is consistent with them having important functions in these organisms.

Bacterial homologues of the proteins Nad2 (NuoN), Nad4 (NuoM) and Nad5 (NuoL), protein subunits of ETC Complex I, are thought to be distant homologues of a class of bacterial antiporters and are therefore inferred to be the main components facilitating proton translocation in this complex (Mathiesen and

Hägerhäll, 2002; Efremov et al., 2010). Some of these subunits are encoded by the hydrogenosome genomes of *Metopus contortus, Metopus es* and *Metopus striatus*, and *Nyctotherus ovalis* seems to contain all three. The presence of these proteins would suggest that the ETC Complex I of *Metopus contortus, Metopus es, Metopus striatus* and *Nyctotherus ovalis* are actively involved in proton translocation. Since none of these species appear to have F_1F_0 ATP synthase to make ATP, proton translocation in these hydrogenosomes likely serve a purpose other than to facilitate ATP production. An electrochemical gradient is thought to be required for protein import across the inner mitochondrial membrane (Gasser et al., 1982), this could be one explanation for why these Complex I subunits are retained in species without F_1F_0 ATP synthase. *Cyclidium porcatum* for which we have the least data is the only species from the present study from which none of these subunits were found. Given that the hydrogenosome genome of *Cyclidium porcatum* encodes a purported component of F_1F_0 ATP synthase, which requires proton translocation by ETC complexes in order to function, it could be predicted that this species does contain the genes *nad2*, *nad4* and *nad5,* encoding proton-pumping subunits of ETC Complex I, but the region of the hydrogenosome genome that encodes them was not sequenced.

Similar patterns of gene retention were observed for ribosomal proteins encoded by the hydrogenosome genomes of *Metopus contortus, Metopus es, Metopus striatus* and *Nyctotherus ovalis*, the most apparent similarity being that they all encode *rps12*. Additionally all the ribosomal protein genes encoded by the hydrogenosome genomes of *Metopus contortus, Metopus es* and *Metopus striatus* are a subset of the ribosomal protein genes encoded by the more completely sequenced hydrogenosome genomes of *Nyctotherus ovalis*, with the exception of *rps13* identified from *Metopus striatus*. *rps13* has not been identified from the hydrogenosome genomes of any other anaerobic ciliate species but has been found encoded by the mitochondrial genomes of aerobic ciliates. The only ribosomal protein gene that has been detected from the hydrogenosome genome of *Cyclidium porcatum* so far is *rpl2*. Compared to aerobic ciliate mitochondrial genomes, the three most complete ciliate hydrogenosome genomes sequenced in the present study from *Metopus es, Metopus contortus* and *Nyctotherus* ovalis, appear to have lost some genes encoding ribosomal proteins. Although more protein genes could still be found on unsequenced regions of these three

genomes, given that the combined length of regions that are sequenced are similar to the length of other ciliate mitochondrial genomes, it could be predicted that they are mostly complete. If this is the case then it is surprising that the *Metopus contortus* hydrogenosome genome appears to be missing such a large number of genes encoding ribosomal proteins compared to the genomes of other ciliate mitochondria and hydrogenosomes. Given that only *rps12* and *rpl16* were detected from the hydrogenosome genome of *Metopus contortus* it is possible that other genes encoding ribosomal proteins may have diverged so much in sequence that they were not detected or that they could have been transferred to the nuclear genome.

Some genes that are typically encoded by one ORF in some organisms are often found to be split and encoded by two ORFs in ciliate mitochondrial genomes. Such splits were identified in the protein encoding genes for *nad1*, *nad2* and *rps3* in mitochondrial genomes of most ciliate species sequenced so far (Swart et al., 2012). The *nad1* and *nad2* genes identified from the hydrogenosome genome of *Nyctotherus ovalis* in the present study are split and encoded by two ORFs and a split *nad2* gene encoded by two ORFs was identified in *Metopus contortus*. Three *nad1a* genes were identified from *Metopus contortus*, encoded by three separate ORFs (labelled *nad1a_i-nad1a_iv* in Figures 3.10 and 3.15). A *nad1b* gene from *Metopus contortus* was not identified and has probably either been lost or is encoded by an unsequenced region of the genome. The three ORFs encoding *nad1a* in *Metopus contortus* show similarity to one another and appear to have arisen through gene duplication. Furthermore, *Metopus contortus* also appears to have four ORFs encoding *nad3* (labelled *nad3_i-nad3_iv* in Figures 3.10 and 3.15). These ORFs are similar but not identical and have presumably also arisen through gene duplication events. In the case of both *nad1a* and *nad3* from *Metopus contortus* the duplicates are located adjacent to each other on the genome and it is unknown to what extent these genes are functional. Gene duplications are not uncommon in ciliate mitochondrial genomes and other examples were previously identified in aerobic ciliates including *nad9* in *Tetrahymena thermophila* (Brunk et al., 2003) and *nad5*, *nad4*, and *nad9* in *Sterkiella histriomuscorum* (Swart et al., 2012).

3.3.11 Ribosomal RNA genes from ciliate hydrogenosome genomes

The genes *rns* and *rnl*, encoding SSU and LSU rRNA respectively, are split and are encoded by two independent regions of the mitochondrial genomes from several ciliates, including *Tetrahymena, Paramecium* and *Sterkiella histriomuscorum*. In the present study the *rns* and *rnl* genes of *Cyclidium porcatum*, appear to be split and transcripts were identified for each of these four genes. This is consistent with these split genes being a conserved feature of Oligohymenophorea mitochondrial genomes. Splits were not observed for the rRNA genes of *Metopus* species or *Nyctotherus ovalis*, which seem to be encoded by a single region of their hydrogenosome genomes and this is consistent with what was observed in previous studies of the hydrogenosome genome of *Nyctotherus ovalis* (de Graaf et al., 2011).

3.3.12 Hydrogenosome genome transfer RNA genes

Prediction of tRNA genes in ciliate mitochondrial genomes seems to vary between studies and despite using the same software the results reported here sometimes differ from what has been observed in another study (Swart et al., 2012). In the present study tRNA genes were predicted using the programs tRNAscan-SE 1.3.1 and tRNAscan-SE On-line (Lowe and Chan, 2016), which uses tRNAscan-SE 2.0, the most recent version of the tRNAscan software. From the genomes analysed in Table 3.3, both versions of the software predicted the same tRNA genes from each genome. Several tRNA genes were predicted to be encoded by the hydrogenosome genome contigs that were sequenced in the present study from *Nyctotherus ovalis, Metopus contortus* and *Metopus es,* but no tRNA genes were predicted to be encoded by the hydrogenosome genome contigs of *Metopus striatus*. The mitochondrial genomes of *Sterkiella histriomuscorum*, *Paramecium aurelia* and *Tetrahymena pyriformis* were also reanalysed using the same methods. The main difference between the present study and what has been predicted previously (de Graaf et al., 2011; Swart et al., 2012), is that all genes which were previously predicted as being *trnW* were now predicted to be *trnU* genes. The reasons for this are unclear, however remoulding of tRNA genes appears to be common in many metazoan mitochondrial genomes (Cantatore et

al., 1987; Rawlings et al., 2003), although tRNA remoulding does not seem to be well documented in other organisms. It might simply be the case that ciliate mitochondrial *trnW* and *trnU* genes are difficult to distinguish. The tRNA genes predicted for hydrogenosome genomes in the present study from *Nyctotherus ovalis, Metopus contortus, Metopus es* and *Metopus striatus* are a subset of the tRNA genes predicted to be encoded by aerobic ciliate mitochondrial genomes (Table 3.4) and *Nyctotherus ovalis* appears to have the same tRNA genes as *Paramecium aurelia*. Some ciliate hydrogenosome and mitochondrial genomes seem to have several copies of the same tRNA genes. This particularly seems to be true for *Metopus es*, which three copies of each of the tRNA genes *trnF*, *trnM* and *trnY*. Since many of these copies appear to be located relatively close to one another on the hydrogenosome genome contigs of *Metopus es* (Figure 3.11), it is possible that they have arose through species-specific gene duplications.

Table 3.4. The number of predicted tRNA genes identified from ciliate hydrogenosome genomes sequenced in the present study and the reanalysis of aerobic ciliate mitochondrial genomes from previous studies (Pritchard et al., 1990; Burger et al., 2000; Swart et al., 2012). Numbers in brackets indicate the tRNA genes that were predicted for the same genomes by Swart et al. (2012).

		Metopus striatus	*Metopus es*	*Metopus contortus*	*Nyctotherus ovalis*	*Sterkiella histrimuscorum*	*Paramecium aurelia*	*Tetrahymena pyriformis*
Cysteine	*trnC*					(1)		
Glutamic acid	*trnE*					1(1)		1(1)
Phenylalanine	*trnF*		3		1(1)	1(1)	1(1)	1(1)
Histidine	*trnH*					1(1)		1(1)
Lysine	*trnK*					1(1)		
Leucine	*trnL*					1(1)		2(1)
Methionine	*trnM*		3			1(2)	(1)	(1)
Glutamine	*trnQ*					1(1)		
Selenocysteine	*trnU*		1	1	1	1	1	1
Tryptophan	*trnW*				(1)	1(1)	(1)	(1)
Tyrosine	*trnY*		3	1	1(1)	(1)	1(1)	1(1)

3.3.13 Comparison of hydrogenosome genomes from Nyctotherus ovalis and Metopus contortus

Including in the present study, the hydrogenosome genome of *Nyctotherus ovalis* has been sequenced three times (Boxma et al., 2005; de Graaf et al., 2011). Although some obvious similarities between these three genomes can be observed, their primary sequences are different from one another. Figure 3.14 shows an alignment of the most conserved region of *rns* genes from the hydrogenosome genome of *Nyctotherus ovalis* sequenced in the present study and others (Boxma et al., 2005; de Graaf et al., 2011). Some regions of this gene appear highly conserved, whilst others appear to have insertions, deletions and point mutations. This demonstrates that even the most conserved gene on these genomes are different from one another. It is probably unlikely that these differences are all due to sequencing or assembly errors, since at the level of sequence identity, some regions of the gene appear to be more conserved than others. Differences caused by sequencing errors might be expected to be distributed more randomly across the gene.

It is most likely that these genomes have been sequenced from several sub-species of *Nyctotherus ovalis*, given that they were isolated from several cockroach genera. A previous study showed that *Nyctotherus ovalis* 18S rRNA gene sequences differed even between isolates from the same species of cockroach (Van Hoek et al., 1998). Hydrogenosome genomes therefore could have been sequenced from different sub-species of *Nyctotherus ovalis*, especially since in the present study it was isolated from the cockroach *Blaptica dubia*, whereas in previous studies it was isolated from strains of the cockroach *Blaberus* sp. (Boxma et al., 2005; de Graaf et al., 2011). In nature, it has been suggested that *Nyctotherus ovalis* lineages can swap species of cockroach hosts but is likely to only occur rarely (Van Hoek et al., 1998). *Nyctotherus* living commensally in other organisms, such as *Nyctotherus cordiformis* in frogs and *Nyctotherus velox* in millipedes, have evolved into distinct species, having probably diverged due to the behaviour and distributions of their hosts (Van Hoek et al., 1998). Although not completely isolated from each other, it could be argued that populations of these commensal ciliates are more isolated than free-living ciliates. Isolated populations are known to diverge from one another as they are subject to different selective

pressures and gene flow between populations is limited, as occurs under allopatric speciation (Dobzhansky, 1937). This could explain why the hydrogenosome genomes of different *Nyctotherus ovalis* isolates appear to be different from one another.

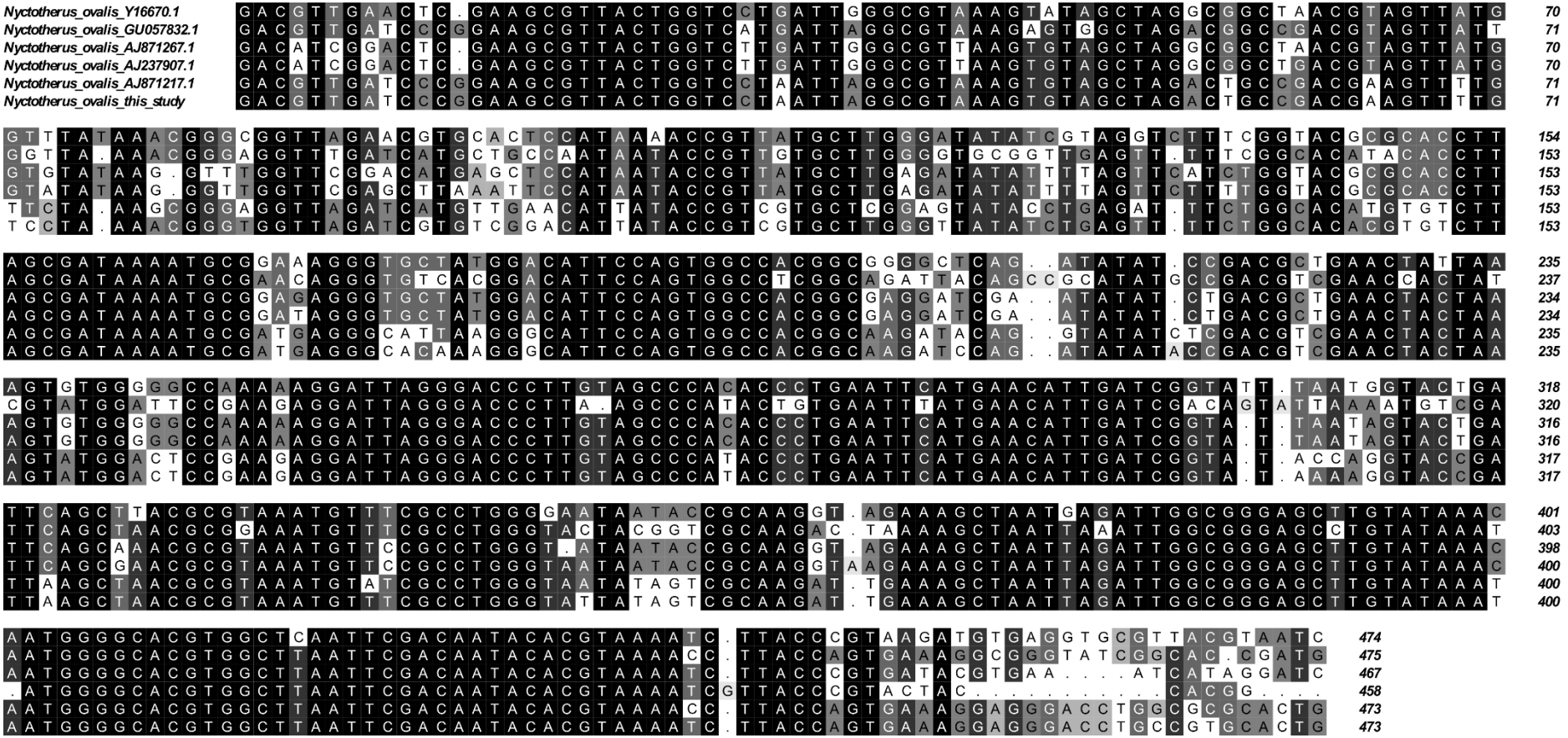


Figure 3.14 Alignment of *rns* (12S rRNA) genes from *Nyctotherus ovalis* sequenced in the present study by Anders Lind (Ettema-lab, Uppsala University) and others obtained from the NCBI nucleotide collection. Grey-scale shading indicates the most conserved nucleotide at each site with darker shades indicating nucleotides that are more conserved. Sequence accession numbers are displayed within the figure. In the region shown 66.9% of sites are conserved across all seven sequences.

The *Nyctotherus ovalis* hydrogenosome genome in the present study was sequenced by Anders Lind (Ettema-lab, Uppsala University) and is the longest assembled to date (48,118bp). The next largest *Nyctotherus ovalis* hydrogenosome genome sequence is 41,666 bp in length (Accession: GU057832) (de Graaf et al., 2011), and the third is 14,928 bp in length (Accession: AJ871267.1). The most obvious structural difference between the two largest genomes is that there appears to have been a large-scale rearrangement, such that in order to recreate the gene order of one genome from the other, it must be divided into two sections and the first section be moved in front of the second (Figure 3.15). The corresponding sections between the two largest *Nyctotherus ovalis* genomes appear to have a gene order that is largely co-linear but the order of these two sections is swapped. It is of course possible that these differences are the results of erroneous assembly and this could be tested by performing PCR to attempt to amplify and then sequence the disputed regions from each isolate, in order to confirm whether they exist.

The mitochondrial genomes of closely related ciliates typically share large sections of collinear gene order, as shown for the oligohymenophoreans *Tetrahymena thermophila* and *Paramecium aurelia* (Burger et al., 2000), and the spirotrichs *Sterkiella histriomuscorum* and *Euplotes minuta* (Swart et al., 2012). This was not observed between the hydrogenosome genomes of *Metopus contortus* and *Nyctotherus ovalis* in the present study and regions of genome synteny between these species is lacking. Just one collinear region encoding *nad3*, *nad9* and *nad2* is conserved between them (Figure 3.15).

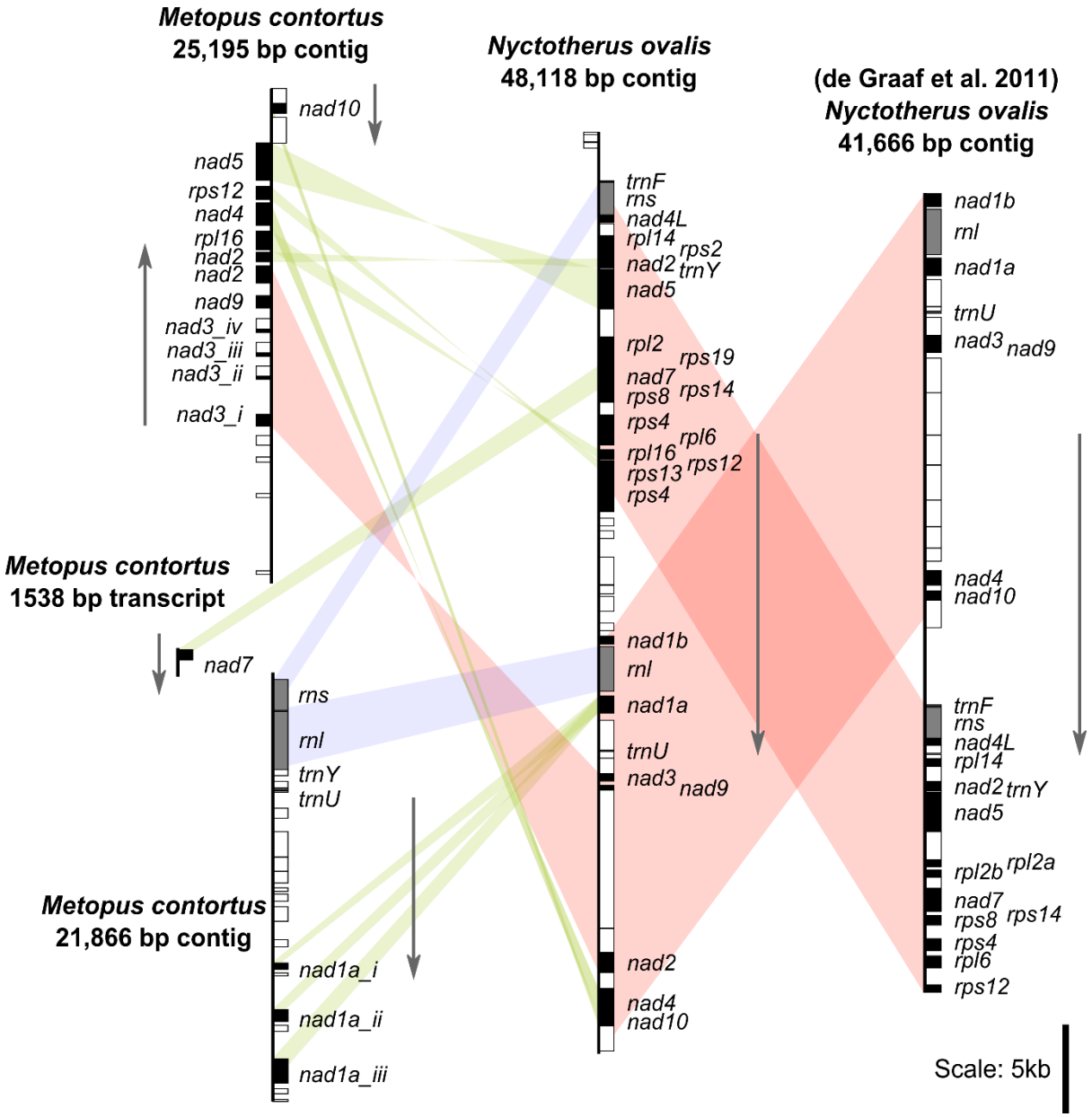


Figure 3.15. Three hydrogenosome genomes from anaerobic ciliates with positions of main coding features mapped. ORFs with similarity to genes of known function are represented by black boxes, ORFs of unknown function are represented by white boxes and RNA encoding genes are represented by grey boxes. Arrows indicate direction of transcription. Similarity between protein coding and RNA coding genes between *Metopus contortus* and *Nyctotherus ovalis* hydrogenosome genomes are indicated by green and blue bands, respectively. Genome regions of collinear gene order are indicated by red bands.

3.3.14 Assembly and analysis of transcriptome datasets

In addition to analysing hydrogenosome genome sequences, nuclear-encoded genes were also analysed to identify proteins that are likely to function in hydrogenosomes. To do this transcriptomic datasets were assembled by Henning Onsbring Gustafson (Ettema-lab, Uppsala University) using methods described in section 2.5.1. Samples of cell isolates were sequenced from *Metopus contortus*, *Metopu*s es, *Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp., whereas the dataset from *Nyctotherus ovalis* was assembled from pooled data of three samples that were sequenced individually, also by Henning Onsbring Gustafson (Ettema-lab, Uppsala University). The final assembled transcriptomic datasets were assessed in order to estimate the numbers of unique transcripts from each that appeared to have been sequenced from Bacteria, Archaea and eukaryotes (including ciliates). This was done by searching each unique transcript from each dataset against the Nr database (NCBI) using blastx. A taxonomic classification was assigned to each transcript based on the taxonomy ID of the most similar protein from the Nr database (NCBI).

The results of this analysis are displayed in Table 3.5. The transcripts that were assigned as ciliate transcripts in Table 3.5. are a subset of those transcripts that were most similar to eukaryote proteins, rather than Bacteria or Archaea. From this it appears that there were a large number of transcripts in the datasets that came from eukaryotes but were not ciliates, but we think this is unlikely for the following reasons. Firstly, the cells in each sequenced sample were isolated from monoxenic cultures, in each of which only one ciliate species and no other eukaryotes were known to be present. Secondly, 18S rRNA and 28S rRNA sequences were recovered from *Nyctotherus ovalis, Metopus contortus*, *Metopu*s es, *Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. in each of their respective datasets but no other 18S rRNA or 28S rRNA sequences were recovered from any other eukaryotes.

Sequence data were obtained from three biological replicate samples of isolated *Nyctotherus ovalis* cells, which were pooled prior to data assembly. This may explain why there were a greater number of unique transcripts recovered for *Nyctotherus ovalis*, since only one sample of isolated cells was sequenced from

Metopus contortus, *Metopus* es, *Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* or *Trimyema* sp. The total number of unique transcripts recovered varied greatly for each species, but generally similar numbers of transcripts were recovered for species from the same taxonomic class. A large number (3915) of bacterial transcripts were recovered from the *Cyclidium porcatum* dataset. This is possibly due to the cells of *Cyclidium porcatum* being smaller (Section 3.3.1, Figure 3.1.) than the other species for which data was generated, which meant that they were more difficult to isolate and wash. *Trimyema* sp. is also smaller relative to most of the other ciliates investigated, however this species was easier to isolate than *Cyclidium porcatum*, as it moves more quickly, making it more visible to the eye.

The whole genome sequences of *Tetrahymena thermophila* and *Paramecium tetraurelia* are predicted to encode 26996 and 39580 proteins, respectively (Aury et al., 2006; Eisen et al., 2006). Therefore the numbers of eukaryote transcripts that were obtained from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. suggest that a large number of genes from the genomes of these species have not been sequenced in the transcriptome datasets.

	Total number of unique transcripts	Number of transcripts with significant blastx hits *b*	Percentage of transcripts with significant blastx hits (%) *b*	Bacterial transcripts	Bacterial transcripts (%) *a*	Archaeal transcripts	Archaeal transcripts (%) *a*	Eukaryotic transcripts	Eukaryotic transcripts (%) *a,d*	Ciliate transcripts *c*	Ciliate transcripts (%) *a*
Nyctotherus ovalis	47592	24672	51.8	2345	9.5	156	0.6	20802	84.3	10696	43.4
Metopus contortus	28796	16032	55.7	1249	7.8	435	2.7	13208	82.4	6597	41.1
Metopus es	28753	15939	55.4	1349	8.5	170	1.1	13160	82.6	7134	44.8
Metopus striatus	29350	17163	58.5	2480	14.4	960	5.6	12489	72.8	5588	32.6
Cyclidium porcatum	31143	10769	34.6	3915	36.4	150	1.4	6192	57.5	4495	41.7
Plagiopyla frontata	8872	5247	59.1	496	9.5	162	3.1	4272	81.4	2737	52.2
Trimyema sp.	14577	6944	47.6	1716	24.7	490	7.1	4227	60.9	2186	31.5

Table 3.5. Summary of the taxonomic affiliations of unique transcripts from assembled transcriptome datasets based on the results of blastx searches against the Nr database (NCBI) using each unique transcript as queries. The taxonomic affiliation of each transcript was assigned based on the taxonomic ID of its best (most significant) blast hit from the Nr database (NCBI) and was calculated using the software MEGAN5 (Huson et al., 2016). The percentage values displayed in the table from different taxonomic groups (*a*) are calculated from the total number of unique transcripts that had significant blastx hits (*b*), therefore any transcripts that did not have significant blastx hits to any proteins in the Nr database (NCBI) were not included in this total. The transcripts that hit ciliates as top hits in the Nr database (NCBI) (*c*) were a subset of the transcripts that hit eukaryotes as top hits in the Nr database (NCBI) (*d*). Transcriptome datasets were produced from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp.

3.3.15 Reconstructing the hydrogenosome metabolisms of anaerobic ciliates from sequence data

The hydrogenosome metabolisms of *Cyclidium porcatum*, *Metopus contortus* and *Plagiopyla frontata*, as representatives of the Oligohymenophorea, Armophorea and Plagiopylea, respectively, were reconstructed on the basis of the data generated in this thesis. The major metabolic processes and pathways that were investigated were glycolysis, pyruvate oxidation, substrate-level phosphorylation by ASCT and SCS, the TCA cycle, the ETC, and the Fe-S cluster biogenesis pathway. Subunits of the protein-importing TIM, TOM and MPP complexes, as well as MCF proteins, were also identified as these are important for the function of mitochondria and hence hydrogenosomes. Hydrogenosome metabolic reconstructions from *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata* are shown in Figures 3.16—3.18 and are discussed in the following sections. Although the dataset from *Nyctotherus ovalis* appeared to be the most complete, we did not carry out a detailed hydrogenosome metabolic reconstruction of this species as this has been done previously (de Graaf et al., 2011) and therefore species were prioritised for which little was previously known. In several cases, some genes that were expected to be found for certain pathways and complexes were not detected. With partial data it is not possible to be sure whether this reflects actual gene loss or missing data, and both of these possibilities should be kept in mind with regards to the results discussed below. Some proteins were predicted to have N-terminal mitochondrial-like targeting signals by prediction programs (summarised in Appendix B), providing additional evidence for them being localised to hydrogenosomes. A large number of proteins were not predicted to have N-terminal targeting signals by prediction programs but it is not possible to be sure whether this reflects their true absence, or whether the 5-prime end of the transcript was incompletely sequenced and therefore the N-terminal of the protein is missing.

Glycolysis

The genomes of well-studied aerobic ciliates such as *Tetrahymena* and *Paramecium* appear to encode nine of the ten classical enzymes of the glycolysis

pathway: GPI, PFK, ALDO, TPI, GAPDH, PGK, PGAM, ENO and PK. These nine enzymes were also detected from *Cyclidium porcatum*, *Metopus contortus* and *Plagiopyla frontata*. Genes for hexokinase, which is classically the first enzyme in the glycolysis pathway, were not detected in the genomes of *Tetrahymena* and *Paramecium* (Smith et al., 2007). Other eukaryotes that lack typical aerobic mitochondria, such as *Giardia intestinalis, Spironucleus barkhanus* and *Trichomonas vaginalis*, appear to use glucokinase (GCK) as the first enzyme in the glycolysis pathway instead of hexokinase (Henze et al., 2001) and GCK was detected in the macronuclear genome sequence of *Tetrahymena thermophila* (Eisen et al., 2006), suggesting that GCK may have also replaced hexokinase in ciliates. GCK was detected in the data for *Metopus contortus* and *Plagiopyla frontata* but not in the less complete set for *Cyclidium porcatum*.

Pyruvate metabolism

Pyruvate is typically oxidised to acetyl-CoA by PDH in aerobic eukaryotes, including ciliates such as *Tetrahymena thermophila* (Smith et al., 2007). Each of the four subunits of the PDH complex, PDH E1α, PDH E1β, PDH E2 and PDH E3, were detected from *Metopus contortus* and *Plagiopyla frontata.* Only one PDH subunit, PDH E1α, was detected from *Cyclidium porcatum*, which has a role in pyruvate-binding. Many anaerobes use the homologous proteins PFO and PNO, or PFL for pyruvate oxidation. Here PNO was detected from *Cyclidium porcatum* but not PFL. Neither PFO/PNO nor PFL were detected from *Metopus contortus* or *Plagiopyla frontata* indicating that in these species the PDH complex is the principal enzyme used for pyruvate oxidation.

The tricarboxylic acid (TCA) cycle

There are eight classical enzymes and enzymes complexes that function together as the TCA cycle, CS, ACO, IDH, OGDC, SCS, SDH, FH and MDH, and each of these were identified in *Tetrahymena thermophila* (Rivière et al., 2004). Of the TCA cycle enzymes neither *Cyclidium porcatum, Metopus contortus* nor *Plagiopyla frontata* appear to have the complete set found in *Tetrahymena*

thermophila. ACO, IDH, FH and MDH, as well as the E3 subunit of OGDC and the SdhA subunit of SDH were detected from *Cyclidium porcatum*; OGDC, SCS, SDH, FH and MDH were detected in *Metopus contortus*; and IDH, OGDC, SCS, FH and MDH were detected from *Plagiopyla frontata*. Seven of the TCA cycle proteins from *Metopus contortus* were strongly predicted as containing N-terminal targeting signals (Appendix B), whereas only one protein from *Cyclidium porcatum* was strongly predicted and one protein from *Plagiopyla frontata* was moderately predicted as containing an N-terminal targeting signal. This suggests that TCA cycle proteins from *Metopus contortus* are likely to be located within their hydrogenosomes, whereas the location of TCA cycle proteins in *Cyclidium porcatum* and *Plagiopyla frontata* is unclear. The only TCA cycle enzymes shared by all three of these species therefore are FH and MDH, whereas CS is absent in all three. Expression of CS in the rhizarian *Brevimastigamonas vehiculus* has been suggested to be O_2-dependant and not expressed in low O_2 conditions (Gawryluk et al., 2016), which could also explain why it was not detected from anaerobically cultured *Cyclidium porcatum, Metopus contortus* or *Plagiopyla frontata*, but it is possible to speculate that the remaining enzymes could be expressed during periods of oxygen exposure. Additionally the hydrogenosomes of some anaerobic protists are suggested to use a subset of the TCA cycle enzymes running in reverse, enabling the use of fumarate as an electron acceptor, producing succinate. This reaction can be catalysed by ETC Complex II, SDH, functioning as a fumarate reductase.

The electron transport chain

The *Tetrahymena thermophila* ETC features several components of the main complexes, ETC Complexes I, II, III and IV, and F_1F_0 ATP synthase, which are typically found in the aerobic mitochondria of most eukaryotes (Smith et al., 2007). Individual subunits from each of the main ETC complexes were detected from *Cyclidium porcatum*, whereas subunits of ETC Complexes I, II and III, but not from ETC Complex IV or F_1F_0 ATP synthase, were detected from *Metopus contortus*. *Plagiopyla frontata* appears to have completely lost the ETC as no subunits from any of the ETC complexes were detected in the genomic or transcriptomic datasets.

ETC Complex I

In aerobic eukaryotes, including *Tetrahymena thermophila* (Smith et al., 2007), ETC Complex I subunits are either encoded by the nuclear or mitochondrial genome. Only a small portion of the *Cyclidium porcatum* hydrogenosome genome has been sequenced, and genes for only three ETC Complex I subunits were identified of those which are typically encoded by mitochondrial genomes in ciliates. These are *nad7*, *nad9* and *nad10*, which along with nuclear-encoded *nad8* that was also detected from *Cyclidium porcatum*, are the core subunits of the Q-module of ETC Complex I. Out of the two core N-module subunits of ETC Complex I; the 24 kDa and 51 kDa subunits, which are nuclear encoded in aerobic ciliates, only the 51 kDa subunit was detected from *Cyclidium porcatum*. Two nuclear-encoded subunits involved in electron transfer in ETC Complex I, Nad11 and the 24 kDa subunit, were not detected from transcriptome analysis. These two subunits are thought to be required for transferring electrons from the NADH oxidising 51 kDa subunit to the Q-module (Brandt, 2006), which was detected from *Cyclidium porcatum*.

None of the core P-module subunits involved in proton translocation, Nad1, Nad2, Nad3, Nad4, Nad4L, Nad5 and Nad6, were identified in *Cyclidium porcatum* and it is therefore unclear whether ETC Complex I can translocate protons. *Cyclidium porcatum* appears to have several core F_1F_0 ATP synthase subunits (discussed below), which would require a proton gradient for ATP synthesis. Since only one core subunit could be detected from each of ETC Complexes III and IV, it may not be possible for these complexes to translocate protons across the inner hydrogenosome membrane, as they do in aerobic mitochondria. If this is the case then ETC Complex I might be entirely responsible for proton translocation across the inner hydrogenosome membrane in *Cyclidium porcatum*. This has also been suggested to be the case in hydrogenosomes that have retained a partial ETC in other organisms (Gawryluk et al., 2016).

More of the fourteen ETC Complex I core subunits were detected from *Metopus contortus* than *Cyclidium porcatum*, partially owing to the more complete hydrogenosome genome sequence that was obtained from this species, and

these include Nad1, Nad2, Nad3, Nad4, Nad5, Nad7, Nad8, Nad9, Nad10, Nad11 and the 24 kDa and 51 kDa subunits. The only core subunits that were not detected were Nad4L and Nad6, which are subunits of the proton pumping P-module. Since the majority of core subunits that make up the three core modules of ETC Complex I are present in *Metopus contortus*, it is likely that this complex is functioning in a similar way to its homologues in the mitochondria of aerobic ciliates and other organisms.

ETC Complex II

Of the two ETC Complex II catalytic subunits, both SdhA and SdhB were identified in *Metopus contortus*, whereas only SdhA was identified from *Cyclidium porcatum*. Similar findings were made from transcriptome analysis of hydrogenosomes from *Pygsuia biforma* (Stairs et al., 2014), from which SdhA was detected but not SdhB. This might indicate either that SdhB is not required for the function of ETC Complex II in some anaerobic eukaryotes or simply has just not been sequenced. The two hydrophobic anchor proteins SdhC and SdhD were not identified in *Metopus contortus* or *Cyclidium porcatum* but these proteins are typically not well conserved and were not identified from the analysis of mitochondria from the aerobic ciliate *Tetrahymena thermophila* either (Smith et al., 2007).

ETC Complex III

The conserved catalytic core of ETC Complex III is composed of three subunits, cytochrome *b,* cytochrome *c*1 and Rieske (Yang and Trumpower, 1986; Iwata et al., 1998). In *Tetrahymena thermophila*, cytochrome *b* is encoded by the mitochondrial genome (Brunk et al., 2003), whereas cytochrome *c*1 and the Fe-S protein, Rieske, are thought to be encoded by the macronuclear genome (Smith et al., 2007). In the present study, cytochrome *c* and Rieske protein were detected from *Metopus contortus* but not cytochrome *b.* Rieske protein was also detected from *Cyclidium porcatum*, but cytochrome *c*1 and cytochrome *b* were not.

ETC Complex IV

Of the thirteen ETC Complex IV protein subunits found in mammals (Kadenbach et al., 1983), seven were previously identified in *Tetrahymena thermophila* (Smith et al., 2007). Cox1 and Cox2 are encoded by the mitochondrial genome and Cox5b, Cox15, Cox19, Sco1 and Sco2 are encoded by the nuclear genome. In the present study no subunits of ETC complex IV were detected from *Metopus contortus* and only Cox15 was detected from *Cyclidium porcatum*. In yeast Cox15 is thought to have a role in the assembly of ETC Complex IV and it also synthesizes heme *a*, a cofactor that is inserted into Cox1 in the assembled complex (Bareth et al., 2013).

F_1F_0 ATP synthase

The mitochondrial F_1F_0 ATP synthase complex is made up of two main components. The F_0 component forms a channel that allows the movement of protons across the inner membrane, which drives ATP synthesis by the F_1 component (reviewed in Junge and Nelson (2015)). The F_1F_0 ATP synthase complex of *Tetrahymena thermophila* has been studied in detail and appears to be highly divergent, containing several proteins for which no homologues can be identified in other organisms (Nina et al., 2010). One of these proteins, Ymf66, is encoded by the mitochondrial genome of *Tetrahymena thermophila*. A transcript encoding Ymf66 was detected from *Cyclidium porcatum* and is therefore likely to be encoded by the hydrogenosome genome from this species. A small number of well conserved eukaryotic F_1F_0 ATP synthase proteins were identified in *Tetrahymena thermophila* however, including ATPα, ATPβ and ATPγ subunits of the F_1 component and ATPδ, ATP9 and ATP12 of the F_0 component (Nina et al., 2010). Each of these subunits are encoded by the macronuclear genome in *Tetrahymena thermophila,* except ATP9 which is encoded by the mitochondrial genome. Subunits ATPα, ATPβ, ATPγ and ATPδ were detected from *Cyclidium porcatum* in the present study but ATP9 and ATP12 were not detected. No F_1F_0 ATP synthase subunits were detected from *Metopus contortus*, which indicates that the F_1F_0 ATP synthase complex may have been lost from this species.

Rhodoquinone, fumarate reduction and alternative oxidase (AOX)

The replacement of ubiquinone with rhodoquinone in the ETC is thought to be a necessary requirement in order for the fumarate reduction pathway to function in the hydrogenosomes of some anaerobic eukaryotes (Tielens and Van Hellemond, 1998). The anaerobic eukaryotes *Pygsuia biforma* (Stairs et al., 2014) and *Brevimastigamonas vehiculus* (Gawryluk et al., 2016) most likely use rhodoquinone since the gene *rquA* was detected from both of these species. *rquA* was first discovered in *Rhodospirillum rubrum* and is currently the only known gene that is thought to be required for the biosynthesis of rhodoquinone from ubiquinone (Lonjers et al., 2012). *rquA* was not detected in the present study from *Cyclidium porcatum* or *Metopus contortus*, which is somewhat surprising given that both these species appear to have a partial ETC, which could possibly benefit from the use of fumarate as an electron sink. Interestingly, this gene appears to have been acquired by lateral gene transfer in *Pygsuia biforma* (Stairs et al., 2014) and *Brevimastigamonas vehiculus* (Gawryluk et al., 2016), therefore *Cyclidium porcatum* and *Metopus contortus* might not have acquired it, or it was not detected in our screen for ciliate genes.

The main enzymes required for fumarate reduction are the TCA cycle enzymes FH and SDH. Both of these enzymes were identified from *Nyctotherus ovalis* previously (de Graaf et al., 2011) and *Nyctotherus ovalis* has also been shown to excrete succinate, the end product of fumarate reduction (Boxma et al., 2005). These findings are consistent with the hypothesis that fumarate reduction occurs in the hydrogenosomes of *Nyctotherus ovalis*. Succinate was also detected from the ciliate *Trimyema compressum* but in very low quantities (Goosen et al., 1990; Holler and Pfennig, 1991). FH and SDH were also detected from *Cyclidium porcatum* and *Metopus contortus*, with the exception of the SdhB subunit which was not detected in *Cyclidium porcatum*, indicating that these species have at least some of the necessary enzymes required for fumarate reduction. Another potential electron sink for the partial ETC of *Cyclidium porcatum* and *Metopus contortus* could be AOX, which was detected in both of these species. AOX typically uses O_2 as a substrate for ubiquinol oxidation in aerobic eukaryotes but has also been detected in some anaerobic eukaryotes, including *Blastocystis* sp.

(Stechmann et al., 2008), *Pygsuia biforma* (Stairs et al., 2014) and *Cantina marsupialis* (Noguchi et al., 2015).

Substrate phosphorylation by acetate:succinate CoA-transferase (ASCT) and succinyl-CoA synthetase (SCS)

ASCT and SCS are the enzymes used to produce ATP from acetyl-CoA via substrate-level phosphorylation, the typical energy-generating pathway in hydrogenosomes (Müller and Lindmark, 1978). Both ASCT and SCS were detected from *Metopus contortus* and *Plagiopyla frontata*, suggesting that this pathway could be used by both of these species use to produce energy. Only ASCT and not SCS was detected from *Cyclidium porcatum*. Since SCS has important roles in the TCA cycle, as well as substrate-level phosphorylation, it is most likely that failure to detect this protein in *Cyclidium porcatum* means that it was not sequenced in the present study, rather than it being absent from this species. However, given that *Cyclidium porcatum* appears to have F_1F_0 ATP synthase, it could be possible that it does not require SCS for ATP production and therefore does not produce energy by substrate-level phosphorylation. The ASCT enzymes detected from *Cyclidium porcatum*, *Metopus* contortus, and *Plagiopyla frontata* each belong to the sub-family 1A of these enzymes, which are homologues of the succinyl-CoA:3-ketoacid CoA-transferase enzymes that are found in aerobic eukaryotes (Rivière et al., 2004; Tielens et al., 2010).

The mitochondrial Fe-S cluster biogenesis (ISC) pathway

The ISC pathway is an essential pathway of yeast mitochondria and is used for the maturation of functional Fe-S proteins (Lill, 2009). The core ISC proteins required for assembly of [2Fe-2S] clusters are Nfs1 and Isd11 (cysteine desulfurase complex), Yah1 (ferredoxin), Arh1 (ferredoxin reductase), Yfh1 (frataxin), and the scaffold protein Isu1 (Freibert et al., 2017). The proteins Ssq1 (mtHsp70), Mge1 and Jac1, transfer [2Fe-2S] clusters from Isu1 to Grx5 (glutaredoxin), which inserts them into apoproteins, forming functional [2Fe-2S] proteins, including Yah1 (ferredoxin) (Dutkiewicz et al., 2003; Freibert et al., 2017).

Of these core proteins, Nfs1, Isu1, Yfh1, Yah1, Arh1, Grx5, Ssq1, Jac1 and Mge1 have previously been identified in *Tetrahymena thermophila* (Smith et al., 2007). In the present study, Nfs1, Isu1, Yfh1, Yah1, Arh1, Ssq1, Jac1, Mge1 and Grx5 were detected from *Cyclidium porcatum* and seven of these proteins were predicted as containing N-terminal targeting signals; Isu1, Isd11, Nfs1, Yfh1, Yah1, Jac1, Ssq1 and Mge1 were detected from *Metopus contortus* and eight of these proteins were predicted as containing N-terminal targeting signals; and Isu1, Nfs1 and Ssq1 were detected from *Plagiopyla frontata* and only one of these proteins was weakly predicted as containing N-terminal targeting signals. Although some core ISC components were not detected from each of these species, the presence of other ISC components is consistent with their hydrogenosomes using this pathway to assemble [2Fe-2S] proteins in a similar way to the mitochondria of their aerobic relatives, such as *Tetrahymena thermophila* (Smith et al., 2007).

A secondary stage of the ISC pathway assembles [4Fe-4S] proteins and in yeast involves Isa1, Isa2, Iba57, Nfu1, Bol1, Bol3 and Ind1 (Melber et al., 2016; Uzarska et al., 2016; Freibert et al., 2017), of which Nfu1 and Isa1 were previously identified in *Tetrahymena thermophila* (Smith et al., 2007). In the present study Isa1, Iba57, Nfu1 and Ind1 were detected from *Cyclidium porcatum*; Isa1, Nfu1 and Ind1 were detected from *Metopus contortus*; and Nfu1 was detected from *Plagiopyla frontata*. The presence of these components indicates that *Cyclidium porcatum*, *Metopus contortus* and possibly *Plagiopyla frontata* can produce [4Fe-4S] clusters using the ISC pathway. The detection of Fe-S cluster proteins from these ciliates, including FeFe-hydrogenase (Discussed in detail in Chapter 4.), is consistent with the detection of components of the Fe-S cluster biogenesis pathway. In ciliates FeFe-hydrogenases contain one [2Fe-2S] cluster- and multiple [4Fe-4S] cluster-binding domains (Akhmanova et al., 1998; Horner et al., 2000) and presumably require the Fe-S cluster biogenesis pathway for the insertion of these clusters, in order to make them functional.

Hydrogenosome import mechanisms

The translocase of the outer membrane (TOM) complex

Import of proteins across the outer mitochondrial membrane is achieved via the TOM complex, which consists of core proteins that form the translocation channel, Tom70, Tom40, Tom22 and Tom7, and four additional subunits that assist in protein recognition, binding and transfer, Tom70, Tom20, Tom6 and Tom5 (Dolezal et al., 2006). In *Tetrahymena thermophila* only the three core proteins, Tom40, Tom22 and Tom7 were previously identified (Smith et al., 2007). Some of the proteins detected in the present study that are thought to function in the hydrogenosomes of *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata* have N-terminal mitochondrial-type targeting signals (Appendix B). This indicates that a classical mitochondrial TOM complex is present in the hydrogenosomes of these species, particularly the Tom20 receptor component, which recognises N-terminal targeting signals (Söllner et al., 1989). In the present study a Tom40-like porin was detected from *Cyclidium porcatum* and *Metopus contortus,* which is likely to facilitate the translocation of proteins across the outer hydrogenosome membrane but no TOM complex proteins were detected from *Plagiopyla frontata*.

The translocase of the inner membrane (TIM) complexes: TIM22, TIM23 and tiny Tim proteins

The TIM22 complex appears not to be well conserved in ciliates, and only the Tim22 and Tim54 subunits can be identified from the *Tetrahymena thermophila* genome sequence (Eisen et al., 2006). Neither of these proteins were detected in *Cyclidium porcatum*, *Metopus contortus* or *Plagiopyla frontata*. Eight of the known TIM23 complex proteins were identified previously in *Tetrahymena thermophila* (Smith et al., 2007), these are Tim50, Tim44, Tim23, Tim17, Tim16 (Pam16), Tim14 (Pam18), mtHsp70 and Mge1. Four of these TIM23 complex proteins were detected from *Cyclidium porcatum*, these were Tim17, Tim14 (Pam18), mtHsp70 and Mge1; five were detected from *Metopus contortus*, these were Tim17, Tim16 (Pam16), Tim14 (Pam18), mtHsp70 and Mge1; and just one was detected from

Plagiopyla frontata, which was mtHsp70. Three tiny Tim proteins, which shuttle substrates from the TOM complex to the TIM22 complex, were previously identified from *Tetrahymena thermophila*, these are Tim8, Tim9 and Tim10. No tiny Tim proteins however, were detected from *Cyclidium porcatum, Metopus contortus* or *Plagiopyla frontata*.

The mitochondrial processing peptidase (MPP) complex and the mitochondrial inner membrane protease (IMP) complex

The MPP complex consists of two separate subunits, Masα (Mas1) and Masβ (Mas2) and the IMP complex also consists of two proteins, Imp1 and Imp2 (Dolezal et al., 2006). The MPP complex cleaves N-terminal targeting presequences from mitochondrial proteins once they reach the mitochondrial matrix (Gakh et al., 2002; Hoogenraad et al., 2002). A subset of mitochondrial proteins are targeted to the inner mitochondrial membrane by a secondary downstream N-terminal targeting signal, which is exposed once the first targeting signal is cleaved by the MPP complex (Ieva et al., 2013). The secondary targeting signal in these proteins is recognised by the TIM23 complex, which guides the protein into the inner mitochondrial membrane, where the targeting signal is then cleaved by the IMP complex (Ieva et al., 2013).

Both MPP complex proteins, Masα (Mas1) and Masβ (Mas2), were identified in *Tetrahymena thermophila* previously (Smith et al., 2007) and both were also detected in the present study from *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata*. Only Imp1 of the IMP complex has previously been identified from *Tetrahymena thermophila* and not Imp2. In the present study Imp1 was detected only from *Metopus contortus* and *Plagiopyla frontata* but not from *Cyclidium porcatum*, and Imp2 was not detected in any of these species. The presence of MPP and IMP subunits from *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata* is consistent with the prediction that some of the purported hydrogenosome proteins from these species contain N-terminal targeting sequences (Appendix B), which are presumably processed by these complexes.

Mitochondrial carrier family (MCF) proteins

MCF proteins transport different substrates needed or produced by mitochondria across the inner mitochondrial membrane (Ferramosca and Zara, 2013). Fifty-three MCF proteins were previously identified from *Tetrahymena thermophila* (Smith et al., 2007). In the present study nineteen MCF proteins were detected from *Cyclidium porcatum*, fifteen were detected from *Metopus contortus* and six were detected from *Plagiopyla frontata*. These MCF proteins and their predicted substrate specificities are listed in Figures 3.16—3.18. The number of MCF proteins in these ciliate species is correlated with their metabolisms; the greatest number of MCF proteins were predicted from *Cyclidium porcatum*, which seems to have the most complete hydrogenosome metabolic pathways, whereas *Plagiopyla frontata* has the least number of MCF proteins and appears to have the most reduced hydrogenosome metabolism. Of course it is difficult to know whether this truly reflects the real nature of these hydrogenosomes, since *Plagiopyla frontata* is also the species that seems to have the least complete dataset. The number of MCF proteins that were predicted for these ciliates is similar to the number predicted for other organisms with hydrogenosomes. Thirteen MCF proteins were predicted from *Brevimastigamonas vehiculus* (Gawryluk et al., 2016) and twenty two were predicted from *Pygsuia biforma (Stairs et al., 2014)*. Notably, ATP/ADP translocases were identified from *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata*, indicating that ATP and ADP is likely exchanged between the hydrogenosomes and the cytosol of these species.

The oxidase assembly (OXA) translocase complex

The OXA complex mediates the insertion of proteins into the inner mitochondrial membrane and consists of four proteins, Oxa1, Mdm38, Mba1 and Y1h47 (Dolezal et al., 2006; Bohnert et al., 2010). Of these, only Oxa1 and Mdm38 were previously detected in *Tetrahymena thermophila* (Smith et al., 2007). In yeast, Mdm38 interacts with mitochondrial ribosomes and exports proteins across the inner mitochondrial membrane (Frazier et al., 2006). Mdm38 was detected from *Metopus contortus*, consistent with ribosomal proteins and rRNA genes subunits being detected from the macronuclear and hydrogenosome genome of this

species, but was not detected from *Cyclidium porcatum* or *Plagiopyla frontata.* Neither of the components Oxa1, Mba1 nor Y1h47 were detected from *Cyclidium porcatum, Metopus contortus* or *Plagiopyla frontata*.

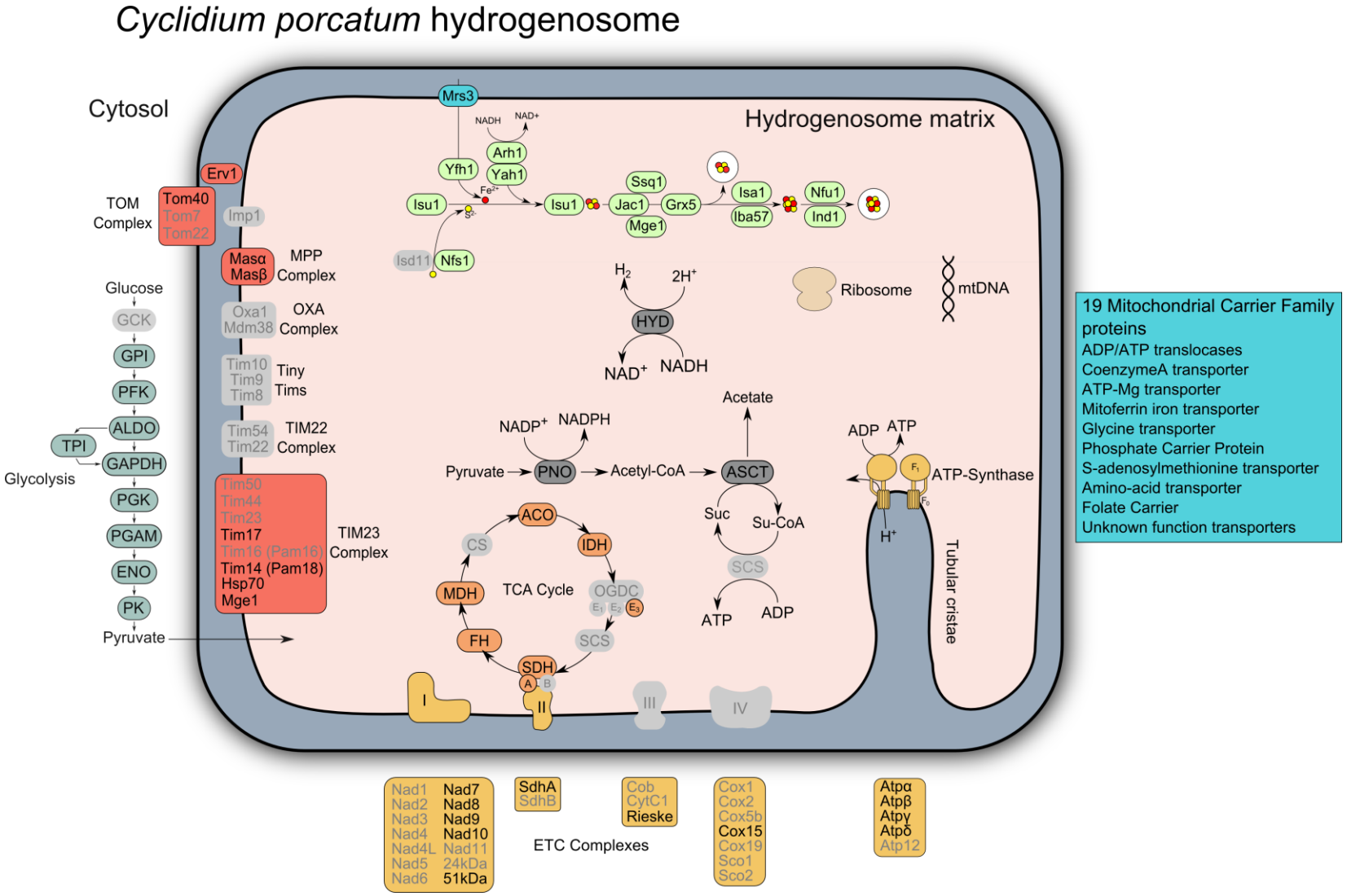


Figure 3.16. Metabolic reconstruction of the hydrogenosomes of *Cyclidium porcatum* based on analyses of genomic and transcriptomic datasets. Grey features without black outlines indicate components that were not identified but are present in mitochondria of the aerobic ciliate *Tetrahymena thermophila* (Smith et al., 2007).

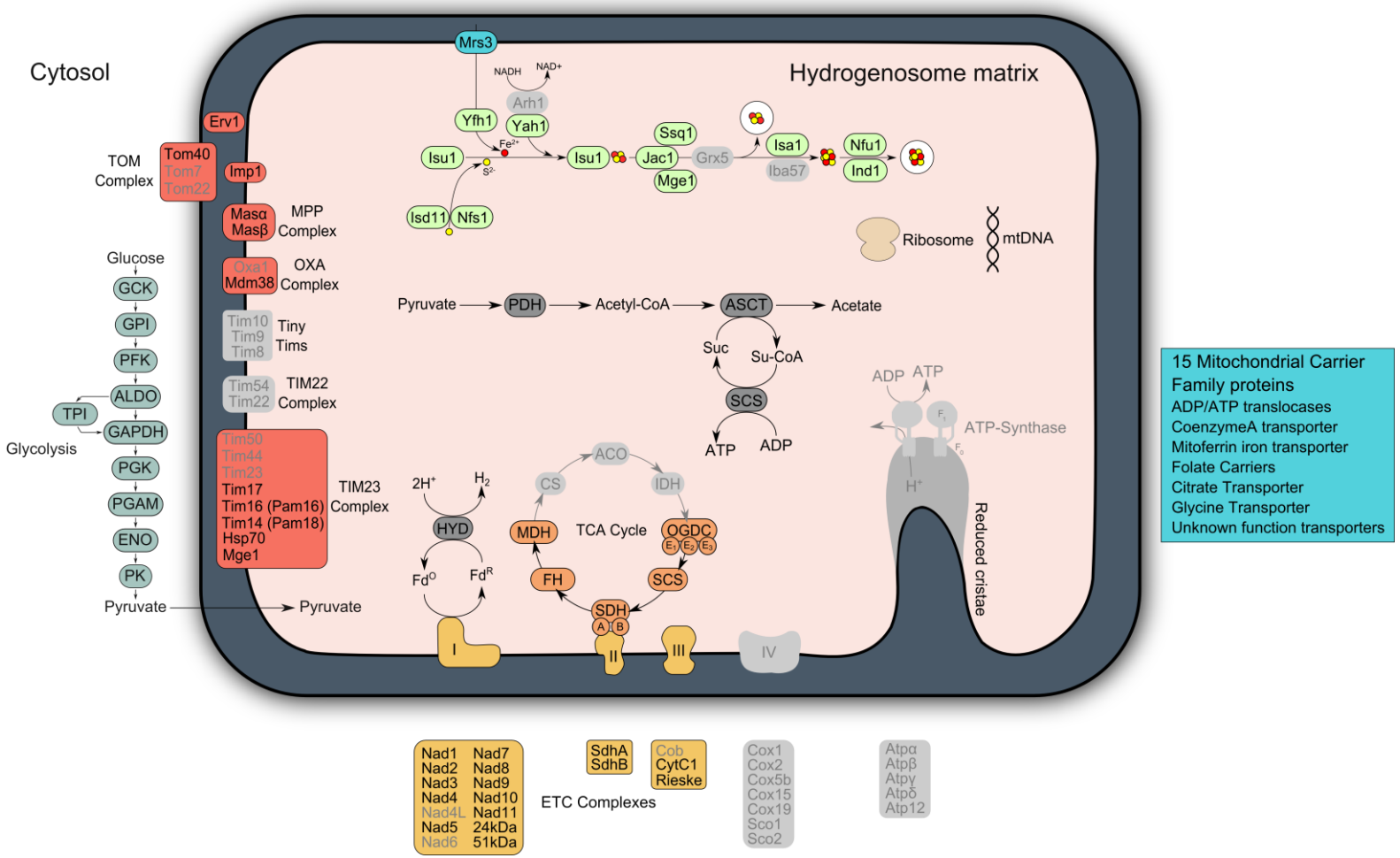


Figure 3.17. Metabolic reconstruction of the hydrogenosomes of *Metopus contortus* based on analyses of genomic and transcriptomic datasets. Grey features without black outlines indicate components that were not identified but are present in mitochondria of the aerobic ciliate *Tetrahymena thermophila* (Smith et al., 2007).

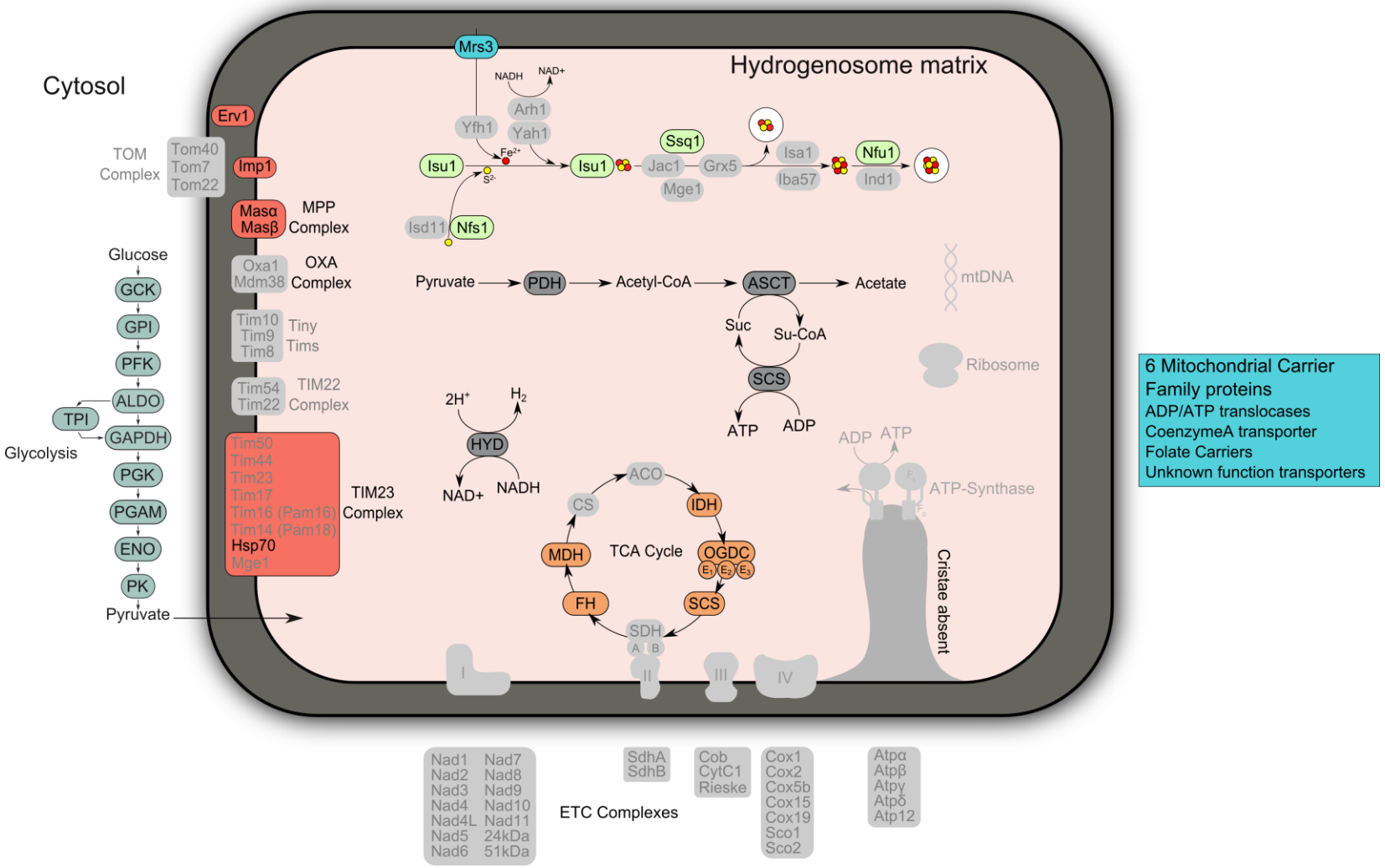


Figure 3.18. Metabolic reconstruction of the hydrogenosomes of *Plagiopyla frontata* based on analyses of genomic and transcriptomic datasets. Grey features without black outlines indicate components that were not identified but are present in mitochondria of the aerobic ciliate *Tetrahymena thermophila* (Smith et al., 2007).

3.3.16 Testing the completeness of transcriptome datasets based on Fe-S cluster biogenesis and glycolysis

The number of unique transcripts that were sequenced from *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata* were lower than the number of genes encoded by the macronuclear genomes of *Tetrahymena thermophila* and *Paramecium tetraurelia* (Discussed in Section 3.3.14). It is therefore likely that the transcriptome datasets from these three ciliates are incomplete representations of the ciliate nuclear genomes. To investigate this further the completeness of two important eukaryotic pathways - for mitochondrial Fe-S cluster biogenesis and for cytosolic glycolysis - was assessed for each ciliate.

All eukaryotes contain essential cytosolic and nuclear Fe-S proteins including DNA polymerase and Rli1 that depend on the mitochondrial Fe-S cluster biogenesis pathway for their maturation (Freibert et al., 2017). This suggests that a functional version of this pathway is needed by all of the studied ciliates. The Fe-S cluster biogenesis pathways of microsporidia appears to be the most reduced in eukaryotes, consisting of the proteins Isd11, Nfs1, Yfh1, Isu1, Yah1, Arh1, mtHsp70 and Jac1 (Freibert et al., 2017). This provides a model for the minimum set of mitochondrial proteins that is required to make Fe-S clusters and support cytosolic Fe-S protein biogenesis. An almost complete set of these proteins was identified for *Cyclidium porcatum* (missing Isd11 - which is very short) and *Metopus contortus* (missing Arh1). By contrast the pathway for *Plagiopyla frontata* was much less complete (missing Isd11, Yfh1, Yah1, Arh1 and Jac1). These findings suggest that the sequence data for these ciliates are incomplete to varying degrees but also suggest that coverage is not so poor that some useful insights into their metabolism cannot be made. This is supported by the detection of a complete glycolytic pathway for *Metopus contortus* and *Plagiopyla frontata*, and a pathway missing only one component, GCK, for *Cyclidium porcatum*. Based upon these data the glycolysis pathway in these three hydrogenosomal ciliates is the same as previously reported for the aerobic ciliate *Tetrahymena thermophila* (Smith et al., 2007).

In some anaerobic microbial eukaryotes, the mitochondrial Fe-S cluster biogenesis pathway has been replaced by alternative pathways through lateral gene transfer. These include an archaeal sulphur mobilisation (SUF) system in

Pygsuia biforma (Stairs et al., 2015) and a bacterial nitrogen fixation (NIF) system in *Mastigamoeba balamuthi* (Nývltová et al., 2013). We searched for these proteins in the datasets from *Cyclidium porcatum*, *Metopus contortus* and *Plagiopyla frontata* but found no evidence that these proteins have replaced the mitochondrial pathway in ciliates. Moreover, several proteins of the Fe-S cluster biogenesis pathway identified from *Cyclidium porcatum* and *Metopus contortus* were predicted to contain N-terminal targeting signals (Appendix B), consistent with their location inside the hydrogenosomes of these species.

3.3.17 Morphology of hydrogenosomes

The morphology of hydrogenosomes from *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata* were investigated using TEM, performed by Benoît Zuber and Beat Haenni (Microscopy Imaging Centre, Institute of Anatomy, University of Bern, Switzerland) (Figure 3.19). This was investigated in order to facilitate comparisons with aerobic mitochondria and understand how the changes in hydrogenosome metabolism influenced aspects of their morphology. Cristae, a hallmark of aerobic mitochondria, were observed within hydrogenosomes from *Cyclidium porcatum* and *Metopus contortus*. These observations are consistent with previous published images of hydrogenosomes from these species (Embley and Finlay, 1994). The cristae in the hydrogenosomes of *Cyclidium porcatum* appear to be widely distributed throughout the matrix of the organelle, whereas the cristae in the hydrogenosomes of *Metopus contortus* are less obvious and there appear to be far fewer in each hydrogenosome. Cristae morphology can differ dramatically between mitochondrial homologues of various organisms (Zick et al., 2009) and it is therefore not too surprising that the morphology of cristae in the hydrogenosomes of *Cyclidium porcatum* and *Metopus contortus* also differ in appearance from one another .

No cristae were observed in the hydrogenosomes of *Plagiopyla frontata* and this is consistent with other published images in which the hydrogenosomes of this species do not appear to have cristae (Fenchel and Finlay, 1991b; Fenchel and Finlay, 1995). Other well studied hydrogenosomes, in species such as *Trichomonas vaginalis*, also do not contain cristae (Bradley et al., 1997). Studies

have previously indicated that hydrogenosomes of *Nyctotherus ovalis* have cristae (Gijzen et al., 1991) and these seem to be similar in appearance to the cristae of the hydrogenosomes found in *Metopus contortus* shown in Figure 3.19, b. and published previously (Finlay and Fenchel, 1989). These two species are closely related (Figure 3.4), indicating that this feature has been conserved from their common ancestors which is predicted to have already had hydrogenosomes.

The structure of F_1F_0 ATP synthase dimers were described in two species of ciliate: *Tetrahymena thermophila* (Nina et al., 2010) and *Paramecium tetraurelia* (Mühleip et al., 2016). In these species the two monomers are arranged in parallel forming U-shaped dimers, whereas in other species the monomers have a more angular arrangement, as such they form V-shaped dimers (Chaban et al., 2014). This structural difference is thought to be due to F_1F_0 ATP synthase complexes being highly divergent in ciliates, particularly the F_0 sub-complex, which typically mediates dimer formation (Nina et al., 2010; Chaban et al., 2014). Nina et al. (2010) identified 13 novel subunits in *Tetrahymena thermophila* for which they could not discern any orthologues in any organisms other than ciliates. U-shaped dimers are thought to form helical tubular cristae (Mühleip et al., 2016), whereas species with V-shaped dimers, such as yeast and mammals, typically have lamellar cristae (Strauss et al., 2008; Davies et al., 2011; Davies et al., 2012). The hydrogenosomes of *Cyclidium porcatum* appear to have more defined cristae (Figure 3.19, a.) and appear to be similar to the cristae observed from TEM images in mitochondria of *Paramecium tetraurelia* and *Tetrahymena thermophila*, which have also been studied in detail (Nina et al., 2010; Mühleip et al., 2016). Since *Cyclidium porcatum*, *Paramecium tetraurelia* and *Tetrahymena thermophila* all belong to the Oligohymenophorea (Figure 3.4) and they appear to have a somewhat conserved cristae structure, it would be reasonable to predict that the F_1F_0 ATP synthase complexes of *Cyclidium porcatum* also form U-shaped dimers, but further research would be necessary to investigate whether this is the case.

Dimers are thought to be formed by protein-protein interactions between monomer complexes but exactly which protein subunits are involved in this process appears to differ between species (Chaban et al., 2014). In yeast F_1F_0 ATP synthase subunits *a* and *e* are thought to have key roles in dimer formation (Paumard et al., 2002; Arselin et al., 2004). TEM data indicates that cristae in

mitochondria of yeast mutants devoid of these two subunits form concentric layers of membranes, described as onion-like, much different in appearance to the tubular shape cristae of the wildtype (Paumard et al., 2002; Arselin et al., 2004). The cristae observed in *Metopus contortus* hydrogenosomes (Figure 3.19, b.) are arguably more similar in appearance to these onion-like cristae, forming a membranous layer, parallel to the outer membrane of the organelle, whereas the cristae in *Cyclidium porcatum* appear to have a more typical tubular shape (Figure 3.19, a.). This is consistent with the finding that *Cyclidium porcatum* has retained F_1F_0 ATP synthase complex, whereas no evidence for this complex was found in *Metopus contortus*.

Another protein complex that is thought to have a role in cristae formation is the MICOS complex (John et al., 2005). The MICOS complex was most likely acquired from the mitochondrial endosymbiont as some subunits of this complex were discovered in alpha-proteobacteria where they are thought to perform a similar function in development of intracytoplasmic membranes that resemble cristae (Muñoz-Gómez et al., 2015). Eight subunits of the MICOS complex, Mic10, Mic12, Mic13 (QIL1), Mic19, Mic23, Mic25, Mic27 and Mic60, were described in Opisthokonts (Zerbes et al., 2012; Guarani et al., 2015) and at least three of these, Mic10, Mic19 and Mic60 have also appear to be present in other eukaryotic lineages (Muñoz-Gómez et al., 2015). In ciliates however, of these subunits, only Mic10 has previously been identified from *Paramecium tetraurelia, Tetrahymena thermophila* and *Oxytricha trifallax*, which were found bioinformatically by searching sequences from publically available databases by two separate studies (Muñoz-Gómez et al., 2015; Huynen et al., 2016). The proteins identified as being Mic10 from ciliates by Muñoz-Gómez et al. (2015) however, do not appear to be homologues of the proteins identified as Mic10 in ciliates by Huynen et al. (2016). Knowing which of the two studies, if either, has most likely identified authentic Mic10 proteins is difficult due to their short lengths and low sequence identity to other proteins. That being said, probable homologues of the ciliate Mic10 proteins identified by Muñoz-Gómez et al. (2015) were detected in the present study from *Metopus contortus*, *Metopus es* and *Metopus striatus* but no proteins with significant sequence similarity to the Mic10 proteins identified by Huynen et al. (2016) could be detected from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* or *Trimyema* sp.

The presence of the purported Mic10 protein detected from *Metopus contortus* could explain why some reduced cristae appear to be present within the hydrogenosomes of this species despite the probable absence of F_1F_0 ATP synthase (Figure 3.19, b.). Given that the hydrogenosomes of *Cyclidium porcatum* appear to have more defined cristae than *Metopus contortus* however, it could be expected that homologues of this purported Mic10 protein would be detected in *Cyclidium porcatum* too but they were not detected in the limited available data.

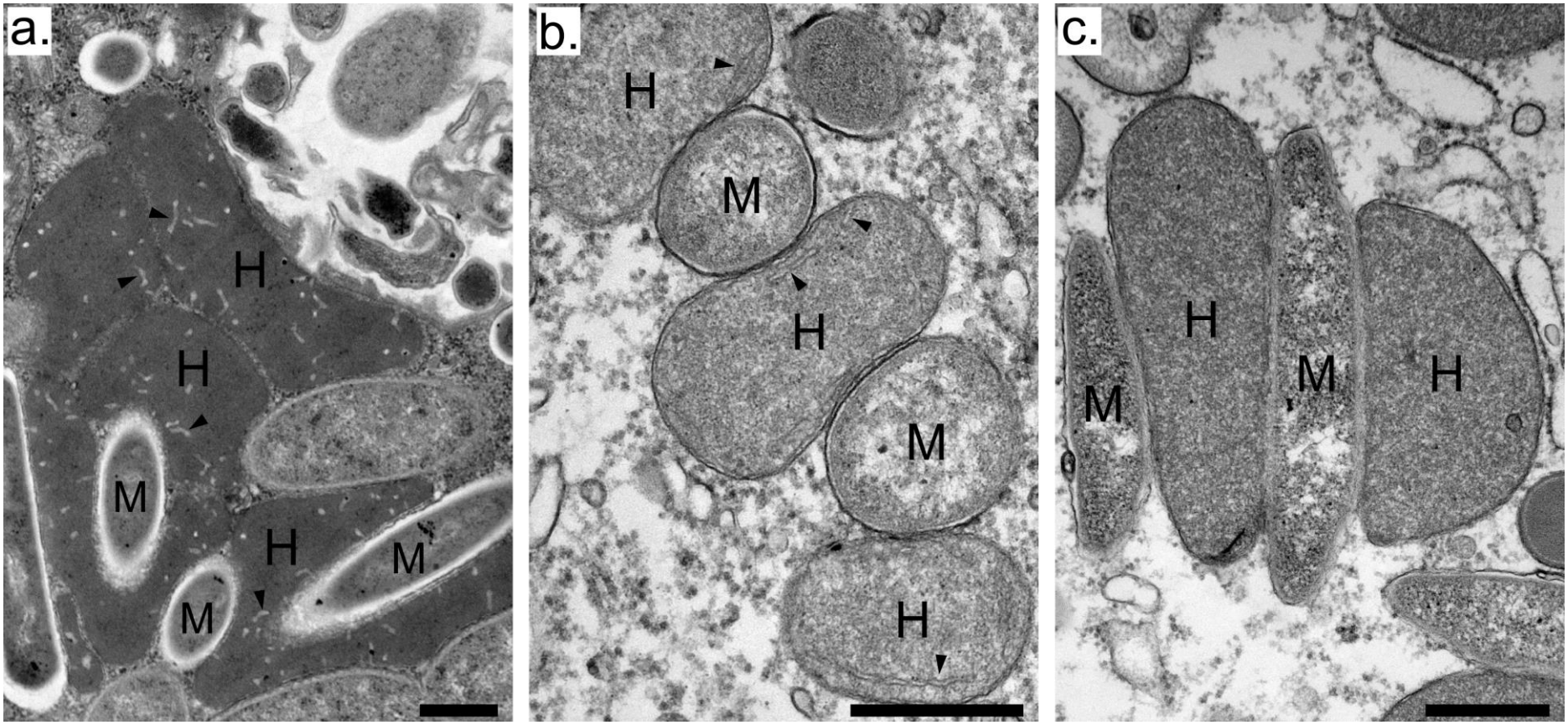


Figure 3.19. Transmission electron micrographs of samples prepared from cells of *Cyclidium porcatum* (a), *Metopus contortus* (b) and *Plagiopyla frontata* (c). Hydrogenosomes (H). Methanogens (M). Cristae (arrowheads). Scale bars represent 500nm.

3.3.18 Identification of endosymbionts in anaerobic ciliates using F420 auto-fluorescence

Intracellular methanogens were tentatively identified, based on their emissions of F420 auto-fluorescence, inside of *Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. (Figure 3.20). Intracellular methanogens have also been demonstrated in *Nyctotherus ovalis* previously (Gijzen et al., 1991). The endosymbionts within cells of *Metopus es, Metopus striatus* and *Cyclidium porcatum* appeared rod-shaped and the endosymbionts of *Metopus contortus, Plagiopyla frontata* and *Trimyema* sp. were more irregularly shaped and did not appear as rods. The presence of these methanogens is taken to be a reliable indicator for H_2 being produced within these ciliates (Fenchel and Finlay, 1992). The methanogenic endosymbionts in *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata* appear to be closely associated with hydrogenosomes (Figure 3.19). This is most likely to maximise their consumption of the H_2 being produced by the hydrogenosomes as hydrogenotrophic methanogens require H_2 to provide electrons, transferred via coenzyme F_{420}, for reduction of CO_2 to CH_4 and H_2O (Ferry, 1992). Previous evidence from electron microscopy data suggests that the cell walls of methanogen endosymbionts in *Plagiopyla frontata, Trimyema* sp. and *Metopus contortus* isolates are modified or reduced and they associate intimately with hydrogenosomes to the point at which their membranes appear fused, presumably to better facilitate transfer of H_2 (Fenchel and Finlay, 1991c; Finlay and Fenchel, 1991; Finlay et al., 1993a). It is unclear from Figure 3.19 whether such fusions occur between hydrogenosomes and endosymbionts in *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata* isolated in the present study, although their close proximity is apparent.

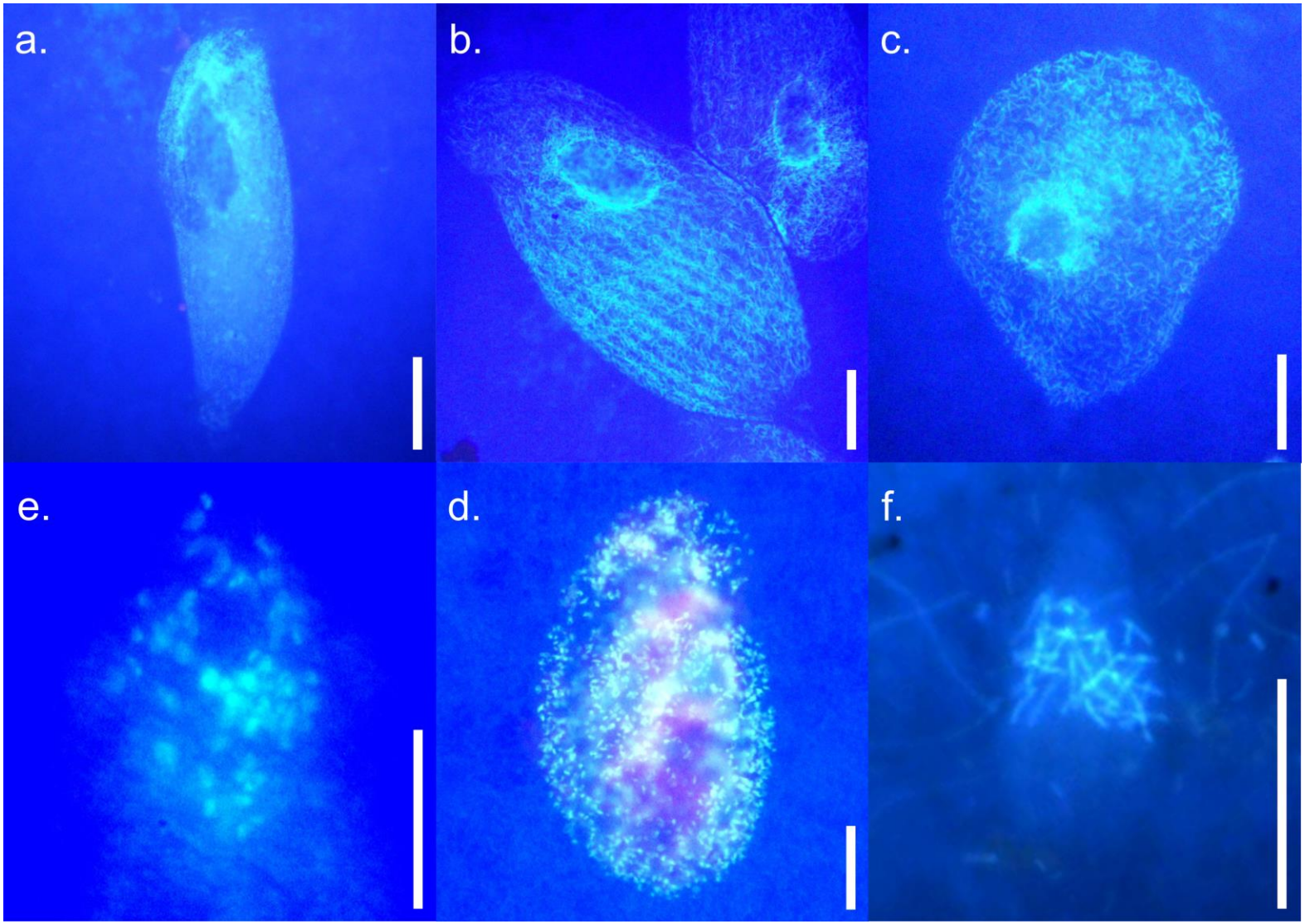


Figure 3.20. F420 auto-fluorescence imaged from methanogen endosymbionts in anaerobic ciliates: *Metopus contortus* (a), *Metopus es* (b), *Metopus striatus* (c), *Trimyema* sp. (d), *Plagiopyla frontata* (e), *Cyclidium porcatum* (f). Methanogens in the figures fluoresce with a relatively high intensity of blue/green emitted light. Scale bars represent 20µm.

3.3.19 Identification of methanogen species by fluorescence in situ hybridisation

Molecular fluorescent probes targeting the specific regions of rRNA from particular methanogen species were used to identify the methanogen species that were endosymbionts of *Metopus contortus* (Figure 3.21)*, Nyctotherus ovalis* (Figure 3.22) and *Trimyema* sp. (Figure 3.23). Based on these experiments, the endosymbionts of *Nyctotherus ovalis* are identified here as being closely related to *Methanobrevibacter arboriphilus*. This is consistent with predictions made earlier based on morphological comparisons (Gijzen et al., 1991) and fluorescent probing (van Hoek et al., 2000b). The endosymbionts of *Metopus contortus* and *Trimyema* sp. are closely related to *Methanocorpusculum labreanum*. This is consistent with the endosymbiont species that were identified from *Metopus contortus* and *Trimyema* sp. in studies previously (Embley et al., 1992a; Finlay et al., 1993b) In these earlier studies, *Metopus contortus* was isolated from marine sands in Denmark (Finlay and Fenchel, 1989; Embley et al., 1992a) and *Trimyema* sp. was isolated from a freshwater pond in Cumbria (UK) (Finlay et al., 1993b), whereas in the present study these species were isolated from locations in Dorset (UK) (methods described in Section 2.1). Given that these species were isolated from geographical locations that are separated by large distances, and more than two decades apart, but are shown to contain the same species of endosymbiont, it seems that the relationships between these ciliates and their endosymbionts are stable in the short to medium term.

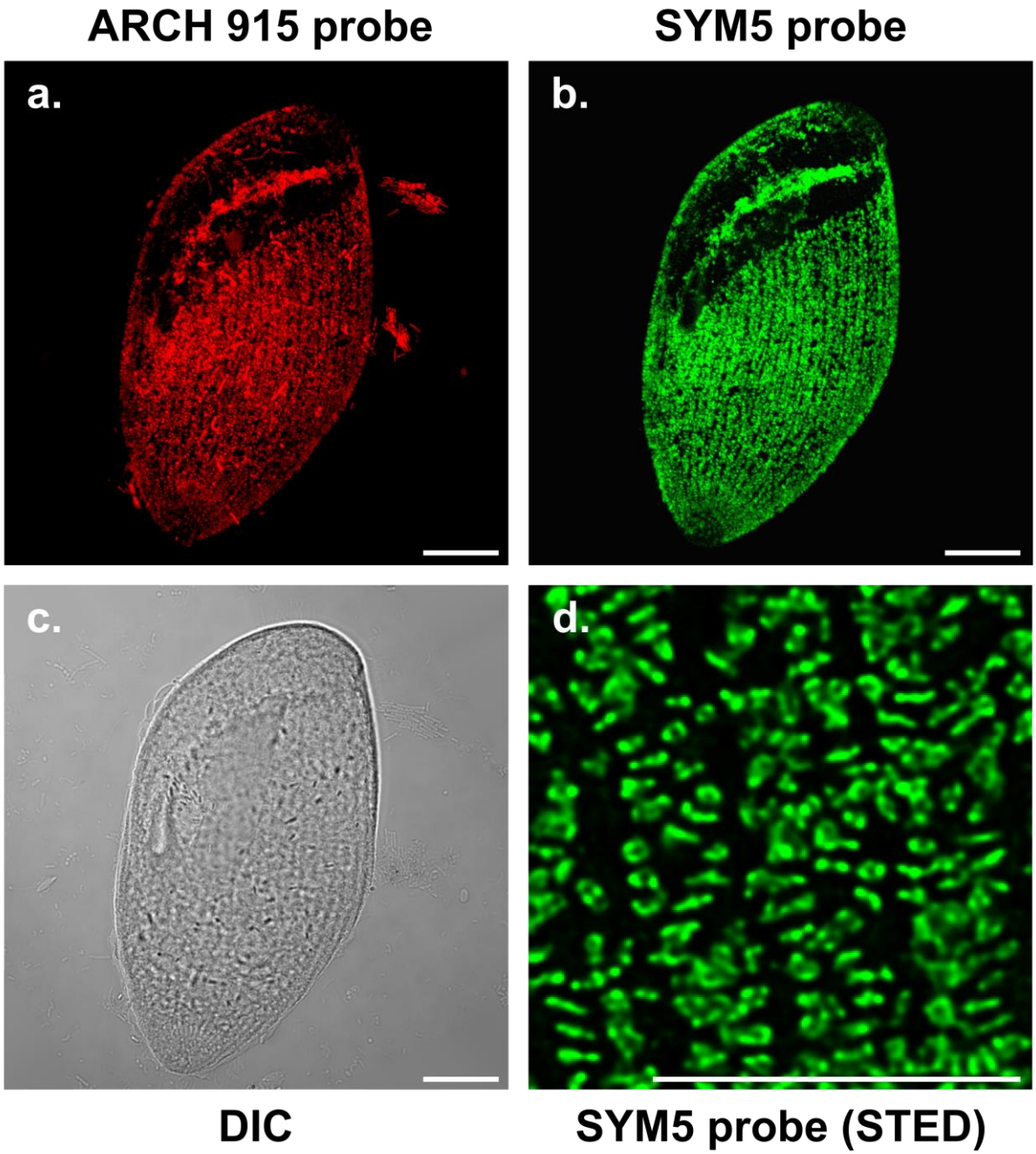


Figure 3.21. Fluorescent probing of *Metopus contortus* cells *in situ*. Archaea-specific probe, ARCH915 (a.), Symbiont-specific probe (b.), DIC image (c.), high-resolution STED image (d). Images a. and b. are maximum intensity z-projections of multiple confocal slices. a., b. and c. were imaged with a 63x objective lens and d. with a 100x objective lens. Scale bars represent 20µm (a., b. and c.) and 10µm (d.)

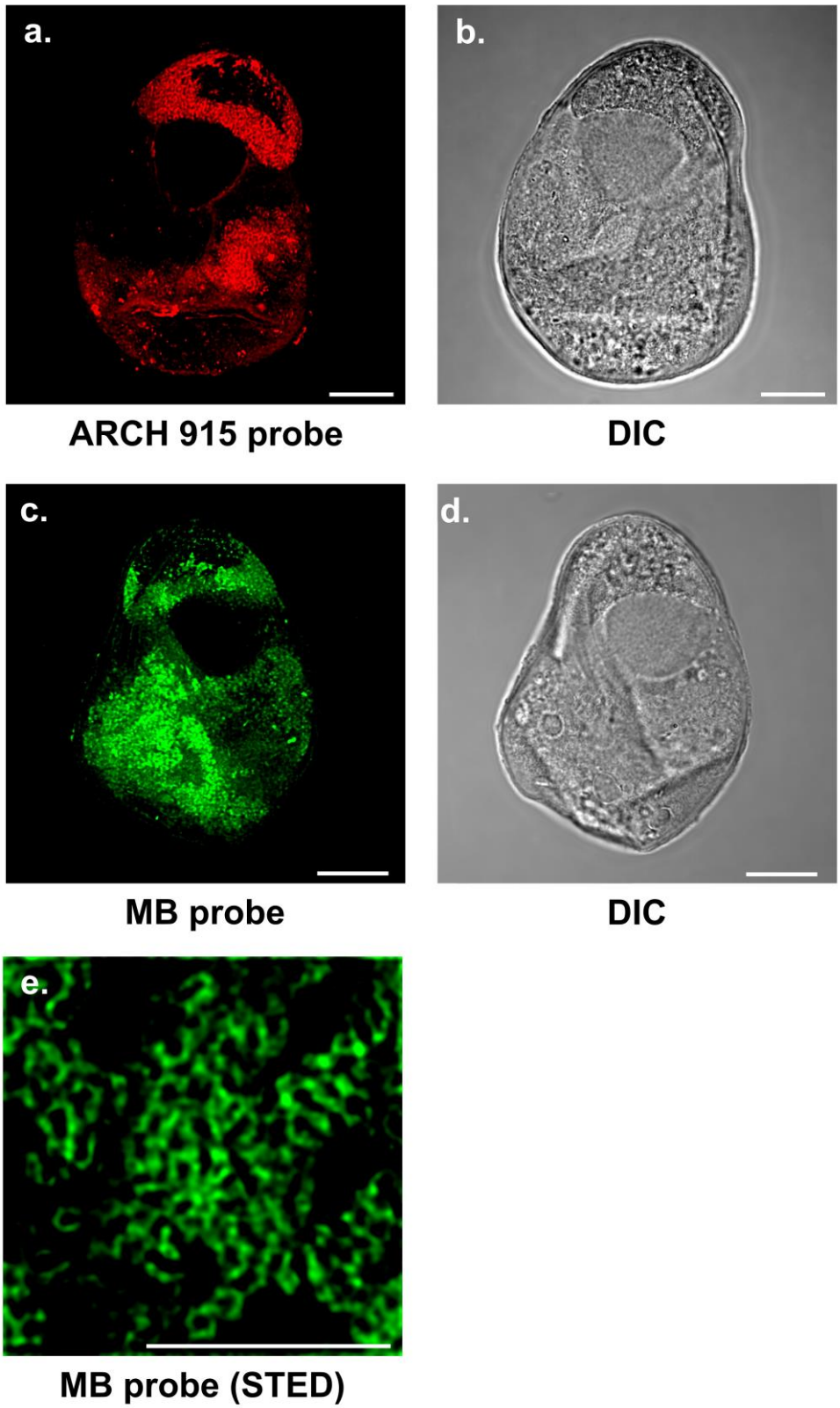


Figure 3.22. Fluorescent probing of *Nyctotherus ovalis* cells *in situ*. Archaea-specific probe, ARCH915 (a.); DIC image of same field shown in a. (b.); symbiont-specific probe, MB (c.); DIC image of same field shown in c. (d.); high-resolution STED image (e); Images a. and c. are maximum intensity z-projections of multiple confocal slices. a., b., c. and d. were imaged with a 63x objective lens and e. was imaged with a 100x objective lens. Scale bars represent 20µm (a., b., c. and d.) and 10µm (e.)

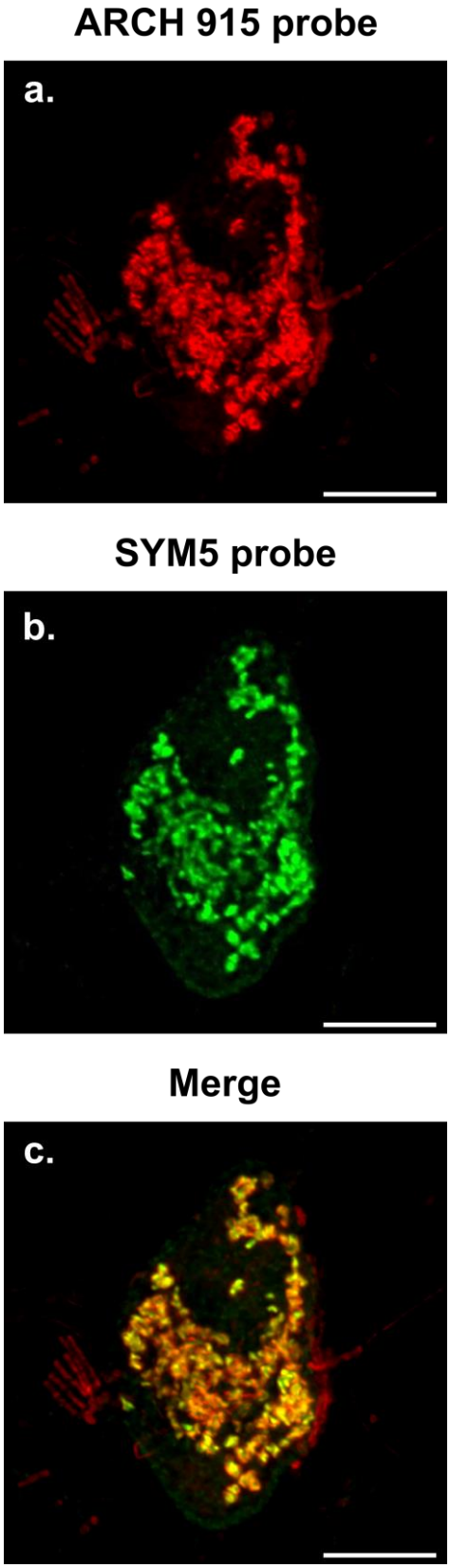


Figure 3.23. Fluorescent probing of *Trimyema* sp. cells *in situ*. Archaea-specific probe, ARCH915 (a.); Symbiont-specific probe, SYM5 (b.); merged image of a. and b. (c.). Images a., b. and c. are maximum intensity z-projections of multiple confocal slices, imaged with a 63x objective lens. Scale bars represent 10µm.

3.3.20 Phylogenetic relationships between methanogenic endosymbionts and their ciliate hosts

The associations between anaerobic ciliates and their methanogenic endosymbionts, and whether they are preserved over large evolutionary timescales, were investigated further. This was done by comparing species phylogenies inferred from conserved regions of endosymbiont and free-living methanogen Archaea 16S rRNA gene sequences with a tree based upon ciliate 18S rRNA gene sequences using the CAT+GTR model (Lartillot and Philippe, 2004) (Figure 3.24). The 16S rRNA gene sequences from the endosymbionts of *Trimyema* sp. (Finlay et al., 1993b), *Trimyema compressum* (Shinzato et al., 2007) and *Metopus palaeformis* (Embley et al., 1992b) were obtained from studies that had previously identified these endosymbiont species, which were confirmed using *in situ* probing. The 16S rRNA gene sequences from the endosymbionts of *Metopus contortus* and *Nyctotherus ovalis* recorded in this thesis were obtained from their whole genome sequences, which were recovered from the hydrogenosome genome sequencing datasets.

Based on the phylogenies in Figure 3.24, it is clear that although *Nyctotheurs ovalis*, *Metopus contortus* and *Metopus palaeformis* form a clade (posterior probability of 1), their endosymbionts are polyphyletic. A similar situation can be observed for the clade containing *Trimyema* sp. and *Trimyema compressum* (posterior probability of 0.99), the endosymbionts of which are also polyphyletic. As predicted based on FISH experiments (Section 3.3.19), the endosymbionts of *Metopus contortus* and *Trimyema* sp. group with the free-living methanogen species *Methanocorpusculum labreanum* (posterior probability of 0.92). The endosymbiont of *Nyctotherus ovalis* forms a clade with the endosymbiont of *Trimyema compressum* and *Methanobrevibacer arboriphilus* (posterior probability of 1), and the endosymbiont of *Metopus palaeformis* is closely related to *Methanobacterium lacus* (posterior probability of 1).

This analysis suggests that ciliates and methanogenic endosymbionts have not co-speciated in agreement with previously published work (Embley and Finlay, 1994). However the expanded sampling does suggest that certain methanogen lineages may form endosymbioses with anaerobic ciliates more frequently than others. This is suggested by the clade that contains the closely related

endosymbionts of *Trimyema* sp. and *Metopus contortus* (posterior probability of 1) and the clade that contains the closely-related endosymbionts of *Nyctotherus ovalis* and *Trimyema compressum*. This could suggest that species from the genera *Methanocorpusculum* and *Methanobrevibacter* have evolved strategies for colonising anaerobic ciliates. Whether this is the case could be investigated by detailed analysis and greater sampling of whole genome sequences from these and other methanogenic endosymbionts of different lineages, and comparisons with their free-living relatives.

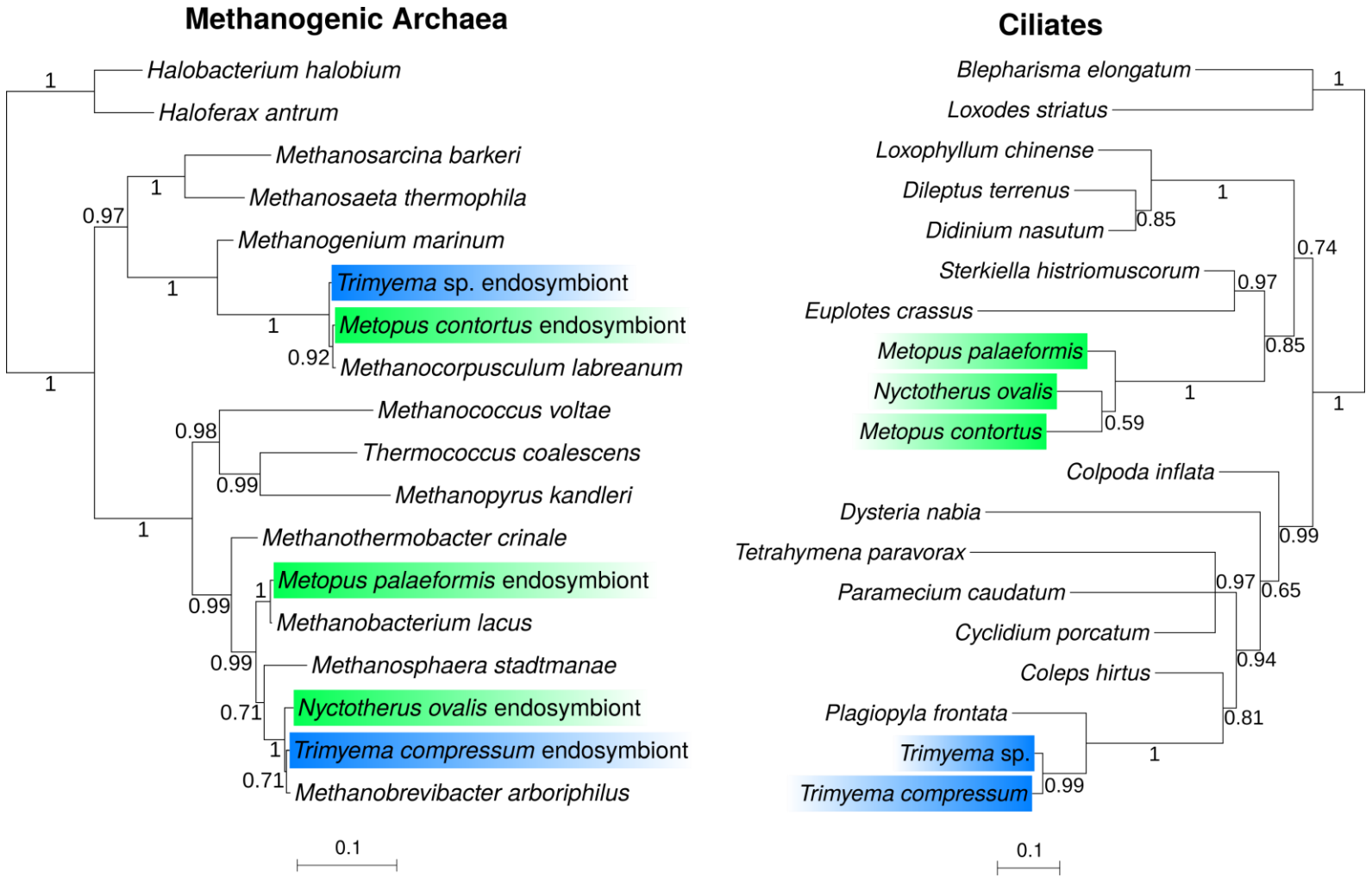


Figure 3.24. Phylogenies of methanogenic Euryarchaeaota Archaea (left) and ciliates (right), inferred from alignments of 16S and 18S rRNA genes, respectively, using the CAT + GTR model (Lartillot and Philippe, 2004). Anaerobic ciliates of the same clade that have had their endosymbiont species identified using *in situ* probing in the present study and others (Embley et al., 1992b; Finlay et al., 1993b; Shinzato et al., 2007), are highlighted in the same colour. Corresponding endosymbiont species are also highlighted in the same colour as their ciliate hosts. Support values indicate posterior probabilities. Scale bars represent estimated number of substitutions per site.

3.4 Discussion

The present study has provided the first experimental evidence for the existence of hydrogenosome genomes in *Metopus contortus, Metopus es, Metopus striatus* and *Cyclidium porcatum*. Detailed analysis of the genes associated with these genomes reveals that they have all evolved from mitochondrial genomes. These data provide strong evidence for a common mitochondrial ancestry of the hydrogenosomes in these particular ciliates and in *Nyctotherus ovalis* (Figures 3.8-3.13). By contrast, the hydrogenosomes of *Plagiopyla frontata* and *Trimyema* sp. appear to have lost their mitochondrial genomes but metabolic reconstructions for the hydrogenosomes from *Plagiopyla frontata* (see below) nevertheless demonstrate that it has retained proteins that are typically found in canonical mitochondria. Taken together the data provide compelling evidence for the hypothesis that ciliate hydrogenosomes have evolved from mitochondria on at least three separate occasions in the ciliate tree.

Although the amount of sequence data varies for the different ciliates, it is apparent that the hydrogenosome genomes from *Nyctotherus ovalis*, *Metopus contortus*, *Metopus es*, *Metopus striatus* and *Cyclidium porcatum* share significant similarities in the genes they contain. All of the genomes have retained the genes *nad7* and *nad10*, and genes *nad1, nad3, nad4, nad5, nad9, rps12* and *rpl2* are also commonly present. Most of the protein-coding genes identified from the hydrogenosome genomes are for ribosomal proteins and for subunits of ETC Complex I. This is similar to the reported gene content of hydrogenosome genomes from other eukaryotes, including *Blastocystis* (Jacob et al., 2016). This suggests that an important function of hydrogenosomal ribosomes is in the production of organelle-genome encoded proteins that make up the subunits of ETC Complex I. In turn this suggests that there is selection for retaining some functions of ETC Complex I, possibly including proton pumping to maintain a membrane gradient, inside of these hydrogenosomes. Only one gene was identified that does not encode a subunit of ETC Complex I or ribosomal proteins, this was a gene for ymf66 which appeared among the partial data for the *Cyclidium porcatum* organelle genome. This protein was identified previously in species of aerobic ciliates and based on the proteomic analysis of F_1F_0 ATP

synthase from *Tetrahymena thermophila* (Smith et al., 2007; Nina et al., 2010) it appears to be part of the F_0 sub-complex.

Genes for 12S and 16S mitochondrial rRNA were identified in all of the hydrogenosome genomes. Interestingly, the hydrogenosomal genome of *Cyclidium porcatum* is predicted to have two copies of each gene: this is also a feature of the mitochondrial genomes of other Oligohymenophorea including *Tetrahymena thermophila* and *Paramecium tetraurelia* (Pritchard et al., 1990; Brunk et al., 2003). The presence of mitochondrial rRNA genes is consistent with the presence of genes for ribosomal proteins, and together with nuclear encoded proteins they are likely to form a functional mitochondrial ribosome. A number of tRNA genes were also identified in each of the ciliate hydrogenosome genomes and these would be available to transfer amino acids to the hydrogenosome ribosomes during protein synthesis. Consistent with this, a predicted MCF amino acid transporter was identified from the transcriptome analysis of *Cyclidium porcatum*, suggesting that the hydrogenosomes of this species could import amino acids needed for protein synthesis from the cytosol.

The genome data was supplemented with transcriptomic data and together they were used to reconstruct the hydrogenosome metabolisms for *Cyclidium porcatum*, *Metopus contortus* and *Plagiopyla frontata*. The hydrogenosomes of *Cyclidium porcatum* appears to have the most complete ETC as at least some nuclear genes for subunits of ETC Complexes I, II, III and IV, as well as for F_1F_0 ATP synthase, were detected from this species. Consistent with the predicted role of F_1F_0 ATP synthase in cristae formation (Strauss et al., 2008), the cristae of *Cyclidium porcatum* hydrogenosomes are clearly visible in the EM images generated here and in published data (Esteban et al., 1993). *Cyclidium porcatum* was also the only species from which genes for PFO and PNO were detected, which it can potentially use for the oxidation of pyruvate and the reduction of the substrates ferredoxin and $NADP^+$ respectively, as was shown for these enzymes from other organisms (Gorrell et al., 1984; Inui et al., 1987). *Cyclidium porcatum* also has some of the genes for substrate-level phosphorylation, so it could potentially use these to make some ATP. With the exception of GCK a complete pathway for glycolysis was detected.

The hydrogenosome metabolism of the two closely related anaerobic ciliates *Metopus contortus*, reconstructed in the present study, and *Nyctotherus ovalis*, reconstructed from a transcriptome analysis previously (de Graaf et al., 2011), share several similarities. Both these species have some genes for ETC Complexes I, II and III and lack genes for ETC Complex IV and F_1F_0 ATP synthase. Both of these species also have some of the genes needed for fumarate reduction, as well as ATP production by substrate-level phosphorylation. In the present study a number of genes were also identified from *Metopus contortus* that were not previously identified in *Nyctotherus ovalis* (de Graaf et al., 2011). These include a complete set of genes for the glycolysis pathway, consistent with the production of pyruvate and with the detection of all of the genes needed to make PDH. Additionally a coenzyme A transporter MCF protein was detected from *Metopus contortus*, which could potentially import cytosolic coenzyme A required for the decarboxylation of pyruvate by PDH.

Fewer genes for hydrogenosome metabolism proteins were detected for *Plagiopyla frontata* compared to *Cyclidium porcatum* and *Metopus contortus*. This is partly due to less data being generated for this species overall, but – given the absence of an organelle genome - it likely also reflects further organelle reductive evolution. For example, neither the genomic or transcriptomic data provided any evidence for components of the ETC suggesting that *Plagiopyla frontata* has completely lost this pathway. Consistent with this, and in agreement with published data for *Plagiopyla frontata* (Fenchel and Finlay, 2010) no cristae were visible in the EM images taken for the hydrogenosomes of *Plagiopyla frontata* in this thesis. A complete set of genes for the glycolysis pathway were detected from *Plagiopyla frontata* as well as the genes for PDH used to oxidise pyruvate. Like *Cyclidium porcatum*, a coenzyme A transporter MCF protein was detected in *Plagiopyla frontata*, consistent with its hydrogenosomes importing the coenzyme A substrate required for pyruvate decarboxylation. Some of the genes required for ATP production by substrate-level phosphorylation were also detected from this species.

At least some of the genes for proteins of the major mitochondrial protein import pathways were detected from *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata*. Both subunits of the MPP complex were detected from each of

these three species and this is consistent with N-terminal targeting signals being predicted for some proteins for each of these three ciliates. The largest number of MCF proteins was detected from *Cyclidium porcatum*, which suggests that it may have retained more of the metabolic functions associated with canonical mitochondria than the other hydrogenosomes. Consistent with the idea that loss of MCF diversity is correlated with degree of metabolic reduction, the smallest number of MCF proteins was detected from *Plagiopyla frontata*.

Genes for core components of the essential mitochondrial pathway for Fe-S cluster biogenesis (Freibert et al., 2017) were detected in the data for all three ciliates. The most complete set of proteins were found for *Cyclidium porcatum* and the least complete pathway was from *Plagiopyla frontata*. The presence of this pathway is consistent with the detection of genes for Fe-S proteins including ferredoxin and the FeFe-hydrogenases, which play a central role in hydrogen production. Given the vital role of the mitochondrial Fe-S cluster biosynthesis pathway in making essential cytosolic and nuclear proteins in model eukaryotes (Freibert et al., 2017), it appears likely that the incomplete nature of the ciliate pathways reflects missing data rather than gene loss. There are cases in anaerobic protists where the mitochondrial ISC pathway has been replaced by cytosolic bacterial-type pathways, that were likely acquired by lateral gene transfer (Stairs et al., 2014). This possibility was also explored by searching the datasets generated from the anaerobic ciliates in the present study, but no potential functional replacements for the ISC pathway could be identified.

The relationships between the ciliates and their methanogenic endosymbionts were investigated and revealed apparent differences in their short and long term co-evolution. On one hand, comparison of host and symbiont species trees provide little evidence that ciliates and their endosymbionts have co-speciated over the long term, since closely related ciliates do not appear to have closely related endosymbionts. On the other hand, the ciliate species *Metopus contortus* and *Trimyema* sp. were re-isolated in the present study, over two decades since their endosymbionts were first identified (Embley et al., 1992a; Finlay et al., 1993b) and from a completely different geographical location, yet they still contain the same species of endosymbionts. Similar patterns are also observed in *Nyctotherus ovalis* which are shown to contain the same endosymbiont species as

earlier studies (van Hoek et al., 2000b). At face value, these observations would suggest that the associations between methanogenic endosymbionts are stable over shorter time periods but not over longer periods of evolution.

Chapter 4. The evolution of enzymes involved in the hydrogenosome metabolism of ciliates

4.1 The evolution of enzymes with roles in H_2 production in the hydrogenosomes of ciliates

Ciliates appear to have evolved hydrogenosomes from aerobic mitochondria repeatedly and independently in several lineages (Embley et al., 1995). Hydrogenosomes oxidise pyruvate, using enzymes such as PFO/PNO and PFL, and produce H_2 using FeFe-hydrogenases, yet these enzymes are generally not found in aerobic mitochondria (Martin and Müller, 1998). In the present study anaerobic metabolism enzymes were analysed from the anaerobic hydrogenosome-containing ciliates *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. and their origins inferred phylogenetically. It is unknown whether the aerobic ancestors of these anaerobic ciliates already had all of the necessary enzymes that they needed to 'make' a hydrogenosome, or whether they acquired them from alternative sources by lateral gene transfer. The origins of these enzymes are therefore most important for understanding how anaerobic ciliates evolved hydrogenosomes.

4.1.1 FeFe-hydrogenases in anaerobic eukaryotes

There are three non-homologous classes of hydrogenases that are capable of reducing protons to H_2, these are defined as NiFe-hydrogenases (found in Bacteria and Archaea), Fe-hydrogenases (found in methanogenic Archaea) and FeFe-hydrogenases (found in Bacteria and eukaryotes), according to the most recent nomenclature (Vignais et al., 2001; Shima et al., 2008; Tard and Pickett, 2009; Peters et al., 2015). Hydrogenases are typically thought to enable the use of protons as a terminal electron acceptor in oxidative pathways (Peters et al., 1998). FeFe-hydrogenases are the only hydrogenases found in eukaryotes (Horner et al., 2002) but these proteins were initially characterised from the anaerobic Bacteria, *Clostridium pasteurianum* (Adams et al., 1989; Hildebrand et al., 1991) and *Desulfovibrio vulgaris* (Voordouw and Brenner, 1985; Voordouw et al., 1989). The

first eukaryote from which a FeFe-hydrogenase enzyme was identified and characterised was the excavate *Trichomonas vaginalis* and was also shown to localise to purified hydrogenosomes (Bui and Johnson, 1996). FeFe-hydrogenases have also been shown to localise to hydrogenosomes *in situ* for the anaerobic ciliate *Nyctotherus ovalis* (Akhmanova et al., 1998), as well as other organisms including *Neocallimastix* sp. (Voncken et al., 2002) and *Blastocystis* sp. (Stechmann et al., 2008). In species with hydrogenosomes, FeFe-hydrogenases usually contain N-terminal mitochondrial-like targeting signals (Akhmanova et al., 1998; Horner et al., 2000; Davidson et al., 2002; Nývltová et al., 2015), which is consistent with a hydrogenosomal sub-cellular localisation. *Giardia intestinalis* however is an example of a species that lacks hydrogenosomes, instead possessing mitosomes, and was also shown to produce cytosolic H_2 (Lloyd et al., 2002). The FeFe-hydrogenase from this species appears to lack N-terminal targeting signals and was shown to localise to the cytosol (Emelyanov and Goldberg, 2011).

All FeFe-hydrogenases share a conserved H-cluster binding domain which contains cysteine ligands that are required for coordination of the Fe-S cluster active site in the functional protein (Peters et al., 1998). The FeFe-hydrogenase active site consists of a [4Fe-4S] cluster coupled to a [2Fe] sub-cluster by a cysteine (Peters et al., 1998). Activation of the complete functional enzyme requires precise assembly by three maturase enzymes, HydE, HydF and HydG, which synthesise and insert the [2Fe] sub-cluster into the active site containing a pre-existing [4Fe-4S] cluster (Posewitz et al., 2004; Posewitz et al., 2005; Mulder et al., 2009). The [4Fe-4S] cluster is thought to be pre-formed and incorporated into the protein by the Fe-S cluster assembly components (Mulder et al., 2009). HydE, HydF and HydG maturases appear to be absent in some eukaryotes that possess FeFe-hydrogenases (Nicolet and Fontecilla-Camps, 2012). For example, none of the three maturases were identified from the whole genome sequences of *Entamoeba histolytica* (McCoy and Mann, 2004) or *Giardia intestinalis* (Jerlström-Hultqvist et al., 2010), only HydE has been identified from the genome sequence of *Blastocystis* sp. (Denoeud et al., 2011). Previous transcriptomic studies of the anaerobic ciliate *Nyctotherus ovalis* (de Graaf et al., 2011) also failed to detect these enzymes. One possible explanation is that the active site of FeFe-hydrogenases from these species is only partially assembled *i.e.* they

contain the [4Fe-4S] cluster but not the [2Fe] sub-cluster, and can still produce H_2 but at lower levels (Nicolet and Fontecilla-Camps, 2012). This is plausible since other Fe-S cluster enzymes can reportedly make low-levels of H_2, such as CO dehydrogenase and PFO (Menon and Ragsdale, 1996).

Also well conserved are two [4Fe-4S] cluster-binding domains that are positioned immediately upstream from the H-cluster, in the direction of the N-terminus (Figure 4.1). In addition to these main conserved domains, different FeFe-hydrogenases often have various numbers of accessory [4Fe-4S] and [2Fe-2S] cluster-binding N-terminal domains that resemble ferredoxins (Vignais et al., 2001). The role of the accessory Fe-S clusters in the mature protein is thought to be the transfer of electrons to and from the active site of the H-cluster (Mulder et al., 2011). A variety of other cofactor and substrate binding domains, thought to be involved in the oxidation of different substrates, are often found fused to the C-terminus of the H-cluster domain in different species.

The only complete FeFe-hydrogenases from an anaerobic ciliate were sequenced from *Nyctotherus ovalis* (Akhmanova et al., 1998; Boxma et al., 2007) and these enzymes are different to those found in other eukaryotes as they have two additional C-terminal domains that have similarity to bacterial NuoE/24 kDa subunit of eukaryote ETC Complex I and bacterial NuoF/51 kDa subunit of eukaryote ETC Complex I (Figure 4.1). Previous analyses indicate that FeFe-hydrogenases of anaerobic ciliates may have a bacterial origin different from the FeFe-hydrogenases of other anaerobic eukaryotes (Horner et al., 2000; Embley et al., 2003). In the current study FeFe-hydrogenases were sequenced and analysed from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp., and the relationships and evolution of these enzymes to those from Bacteria and other eukaryotes are analysed and discussed.

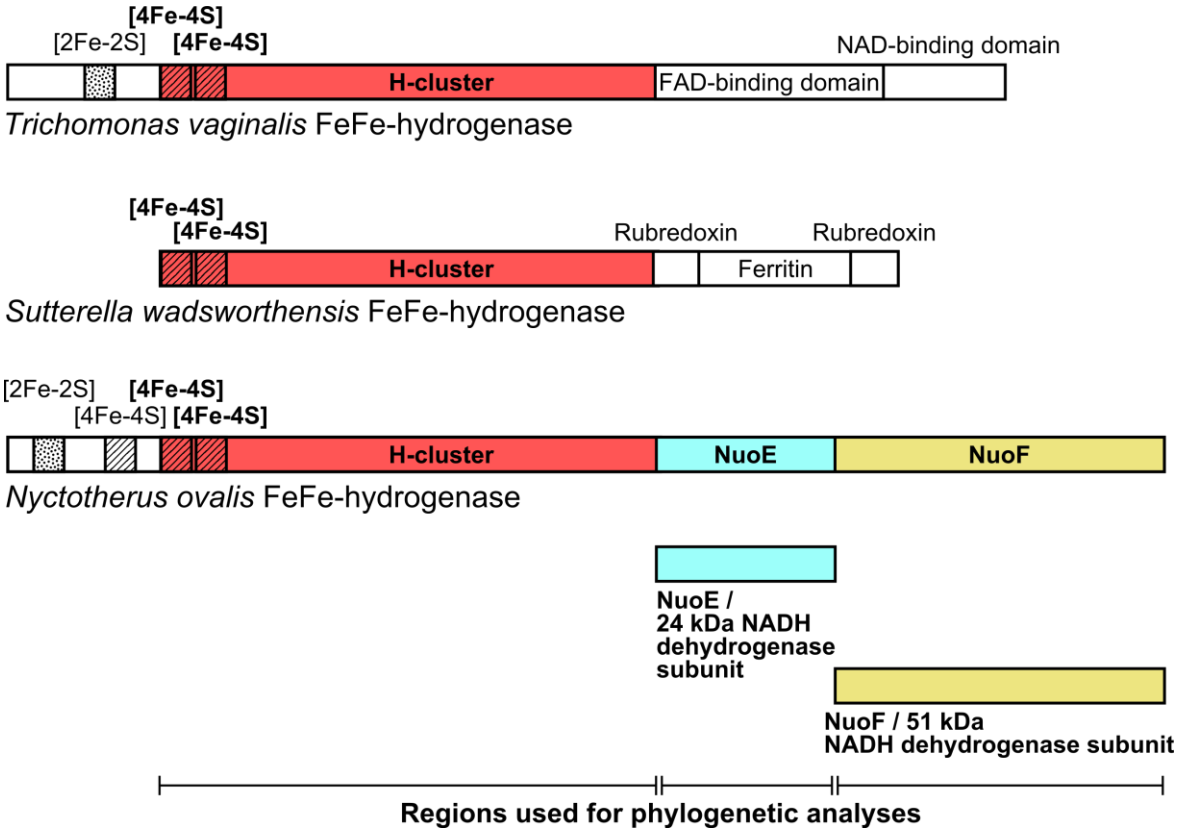


Figure 4.1. Diagram showing the domain structure of the FeFe-hydrogenases from the excavate *Trichomonas vaginalis*, the ciliate *Nyctotherus ovalis,* and the beta-proteobacteria *Sutterella wadsworthensis* (this sequence appears to be the most closely related to ciliate FeFe-hydrogenase sequences based on the analysis shown in Figure 4.5. The well conserved H-cluster and two [4Fe-4S] cluster-binding domains are shown (red) as well as various N-terminal and C-terminal accessory binding domains. The C-terminal domains of the FeFe-hydrogenase from *Nyctotherus ovalis* with homology to NuoE (blue) and NuoF (yellow) are highlighted. The protein regions used to infer phylogenies in Sections 4.3.4 and 4.3.5 of the present study are shown. Based on Figure 1 from Horner et al. (2000).

4.1.2 Pyruvate:ferredoxin oxidoreductase (PFO) and pyruvate: NADP+ oxidoreductase (PNO)

PFO oxidises pyruvate and has previously been found in some anaerobic species of Bacteria, eukaryotes and Archaea (Müller, 1993; Kletzin and Adams, 1996). In the hydrogenosomes of eukaryotes, PFO is thought to replace the role of pyruvate dehydrogenase (PDH), which typically oxidises pyruvate to acetyl-CoA in aerobic mitochondria (Kerscher and Oesterhelt, 1982; Horner et al., 1999; Embley, 2006). Like PDH, PFO also oxidises pyruvate to acetyl-CoA but can also reduce ferredoxin (Chabrière et al., 1999; Furdui and Ragsdale, 2000), which can then be reoxidised by FeFe-hydrogenase (Demuez et al., 2007). The PFO found in eukaryotes and Bacteria are similar as they consist of a single protein, whereas

the PFO found in Archaea are composed of multiple independent protein subunits, each of which are homologous to different domains of the PFO from eukaryotes and Bacteria (Hrdý and Müller, 1995a; Kletzin and Adams, 1996). Early phylogenetic analysis of a limited number of eukaryote PFO sequences that were available at that time recovered them as monophyletic (Horner et al., 1999). Since then the sampling of PFO from eukaryotes has been expanded and more recent analyses provide only weak support for the monophyly of eukaryotic PFO sequences and also indicated that the topology of the eukaryotic sequences do not agree with the species tree of the same organisms, leading to the suggestion that PFO evolution has been affected by extensive lateral gene transfers between eukaryotes (Hug et al., 2010).

PNO is similar to PFO except it contains an additional C-terminal NADPH-cytochrome p450 reductase domain (Figure 4.2) (Rotte et al., 2001). PNO from *Euglena gracilis* has been functionally characterised and was shown to oxidise pyruvate and reduce $NADP^+$ but did not react with ferredoxin, making it functionally distinct from PFO (Inui et al., 1987). The relationship of PFO and PNO enzymes were analysed by several phylogenetic studies previously (Kletzin and Adams, 1996; Horner et al., 1999; Rotte et al., 2001) but the evolutionary significance of the fusion that occurred between PFO and NADPH-cytochrome p450 reductase, creating PNO, has only received limited discussion (Rotte et al., 2001).

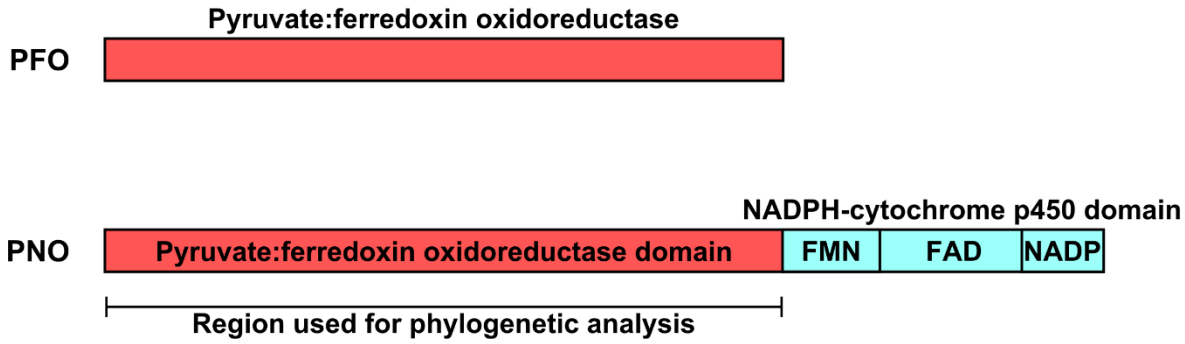


Figure 4.2. Diagram of the general domain structure of PFO and PNO enzymes including the conserved pyruvate: ferredoxin oxidoreductase domain (red) and the fused NADPH-cytochrome p450 domain (blue) that is present in PNO only. The protein region used for phylogenetic inference in Section 4.3.6 of the present study is shown.

In this chapter the identification of FeFe-hydrogenase, PFO and PNO enzymes from the anaerobic ciliates *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. is described. Additionally the subcellular location of these proteins are predicted and phylogenetic analysis is used to infer the evolutionary origins of these proteins.

4.2 Results

4.2.1. Identification of hydrogenosome metabolism enzymes from ciliate transcriptomes

Blast-based methods, described in Section 2.5, were used to search the assembled transcriptomic datasets of *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. for sequences related to FeFe-hydrogenase; the maturase enzymes, HydE, HydF and HydG; PFO/PNO; and PFL. Genes encoding FeFe-hydrogenase were detected from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. and genes encoding PFO and PNO were detected from *Cyclidium porcatum* (listed in Tables 4.1. a—4.1. f.). HydE, HydF, HydG or PFL were not detected from the available data for any of these species.

No PFO or PNO sequences were detected from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Plagiopyla frontata* or *Trimyema* sp. The three subunits of the pyruvate dehydrogenase (PDH) complex, however were detected from each of these species, many of which were predicted as having N-terminal mitochondrial-like targeting signals (Appendix B). This suggests that the hydrogenosomes of these species use PDH to oxidise pyruvate, similar to aerobic ciliate species (Smith et al., 2007). Similar findings were reported from a transcriptome analysis of *Nyctotherus ovalis* previously, where PDH was detected but not PFO, PNO or PFL (de Graaf et al., 2011).

The transcripts identified as encoding FeFe-hydrogenase, PFO and PNO from ciliates in the present study can be found in Appendix C, along with their translated protein sequences.

4.2.2 Codon usage analysis of genes identified for FeFe-hydrogenase, PFO and PNO

Numerous genes from the assembled transcriptomic datasets were identified, based on sequence similarity from blast searches, as being encoded by the genomes of organisms other than *Nyctotherus ovalis, Metopus contortus, Metopus*

es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata and *Trimyema* sp. These organisms could be free-living prokaryotes that were not successfully removed from the samples by washing, endosymbionts living inside the ciliates or possibly undigested prokaryotes within ciliate food vacuoles. In addition to this some genes that are encoded by the ciliate macronuclear genome might have been recently acquired by lateral gene transfer and therefore the closest homologues of these genes would not be genes in other ciliates but rather could be an alternative prokaryotic or eukaryotic donor. All of these factors create a level of uncertainty regarding which genes in the datasets are encoded by which genomes.

To investigate whether the genes for FeFe-hydrogenase, PFO and PNO were likely to be encoded by the macronuclear genomes of the ciliates, the codon usage of these genes was analysed and compared to the inferred codon usage for the ciliate macronuclear genome as a whole. These analyses were performed in collaboration with Dr Tom A. Williams (University of Bristol). The complete genomes from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. have not been sequenced, so the following strategy was used to identify sequences from each ciliate genome. These were identified by selecting the transcripts that were most similar to proteins encoded by ciliates in the nr database (NCBI) using blastx searches. The coding sequences predicted from the ciliate transcripts were used as the sets of 'true'-positive ciliate genes from which codon usage tables were calculated, using the program cusp (Rice et al., 2000) (These tables can be found in Appendix D). Next codon adaptation index (CAI) (Sharp and Li, 1987) scores, which is a method of measuring codon usage bias, were calculated with the program cai (Rice et al., 2000). This was done for each gene in the ciliate 'true'-positive sets and for each gene from the genomes of what appeared to be the most abundant species of Bacteria and Archaea in the transcriptomic datasets. These CAI scores were then plotted as three normalised frequency distributions on the same axis to facilitate comparison of the CAI scores for each genome calculated under the codon usage tables of *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. (Figure 4.3 a—f.).

Each of the ciliate CAI score frequency distributions were unimodal, which is consistent with the genes that were used being encoded by the same genome. The purpose of the distributions plotted for CAI scores from Bacteria and Archaea genes was to investigate whether this method provided sufficient resolution to identify whether genes were more likely to be ciliate macronuclear genes or more likely to have come from a different source. Because there was generally only limited overlap between the distributions of CAI scores for ciliate genes and the distributions for Bacteria and Archaea genes, this suggested that calculating the CAI scores for other genes would provide evidence for whether the genes are likely encoded by the ciliate macronuclear genomes or not. Therefore CAI scores were calculated for each of the genes encoding FeFe-hydrogenase, PFO and PNO.

In the case of all of the genes for FeFe-hydrogenase, PFO and PNO, their CAI scores were within the frequency distributions for ciliate genes. In most cases they also had high values within these distributions, suggesting that they had CAI values that were typical for genes from these ciliate genomes and this therefore provided good evidence that they were real ciliate genes. Furthermore, in the majority of cases the CAI scores of these genes were outside of the distributions for Archaea and Bacteria species. Even in cases where they were within the Archaea and Bacteria distributions, they were in the top ten percent of the distribution suggesting that if they were genes from Archaea or Bacteria then their CAI values would be distant outliers. These results therefore provide strong evidence that each of the FeFe-hydrogenase genes identified in this study came from ciliates.

				Percentile Rank		
Species	Enzyme	Transcript IDs	CAI score	Ciliate	Archaea	Bacteria
Cyclidium porcatum	PFO	c1688_g1_i1	0.785	88.8	100	100
	PFO/PNO	c15969_g1_i1	0.764	75.1	100	100
	PFO/PNO	c3950_g1_i1	0.752	64.9	100	100
	PNO	c4339_g1_i1	0.745	58.3	100	100
	PNO	c4345_g2_i1	0.78	85.9	100	100
	FeFe-hydrogenase	c4008_g1_i1	0.751	64.1	100	100
		c4008_g2_i1	0.781	86.5	100	100

Table 4.1. a. CAI scores calculated for coding sequences from *Cyclidium porcatum* transcripts that were identified as encoding PNO and FeFe-hydrogenase enzymes based on blast searches. CAI scores were calculated using a codon usage table that was calculated from a set of 'true'-positive ciliate genes from *Cyclidium porcatum*. The percentile rank of CAI scores for three distributions are shown. Values of 100 in the Archaea and Bacteria columns means that the CAI scores for these genes fell outside of the distribution of CAI scores for all Archaea and Bacteria genes, providing strong evidence they are from *Cyclidium porcatum*.

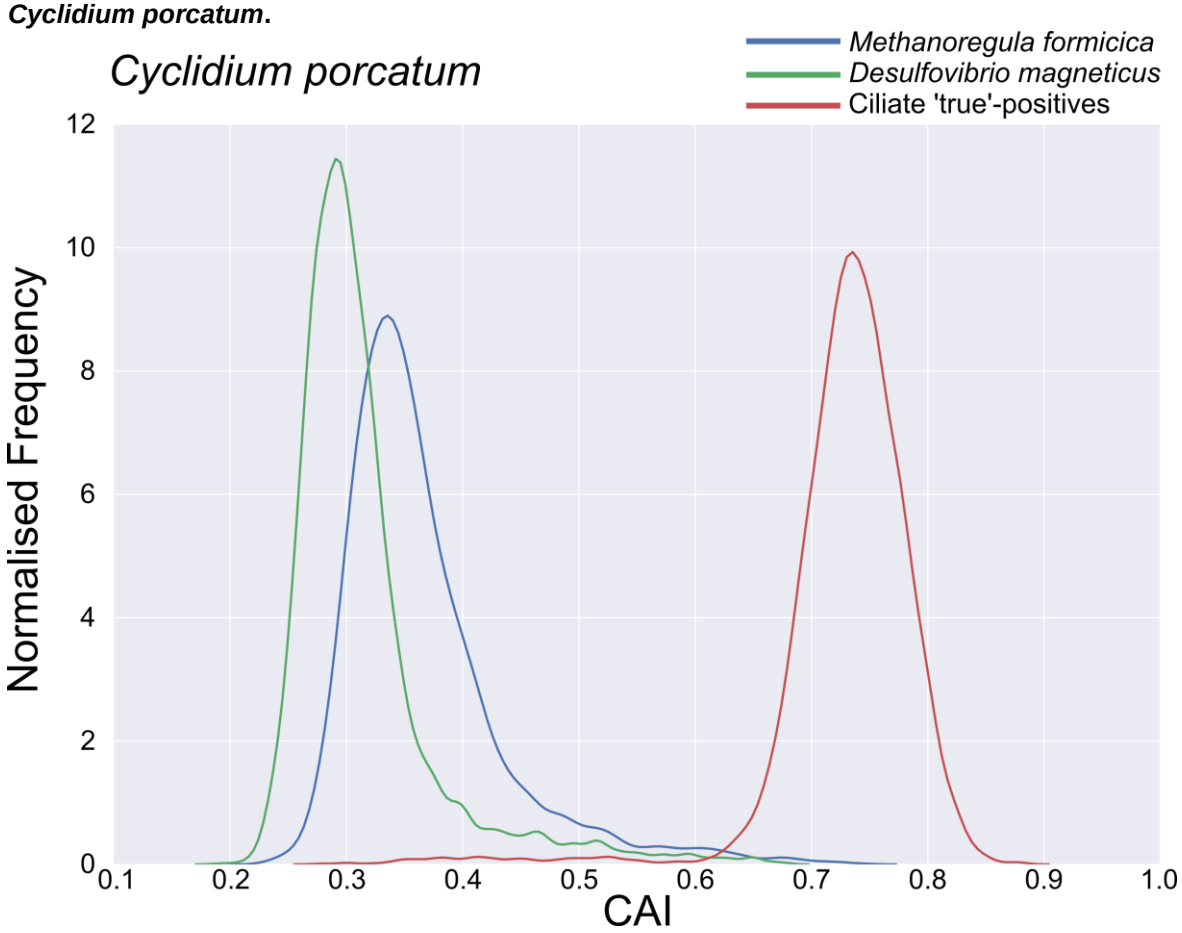


Figure 4.3. a. Normalised distributions of codon adaptation index (CAI) scores calculated for three datasets of genes from *Cyclidium porcatum, Methanoregula formicica* and *Desulfovibrio magneticus*. *Cyclidium porcatum* coding sequences were predicted from transcripts that had most significant similarity to ciliate proteins in the nr database (NCBI) based on blastx searches and these were considered the 'true'-positive set of ciliate genes. *Methanoregula formicica* (NC_019943.1) and *Desulfovibrio magneticus* (NC_012796.1) coding sequences were retrieved from their genome sequences, as annotated in their NCBI entries. All CAI scores were calculated using a codon usage table, which was calculated from the total set of 'true'-positive ciliate genes from *Cyclidium porcatum*.

				Percentile Rank		
Species	Enzyme	Transcript IDs	CAI score	Ciliate	Archaea	Bacteria
Nyctotherus ovalis	FeFe-hydrogenase	c14134_g1_i2	0.825	85.1	100	100
		c14134_g2_i1	0.841	94.1	100	100
		c14134_g1_i1	0.826	85.9	100	100

Table 4.1. b. CAI scores calculated for coding sequences from *Nyctotherus ovalis* transcripts that were identified as encoding FeFe-hydrogenase enzymes based on blast searches. CAI scores were calculated using a codon usage table that was calculated from a set of 'true'-positive ciliate genes from *Nyctotherus ovalis*. The percentile rank of CAI scores for three distributions are shown. Values of 100 in the Archaea and Bacteria columns mean that the CAI scores for these genes fell outside of the distribution of CAI scores for all Archaea and Bacteria genes, providing strong evidence they are from *Nyctotherus ovalis*.

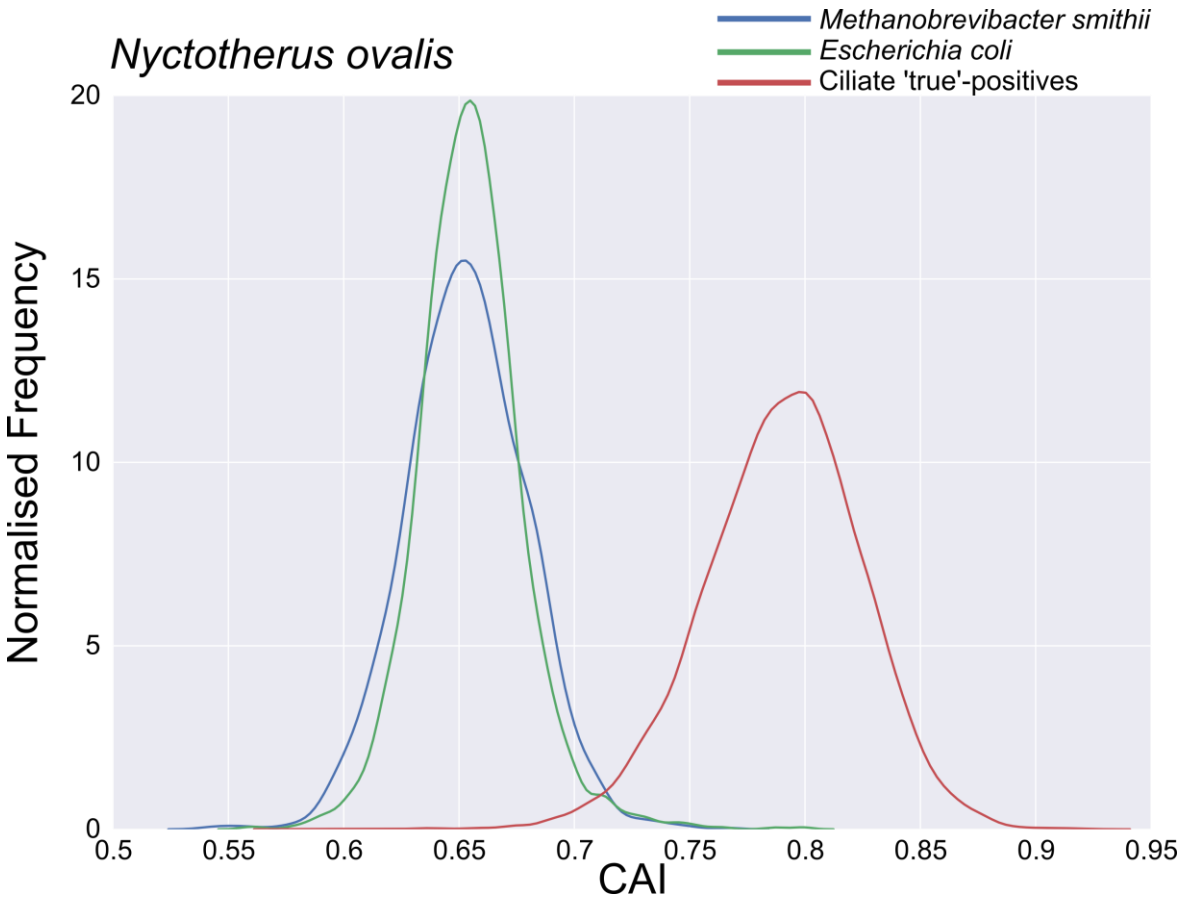


Figure 4.3. b. Normalised distributions of codon adaptation index (CAI) scores calculated for three datasets of genes from *Nyctotherus ovalis, Methanobrevibacter smithii* and *Escherichia coli*. *Nyctotherus ovalis* coding sequences were predicted from transcripts that had most significant similarity to ciliate proteins in the nr database (NCBI) based on blastx searches and these were considered the 'true'-positive set of ciliate genes. *Methanobrevibacter smithii* (NC_009515.1) and *Escherichia coli* (NC_002695.1) coding sequences were retrieved from their genome sequences, as annotated in their NCBI entries. All CAI scores were calculated using a codon usage table, which was calculated from the total set of 'true'-positive ciliate genes from *Nyctotherus ovalis*.

				Percentile Rank		
Species	Enzyme	Transcript IDs	CAI score	Ciliate	Archaea	Bacteria
Metopus contortus	FeFe-hydrogenase	c9793_g1_i1	0.785	54.7	98.6	96.9
		c10068_g1_i1	0.788	59.8	98.8	97.3
		c10161_g1_i1	0.785	54.7	98.6	96.9

Table 4.1. c. CAI scores calculated for coding sequences from *Metopus contortus* transcripts that were identified as encoding FeFe-hydrogenase enzymes based on blast searches. CAI scores were calculated using a codon usage table that was calculated from a set of 'true'-positive ciliate genes from *Metopus contortus*. The percentile rank of CAI scores for three distributions are shown. High values in the Archaea and Bacteria columns mean that these genes are outliers in the distribution of CAI scores for all Archaea and Bacteria genes, providing strong evidence they are from *Metopus contortus*.

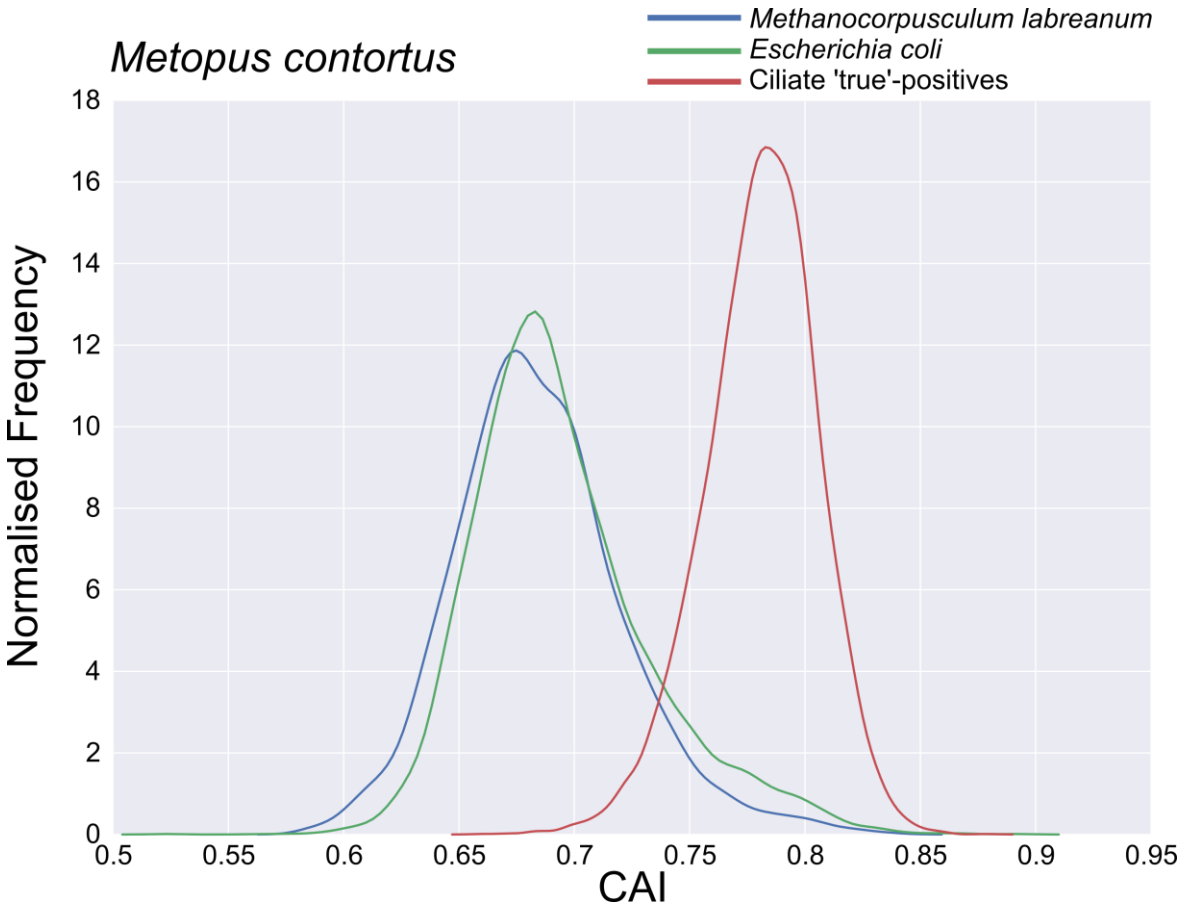


Figure 4.3. c. Normalised distributions of codon adaptation index (CAI) scores calculated for three datasets of genes from *Metopus contortus, Methanocorpusculum labreanum* and *Escherichia coli*. *Metopus contortus* coding sequences were predicted from transcripts that had most significant similarity to ciliate proteins in the nr database (NCBI) based on blastx searches and these were considered the 'true'-positive set of ciliate genes. *Methanocorpusculum labreanum* (NC_008942.1) and *Escherichia coli* (NC_002695.1) coding sequences were retrieved from their genome sequences, as annotated in their NCBI entries. All CAI scores were calculated using a codon usage table, which was calculated from the total set of 'true'-positive ciliate genes from *Metopus contortus*.

				Percentile Rank		
Species	Enzyme	Transcript IDs	CAI score	Ciliate	Archaea	Bacteria
Metopus es	FeFe-hydrogenase	c8491_g1_i1	0.792	91.5	94.9	94.6
		c10311_g1_i1	0.786	85.2	93.6	93.6
		c16173_g1_i1	0.77	55.1	90.5	90.6

Table 4.1. d. CAI scores calculated for coding sequences from *Metopus es* transcripts that were identified as encoding hydrogenase enzymes based on blast searches. CAI scores were calculated using a codon usage table that was calculated from a set of 'true'-positive ciliate genes from *Metopus es*. The percentile rank of CAI scores for three distributions are shown. High values in the Archaea and Bacteria columns mean that these genes are outliers in the distribution of CAI scores for all Archaea and Bacteria genes, providing strong evidence they are from *Metopus es*.

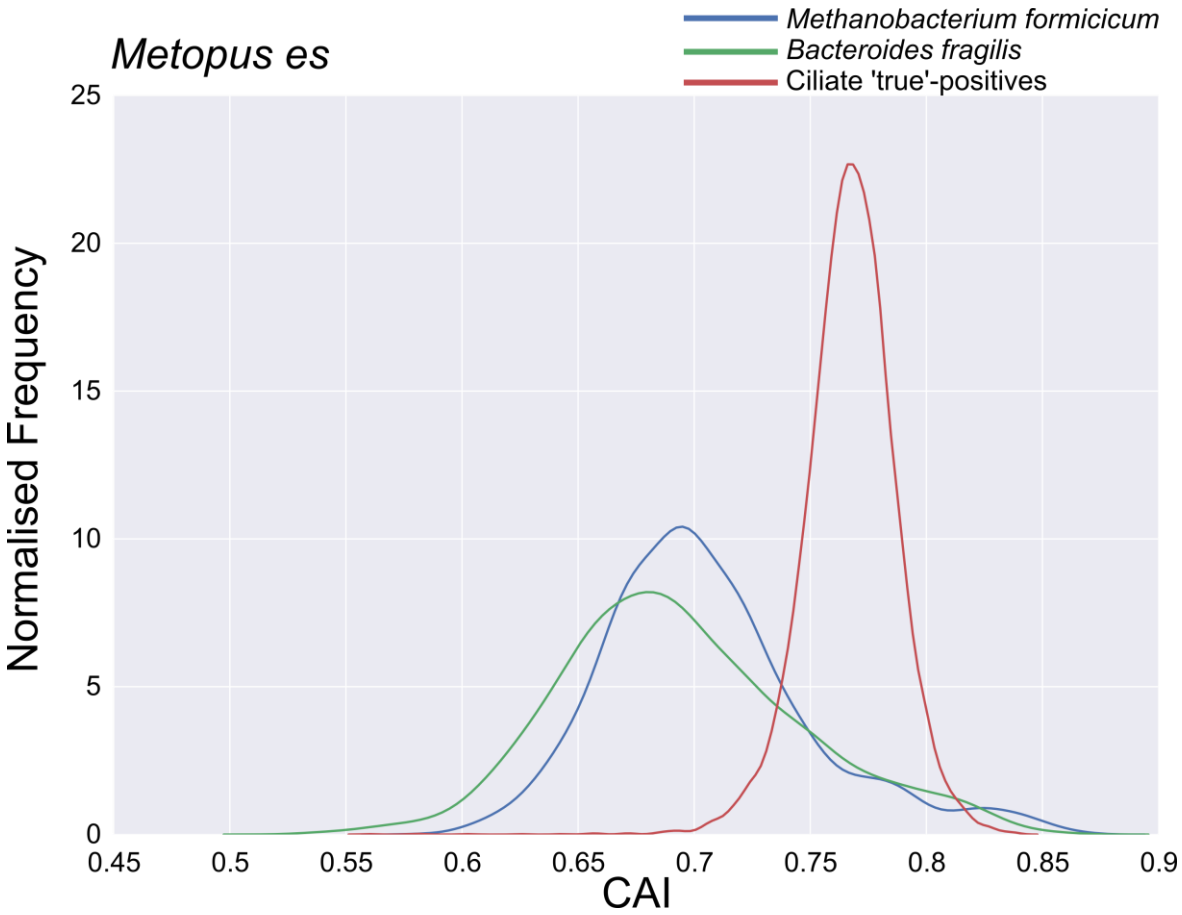


Figure 4.3. d. Normalised distributions of codon adaptation index (CAI) scores calculated for three datasets of genes from *Metopus es, Methanobacterium formicicum* and *Bacteroides fragilis*. *Metopus es* coding sequences were predicted from transcripts that had most significant similarity to ciliate proteins in the nr database (NCBI) based on blastx searches and these were considered the 'true'-positive set of ciliate genes. *Methanobacterium formicicum* (NZ_CP006933.1) and *Bacteroides fragilis* (NC_006347.1) coding sequences were retrieved from their genome sequences, as annotated in their NCBI entries. All CAI scores were calculated using a codon usage table, which was calculated from the total set of 'true'-positive ciliate genes from *Metopus es*.

				Percentile Rank		
Species	Enzyme	Transcript IDs	CAI score	Ciliate	Archaea	Bacteria
Metopus striatus	FeFe-hydrogenase	c8165_g1_i1	0.691	18.6	99.9	91.0
		c9476_g1_i1	0.727	55.7	100	95.8
		c9476_g2_i1	0.716	42.6	100	94.6

Table 4.1. e. CAI scores calculated for coding sequences from *Metopus striatus* transcripts that were identified as encoding FeFe-hydrogenase enzymes based on blast searches. CAI scores were calculated using a codon usage table that was calculated from a set of 'true'-positive ciliate genes from *Metopus striatus*. The percentile rank of CAI scores for three distributions are shown. High values in the Archaea and Bacteria columns mean that these genes are outliers in the distribution of CAI scores for all Archaea and Bacteria genes, providing strong evidence they are from *Metopus striatus*.

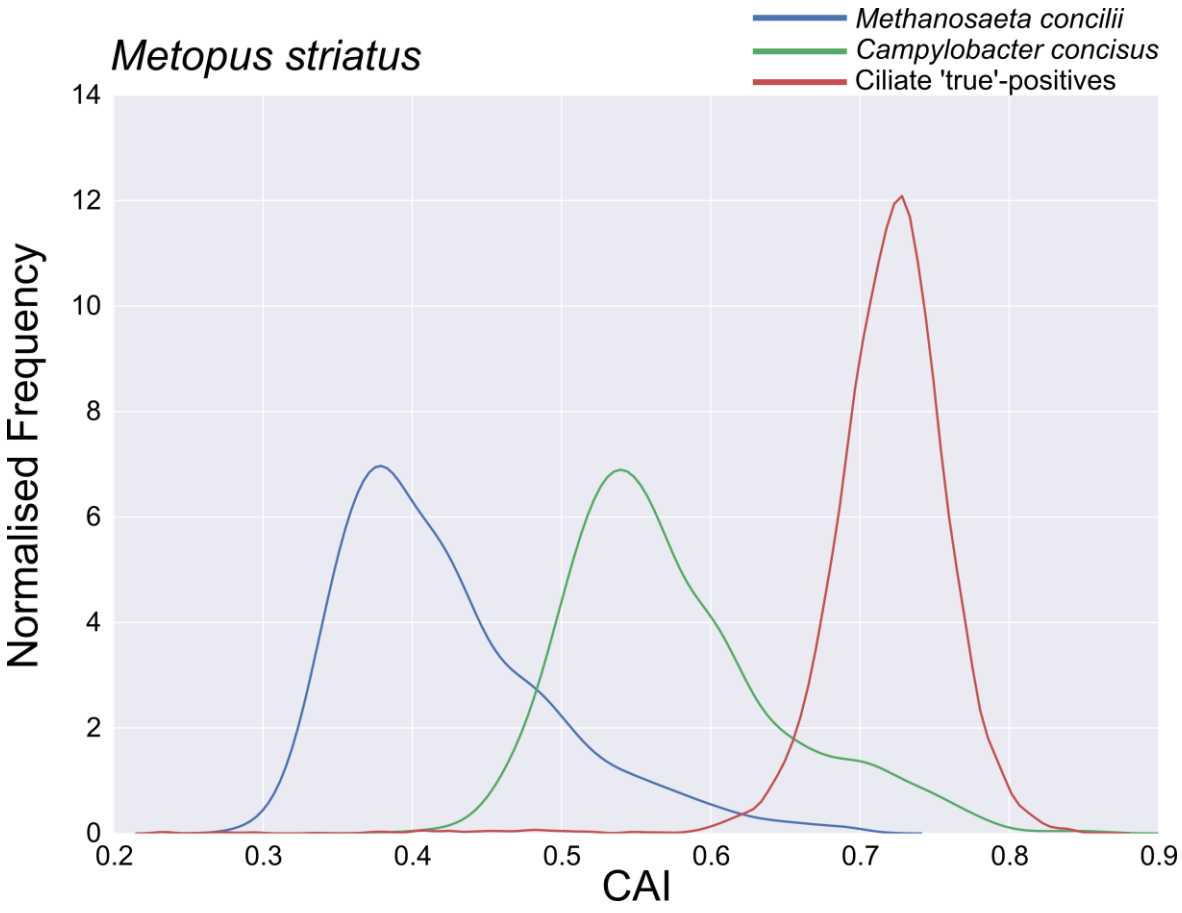


Figure 4.3. e. Normalised distributions of codon adaptation index (CAI) scores calculated for three datasets of genes from *Metopus striatus, Methanosaeta concilii* and *Campylobacter concisus*. *Metopus striatus* coding sequences were predicted from transcripts that had most significant similarity to ciliate proteins in the nr database (NCBI) based on blastx searches and these were considered the 'true'-positive set of ciliate genes. *Methanosaeta concilii* (NC_015416.1) and *Campylobacter concisus* (NZ_CP012541.1) coding sequences were retrieved from their genome sequences, as annotated in their NCBI entries. All CAI scores were calculated using a codon usage table, which was calculated from the total set of 'true'-positive ciliate genes from *Metopus striatus*.

				Percentile Rank		
Species	Enzyme	Transcript IDs	CAI score	Ciliate	Archaea	Bacteria
Plagiopyla frontata	FeFe-hydrogenase	c6085_g1_i1	0.703	48.1	99.9	99.8

Table 4.1. f. CAI scores calculated for coding sequences from *Plagiopyla frontata* transcripts that were identified as encoding FeFe-hydrogenase enzymes based on blast searches. CAI scores were calculated using a codon usage table that was calculated from a set of 'true'-positive ciliate genes from *Plagiopyla frontata*. The percentile rank of CAI scores for three distributions are shown. High values in the Archaea and Bacteria columns mean that these genes are extreme outliers in the distribution of CAI scores for all Archaea and Bacteria genes, providing strong evidence they are from *Plagiopyla frontata*.

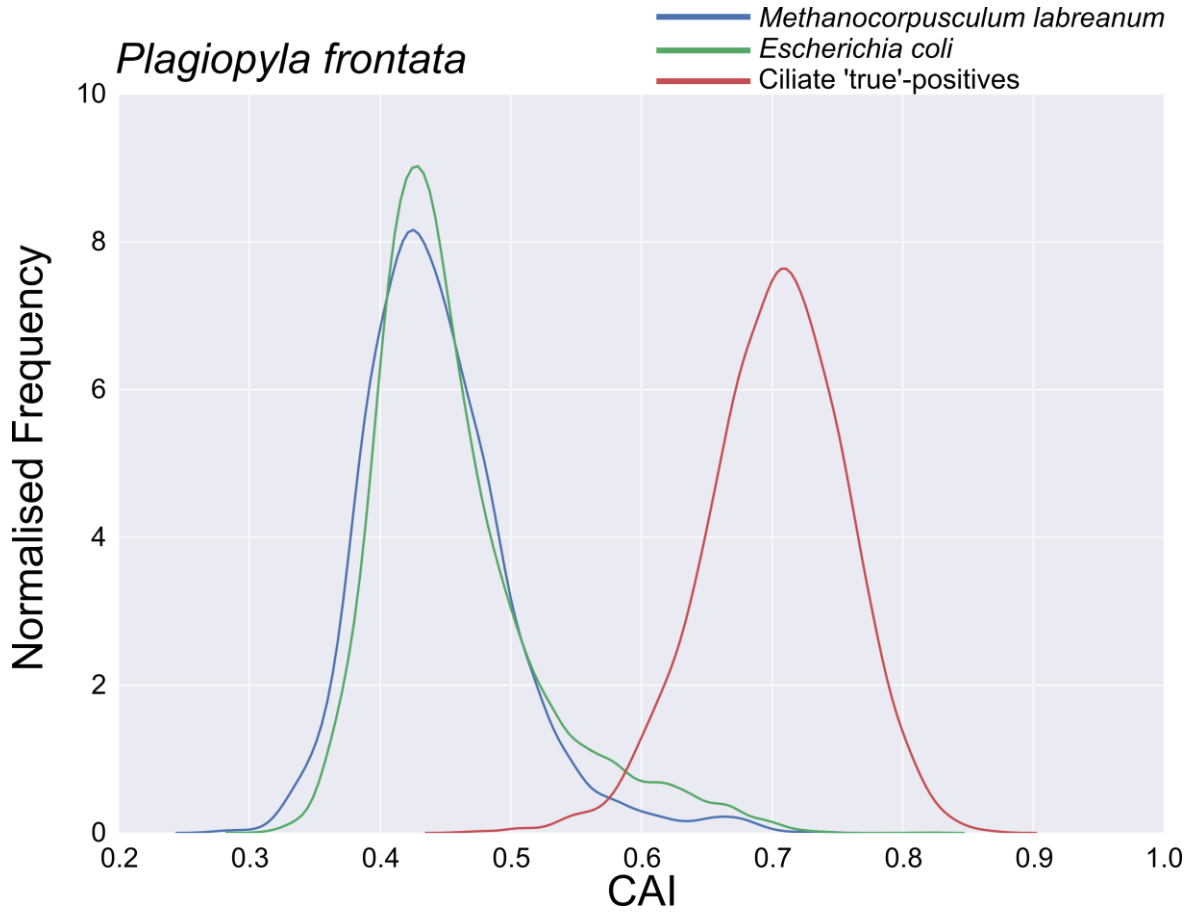


Figure 4.3. f. Normalised distributions of codon adaptation index (CAI) scores calculated for three datasets of genes from *Plagiopyla frontata, Methanocorpusculum labreanum* and *Escherichia coli*. *Plagiopyla frontata* coding sequences were predicted from transcripts that had most significant similarity to ciliate proteins in the nr database (NCBI) based on blastx searches and these were considered the 'true'-positive set of ciliate genes. *Methanocorpusculum labreanum* NC_008942.1 and *Escherichia coli* NC_002695.1 coding sequences were retrieved from their genome sequences, as annotated in their NCBI entries. All CAI scores were calculated using a codon usage table, which was calculated from the total set of 'true'-positive ciliate genes from *Plagiopyla frontata*.

				Percentile Rank		
Species	Enzyme	Transcript IDs	CAI score	Ciliate	Archaea	Bacteria
Trimyema sp.	FeFe-hydrogenase	c8419_g1_i1	0.697	43.3	100	100

Table 4.1. g. CAI scores calculated for coding sequences from *Trimyema* sp. transcripts that were identified as encoding FeFe-hydrogenase enzymes based on blast searches. CAI scores were calculated using a codon usage table that was calculated from a set of 'true'-positive ciliate genes from *Trimyema* sp. The percentile rank of CAI scores for three distributions are shown. Values of 100 in the Archaea and Bacteria columns mean that the CAI scores for these genes fell outside of the distribution of CAI scores for all Archaea and Bacteria genes, providing strong evidence they are from *Trimyema* sp.

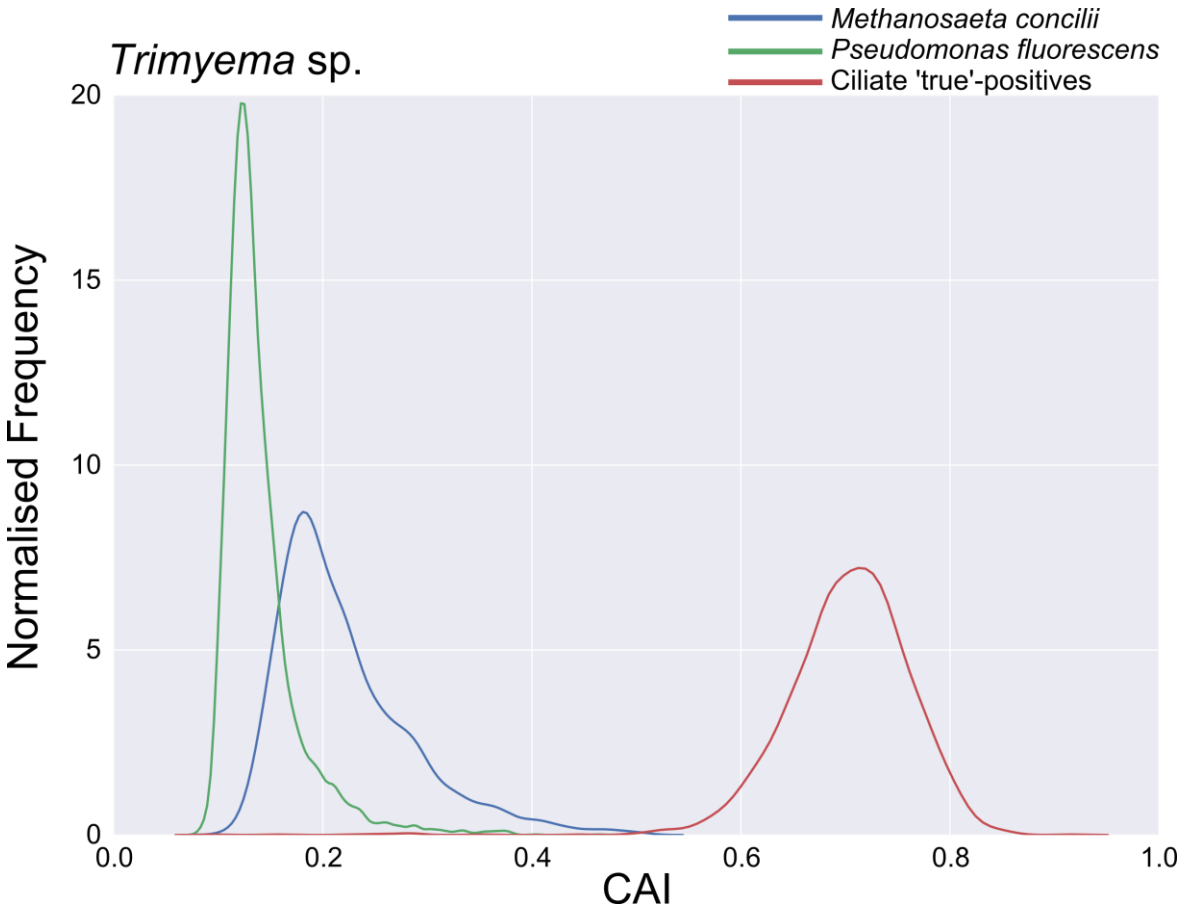


Figure 4.3. g. Normalised distributions of codon adaptation index (CAI) scores calculated for three datasets of genes from *Trimyema* sp., *Methanosaeta concilii* and *Pseudomonas fluorescens*. *Trimyema* sp. coding sequences were predicted from transcripts that had most significant similarity to ciliate proteins in the nr database (NCBI) based on blastx searches and these were considered the 'true'-positive set of ciliate genes. *Methanosaeta concilii* NC_015416.1 and *Pseudomonas fluorescens* NC_016830.1 coding sequences were retrieved from their genome sequences, as annotated in their NCBI entries. All CAI scores were calculated using a codon usage table, which was calculated from the total set of 'true'-positive ciliate genes from *Trimyema* sp.

4.2.3 Detection of N-terminal targeting signals in sequences of hydrogenosomal FeFe-hydrogenases, pyruvate: ferredoxin oxidoreductases and pyruvate: NADP+ oxidoreductases

The mitochondrial processing peptidases (MPP) which cleave N-terminal targeting signals from mitochondrial pre-proteins as they enter the matrix were detected from the transcriptomes of *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata* and this indicates that the hydrogenosomes of these species use N-terminal targeting signals to import proteins. Four different prediction programs (Claros and Vincens, 1996; Small et al., 2004; Emanuelsson et al., 2007; Fukasawa et al., 2015), each of which use a combination of strategies to predict N-terminal targeting signals, were used to analyse the FeFe-hydrogenase and PNO sequences from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. The results of these analyses are shown in Tables 4.2 and 4.3. Due to transcripts being incompletely sequenced, two FeFe-hydrogenase sequences from *Metopus contortus* and *Plagiopyla frontata* and one PNO sequence from *Cyclidium porcatum* had truncated N-termini, which was determined from multiple-sequence alignments with other homologues, and therefore these sequences could not be analysed for targeting signals.

All of the FeFe-hydrogenase sequences from each species, except one from *Cyclidium porcatum* (ID c4008_g2_i1) were predicted to contain N-terminal targeting signals, with a probability greater than 0.5, by the majority of the prediction programs, consistent with them being targeted to the hydrogenosomes of these species. Only one PNO sequence (ID: c4345_g2_i1) from *Cyclidium porcatum* was predicted as having an N-terminal targeting signal with a probability greater than 0.5 by all four programs, whereas the other three PNO sequences (IDs: c15969_g1_i1, c3950_g1_i1, c4339_g1_i1) were predicted to have N-terminal targeting signals with a probability less than 0.5 by most of the programs used. One of these sequences (ID: c4339_g1_i1) was not predicted as having an N-terminal targeting signal with a probability greater than 0.5 by any of the programs.

Since each of the prediction programs use different methods it is difficult to argue which is the most reliable. The most recently developed of these programs,

MitoFates, was previously tested by other authors alongside the other three prediction programs, Mitoprot II, TargetP and Predotar, which were also used in the present study (Fukasawa et al., 2015). The results of this comparison determined that MitoFates predicted N-terminal targeting signals most accurately (Fukasawa et al., 2015), however these comparisons were made only whilst testing the programs on three datasets that contained sequences from metazoa, yeast and plants, separately, and therefore it is unknown whether this would be the most accurate program for predicting N-terminal targeting signals from ciliate sequences. Therefore the proteins shown in Table 4.2 and 4.3 were considered as being targeted to hydrogenosomes if they were predicted as containing an N-terminal targeting signal, with a probability greater than 0.5, by a majority of (*i.e*. three or more) prediction programs.

Using this interpretation of the data, this analysis suggests that each of the FeFe-hydrogenase enzymes sequenced in the present study that have complete N-termini are likely to be targeted to, and therefore function within hydrogenosomes, except the one sequence from *Cyclidium porcatum* (ID c4008_g2_i1). Additionally only one of the PNO enzymes sequenced from *Cyclidium porcatum* is predicted to be targeted to hydrogenosomes (ID c4345_g2_i1), whereas the lack of detectable targeting signals from the other PFO/PNO enzymes suggests that they are more likely to function in the cytosol. These results suggests that *Cyclidium porcatum* has cytosolic and hydrogenosomal copies of both FeFe-hydrogenase and PFO/PNO, although experimental work is needed to be sure. This is similar to what is observed in other anaerobic eukaryotes, such as *Pygsuia biforma* (Stairs et al., 2014) and *Mastigamoeba balamuthi* (Nývltová et al., 2015) each of which seem to possess FeFe-hydrogenase and PFO/PNO copies, with and without N-terminal targeting sequences.

Visual inspection of the well-aligned first eight amino acids of the N-termini of the proteins that are not considered to be targeted to hydrogenosomes (ID c4008_g2_i2 & c4345_g1_i1), revealed some differences to those that are considered to be targeted to hydrogenosomes (Figure 4.4). These differences could be involved in the different targeting predictions of these proteins, although analysis of a much larger number of sequences from the same species would be

required to confidently determine what differences actually affect the targeting location of these proteins. Mitoprot II and TargetP in some cases were able to predict mitochondrial processing peptidase cleavage sites (Figure 4.4), although there were differences between predictions made by the two programs and in some cases neither program could make a prediction. In the case of proteins with and without targeting signals however, both appear to have an N-terminus region (Figure 4.4, highlighted in red) prior to the start of the functional protein domains (Figure 4.4, highlighted in yellow).

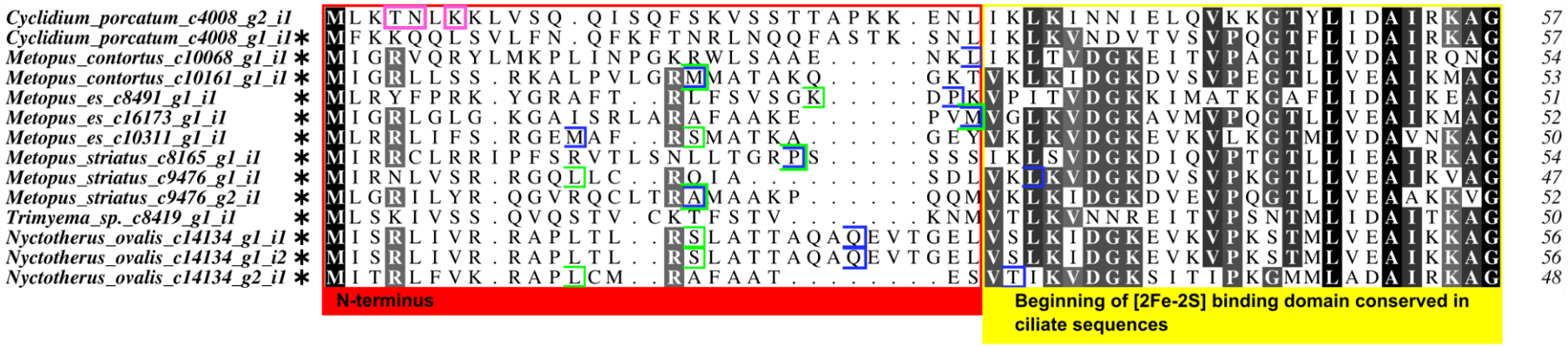
		Targeting signal prediction probabilities			
	Transcript ID	**Mitoprot II**	**TargetP**	**Predotar**	**MitoFates**
Cyclidium porcatum	c4008_g2_i1	**0.9136**	**0.513**	0.5	0.401
	c4008_g1_i1	**0.9883**	**0.771**	0.47	**0.657**
Metopus contortus	c10068_g1_i1	**0.9512**	0.468	**0.88**	**0.758**
	c10161_g1_i1	**0.9701**	**0.759**	**0.8**	**0.984**
Metopus es	c8491_g1_i1	**0.9945**	**0.572**	**0.88**	**0.98**
	c16173_g1_i1	**0.9567**	**0.635**	**0.91**	**0.962**
	c10311_g1_i1	**0.7631**	**0.59**	**0.62**	**0.849**
Metopus striatus	c8165_g1_i1	**0.999**	**0.91**	**0.89**	**0.983**
	c9476_g1_i1	**0.733**	**0.789**	**0.53**	**0.947**
	c9476_g2_i1	**0.9965**	**0.729**	**0.92**	**0.956**
Trimyema sp.	c8419_g1_i1	**0.8815**	0.472	**0.52**	**0.595**
Nyctotherus ovalis	c14134_g1_i1	**0.9937**	**0.73**	**0.91**	**0.988**
	c14134_g1_i2	**0.9952**	**0.73**	**0.91**	**0.987**
	c14134_g2_i1	**0.953**	**0.575**	**0.93**	**0.871**
Metopus contortus	9793_g1_i1	N-terminus missing			
Plagiopyla frontata	c6085_g1_i1				

Table 4.2. Results of N-terminal targeting signal prediction for FeFe-hydrogenase amino acid sequences obtained from ciliates in the present study. The results from the programs Mitoprot II (Claros and Vincens, 1996), TargetP (Emanuelsson et al., 2007), Predotar (Small et al., 2004) and MitoFates (Fukasawa et al., 2015) are shown as probability values. Numbers highlighted in bold indicate a probabilities greater than 0.5 and transcripts highlighted in yellow are predicted as having N-terminal targeting signals based on the majority prediction of the prediction programs.

		Targeting signal prediction probabilities			
	Transcript ID	**Mitoprot II**	**TargetP**	**Predotar**	**MitoFates**
Cyclidium porcatum	c15969_g1_i1	**0.8213**	0.346	0.32	0.049
	c3950_g1_i1	0.3049	0.417	**0.86**	**0.559**
	c4339_g1_i1	0.2037	0.366	0.26	0.374
	c4345_g2_i1	**0.925**	**0.528**	**0.56**	**0.688**
Cyclidium porcatum	c1688_g1_i1	N-terminus missing			

Table 4.3. Results of N-terminal targeting signal prediction for pyruvate: ferredoxin oxidoreductase (PFO) / pyruvate: NADP+ oxidoreductase (PNO) amino acid sequences obtained from *Cyclidium porcatum* in the present study. The results from the programs Mitoprot II (Claros and Vincens, 1996), TargetP (Emanuelsson et al., 2007), Predotar (Small et al., 2004) and MitoFates (Fukasawa et al., 2015) are shown as probability values. Numbers highlighted in bold indicate a probabilities greater than 0.5 and transcripts highlighted in yellow are predicted as having N-terminal targeting signals based on the majority prediction of the prediction programs.

a. Alignment of the N-terminus of ciliate FeFe-hydrogenases



b. Alignment of the N-terminus of *Cyclidium porcatum* PFO/PNO

Figure 4.4. a. Sequence alignment of the N-terminus of FeFe-hydrogenases from *Cyclidium porcatum, Metopus contortus, Metopus es, Metopus striatus, Trimyema* sp. and *Nyctotherus ovalis*. The N-terminus (red) and beginning of the [2Fe-2S] cluster-binding domain (yellow) are highlighted. b. PFO/PNO N-termini from *Cyclidium porcatum*. The N-terminus (red), beginning of the PFO domain (yellow) are highlighted. a & b. Sequences that were predicted as having an N-terminal mitochondrial-like targeting signal are represented by '✱' . Differences in the first eight amino acids that are unique to sequences which were not predicted as having N-terminal targeting signals are highlighted in pink. Mitochondrial processing peptidase cleavage sites as predicted by Mitoprot II (blue) and TargetP (green) are highlighted and where absent indicates that the program was unable to make a prediction. Site conservation between sequences is represented in grey-scale shading, with the most conserved amino acid shaded black if it is well conserved, or white if there is no conservation between sequences.

4.2.4 Phylogenetic analysis of FeFe-hydrogenases from anaerobic ciliates

The FeFe-hydrogenases obtained from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. all have a domain structure that is similar to the FeFe-hydrogenases that were detected from *Nyctotherus ovalis* previously (Akhmanova et al., 1998; Boxma et al., 2007) (Figure 4.1). Each has the conserved H-cluster domain, with one [2Fe-2S] and three [4Fe-4S] cluster binding domains at the N-terminus, and a NuoE and a NuoF domain at the C-terminus. To investigate the relationships between FeFe-hydrogenase enzymes an amino acid sequence alignment was constructed of the conserved regions of the H-cluster domain, including the two well conserved [4Fe-4S] cluster binding domains (Figure 4.1). A phylogeny was then inferred from this alignment using the LG+C60 model (Quang and Gascuel, 2008; Quang et al., 2008), which was chosen as the best fitting model for each of these datasets by IQ-Tree (Figure 4.5).

This analysis resolved the ciliate FeFe-hydrogenase sequences as being monophyletic (100% bootstrap support), forming a sister group to a sequence from the beta-proteobacterium *Sutterella wadworthensis* (90% bootstrap support). *Nyctotherus ovalis* sequences from other studies (Akhmanova et al., 1998; Boxma et al., 2007) also grouped with the *Nyctotherus ovalis* sequences from the present study. The branching pattern of sequences within the ciliate clade is similar to the species tree of these organisms, inferred by 18S rRNA gene sequences (Figure 3.4, Section 3.3.3). This is consistent with a single acquisiton of FeFe-hydrogenases in the common ancestor of these species. There also appears to have been a number of duplications of this enzyme in lineages of some ciliate species, such as *Cyclidium porcatum*, and *Nycotherus ovalis, Metopus contortus, Metopus es* and *Metopus striatus* also seem to contain several copies that are most closely related to one another (bootstrap support 100%). Most of the other eukaryotic sequences included in the alignment form a single clade (79% bootstrap support) that does not appear to be closely related to the clade of ciliate sequences, and this is similar to what other studies have found previously (Horner et al., 2000; Embley et al., 2003). Neither of these eukaryote clades group with any sequences from alpha-proteobacteria, which suggests that there is no evidence that FeFe-hydrogenases in these two clades were acquired from the

mitochondrial endosymbiont despite recent claims that they were (Esposti et al., 2016). A third eukaryotic clade contains sequences from *Paratrimastix pyriformis* and *Monocercomonoides* sp. (98% bootstrap support), which cluster with a sequence from the alpha-proteobacterium *Azospirillum* sp. on a long branch (100% bootstrap support).

The remaining alpha-proteobacteria sequences cluster with sequences from other bacterial groups with high support (91-100% bootstrap support) but are not monophyletic, suggesting lateral gene transfer has affected the relationships between FeFe-hydrogenases from this bacterial lineage and others. Lateral gene transfer is generally accepted to occur frequently in Bacteria (Ochman et al., 2000) and therefore the gene content of modern-day alpha-proteobacteria is likely to be different to that of the mitochondrial endosymbiont (Martin, 1999). This means that if one of the eukaryote clades in the present analysis had acquired FeFe-hydrogenase via the mitochondrial endosymbiont then this clade would not necessarilly group with alpha-proteobacteria. The position of eukaryote sequences in tree topologies in relation to alpha-proteobacteria sequences therefore is less important in analyses such as this one and what is more significant is that eukaryote sequences do not form a single clade and this suggests that they were acquired on more than one occasion from Bacteria in multiple lineages. Therefore this analysis has provided no support for an alpha-proteobacterial origin of the FeFe-hydrogenases found in eukaryotes. It is difficult to reject the possiblity that an FeFe-hydrogenase was present on the genome of the mitochondrial endosymbiont but there is no evidence to suggest it was.

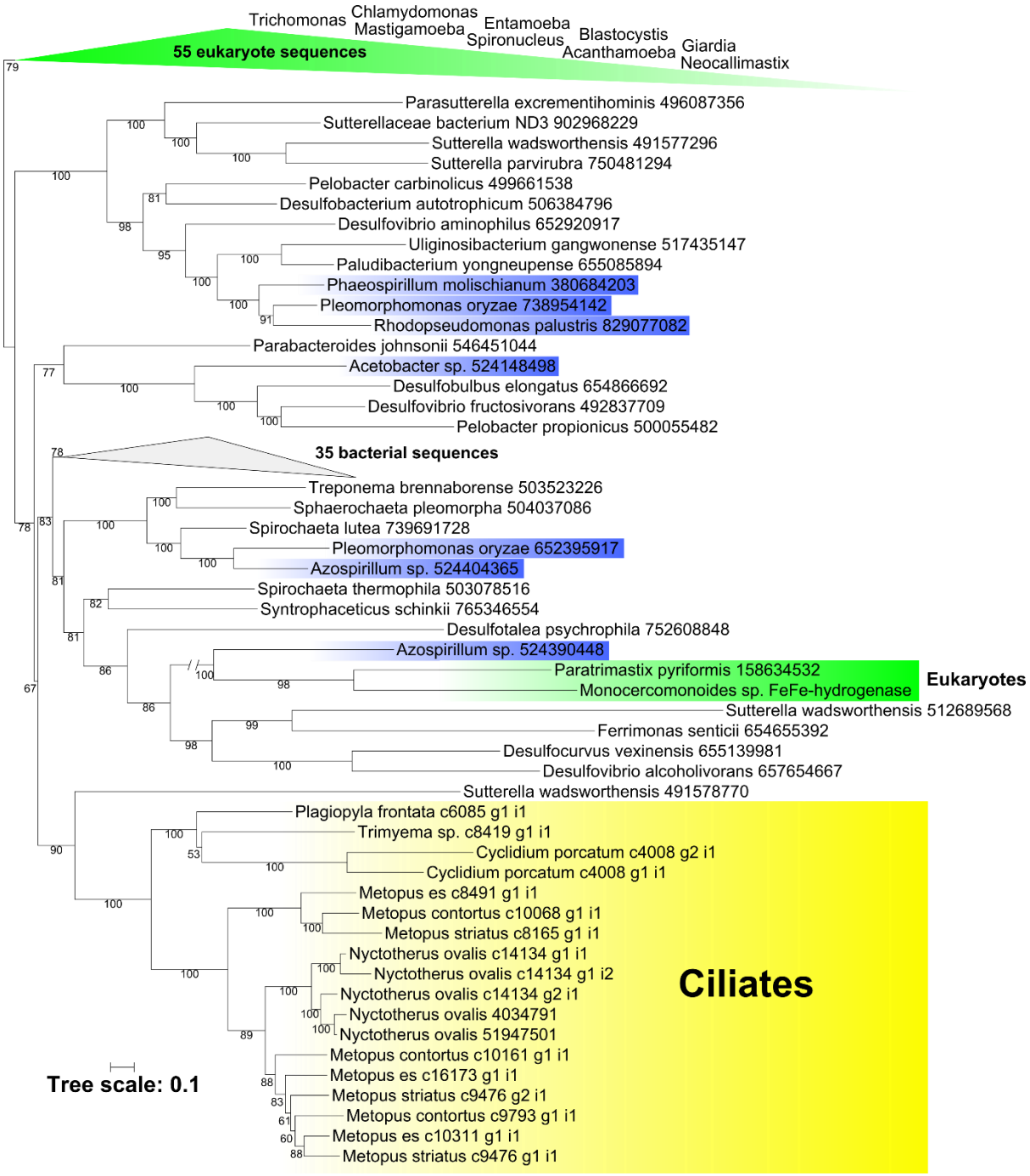


Figure 4.5. Phylogeny of the conserved region of FeFe-hydrogenases sampled from Bacteria and eukaryotes, inferred from an alignment of 464 amino acid sites using the model LG+C60 (Quang and Gascuel, 2008; Quang et al., 2008). Support values were obtained using ultrafast bootstrap (Minh et al., 2013) in the program IQ-TREE (Nguyen et al., 2015). Eukaryotes are highlighted in green, alpha-proteobacteria are highlighted in blue and sequences obtained from Ciliates in the present study are highlighted in yellow. Scale bar represents estimated number of substitutions per site.

4.2.5 Phylogenetic inference of the NuoE and NuoF domains of FeFe-hydrogenases from anaerobic ciliates

Blastp searches using the sequences of the NuoE and NuoF C-terminal domains of the FeFe-hydrogenases from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. as queries suggested that these domains might have a different evolutionary history to the conserved H-cluster of the FeFe-hydrogenase enzyme to which they are fused. Therefore the relationships of these domains with homolgoues from bacteria and eukaryotes were investigated by inferring phylogenies using the LG+C60 model (Quang and Gascuel, 2008; Quang et al., 2008) (Figures 4.6 and 4.7). FeFe-hydrogenase sequences were included in this analysis that were recently discovered from alpha-proteobacteria and from which it was inferred previously that FeFe-hydrogenases in eukaryotes have an alpha-proteobacterial ancestry (Esposti et al., 2016).

In the NuoE analysis (Figure 4.6), the FeFe-hydrogenase NuoE domains sequenced in the present study from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. formed a clade with other *Nyctotherus ovalis* FeFe-hydrogenase NuoE domain sequences in other studies (Boxma et al., 2005; Boxma et al., 2007) (100% bootstrap support). This clade grouped with NuoE sequences from various bacterial species to the exclusion of any sequences of eukaryote 24 kDa subunits of ETC Complex I. By contrast 24 kDa subunits of ETC Complex I from *Nyctotherus ovalis, Metopus es, Metopus contortus* and *Metopus striatus* formed a clade with 24 kDa subunits of ETC Complex I from other ciliates (98% bootstrap support) and this ciliate clade emerged from within a larger clade of eukaryote 24 kDa subunits of ETC Complex I (72% bootstrap support). The clade of eukaryotic 24 kDa subunits of ETC Complex I emerged from within a clade of NuoE sequences from alpha-proteobacteria (100% bootstrap support), consistent with ancestry from the mitochondrial endosymbiont. This suggests that the NuoE domains of ciliate FeFe-hydrogenases have a different origin from the 24 kDa subunits of ETC Complex I.

The NuoF analysis (Figure 4.7) shows a similar pattern to the NuoE analysis. The FeFe-hydrogenase NuoF domains sequenced in the present study from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. formed a clade with other *Nyctotherus ovalis* FeFe-hydrogenase NuoF domain sequences from other studies (Boxma et al., 2005; Boxma et al., 2007) (100% bootstrap support). This clade of ciliate FeFe-hydrogenase NuoF domains grouped with a clade of sequences from species of Bacteroidetes (100% bootstrap support). By contrast the 51 kDa subunits of ETC Complex I sequenced in the present study from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus* and *Cyclidium porcatum* form a clade with other ciliate species (100% bootstrap support) which forms a larger clade with the 51 kDa subunits of ETC Complex I sequences from other eukaryotes (100% bootstrap support). This clade emerges from within a group of NuoF sequences from alpha-proteobacteria (51% bootstrap support).

The main finding of these analyses is that the C-terminal NuoE and NuoF domains of ciliate FeFe-hydrogenase enzymes appear to be distant homologues of the 24 kDa and 51 kDa subunits of ETC Complex I that are found in ciliates and other eukaryotes. Previously it has been claimed that these NuoE and NuoF domains of FeFe-hydrogenases from *Nyctotherus ovalis* were most similar to and assembled from the 24 kDa and 51 kDa subunits of *Nyctotherus ovalis* ETC Complex I, that had become fused to the C-terminus of the FeFe-hydrogenase H-cluster (Akhmanova et al., 1998; Andersson and Kurland, 1999). It has been shown previously by Horner et al. (2000) that this is not the case however, but rather the NuoE and NuoF domains of ciliate FeFe-hydrogenases appear to be more closely related to NuoE and NuoF proteins found in Bacteria. The results of the current study confirm this conclusion and in addition show that the NuoE and NuoF domains of ciliate FeFe-hydrogenases are distant homologues of bona fide 24 kDa and 51 kDa subunits of ciliate ETC Complex I.

Figure 4.6. Phylogenetic inference of the NuoE domain of ciliate FeFe-hydrogenase enzymes, bacterial NuoE and eukaryotic 24 kDa subunits of ETC Complex I, inferred from an alignment of 124 amino acid sites using the model LG+C60 (Quang and Gascuel, 2008; Quang et al., 2008). Support values were obtained using ultrafast bootstrap (Minh et al., 2013) in the program IQ-TREE (Nguyen et al., 2015). Eukaryotes are highlighted in green, alpha-proteobacteria are highlighted in blue and the sequences obtained from Ciliates in the present study are highlighted in yellow. Scale bar represents estimated number of substitutions per site.

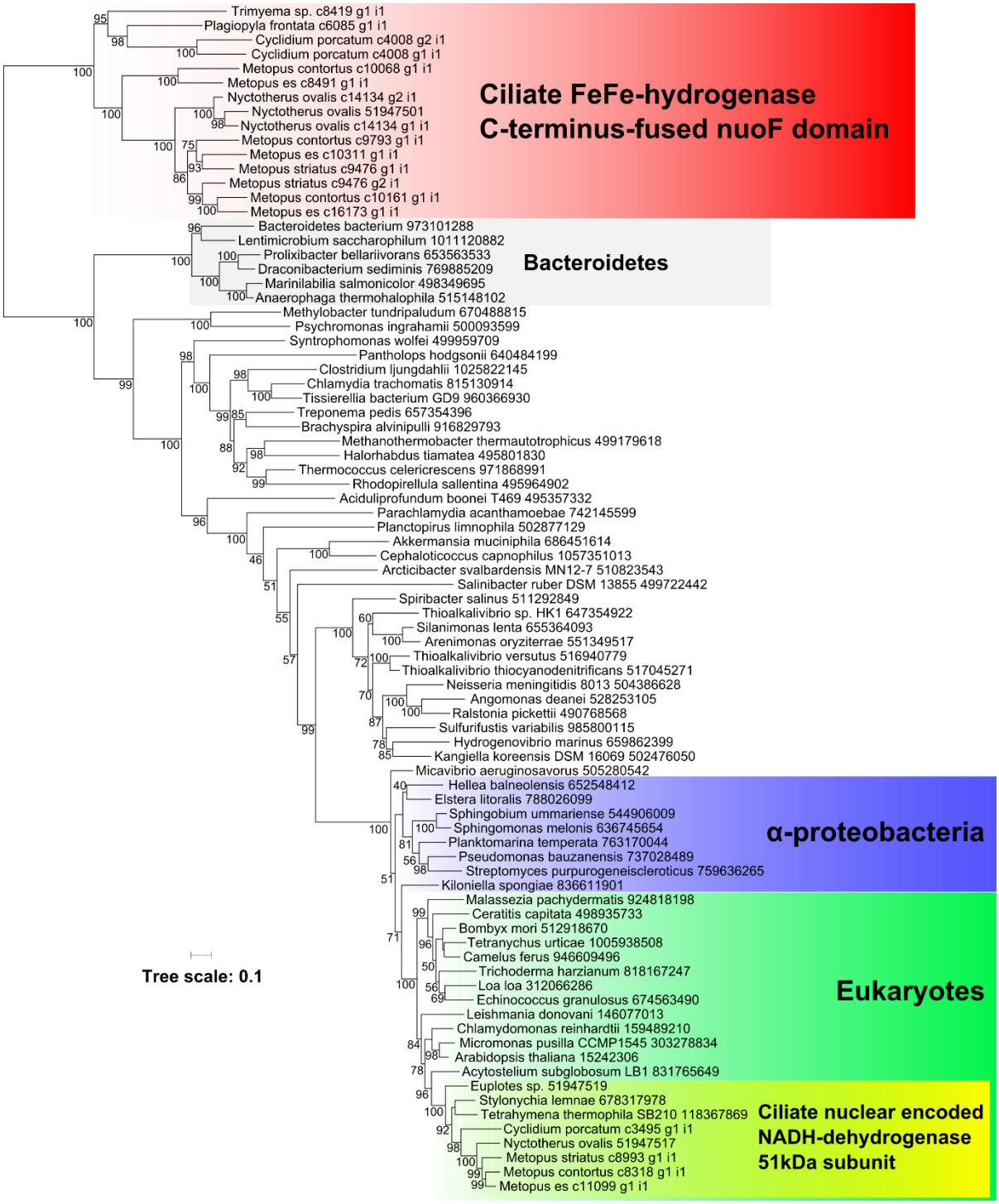


Figure 4.7. Phylogenetic inference of the NuoF domain of ciliate FeFe-hydrogenase enzymes, bacterial NuoF and eukaryotic 51 kDa subunits of ETC Complex I, inferred from an alignment of 352 amino acid sites using the model LG+C60 (Quang and Gascuel, 2008; Quang et al., 2008). Support values were obtained using ultrafast bootstrap (Minh et al., 2013) in the program IQ-TREE (Nguyen et al., 2015). Eukaryotes are highlighted in green, alpha-proteobacteria are highlighted in blue and the sequences obtained from Ciliates in the present study are highlighted in yellow. Scale bar represents estimated number of substitutions per site.

4.2.6 Phylogenetic analysis of pyruvate: ferredoxin oxidoreductase from Cyclidium porcatum

The relationships between the PFO and PNO enzymes sequenced in the present study from *Cyclidium porcatum* and homologues from other eukaryotes and bacteria were investigated by phylogenetic inference from an alignment of amino acid sequences using the LG+C60 model (Quang and Gascuel, 2008; Quang et al., 2008) (figure 4.8). This recovered all of the PFO and PNO sequences obtained from *Cyclidium porcatum* in the present study as a single clade (100% bootstrap support) suggesting that these genes have arisen through multiple duplications after a single origin in this lineage. The *Cyclidium porcatum* clade did not group with sequences from other alveolates in the phylogeny, including *Cryptosporidium parvum* and *Perkinsus marinus*.

The *Cyclidium porcatum* sequences did form part of a larger group which contained all of the eukaryotic PFO and PNO sequences (86% bootstrap support). Although this major PNO clade is not well supported, it is consistent with the hypothesis that the fusion between PFO and a NADPH-cytochrome p450 reductase, which formed PNO happened on just one occasion as has been suggested by previous studies (Rotte et al., 2001). The PNO clade emerges from within the major eukaryote PFO clade, which is separated from the major bacterial PFO clade by a branch with just 25% bootstrap support. This branch however represents the eukaryotic lineage in which the PNO fusion event is likely to have occurred and since PNO has only been reported from eukaryotes (Rotte et al., 2001; Nývltová et al., 2015; Gawryluk et al., 2016), this is consistent with this fusion protein being a eukaryotic invention. The sequences in the major eukaryote PFO clade do not have any significant sequence similarity to the NADPH-cytochrome p450 reductase domain of the PNO enzymes and hence they appear to lack the fusion and have diverged before the PNO fusion event.

Branching from within the major eukaryote PNO clade there are a total of four PFO sequences from *Thalassiosira pseudonana*, *Mastigamoeba balamuthi, Entamoeba histolytica* and *Cyclidium porcatum*. To investigate whether these sequences were truncated PNO sequences, each of their coding sequences were analysed and identified as containing a stop codon, the position of which corresponded to the end of the PFO C-terminus based on multiple sequence

alignments. The PFO sequence from *Thalassiosira pseudonana* within the PNO clade is deep-branching and not well supported (43% bootstrap support) and because of this it is possible that this sequence may actually branch within the major eukaryote PFO clade, and could have diverged before the PNO fusion event. However, the branches leading to the PFO sequences from *Mastigamoeba balamuthi, Entamoeba histolytica* and *Cyclidium porcatum* within the PNO clade have much better support (96%, 99% and 97% bootstrap supports, respectively). The most likely explaination for the presence of these PFO enzymes in the PNO clade is that these enzymes have undergone a secondary loss of the NADPH-cytochrome p450 reductase domains and in doing so have reverted back to an ancestral domain structure similar to that of the PFO enzymes in the major eukaryote PFO domain. It could be envisaged that such a PNO to PFO-reversion is more likely to occur in an organism that has multiple copies of genes encoding PNO. This would allow such an organism to retain the enzymatic abilities of PNO encoded by one of the gene copies, using $NADP^+$ as an electron acceptor for pyruvate oxidation (Inui et al., 1987), thereby not suffering any selective costs that would be associated with its loss. Then, losing the NADPH-cytochrome p450 reductase domain from another gene copy could be of selective benefit if the resulting protein could function in the same way as an ancestral PFO, thereby expanding the repertoir of substrates that can used as an electron sink to include ferredoxin (Uyeda and Rabinowitz, 1971). Evidence in this analysis from *Mastigamoeba balamuthi* and *Cyclidium porcatum* support this, since they both possess at least one copy of PNO, as well as what appear to be reverted PFO.

Emerging from within the major bacterial PFO clade are two smaller eukaryote clades, one of which contains just a single sequence from *Paratrimastix pyriformis* (92% bootstrap support). The second of these clades contains sequences from *Beauveria bassiana* and *Trichuris trichiura* (100% bootstrap support) and emerges from within a bacterial clade that contains five sequences from alpha-proteobacteria species, although they are separated by a small number of branches leading to other bacterial sequences. The sequences in the *Paratrimastix pyriformis* PFO clade and the major eukaryote PFO clade appear to be distantly related to the sequences from alpha-proteobacteria and this therefore suggests that PFO enzymes were acquired in eukaryotes by lateral gene transfer from Bacteria on more than one occassion. Early phylogenetic studies of PFO

suggested that this enzyme is monophyletic in eukaryotes but found no support that they were acquired via the mitochondria endosymbiont since they did not group with PFO from alpha-proteobacteria (Horner et al., 1999). The present study also finds no support for an alpha-proteobacterial ancestry of eukaryote PFO but due to expanded sampling it appears that PFO in eukaryotes may not be monophyletic. This agrees with a recent study which suggested that eukaryote PFO sequences may resolve as multiple clades, although monophyly of eukaryote PFO was unable to be rejected (Hug et al., 2010).

The relationships between sequences that diverge within the major eukaryotic PFO and PNO clades are not consistent with accepted relationships between the species from which they were identified (Hampl et al., 2009; Burki, 2014), although many of the deeper branches are weakly supported. For example the strongly supported sister group relationship between the PNO sequences from *Cyclidium porcatum* and the opisthokonts *Lingula anatina*, *Priapulus caudatus* and *Capitella teleta* (100% bootstrap support), is not reflected by a close relationship between alveolates and opisthokonts (Burki, 2014), suggesting that lateral gene transfer has had a role in the origin of PFO/PNO in these clades. In contrast to this, the clade consisting of sequences from chlorophyte species, *Monoraphidium neglectum*, *Chlorella variabilis*, *Volvox carteri*, *Gonium pectorale* and *Chlamydomonas reinhardtii* is well supported (100% bootstrap support) and this is suggestive of PFO being acquired in the common ancestor of these species and inherited vertically. This can also be observed in the PNO clade that consists of the amoebezoa *Mastigamoeba balamuthi* and *Entamoeba histolytica* too, suggesting a common ancestry of these sequences. Together these patterns provide evidence for a lateral mode of transfer for genes encoding PFO and PNO in some eukaryote lineages. Further evidence for this is provided by the two PFO and one PNO sequences from *Mastigamoeba balamuthi* that were included in this phylogenetic analysis. The relationships between these three sequences and those from other species suggest that one of them evolved before the PNO fusion event, whereas other two seem to have evolved after it. If this is the case then it is most likely that at least one of these sequences was acquired by *Mastigamoeba balamuthi*, or one of its ancestors, by lateral gene transfer. This would be true for any eukaryote species that possesses an ancestral-type PFO, diverging before the PNO fusion event, as well as a PNO or PFO that diverged after the PNO

fusion event. Therefore the PNO fusion event can act as a marker-point in identifying some lateral gene transfers in the evolution of these enzymes.

Other evidence for lateral gene transfer of these enzymes can be observed from the PNO sequences from the excavates *Euglena gracilis* and *Peranema trichophorum*, which together form a clade (100% bootstrap support) suggesting that PNO has been acquired once by the euglenozoa lineage. This clade however, is not closely related to the PFO sequences from *Giardia, Trichomonas vaginalis*, and *Spironucleus,* which are also excavates. By contrast, if these enzymes had evolved by vertical inheritence in the excavates then all excavate species would be expected to form a single clade.

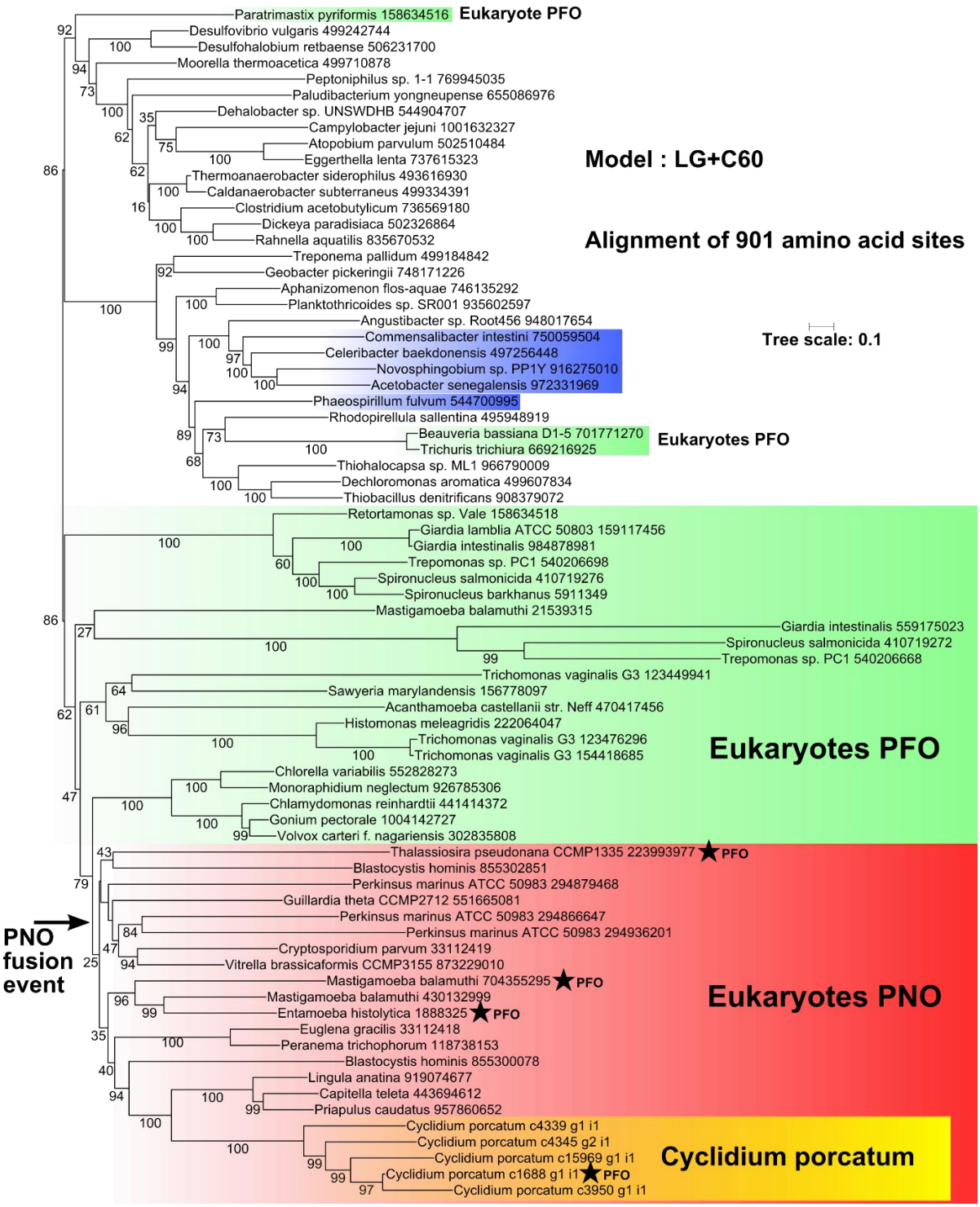


Figure 4.8. Phylogeny of PFO and the PFO domain of PNO enzymes (without the NADPH-cytochrome p450 domain), sampled from Bacteria and eukaryotes, inferred from an alignment of 901 amino acid sites using the model LG+C60 (Quang and Gascuel, 2008; Quang et al., 2008). Support values were obtained using ultrafast bootstrap (Minh et al., 2013) in the program IQ-TREE (Nguyen et al., 2015). Eukaryote PFO sequences are highlighted in green, Eukaryote PNO sequences are highlighted in red, alpha-proteobacteria are highlighted in blue and sequences obtained from *Cyclidium porcatum* in the present study are highlighted in yellow. The branch on which the PNO fusion event is likely to have occurred is shown and PFO sequences that are thought to have evolved after this event by loss of NADPH-cytochrome p450 reductase domains are highlighted (★ PFO). Scale bar represents estimated number of substitutions per site.

4.3 Discussion

This study has expanded the sampling of genes encoding key anaerobic metabolism enzymes from anaerobic ciliates and shown that they have a codon usage that is typical of other genes that are encoded by the ciliate genomes. FeFe-hydrogenase sequences that contain N-terminal targeting signals were sequenced from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. These sequences appear to share a common domain structure not found in other eukaryotes and are monophyletic in our analyses. The ciliate sequences do not group with the FeFe-hydrogenases of other eukaryotes, in agreement with findings from previous studies (Horner et al., 2000; Embley et al., 2003). We found no evidence that ciliate FeFe-hydrogenases originated from alpha-proteobacteria. This supports a more recent acquisition of FeFe-hydrogenase enzymes in ciliates from Bacteria by lateral gene transfer rather than them having been retained in the ancestors of ciliates since the mitochondrial endosymbiosis. The C-terminal NuoE and NuoF domains of the FeFe-hydrogenases obtained from these ciliates are closely related to homologues in Bacteria but are only distantly related to homologues from alpha-proteobacteria and the 24 kDa and 51 kDa subunits of ETC Complex I found in eukaryotes, including those from ciliates.

PFO and PNO sequences were also sequenced and analysed from *Cyclidium porcatum* and they appear to be monophyletic, indicating that they have evolved by gene duplications in this lineage. An N-terminal targeting signal could only be predicted for one PNO homologue from *Cyclidium porcatum* and appeared to be absent from others, suggesting that PNO may function in the cytosol as well as the hydrogenosomes of this species. Furthermore, the *Cyclidium porcatum* PFO appears to have evolved from a PNO by loss of the NADPH-cytochrome p450 reductase domain, a processes that appears to have occurred in other eukaryotes but since the N-terminus of this PFO is truncated it is unclear whether it is located to the cytosol or hydrogenosomes of this species. The PFO/PNO sequences from *Cyclidium porcatum* do not appear to be closely related to sequences from other alveolates, including *Cryptosporidium parvum* and *Perkinsus marinus*, and instead cluster strongly with opisthokonts, *Lingula anatina*, *Capitella teleta* and *Priapulus caudatus*, suggesting that this enzyme was acquired

by lateral gene transfer in the ancestor of *Cyclidium porcatum* and possibly from an opisthokont.

Recently a number of FeFe-hydrogenase sequences were identified from several alpha-proteobacteria species and the discovery of these has been used to suggest an alpha-proteobacteria ancestry of FeFe-hydrogenases in eukaryotes (Esposti et al., 2016). This was used to infer that FeFe-hydrogenases were acquired once in eukaryotes, from the alpha-proteobacterium mitochondrial endosymbiont and then underwent differential loss in various lineages (Esposti et al., 2016). This interpretation would therefore be consistent with the Hydrogen Hypothesis (Martin and Müller, 1998). The relevant sequences were included in the analyses in the present study, however in contrast to Esposti et al. (2016) no evidence was found to suggest that they are the sister group of eukaryotes. Rather, since eukaryotes resolved into three separate clades, this indicates that they have probably been acquired on at least three occasions in eukaryotes. Additionally, only one of these clades, which contains sequences from two excavates, *Paratrimastix pyriformis* and *Monocercomonoides* sp., branches with a sequence from an alpha-proteobacteria. Yet FeFe-hydrogenases from other excavates like *Trichomonas vaginalis*, do not group within this clade which is inconsistent with a single acquisition in this lineage.

Previously it has been suggested that lateral gene transfer in bacteria has obscured our view of the ancestry of anaerobic metabolism enzymes from the mitochondrial endosymbiont (Martin, 1999; Rotte et al., 2001; Muller et al., 2012). This however does not explain why in phylogenies of anaerobic metabolic enzymes, such as FeFe-hydrogenases and PFO/PNO, sequences from eukaryotes are polyphyletic, as has been shown by the present study and others (Horner et al., 2000; Hug et al., 2010). A more parsimonious explanation for this is that some of these sequences were acquired from Bacteria by lateral gene transfer and therefore have evolved in eukaryotes by a combination of vertical and lateral inheritance.

To explain the patchy distribution of anaerobic metabolism enzymes in eukaryotes in a manner that is consistent with the Hydrogen Hypothesis (Martin and Müller, 1998) it has been suggested that these enzymes were acquired from the mitochondrial endosymbiont and were then differentially lost in different

eukaryote lineages (Muller et al., 2012). Given the disagreement between the phylogenies of these enzymes and the species-phylogeny of the organisms they are found in, it seems possible that anaerobic metabolism enzymes are being inherited by eukaryote-to-eukaryote lateral gene transfer in some lineages, which could also explain their patchy distribution. In sum, the evidence appears to favour multiple acquisitions of anaerobic metabolic enzymes in eukaryotes from Bacteria, and a lateral mode of inheritance between eukaryotes, and this is ultimately inconsistent with the Hydrogen Hypothesis (Martin and Müller, 1998).

Chapter 5. General Discussion and conclusions

Anaerobic ciliates from three separate clades, Armophorea, Oligohymenophorea and Plagiopylea, were isolated into culture and investigated using whole genome sequencing and transcriptomics to provide the first detailed molecular insights into the metabolism of their hydrogenosomes. Mitochondrial (hydrogenosome) genomes were detected and partially sequenced from *Metopus contortus, Metopus es, Metopus striatus* and *Cyclidium porcatum*, and a new hydrogenosome genome was sequenced from a new isolate of *Nyctotherus ovalis* that is distinct from those studied previously (Boxma et al., 2005; de Graaf et al., 2011). These data provide strong evidence for a common mitochondrial ancestry of the hydrogenosomes in these particular ciliates. The hydrogenosomes of *Plagiopyla frontata* and *Trimyema* sp. appear to have lost their mitochondrial genomes but metabolic reconstructions for the hydrogenosomes from *Plagiopyla frontata* revealed that they have retained features of canonical mitochondria. These data provide strong evidence that ciliate hydrogenosomes have evolved from mitochondria on at least three separate occasions in the ciliate tree.

The gene content of the hydrogenosome genome sequences from *Nyctotherus ovalis, Metopus contortus, Metopus es* and *Metopus striatus* is reduced compared to the mitochondrial genomes of aerobic ciliates. Thus, no genes were detected for components of F_1F_0 ATP synthase or ETC complexes III and IV for any species, but genes were detected for components of ETC Complex I, some mitochondrial ribosomal proteins and rRNA. The retention of the latter genes, which are needed for the biogenesis of mitochondrial ribosomes, provides direct evidence for a capacity for protein synthesis inside the organelles. The ETC Complex I subunits encoded by the hydrogenosome genomes of these species are predicted to form part of the P-module, the inner membrane-embedded arm of the complex that transports protons across the inner mitochondrial membrane (Brandt, 2006). This suggests that a transmembrane potential can be maintained across the inner membrane of these hydrogenosomes. In canonical mitochondria the membrane potential is needed for ATP production using F_1F_0 ATP synthase, which these ciliates appear to have lost, but also for the import of nuclear encoded proteins into the organelle. It appears likely that this requirement provides the

selective pressure for the retention of genes for Complex I on the hydrogenosome genomes of these ciliates.

The reconstructed hydrogenosome metabolism of *Metopus contortus*, where we recovered an almost complete organelle genome, is similar to the metabolism previously inferred for the hydrogenosomes of its close relative *Nyctotherus ovalis* (de Graaf et al., 2011). Both ciliates have most of the same genes for ETC Complex I, ETC Complex II and the cytochrome *c*1 subunit of ETC Complex III and both lack genes for ETC Complex IV and F_1F_0 ATP synthase. *Nyctotherus ovalis* and *Metopus contortus* also retained genes needed for fumarate reduction and ATP production by substrate-level phosphorylation. Both have retained nuclear encoded genes for all four subunits of PDH, suggesting that they can metabolise pyruvate in a similar way to aerobic ciliates such as *Tetrahymena thermophila* (Smith et al., 2007). More genes for the essential mitochondrial Fe-S cluster biogenesis pathway were identified from *Metopus contortus* than were identified previously from *Nyctotherus ovalis*, as well as more genes for the mitochondrial import complexes and for the mitochondrial carrier family (MCF) proteins that mediate transport of substrates including ATP, into and out of mitochondria. These differences may be due in part, for example to explain the missing essential components of the Fe-S cluster biogenesis pathway, to the incomplete nature of the sequence data for both species.

Despite generating less nuclear and organelle genome data for *Cyclidium porcatum*, the sequences we detected are consistent with it retaining the most complete ETC. Thus, genes were detected from this species for some of the proteins from mitochondrial ETC Complexes I, II, III and IV and the F_1F_0 ATP synthase. *Cyclidium porcatum* was the only anaerobic ciliate for which any components of the F_1F_0 ATP synthase was identified, and it is likely that *Cyclidium porcatum* can use this to make ATP. We also detected some of the genes that would be required for ATP production by substrate-level phosphorylation. In mitochondria the production of ATP is powered by a proton gradient across the inner membrane, generated by the transfer of electrons through proton-pumping components of the ETC (Mitchell, 1961). The proton pumping ETC complex that seems to be most complete in *Cyclidium porcatum* is ETC Complex I and, as discussed above, this could create a transmembrane proton gradient for use by an

F_1F_0 ATP synthase. *Cyclidium porcatum* also has some of the genes required for the reduction of fumarate and an AOX, which could both potentially act as terminal electron acceptors for its partial ETC, under favourable conditions. An almost complete set of genes were identified from *Cyclidium porcatum* for the mitochondrial Fe-S cluster biogenesis pathway and the largest number (19) of MCF genes among the studied species. The relative abundance of MCF genes, which generally have different transport properties from each other, is consistent with the hypothesis that the hydrogenosomes of *Cyclidium porcatum* are the least reduced of the hydrogenosomes studied here.

Relatively few genes for mitochondrial proteins were detected from *Plagiopyla frontata*, although it should be noted that the transcriptome data from this species was also the least abundant. Nevertheless the data suggests that there is no ETC present in this species, since not a single subunit of any of the ETC complexes were found, and this is consistent with our inability to detect any evidence for the retention of a mitochondrial (hydrogenosome) genome. *Plagiopyla frontata* has retained some of the genes required for substrate-level phosphorylation, which it may use to make ATP. Consistent with this, one of the small number of MCF genes detected was an ATP/ADP translocase. *Plagiopyla frontata* has genes for PDH, as do the anaerobic ciliates *Metopus contortus* and *Nyctotherus ovalis*, which it could use to oxidise pyruvate produced by the complete glycolysis pathway found in this species. An incomplete Fe-S cluster biogenesis pathway and failure to detect important components of the mitochondrial-type import machinery, including the outer and inner membrane translocases, are consistent with incomplete coverage of the proteome of this species.

Cyclidium porcatum, *Metopus contortus* and *Plagiopyla frontata* have all retained components of the mitochondrial processing peptidase (MPP) complex, which in mitochondria cleaves N-terminal targeting signals from proteins targeted to the mitochondrial matrix (Ieva et al., 2013). This is consistent with the retention of other components of the mitochondrial import pathway. It is also consistent with the detection of N-terminal targeting signals on the FeFe-hydrogenases and other nuclear encoded proteins that must function inside the organelles if they are to make H_2. It thus appears that the hydrogenosomes of these ciliates use similar

mechanisms as classical mitochondria to import proteins from the cytosol. The MCF proteins found in *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata* suggests that the hydrogenosomes of these species can exchange a range of substrates with the cytosol, their diversity suggesting that there are still gaps to fill in our knowledge of hydrogenosome function. All three species have an ATP/ADP translocase that would be able to import or export ATP to or from the cytosol.

The FeFe-hydrogenases detected from all of the ciliates have similar domain structures, each containing C-terminal NuoE and NuoF domains that are not of mitochondrial origin. This structure is so far unique among eukaryotic hydrogenases. These domains may enable the ciliate FeFe-hydrogenases to use NADH as a substrate for the reduction of protons to H_2, as has been shown previously for the FeFe-hydrogenase of the bacterium *Thermotoga maritima* (Schut and Adams, 2009). In contrast to the other ciliates, which have retained PDH, *Cyclidium porcatum* has acquired PFO genes and it also has a predicted mitochondrial ferredoxin. In *Trichomonas vaginalis* PFO has been shown to reduce ferredoxin (Gorrell et al., 1984). *Metopus contortus* and *Nyctotherus ovalis* have retained the 24 kDa and 51 kDa subunits of ETC Complex I which were also shown to be able to reduce ferredoxin in *Trichomonas vaginalis* (Hrdy et al., 2004) and we detected mitochondrial ferredoxin for *Metopus contortus* and *Cyclidium porcatum*. Based upon these data it appears possible that the FeFe-hydrogenases of *Cyclidium porcatum, Metopus contortus* and *Nyctotherus ovalis* can all use reduced ferredoxin as a substrate for H_2 production. The source of reducing power used for H_2 production is less clear for *Plagiopyla frontata* because no genes for PFO or the 24 kDa and 51 kDa subunits of ETC complex I were detected for this species.

The evolutionary origins of the FeFe-hydrogenases from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus, Cyclidium porcatum, Plagiopyla frontata* and *Trimyema* sp. and the PFO and PNO sequences from *Cyclidium porcatum* were investigated using phylogenetic trees. The individual domains of the FeFe-hydrogenases, namely the H-cluster, NuoE and NuoF domains, were analysed separately. The phylogeny of the H-cluster domain, which is well conserved in FeFe-hydrogenases from both Bacteria and Eukaryotes

recovered all the ciliate sequences as a monophyletic group. This clade was the sister group to a sequence from the beta-proteobacterium *Sutterella wadsworthensis*, suggesting a gene transfer from this group of bacteria to ciliates. Interestingly, the *Sutterella wadsworthensis* sequence does not contain the C-terminal NuoE and NuoF domains found in the ciliate FeFe-hydrogenases. This suggests that the domains were added to the ciliate sequences in a single event postdating the split from *Sutterella wadsworthensis*, or that *Sutterella wadsworthensis* has lost them. Investigating the occurrence and structure of the hydrogenases of the close relatives of *Sutterella wadsworthensis* may help to resolve this. The observation that the ciliate FeFe-hydrogenases are monophyletic and robustly separated from the other eukaryote sequences strongly suggests that they were acquired through a single independent lateral gene transfer. Since there is no evidence that the topology of the ciliate FeFe-hydrogenase tree is incompatible with ciliate species relationships, it appears that acquisition was followed by vertical inheritance. These data provide no direct support for an origin of ciliate FeFe-hydrogenases from the mitochondrial endosymbiont as predicted by the Hydrogen Hypothesis (Martin and Müller, 1998).

Phylogenetic analysis of the ciliate FeFe-hydrogenase NuoE and NuoF domains also recovered these sequences as monophyletic suggesting that this tripartite domain structure was present in their common ancestor. The NuoE and NuoF domains group with sequences from different Bacteria, so potentially have different origins, and they do not group with mitochondrial homologues. By contrast, the 24 kDa and 51 kDa subunits of ETC Complex I from the anaerobic ciliates did group with mitochondrial homologues from other eukaryotes. This clade emerges from within the alpha-proteobacteria, consistent with the genes for the 24 kDa and 51 kDa subunits of ETC Complex I being acquired from the mitochondrial endosymbiont.

All of the PFO and PNO sequences from *Cyclidium porcatum* were recovered as a monophyletic group, consistent with gene duplication being the source of their diversity. The *Cyclidium porcatum* sequences seem to have been acquired by lateral gene transfer, possibly from an opisthokont. Consistent with this idea, the PFO and PNO sequences from *Cyclidium porcatum* are not closely related to those found in other alveolates, including *Cryptosporidium parvum* and

Perkinsus marinus, suggesting that they do not share a common origin. The *Cyclidium porcatum* PFO sequence seems to have evolved from a PNO by loss of the NADPH-cytochrome p450 reductase domain. Similar reversions appear to have occurred in other eukaryote species that contain a mixture of PFO and PNO sequences.

The production of H_2 by ciliate hydrogenosomes facilitates the growth of endosymbiotic methanogen Archaea, which were detected in all of the ciliates studied. The endosymbionts were identified from isolates of *Metopus contortus*, *Nyctotherus ovalis* and *Trimyema* sp. using a combination of sequencing and *in situ* probing. These data show that different types of methanogens can colonise anaerobic ciliates. Intriguingly, the endosymbiont sequences for *Metopus contortus* and *Trimyema* sp. were the same as those detected in the same species isolated from different locations over two decades ago. This suggests a degree of stability in the associations over this time period. By contrast, comparison of the phylogenies for the endosymbionts and host ciliates provided no compelling evidence for long-term co-speciation between endosymbionts and hosts.

The present study has investigated the molecular basis of hydrogenosome metabolism in anaerobic ciliates. Future work would benefit from more complete sequence data for the studied ciliates, including whole genome sequences, to provide a more complete picture of the metabolic processes in these organisms. Due to the relative ease at which single-cell transcriptomes can now be produced for anaerobic ciliates and other single-celled eukaryotes, as is shown by the data in this thesis, it is possible to gain extensive molecular data without the need for culturing, which in many cases can be labour intensive and time consuming. It is therefore likely that such methods will be used in the future, if not already, to facilitate large-scale sampling efforts and production of transcriptomes for a varied and taxonomically diverse of single-celled eukaryotes. This would provide a wealth of knowledge for better understanding the relationships between the main eukaryote phyla and refine our current placement for the root of the eukaryote tree. With regards to anaerobic ciliates, improved large-scale sampling would likely help to resolve the origin of FeFe-hydrogenase in this group of organisms and additional experimental work is needed to localise proteins and to further characterise the organelle proteomes. Such data could have significant

implications for our understanding of eukaryote evolution and may help to revise our ideas of how they first evolved.

It is already clear from the present data that ciliate hydrogenosomes have evolved convergently in different anaerobic clades from aerobic mitochondria, and that they represent different stages of reductive evolution. This thesis has also provided some strong evidence that lateral gene transfer has played an important role in the evolution of ciliate hydrogenosomes, particularly with regards to their FeFe-hydrogenases and the PFO found in *Cyclidium porcatum*. This is seemingly not unusual for eukaryotes adapting to low oxygen environments (Nývltová et al., 2015; Eme et al., 2017), and provides further evidence against the opinions of some researchers that believe that lateral gene transfer has not had a significant impact on the evolution of eukaryotes (Ku and Martin, 2016). Understanding, the apparently contradictory data on the stability of the associations between endosymbiotic methanogens and their ciliate hosts would benefit from sampling of a greater number of such consortia, particularly among the Armophorea, which contains a large number of closely related and exclusively anaerobic genera (including *Metopus* and *Nyctotherus*), which also contain endosymbiotic methanogens, the species identities of which remain unknown for the vast majority of them. In particular, sequencing the genomes of endosymbiotic methanogens from diverse anaerobic ciliates, would facilitate comparisons both between them, as well as with other free-living methanogens. Such studies might reveal both generalised and lineage-specific adaptations to endosymbiotic lifestyles in methanogenic Archaea, which currently remain less studied than their endosymbiotic Bacteria counter-parts and otherwise seems to be a relatively rare trait in Archaea.

Appendices

The data in the following appendices are made available on the CD provided with this thesis.

Appendix A

Annotated hydrogenosome genome contigs from *Nyctotherus ovalis, Metopus contortus, Metopus es, Metopus striatus* and *Cyclidium porcatum* in GenBank format, corresponding to Figures 3.9—3.13.

Appendix B

Tables of transcripts identified in metabolic reconstructions as genes functioning in the hydrogenosome metabolic pathways of *Cyclidium porcatum, Metopus contortus* and *Plagiopyla frontata*, summarised in Figures 3.16—3.18.

These include raw transcript sequences, predicted coding sequences, translated protein sequences, probability values of predicted N-terminal targeting signals and codon adaptation index (CAI) scores, which correspond to ciliate CAI distributions in Figure 4.3.

Appendix C

Transcript and translated protein sequences of FeFe-hydrogenases, Pyruvate: ferredoxin oxidoreductase and Pyruvate: $NADP^+$ oxidoreductases discussed in Chapter 4.

Appendix D

Codon usage tables that were calculated from the coding sequences of transcripts with best blast hits to ciliate genes. These tables were used to calculate the CAI scores that are plotted as distributions in Figure 4.3.

References

Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Molecular phylogenetics and evolution **29**: 380-395.

Adams M, Eccleston E, Howard JB (1989) Iron-sulfur clusters of hydrogenase I and hydrogenase II of *Clostridium pasteurianum*. Proceedings of the National Academy of Sciences **86**: 4932-4936.

Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V (2012) The revised classification of eukaryotes. Journal of Eukaryotic Microbiology **59**: 429-514.

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

Akhmanova A, Voncken FG, Hosea KM, Harhangi H, Keltjens JT, Op Den Camp HJ, Vogels GD, Hackstein JH (1999) A hydrogenosome with pyruvate formate-lyase: Anaerobic chytrid fungi use an alternative route for pyruvate catabolism. Molecular Microbiology **32**: 1103-1114.

Albracht SP (1993) Intimate relationships of the large and the small subunits of all nickel hydrogenases with two nuclear-encoded subunits of mitochondrial NADH: ubiquinone oxidoreductase. Biochimica et Biophysica Acta (BBA)-Bioenergetics **1144**: 221-224.

Allen JF (1993) Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. Journal of Theoretical Biology **165**: 609-631.

Allen JF (2003) The function of genomes in bioenergetic organelles. Philosophical Transactions of the Royal Society of London B: Biological Sciences **358**: 19-38.

Allen JF (2015) Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. Proceedings of the National Academy of Sciences **112**: 10231-10238.

Allen JF, Martin WF (2016) Why Have Organelles Retained Genomes? Cell Systems **2**: 70-72.

Allen RD, Schroeder CC, Fok AK (1989) An investigation of mitochondrial inner membranes by rapid-freeze deep-etch techniques. The Journal of Cell Biology **108**: 2233-2240.

Alsmark C, Foster PG, Sicheritz-Ponten T, Nakjang S, Embley TM, Hirt RP (2013) Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. Genome Biology **14**: 1.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Applied and Environmental Microbiology **56**: 1919-1925.

Anderson S, Bankier AT, Barrell BG, De Bruijn M, Coulson AR, Drouin J, Eperon I, Nierlich D, Roe BA, Sanger F (1981) Sequence and organization of the human mitochondrial genome. Nature.

Andersson JO (2009) Gene transfer and diversification of microbial eukaryotes. Annual Review of Microbiology **63**: 177-193.

Andersson SG, Kurland CG (1999) Origins of mitochondria and hydrogenosomes. Current Opinion in Microbiology **2**: 535-541.

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

Andrews S (2010) FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arnold I, Pfeiffer K, Neupert W, Stuart RA, Schägger H (1998) Yeast mitochondrial F1F0-ATP synthase exists as a dimer: identification of three dimer-specific subunits. The EMBO Journal **17**: 7170-7178.

Arselin G, Vaillier J, Salin B, Schaeffer J, Giraud M-F, Dautant A, Brèthes D, Velours J (2004) The modulation in subunits e and g amounts of yeast ATP synthase modifies mitochondrial cristae morphology. Journal of Biological Chemistry **279**: 40392-40399.

Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature **444**: 171-178.

Balk J, Pierik AJ, Netz DJA, Mühlenhoff U, Lill R (2004) The hydrogenase-like Nar1p is essential for maturation of cytosolic and nuclear iron–sulphur proteins. The EMBO Journal **23**: 2105-2115.
Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology **19**: 455-477.
Barberà MJ, Ruiz-Trillo I, Tufts JY, Bery A, Silberman JD, Roger AJ (2010) *Sawyeria marylandensis* (Heterolobosea) has a hydrogenosome with novel metabolic properties. Eukaryotic Cell **9**: 1913-1924.
Bareth B, Dennerlein S, Mick DU, Nikolov M, Urlaub H, Rehling P (2013) The heme a synthase Cox15 associates with cytochrome c oxidase assembly intermediates during Cox1 maturation. Molecular and Cellular Biology **33**: 4128-4137.
Barth D, Berendonk TU (2011) The mitochondrial genome sequence of the ciliate *Paramecium caudatum* reveals a shift in nucleotide composition and codon usage within the genus *Paramecium*. BMC Genomics **12**: 272.
Bauchop T (1971) Mechanism of hydrogen formation in *Trichomonas foetus*. Microbiology **68**: 27-33.
Bender T, Pena G, Martinou JC (2015) Regulation of mitochondrial pyruvate uptake by alternative pyruvate carrier complexes. The EMBO Journal **34**: 911-924.
Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell **26**: 167-180.
Bohnert M, Rehling P, Guiard B, Herrmann JM, Pfanner N, van der Laan M (2010) Cooperation of stop-transfer and conservative sorting mechanisms in mitochondrial protein transport. Current Biology **20**: 1227-1232.
Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics: btu170.
Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Nishimura EO, Tintori SC, Li Q, Jones CD, Yandell M (2015) Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. Proceedings of the National Academy of Sciences **112**: 15976-15981.
Boxma B, de Graaf RM, van der Staay GWM, van Alen TA, Ricard G, Gabaldon T, van Hoek AHAM, Moon-van der Staay SY, Koopman WJH, van Hellemond JJ, Tielens AGM, Friedrich T, Veenhuis M, Huynen MA, Hackstein JHP (2005) An anaerobic mitochondrion that produces hydrogen. Nature **434**: 74-79.
Boxma B, Ricard G, van Hoek AH, Severing E, Moon-van der Staay S-Y, van der Staay GW, van Alen TA, de Graaf RM, Cremers G, Kwantes M (2007) The [FeFe] hydrogenase of *Nyctotherus ovalis* has a chimeric origin. BMC Evolutionary Biology **7**: 230.
Boyer PD (1997) The ATP synthase—a splendid molecular machine. Annual review of biochemistry **66**: 717-749.
Bradley PJ, Lahti CJ, Plümper E, Johnson PJ (1997) Targeting and translocation of proteins into the hydrogenosome of the protist *Trichomonas*: similarities with mitochondrial protein import. The EMBO Journal **16**: 3484-3493.
Brandt U (2006) Energy converting NADH: quinone oxidoreductase (complex I). Annual Review of Biochemistry **75**: 69-92.
Bremmer J (1969) Pyruvate dehydrogenase, substrate specificity and product inhibition. European Journal of Biochemistry **8**: 535-540.
Brunk CF, Lee LC, Tran AB, Li J (2003) Complete sequence of the mitochondrial genome of *Tetrahymena thermophila* and comparative methods for identifying highly divergent genes. Nucleic Acids Research **31**: 1673-1682.
Bui E, Bradley P, Johnson P (1996) A common evolutionary origin for mitochondria and hydrogenosomes. Proceedings of the National Academy of Sciences **93**: 9651-9656.
Bui ET, Johnson PJ (1996) Identification and characterization of [Fe]-hydrogenases in the hydrogenosome of *Trichomonas vaginalis*. Molecular and Biochemical Parasitology **76**: 305-310.

Burger G, Forget L, Zhu Y, Gray MW, Lang BF (2003a) Unique mitochondrial genome architecture in unicellular relatives of animals. Proceedings of the National Academy of Sciences **100**: 892-897.
Burger G, Gray MW, Forget L, Lang BF (2013) Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biology and Evolution **5**: 418-438.
Burger G, Gray MW, Lang BF (2003b) Mitochondrial genomes: anything goes. Trends in Genetics **19**: 709-716.
Burger G, Zhu Y, Littlejohn TG, Greenwood SJ, Schnare MN, Lang BF, Gray MW (2000) Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. Journal of Molecular Biology **297**: 365-380.
Burki F (2014) The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harbor Perspectives in Biology **6**: a016147.
Burki F, Shalchian-Tabrizi K, Pawlowski J (2008) Phylogenomics reveals a new 'mega group' including most photosynthetic eukaryotes. Biology Letters **4**: 366-369.
Cantatore P, Gadaleta M, Roberti M, Saccone C, Wilson A (1987) Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature **329**: 853-855.
Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**: 1972-1973.
Carroll J, Fearnley IM, Shannon RJ, Hirst J, Walker JE (2003) Analysis of the subunit composition of complex I from bovine heart mitochondria. Molecular & Cellular Proteomics **2**: 117-126.
Cassidy-Hanley DM (2012) *Tetrahymena* in the laboratory: strain resources, methods for culture, maintenance, and storage. Methods in Cell Biology **109**: 237.
Cavalier-Smith T (1987) The origin of eukaryote and archaebacterial cells. Annals of the New York Academy of Sciences **503**: 17-54.
Cavalier-Smith T (1989) Archaebacteria and Archezoa. Nature **339**: l00-101.
Cavalier-Smith T (1993) Kingdom protozoa and its 18 phyla. Microbiological Reviews **57**: 953-994.
Chaban Y, Boekema EJ, Dudkina NV (2014) Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation. Biochimica et Biophysica Acta (BBA)-Bioenergetics **1837**: 418-426.
Chabrière E, Charon MH, Volbeda A, Pieulle L, Hatchikian EC, Fontecilla–Camps JC (1999) Crystal structures of the key anaerobic enzyme pyruvate: ferredoxin oxidoreductase, free and in complex with pyruvate. Nature Structural & Molecular Biology **6**: 182-190.
Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N (2009) Importing mitochondrial proteins: machineries and mechanisms. Cell **138**: 628-644.
Cheeseman P, Toms-Wood A, Wolfe R (1972) Isolation and properties of a fluorescent compound, Factor420, from *Methanobacterium* strain MoH. Journal of Bacteriology **112**: 527-531.
Chen X, Jung S, Beh LY, Eddy SR, Landweber LF (2015) Combinatorial DNA rearrangement facilitates the origin of new genes in ciliates. Genome Biology and Evolution **7**: 2859-2870.
Ciccarelli FD, Doerks T, Von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. Science **311**: 1283-1287.
Clark CG, Roger AJ (1995) Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. Proceedings of the National Academy of Sciences **92**: 6518-6521.
Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. European Journal of Biochemistry **241**: 779-786.
Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM (2008) The archaebacterial origin of eukaryotes. Proceedings of the National Academy of Sciences **105**: 20356-20361.
Coyne RS, Hannick L, Shanmugam D, Hostetler JB, Brami D, Joardar VS, Johnson J, Radune D, Singh I, Badger JH (2011) Comparative genomics of the pathogenic ciliate *Ichthyophthirius multifiliis*, its free-living relatives and a host species provide insights into adoption of a parasitic lifestyle and prospects for disease control. Genome Biology **12**: 1.

Daims H, Stoecker K, Wagner M (2005) Fluorescence in situ hybridization for the detection of prokaryotes, Molecular Microbial Ecology. Taylor & Francis Group, pp. 213-239.
Davidson EA, van der Giezen M, Horner DS, Embley TM, Howe CJ (2002) An [Fe] hydrogenase from the anaerobic hydrogenosome-containing fungus *Neocallimastix frontalis* L2. Gene **296**: 45-52.
Davies KM, Anselmi C, Wittig I, Faraldo-Gómez JD, Kühlbrandt W (2012) Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae. Proceedings of the National Academy of Sciences **109**: 13602-13607.
Davies KM, Strauss M, Daum B, Kief JH, Osiewacz HD, Rycovska A, Zickermann V, Kühlbrandt W (2011) Macromolecular organization of ATP synthase and complex I in whole mitochondria. Proceedings of the National Academy of Sciences **108**: 14121-14126.
de Graaf RM, Ricard G, van Alen TA, Duarte I, Dutilh BE, Burgtorf C, Kuiper JWP, van der Staay GWM, Tielens AGM, Huynen MA, Hackstein JHP (2011) The Organellar Genome and Metabolic Potential of the Hydrogen-Producing Mitochondrion of *Nyctotherus ovalis*. Molecular Biology and Evolution **28**: 2379-2391.
de Graaf RM, van Alen TA, Dutilh BE, Kuiper JW, van Zoggel HJ, Huynh MB, Görtz H-D, Huynen MA, Hackstein JH (2009) The mitochondrial genomes of the ciliates *Euplotes minuta* and *Euplotes crassus*. BMC Genomics **10**: 514.
Demuez M, Cournac L, Guerrini O, Soucaille P, Girbal L (2007) Complete activity profile of *Clostridium acetobutylicum* [FeFe]-hydrogenase and kinetic parameters for endogenous redox partners. FEMS Microbiology Letters **275**: 113-121.
Denoeud F, Roussel M, Noel B, Wawrzyniak I, Da Silva C, Diogon M, Viscogliosi E, Brochier-Armanet C, Couloux A, Poulain J (2011) Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. Genome Biology **12**: R29.
Deppenmeier U, Müller V, Gottschalk G (1996) Pathways of energy conservation in methanogenic archaea. Archives of Microbiology **165**: 149-163.
DiMarco AA, Bobik TA, Wolfe RS (1990) Unusual coenzymes of methanogenesis. Annual Review of Biochemistry **59**: 355-394.
Dobzhansky T (1937) Genetics and the Origin of Species. Columbia University Press.
Doddema H, Vogels G (1978) Improved identification of methanogenic bacteria by fluorescence microscopy. Applied and Environmental Microbiology **36**: 752-754.
Dolezal P, Likic V, Tachezy J, Lithgow T (2006) Evolution of the molecular machines for protein import into mitochondria. Science **313**: 314-318.
Doolittle WF (1999) Phylogenetic classification and the universal tree. Science **284**: 2124-2128.
Dunthorn M, Foissner W, Katz LA (2011) Expanding character sampling for ciliate phylogenetic inference using mitochondrial SSU-rDNA as a molecular marker. Protist **162**: 85-99.
Dutilh BE, Jurgelenaite R, Szklarczyk R, van Hijum SA, Harhangi HR, Schmid M, de Wild B, Stunnenberg HG, Strous M, Jetten MS (2011) FACIL: fast and accurate genetic code inference and logo. Bioinformatics **27**: 1929-1933.
Dutkiewicz R, Schilke B, Knieszner H, Walter W, Craig EA, Marszalek J (2003) Ssq1, a mitochondrial Hsp70 involved in iron-sulfur (Fe/S) center biogenesis Similarities to and differences from its bacterial counterpart. Journal of Biological Chemistry **278**: 29719-29727.
Eddy SR (1998) Profile hidden Markov models. Bioinformatics **14**: 755-763.
Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**: 1792-1797.
Edgcomb VP, Leadbetter ER, Bourland W, Beaudoin D, Bernhard J (2011) Structured multiple endosymbiosis of bacteria and archaea in a ciliate from marine sulfidic sediments: a survival mechanism in low oxygen, sulfidic sediments? Frontiers in microbiology **2**: 55.
Edwards T, McBride B (1975) New method for the isolation and identification of methanogenic bacteria. Applied Microbiology **29**: 540-545.
Efremov RG, Baradaran R, Sazanov LA (2010) The architecture of respiratory complex I. Nature **465**: 441-445.

Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. PLoS Biology **4**: e286.
Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols **2**: 953-971.
Embley T, Finlay B, Brown S (1992a) RNA sequence analysis shows that the symbionts in the ciliate *Metopus contortus* are polymorphs of a single methanogen species. FEMS Microbiology Letters **97**: 57-61.
Embley TM (2006) Multiple secondary origins of the anaerobic lifestyle in eukaryotes. Philosophical Transactions of the Royal Society of London B: Biological Sciences **361**: 1055-1067.
Embley TM, Finlay BJ (1994) The use of small subunit rRNA sequences to unravel the relationships between anaerobic ciliates and their methanogen endosymbionts. Microbiology **140**: 225-235.
Embley TM, Finlay BJ, Dyal PL, Hirt RP, Wilkinson M, Williams AG (1995) Multiple origins of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. Proceedings of the Royal Society of London. Series B: Biological Sciences **262**: 87-93.
Embley TM, Finlay BJ, Thomas RH, Dyal PL (1992b) The use of rRNA sequences and fluorescent probes to investigate the phylogenetic positions of the anaerobic ciliate *Metopus palaeformis* and its archaeobacterial endosymbiont. Journal of General Microbiology **138**: 1479-1487.
Embley TM, Hirt RP (1998) Early branching eukaryotes? Current Opinion in Genetics & Development **8**: 624-629.
Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. Nature **440**: 623-630.
Embley TM, van der Giezen M, Horner D, Dyal P, Bell S, Foster P (2003) Hydrogenosomes, mitochondria and early eukaryotic evolution. IUBMB Life **55**: 387-395.
Eme L, Gentekaki E, Curtis B, Archibald JM, Roger AJ (2017) Lateral gene transfer in the adaptation of the anaerobic parasite Blastocystis to the gut. Current Biology **27**: 807-820.
Emelyanov VV, Goldberg AV (2011) Fermentation enzymes of *Giardia intestinalis*, pyruvate: ferredoxin oxidoreductase and hydrogenase, do not localize to its mitosomes. Microbiology **157**: 1602-1611.
Esposti MD, Cortez D, Lozano L, Rasmussen S, Nielsen HB, Romero EM (2016) Alpha proteobacterial ancestry of the [FeFe]-hydrogenases in anaerobic eukaryotes. Biology Direct **11**: 34.
Esteban G, Guhl BE, Clarke KJ, Embley TM, Finlay BJ (1993) *Cyclidium porcatum* n. sp.: a Free-living anaerobic scuticociliate containing a stable complex of hydrogenosomes, eubacteria and archaeobacteria. European Journal of Protistology **29**: 262-270.
Esteban GF, Fenchel T, Finlay BJ (2010) Mixotrophy in ciliates. Protist **161**: 621-641.
Esteban GF, Finlay BJ, Clarke KJ (2009) Sequestered organelles sustain aerobic microbial life in anoxic environments. Environmental Microbiology **11**: 544-550.
Ettema TJ, Lindås AC, Bernander R (2011) An actin-based cytoskeleton in archaea. Molecular Microbiology **80**: 1052-1061.
Evans PN, Parks DH, Chadwick GL, Robbins SJ, Orphan VJ, Golding SD, Tyson GW (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. Science **350**: 434-438.
Fenchel T, Finlay B (1992) Production of methane and hydrogen by anaerobic ciliates containing symbiotic methanogens. Archives of Microbiology **157**: 475-480.
Fenchel T, Finlay B, J (1995) Ecology and evolution in anoxic worlds. Oxford University Press; Oxford; New York: Oxford University Press, 1995, Oxford.
Fenchel T, Finlay BJ (1991a) Endosymbiotic methanogenic bacteria in anaerobic ciliates: significance for the growth efficiency of the host. Journal of Eukaryotic Microbiology **38**: 18-22.
Fenchel T, Finlay BJ (1991b) Endosymbiotic methanogenic bacteria in anaerobic ciliates: significance for the growth efficiency of the host. The Journal of Protozoology **38**: 18-22.

Fenchel T, Finlay BJ (1991c) Synchronous division of an endosymbiotic methanogenic bacterium in the anaerobic ciliate *Plagiopyla frontata* Kahl. The Journal of Protozoology **38**: 22-28.
Fenchel T, Finlay BJ (2010) Free-living protozoa with endosymbiotic methanogens, (Endo) symbiotic Methanogenic Archaea. Springer, pp. 1-11.
Fenchel T, Perry T, Thane A (1977) Anaerobiosis and Symbiosis with Bacteria in Free-living Ciliates. The Journal of Protozoology **24**: 154-163.
Fernie AR, Carrari F, Sweetlove LJ (2004) Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. Current opinion in plant biology **7**: 254-261.
Ferramosca A, Zara V (2013) Biogenesis of mitochondrial carrier proteins: molecular mechanisms of import into mitochondria. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1833**: 494-502.
Ferry JG (1992) Biochemistry of methanogenesis. Critical Reviews in Biochemistry and Molecular Biology **27**: 473-503.
Finlay B, Embley T, Fenchel T (1993a) A new polymorphic methanogen, closely related to Methanocorpusculum parvum, living in stable symbiosis within the anaerobic ciliate Trimyema sp. Microbiology **139**: 371-378.
Finlay B, Fenchel T (1989) Hydrogenosomes in some anaerobic protozoa resemble mitochondria. FEMS Microbiology Letters **65**: 311-314.
Finlay B, Fenchel T (1991) Polymorphic bacterial symbionts in the anaerobic ciliated protozoon *Metopus*. FEMS Microbiology Letters **79**: 187-190.
Finlay BJ, Embley TM, Fenchel T (1993b) A new polymorphic methanogen, closely related to *Methanocorpusculum parvum*, living in stable symbiosis within the anaerobic ciliate *Trimyema* sp. Journal of General Microbiology **139**: 371-378.
Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research **44**: D279-D285.
Foster PG, Cox CJ, Embley TM (2009) The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. Philosophical Transactions of the Royal Society of London B: Biological Sciences **364**: 2197-2207.
Frazier AE, Taylor RD, Mick DU, Warscheid B, Stoepel N, Meyer HE, Ryan MT, Guiard B, Rehling P (2006) Mdm38 interacts with ribosomes and is a component of the mitochondrial protein export machinery. Journal of Cell Biology **172**: 553-564.
Freibert S-A, Goldberg AV, Hacker C, Molik S, Dean P, Williams TA, Nakjang S, Long S, Sendra K, Bill E, Heinz E, Hirt RP, Lucocq JM, Embley TM, Lill R (2017) Evolutionary conservation and in vitro reconstitution of microsporidian iron-sulfur cluster biosynthesis. Nature Communications **8**.
Fukasawa Y, Tsuji J, Fu S-C, Tomii K, Horton P, Imai K (2015) MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Molecular & Cellular Proteomics **14**: 1113-1126.
Furdui C, Ragsdale SW (2000) The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. Journal of Biological Chemistry **275**: 28494-28499.
Gabaldón T, Huynen MA (2003) Reconstruction of the proto-mitochondrial metabolism. Science **301**: 609-609.
Gakh O, Cavadini P, Isaya G (2002) Mitochondrial processing peptidases. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1592**: 63-77.
Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Computer Applications in the Biosciences **12**: 543-548.
Gao F, Katz LA (2014) Phylogenomic analyses support the bifurcation of ciliates into two major clades that differ in properties of nuclear division. Molecular Phylogenetics and Evolution **70**: 240-243.

Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR (2009) Rfam: updates to the RNA families database. Nucleic Acids Research **37**: D136-D140.
Gasser S, Daum G, Schatz G (1982) Import of proteins into mitochondria. Energy-dependent uptake of precursors by isolated mitochondria. Journal of Biological Chemistry **257**: 13034-13041.
Gaston D, Tsaousis AD, Roger AJ (2009) Predicting proteomes of mitochondria and related organelles from genomic and expressed sequence tag data. Methods in Enzymology **457**: 21-47.
Gawryluk RM, Kamikawa R, Stairs CW, Silberman JD, Brown MW, Roger AJ (2016) The earliest stages of mitochondrial adaptation to low oxygen revealed in a novel rhizarian. Current Biology **26**: 2729-2738.
Germot A, Philippe H, Le Guyader H (1996) Presence of a mitochondrial-type 70-kDa heat shock protein in *Trichomonas vaginalis* suggests a very early mitochondrial endosymbiosis in eukaryotes. Proceedings of the National Academy of Sciences **93**: 14614-14617.
Germot A, Philippe H, Le Guyader H (1997) Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. Molecular and Biochemical Parasitology **87**: 159-168.
Gijzen HJ, Broers C, Barughare M, Stumm CK (1991) Methanogenic bacteria as endosymbionts of the ciliate Nyctotherus ovalis in the cockroach hindgut. Applied and environmental microbiology **57**: 1630-1634.
Gilkerson RW, Selker JM, Capaldi RA (2003) The cristal membrane of mitochondria is the principal site of oxidative phosphorylation. FEBS Letters **546**: 355-358.
Gill EE, Diaz-Triviño S, Barberà MJ, Silberman JD, Stechmann A, Gaston D, Tamas I, Roger AJ (2007) Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*. Molecular Microbiology **66**: 1306-1320.
Goddard JM, Cummings DJ (1975) Structure and replication of mitochondrial DNA from *Paramecium aurelia*. Journal of Molecular Biology **97**: 593-609.
Golding GB, Gupta RS (1995) Protein-based phylogenies support a chimeric origin for the eukaryotic genome. Molecular Biology and Evolution **12**: 1-6.
Goosen NK, Drift C, Stumm CK, Vogels GD (1990) End products of metabolism in the anaerobic ciliate *Trimyema compressum*. FEMS Microbiology Letters **69**: 171-175.
Gorrell TE, Yarlett N, Müller M (1984) Isolation and characterization of Trichomonas vaginalis ferredoxin. Carlsberg Research Communications **49**: 259.
Gouy M, Li W-H (1989) Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. Nature **339**: 145-147.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology **29**: 644-652.
Gray MW (2005) Evolutionary biology: The hydrogenosome's murky past. Nature **434**: 29-31.
Guarani V, McNeill EM, Paulo JA, Huttlin EL, Fröhlich F, Gygi SP, Van Vactor D, Harper JW (2015) QIL1 is a novel mitochondrial protein required for MICOS complex stability and cristae morphology. eLife **4**: e06265.
Gutiérrez J, Callejas S, Borniquel S, Benítez L, Martín-González A (2001) Ciliate cryptobiosis: a microbial strategy against environmental starvation. International Microbiology **4**: 151-157.
Guy L, Ettema TJ (2011) The archaeal 'TACK'superphylum and the origin of eukaryotes. Trends in Microbiology **19**: 580-587.
Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proceedings of the National Academy of Sciences **106**: 3859-3864.
Hampl V, Stairs CW, Roger AJ (2011) The tangled past of eukaryotic enzymes involved in anaerobic metabolism. Mobile Genetic Elements **1**: 71-74.
Handa H (2008) Linear plasmids in plant mitochondria: peaceful coexistences or malicious invasions? Mitochondrion **8**: 15-25.

Hatefi Y (1985) The mitochondrial electron transport and oxidative phosphorylation system. Annual Review of Biochemistry **54**: 1015-1069.
Henze K, Horner DS, Suguri S, Moore DV, Sánchez LB, Müller M, Embley TM (2001) Unique phylogenetic relationships of glucokinase and glucosephosphate isomerase of the amitochondriate eukaryotes *Giardia intestinalis*, *Spironucleus barkhanus* and *Trichomonas vaginalis*. Gene **281**: 123-131.
Hildebrand J, Lee B, Richards F, Matsumura M, Becktel W, Matthews B, Mezei M, Beveridge D (1991) Primary Structure of Hydrogenase I from *Clostridium pasteurianum*. Biochemistry **30**: 9697-9704.
Hinchliffe P, Sazanov LA (2005) Organization of iron-sulfur clusters in respiratory complex I. Science **309**: 771-774.
Hirt RP, Alsmark C, Embley TM (2015) Lateral gene transfers and the origins of the eukaryote proteome: a view from microbial parasites. Current Opinion in Microbiology **23**: 155-162.
Hirt RP, Healy B, Vossbrinck CR, Canning EU, Embley TM (1997) A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. Current Biology **7**: 995-998.
Hirt RP, Logsdon JM, Healy B, Dorey MW, Doolittle WF, Embley TM (1999) Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. Proceedings of the National Academy of Sciences **96**: 580-585.
Hoek AHV, Sprakel VS, Alen TA, Theuvenet AP, Vogels GD, Hackstein JH (1999) Voltage-Dependent Reversal of Anodic Galvanotaxis in *Nyctotherus ovalis*. Journal of Eukaryotic Microbiology **46**: 427-433.
Holler S, Pfennig N (1991) Fermentation products of the anaerobic ciliate *Trimyema compressum* in monoxenic cultures. Archives of Microbiology **156**: 327-334.
Hoogenraad NJ, Ward LA, Ryan MT (2002) Import and assembly of proteins into mitochondria of mammalian cells. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1592**: 97-105.
Horner DS, Foster PG, Embley TM (2000) Iron hydrogenases and the evolution of anaerobic eukaryotes. Molecular Biology and Evolution **17**: 1695-1709.
Horner DS, Heil B, Happe T, Embley TM (2002) Iron hydrogenases–ancient enzymes in modern eukaryotes. Trends in Biochemical Sciences **27**: 148-153.
Horner DS, Hirt RP, Embley TM (1999) A single eubacterial origin of eukaryotic pyruvate: ferredoxin oxidoreductase genes: implications for the evolution of anaerobic eukaryotes. Molecular Biology and Evolution **16**: 1280-1291.
Horner DS, Hirt RP, Kilvington S, Lloyd D, Embley TM (1996) Molecular data suggest an early acquisition of the mitochondrion endosymbiont. Proceedings of the Royal Society of London B: Biological Sciences **263**: 1053-1059.
Hrdy I, Hirt RP, Dolezal P, Bardonova L, Foster PG, Tachezy J, Martin Embley T (2004) *Trichomonas hydrogenosomes* contain the NADH dehydrogenase module of mitochondrial complex I. Nature **432**: 618-622.
Hrdý I, Müller M (1995a) Primary structure and eubacterial relationships of the pyruvate: ferredoxin oxidoreductase of the amitochondriate eukaryote *Trichomonas vaginalis*. Journal of Molecular Evolution **41**: 388-396.
Hrdý I, Müller M (1995b) Primary structure of the hydrogenosomal malic enzyme of *Trichomonas vaginalis* and its relationship to homologous enzymes. Journal of Eukaryotic Microbiology **42**: 593-603.
Huerta-Cepas J, Forslund K, Szklarczyk D, Jensen LJ, von Mering C, Bork P (2016) Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. bioRxiv: 076331.
Huet J, Schnabel R, Sentenac A, Zillig W (1983) Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. The EMBO Journal **2**: 1291.
Hug LA, Stechmann A, Roger AJ (2010) Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes. Molecular Biology and Evolution **27**: 311-324.

Hunte C, Zickermann V, Brandt U (2010) Functional modules and structural basis of conformational coupling in mitochondrial complex I. Science **329**: 448-451.
Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R (2016) MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Computational Biology **12**: e1004957.
Huynen MA, Mühlmeister M, Gotthardt K, Guerrero-Castillo S, Brandt U (2016) Evolution and structural organization of the mitochondrial contact site (MICOS) complex and the mitochondrial intermembrane space bridging (MIB) complex. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1863**: 91-101.
Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**: 1.
Ieva R, Heißwolf AK, Gebert M, Vögtle F-N, Wollweber F, Mehnert CS, Oeljeklaus S, Warscheid B, Meisinger C, Van Der Laan M (2013) Mitochondrial inner membrane protease promotes assembly of presequence translocase by removing a carboxy-terminal targeting sequence. Nature Communications **4**.
Inui H, Ono K, Miyatake K, Nakano Y, Kitaoka S (1987) Purification and characterization of pyruvate: NADP+ oxidoreductase in *Euglena gracilis*. Journal of Biological Chemistry **262**: 9130-9135.
Iwata S, Lee JW, Okada K, Lee JK, Iwata M, Rasmussen B, Link TA, Ramaswamy S, Jap BK (1998) Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. Science **281**: 64-71.
Iwata S, Ostermeier C, Ludwig B, Michel H (1995) Structure at 2.8 angstrom resolution of cytochrome c oxidase from Paracoccus denitrificans. Nature **376**: 660.
Jacob AS, Andersen LOB, Bitar PP, Richards VP, Shah S, Stanhope MJ, Stensvold CR, Clark CG (2016) Blastocystis mitochondrial genomes appear to show multiple independent gains and losses of start and stop codons. Genome Biology and Evolution **8**: 3340-3350.
Jenkins TM, Gorrell TE, Müller M, Weitzman P (1991) Hydrogenosomal succinate thiokinase in *Tritrichomonas foetus* and *Trichomonas vaginalis*. Biochemical and Biophysical Research Communications **179**: 892-896.
Jerlström-Hultqvist J, Einarsson E, Xu F, Hjort K, Ek B, Steinhauf D, Hultenby K, Bergquist J, Andersson JO, Svärd SG (2013) Hydrogenosomes in the diplomonad *Spironucleus salmonicida*. Nature Communications **4**.
Jerlström-Hultqvist J, Franzén O, Ankarklev J, Xu F, Nohýnková E, Andersson JO, Svärd SG, Andersson B (2010) Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. BMC Genomics **11**: 543.
John GB, Shang Y, Li L, Renken C, Mannella CA, Selker JM, Rangell L, Bennett MJ, Zha J (2005) The mitochondrial inner membrane protein mitofilin controls cristae morphology. Molecular Biology of the Cell **16**: 1543-1554.
Johnson MD (2011) Acquired phototrophy in ciliates: a review of cellular interactions and structural adaptations. Journal of Eukaryotic Microbiology **58**: 185-195.
Junge W, Nelson N (2015) ATP synthase. Annual Review of Biochemistry **84**: 631-657.
Kadenbach B, Jarausch J, Hartmann R, Merle P (1983) Separation of mammalian cytochrome c oxidase into 13 polypeptides by a sodium dodecyl sulfate-gel electrophoretic procedure. Analytical Biochemistry **129**: 517-521.
Karlberg O, Canbäck B, Kurland CG, Andersson SG (2000) The dual origin of the yeast mitochondrial proteome. Yeast **17**: 170-187.
Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V, Barlow LD, Herman EK (2016) A eukaryote without a mitochondrial organelle. Current Biology **26**: 1274-1284.
Katz LA (2001) Evolution of nuclear dualism in ciliates: a reanalysis in light of recent molecular data. International journal of systematic and evolutionary microbiology **51**: 1587-1592.

Kawano S, Takano H, Mori K, Kuroiwa T (1991) A mitochondrial plasmid that promotes mitochondrial fusion in *Physarum polycephalum*. Protoplasma **160**: 167-169.
Kennedy EP, Lehninger AL (1949) Oxidation of fatty acids and tricarboxylic acid cycle intermediates by isolated rat liver mitochondria. Journal of Biological Chemistry **179**: 957-972.
Kerscher L, Oesterhelt D (1982) Pyruvate: ferredoxin oxidoreductase—new findings on an ancient enzyme. Trends in Biochemical Sciences **7**: 371-374.
Kletzin A, Adams M (1996) Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*. Journal of Bacteriology **178**: 248-257.
Koonin EV (2015) Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? Philosophical Transactions of the Royal Society B **370**: 20140333.
Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H, Thomas F, Aboobaker AA, Blaxter M (2016) No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. Proceedings of the National Academy of Sciences **113**: 5053-5058.
Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell **31**: 147-157.
Ku C, Martin WF (2016) A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70% rule. BMC Biology **14**: 89.
Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. Mitochondrion **8**: 5-14.
Lake JA, Henderson E, Oakes M, Clark MW (1984) Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proceedings of the National Academy of Sciences **81**: 3786-3790.
Lane N, Martin W (2010) The energetics of genome complexity. Nature **467**: 929-934.
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods **9**: 357-359.
Lantsman Y, Tan KS, Morada M, Yarlett N (2008) Biochemical characterization of a mitochondrial-like organelle from *Blastocystis* sp. subtype 7. Microbiology **154**: 2757-2766.
Lanzén A, Jørgensen SL, Bengtsson MM, Jonassen I, Øvreås L, Urich T (2011) Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. FEMS Microbiology Ecology **77**: 577-589.
Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Molecular Biology and Evolution **21**: 1095-1109.
Lartillot N, Rodrigue N, Stubbs D, Richer J (2013) PhyloBayes MPI. Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Systematic biology: syt022.
Lasken RS, Stockwell TB (2007) Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC Biotechnology **7**: 19.
Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: rates of change and exchange. Journal of Molecular Evolution **44**: 383-397.
Leger MM, Eme L, Hug LA, Roger AJ (2016) Novel hydrogenosomes in the microaerophilic jakobid *Stygiella incarcerata*. Molecular Biology and Evolution: msw103.
Lever MA (2016) A New Era of Methanogenesis Research. Trends in Microbiology **24**: 84-86.
Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12**: 1.
Lill R (2009) Function and biogenesis of iron–sulphur proteins. Nature **460**: 831-838.
Lill R, Diekert K, Kaut A, Lange H, Pelzer W, Prohl C, Kispal G (1999) The essential role of mitochondria in the biogenesis of cellular iron-sulfur proteins. Biological Chemistry **380**: 1157-1166.

Lindmark DG, Müller M (1973) Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. Journal of Biological Chemistry **248**: 7724-7728.
Lindmark DG, Müller M, Shio H (1975) Hydrogenosomes in *Trichomonas vaginalis*. The Journal of Parasitology **61**: 552-554.
Lloyd D, Ralphs JR, Harris JC (2002) *Giardia intestinalis*, a eukaryote without hydrogenosomes, produces hydrogen. Microbiology **148**: 727-733.
Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ (2005) The genome of the protist parasite *Entamoeba histolytica*. Nature **433**: 865-868.
Lonjers ZT, Dickson EL, Chu T-PT, Kreutz JE, Neacsu FA, Anders KR, Shepherd JN (2012) Identification of a new gene required for the biosynthesis of rhodoquinone in *Rhodospirillum rubrum*. Journal of Bacteriology **194**: 965-971.
Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Research **44**: W54-W57.
Loy A, Horn M, Wagner M (2003) probeBase: an online resource for rRNA-targeted oligonucleotide probes. Nucleic Acids Research **31**: 514-516.
Lozupone CA, Knight RD, Landweber LF (2001) The molecular basis of nuclear genetic code change in ciliates. Current Biology **11**: 65-74.
Lynch M, Koskella B, Schaack S (2006) Mutation pressure and the evolution of organelle genomic architecture. Science **311**: 1727-1730.
Lynn D (2008) The ciliated protozoa: characterization, classification, and guide to the literature. Springer Science & Business Media.
Lynn DH (2003) Morphology or molecules: How do we identify the major lineages of ciliates (Phylum Ciliophora). European Journal of Protistology **39**: 356-364.
Magnitsky S, Toulokhonova L, Yano T, Sled VD, Hägerhäll C, Grivennikova VG, Burbaev DS, Vinogradov AD, Ohnishi T (2002) EPR characterization of ubisemiquinones and iron–sulfur cluster N2, central components of the energy coupling in the NADH-ubiquinone oxidoreductase (complex I) in situ. Journal of Bioenergetics and Biomembranes **34**: 193-208.
Mai Z, Ghosh S, Frisardi M, Rosenthal B, Rogers R, Samuelson J (1999) Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("crypton") in the microaerophilic protozoan parasite *Entamoeba histolytica*. Molecular and Cellular Biology **19**: 2198-2205.
Margulis L (1970) Origin of eukaryotic cells: evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth. Yale University Press.
Marques I, Duarte M, Assunção J, Ushakova AV, Videira A (2005) Composition of complex I from *Neurospora crassa* and disruption of two "accessory" subunits. Biochimica et Biophysica Acta (BBA)-Bioenergetics **1707**: 211-220.
Martin W (1999) Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. Bioessays **21**: 99-104.
Martin W, Brinkmann H, Savonna C, Cerff R (1993) Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proceedings of the National Academy of Sciences **90**: 8692-8696.
Martin W, Müller M (1998) The hydrogen hypothesis for the first eukaryote. Nature **392**: 37-41.
Marvin-Sikkema F, Kraak M, Veenhuis M, Gottschal J, Prins R (1993) 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. European Journal of Cell Biology **61**: 86-91.
Massey V, Singer TP (1957) Studies on succinic dehydrogenase III. The fumaric reductase activity of succinic dehydrogenase. Journal of Biological Chemistry **228**: 263-274.

Mathiesen C, Hägerhäll C (2002) Transmembrane topology of the NuoL, M and N subunits of NADH: quinone oxidoreductase and their homologues among membrane-bound hydrogenases and bona fide antiporters. Biochimica et Biophysica Acta (BBA)-Bioenergetics **1556**: 121-132.
McCoy J, Mann B (2004) *Entamoeba histolytica* Genome, The Pathogenic Enteric Protozoa: Giardia, Entamoeba, Cryptosporidium and Cyclospora. Springer, pp. 141-152.
Meinhardt F, Kempken F, Kämper J, Esser K (1990) Linear plasmids among eukaryotes: fundamentals and application. Current Genetics **17**: 89-95.
Melber A, Na U, Vashisht A, Weiler BD, Lill R, Wohlschlegel JA, Winge DR (2016) Role of Nfu1 and Bol3 in iron-sulfur cluster transfer to mitochondrial clients. eLife **5**: e15991.
Menon S, Ragsdale SW (1996) Unleashing hydrogenase activity in carbon monoxide dehydrogenase/acetyl-CoA synthase and pyruvate: ferredoxin oxidoreductase. Biochemistry **35**: 15814-15821.
Milenkovic D, Larsson N-G (2015) Mic10 Oligomerization Pinches off Mitochondrial Cristae. Cell Metabolism **21**: 660-661.
Minauro-Sanmiguel F, Wilkens S, García JJ (2005) Structure of dimeric mitochondrial ATP synthase: novel F0 bridging features and the structural basis of mitochondrial cristae biogenesis. Proceedings of the National Academy of Sciences of the United States of America **102**: 12356-12358.
Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. Molecular Biology and Evolution: mst024.
Mink RW, Dugan PR (1977) Tentative identification of methanogenic bacteria by fluorescence microscopy. Applied and Environmental Microbiology **33**: 713-717.
Mitchell P (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. Nature **191**: 144-148.
Moradian MM, Beglaryan D, Skozylas JM, Kerikorian V (2007) Complete mitochondrial genome sequence of three *Tetrahymena* species reveals mutation hot spots and accelerated nonsynonymous substitutions in Ymf genes. PLoS One **2**: e650.
Morado JF, Small EB (1995) Ciliate parasites and related diseases of Crustacea: a review. Reviews in Fisheries Science **3**: 275-354.
Moreira D, López-García P (1998) Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. Journal of Molecular Evolution **47**: 517-530.
Morin GB, Cech TR (1986) The telomeres of the linear mitochondrial DNA of *Tetrahymena thermophila* consist of 53 bp tandem repeats. Cell **46**: 873-883.
Morin GB, Cech TR (1988) Mitochondrial telomeres: surprising diversity of repeated telomeric DNA sequences among six species of Tetrahymena. Cell **52**: 367-374.
Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ (2007) Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. Science **317**: 1921-1926.
Mühleip AW, Joos F, Wigge C, Frangakis AS, Kühlbrandt W, Davies KM (2016) Helical arrays of U-shaped ATP synthase dimers form tubular cristae in ciliate mitochondria. Proceedings of the National Academy of Sciences **113**: 8442-8447.
Mulder DW, Ortillo DO, Gardenghi DJ, Naumov AV, Ruebush SS, Szilagyi RK, Huynh B, Broderick JB, Peters JW (2009) Activation of HydAΔEFG requires a preformed [4Fe-4S] cluster. Biochemistry **48**: 6240-6248.
Mulder DW, Shepard EM, Meuser JE, Joshi N, King PW, Posewitz MC, Broderick JB, Peters JW (2011) Insights into [FeFe]-hydrogenase structure, mechanism, and maturation. Structure **19**: 1038-1052.
Muller HJ (1964) The relation of recombination to mutational advance. Mutation Research **1**: 2-9.
Müller M (1993) Review Article: The hydrogenosome. Microbiology **139**: 2879-2889.
Müller M, Lindmark D (1978) Respiration of hydrogenosomes of *Tritrichomonas foetus*. II. Effect of CoA on pyruvate oxidation. Journal of Biological Chemistry **253**: 1215-1218.

Muller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu RY, van der Giezen M, Tielens AG, Martin WF (2012) Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiology and Molecular Biology Reviews **76**: 444-495.
Muñoz-Gómez SA, Slamovits CH, Dacks JB, Baier KA, Spencer KD, Wideman JG (2015) Ancient homology of the mitochondrial contact site and cristae organizing system points to an endosymbiotic origin of mitochondrial cristae. Current Biology **25**: 1489-1495.
Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics **29**: 2933-2935.
Nawy T (2014) Single-cell sequencing. Nature Methods **11**: 18-18.
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) Iq-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution **32**: 268-274.
Nicolet Y, Fontecilla-Camps JC (2012) Structure-function relationships in [FeFe]-hydrogenase active site maturation. Journal of Biological Chemistry **287**: 13532-13540.
Nigrelli RF, Pokorny KS, Ruggieri GD (1976) Notes on *Ichthyophthirius multifiliis,* a ciliate parasitic on fresh-water fishes, with some remarks on possible physiological races and species. Transactions of the American Microscopical Society: 607-613.
Nina PB, Dudkina NV, Kane LA, van Eyk JE, Boekema EJ, Mather MW, Vaidya AB (2010) Highly divergent mitochondrial ATP synthase complexes in *Tetrahymena thermophila*. PLoS Biology **8**: e1000418.
Nixon JE, Field J, McArthur AG, Sogin ML, Yarlett N, Loftus BJ, Samuelson J (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.
Noguchi F, Shimamura S, Nakayama T, Yazaki E, Yabuki A, Hashimoto T, Inagaki Y, Fujikura K, Takishita K (2015) Metabolic capacity of mitochondrion-related organelles in the free-living anaerobic stramenopile *Cantina marsupialis*. Protist **166**: 534-550.
Nývltová E, Stairs CW, Hrdý I, Rídl J, Mach J, Pačes J, Roger AJ, Tachezy J (2015) Lateral gene transfer and gene duplication played a key role in the evolution of *Mastigamoeba balamuthi* hydrogenosomes. Molecular Biology and Evolution **32**: 1039-1055.
Nývltová E, Šuták R, Harant K, Šedinová M, Hrdý I, Pačes J, Vlček Č, Tachezy J (2013) NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of Mastigamoeba balamuthi. Proceedings of the National Academy of Sciences **110**: 7371-7376.
Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature **405**: 299-304.
Ohnishi T (1998) Iron–sulfur clusters/semiquinones in complex I. Biochimica et Biophysica Acta (BBA)-Bioenergetics **1364**: 186-206.
Parfrey LW, Lahr DJ, Knoll AH, Katz LA (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proceedings of the National Academy of Sciences **108**: 13624-13629.
Patel MS, Nemeria NS, Furey W, Jordan F (2014) The pyruvate dehydrogenase complexes: structure-based function and regulation. Journal of Biological Chemistry **289**: 16615-16623.
Paumard P, Vaillier J, Coulary B, Schaeffer J, Soubannier V, Mueller DM, Brèthes D, di Rago J-P, Velours J (2002) The ATP synthase is involved in generating mitochondrial cristae morphology. The EMBO Journal **21**: 221-230.
Pazour GJ, Agrin N, Leszyk J, Witman GB (2005) Proteomic analysis of a eukaryotic cilium. The Journal of Cell Biology **170**: 103-113.
Pedersen LB, Schrøder JM, Satir P, Christensen ST (2012) The ciliary cytoskeleton. Comprehensive Physiology.
Peters JW, Lanzilotta WN, Lemon BJ, Seefeldt LC (1998) X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 angstrom resolution. Science **282**: 1853-1858.

Peters JW, Schut GJ, Boyd ES, Mulder DW, Shepard EM, Broderick JB, King PW, Adams MW (2015) [FeFe]-and [NiFe]-hydrogenase diversity, mechanism, and maturation. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1853**: 1350-1369.
Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. Nature Protocols **9**: 171-181.
Pilkington SJ, Skehel JM, Gennis RB, Walker JE (1991) Relationship between mitochondrial NADH-ubiquinone reductase and a bacterial NAD-reducing hydrogenase. Biochemistry **30**: 2166-2175.
Posewitz M, King P, Smolinski S, Smith RD, Ginley A, Ghirardi M, Seibert M (2005) Identification of genes required for hydrogenase activity in *Chlamydomonas reinhardtii*. Biochemical Society Transactions.
Posewitz MC, King PW, Smolinski SL, Zhang L, Seibert M, Ghirardi ML (2004) Discovery of two novel radical S-adenosylmethionine proteins required for the assembly of an active [Fe] hydrogenase. Journal of Biological Chemistry **279**: 25711-25720.
Prescott DM (2000) Genome gymnastics: unique modes of DNA evolution and processing in ciliates. Nature Reviews Genetics **1**: 191-198.
Pritchard A, Seilhamer J, Mahalingam R, Sable C, Venuti S, Cummings D (1990) Nucleotide sequence of the mitochondrial genome of *Paramecium*. Nucleic Acids Research **18**: 173-180.
Pütz S, Dolezal P, Gelius-Dietrich G, Bohacova L, Tachezy J, Henze K (2006) Fe-hydrogenase maturases in the hydrogenosomes of *Trichomonas vaginalis*. Eukaryotic Cell **5**: 579-586.
Quang LS, Gascuel O (2008) An improved general amino acid replacement matrix. Molecular Biology and Evolution **25**: 1307-1320.
Quang LS, Gascuel O, Lartillot N (2008) Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics **24**: 2317-2323.
Rawlings TA, Collins TM, Bieler R (2003) Changing identities: tRNA duplication and remolding within animal mitochondrial genomes. Proceedings of the National Academy of Sciences **100**: 15700-15705.
Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends in Genetics **16**: 276-277.
Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science **257**: 74-76.
Rivière L, van Weelden SW, Glass P, Vegh P, Coustou V, Biran M, van Hellemond JJ, Bringaud F, Tielens AG, Boshart M (2004) Acetyl: succinate CoA-transferase in procyclic *Trypanosoma brucei* Gene identification and role in carbohydrate metabolism. Journal of Biological Chemistry **279**: 45337-45346.
Roger AJ, Clark CG, Doolittle WF (1996) A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. Proceedings of the National Academy of Sciences **93**: 14618-14622.
Roger AJ, Svärd SG, Tovar J, Clark CG, Smith MW, Gillin FD, Sogin ML (1998) A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proceedings of the National Academy of Sciences **95**: 229-234.
Rotte C, Stejskal F, Zhu G, Keithly JS, Martin W (2001) Pyruvate: NADP oxidoreductase from the mitochondrion of Euglena gracilis and from the apicomplexan *Cryptosporidium parvum*: a biochemical relic linking pyruvate metabolism in mitochondriate and amitochondriate protists. Molecular Biology and Evolution **18**: 710-720.
Ryan R, Grant D, Chiang K-S, Swift H (1978) Isolation and characterization of mitochondrial DNA from *Chlamydomonas reinhardtii*. Proceedings of the National Academy of Sciences **75**: 3268-3272.
Rycovska A, Valach M, Tomaska L, Bolotin-Fukuhara M, Nosek J (2004) Linear versus circular mitochondrial genomes: intraspecies variability of mitochondrial genome architecture in *Candida parapsilosis*. Microbiology **150**: 1571-1580.

Sagan L (1967) On the origin of mitosing cells. Journal of Theoretical Biology **14**: 225-274.
Saraste M (1999) Oxidative phosphorylation at the fin de siecle. Science **283**: 1488-1493.
Satir P, Christensen ST (2007) Overview of structure and function of mammalian cilia. Annual Review of Physiology **69**: 377-400.
Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B (2012) Fiji: an open-source platform for biological-image analysis. Nature Methods **9**: 676-682.
Schut GJ, Adams MW (2009) The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. Journal of Bacteriology **191**: 4451-4457.
Selengut JD, Haft DH (2010) Unexpected abundance of coenzyme F420-dependent enzymes in *Mycobacterium tuberculosis* and other actinobacteria. Journal of Bacteriology **192**: 5788-5798.
Shao R, Kirkness EF, Barker SC (2009) The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. Genome Research **19**: 904-912.
Sharp PM, Li W-H (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research **15**: 1281-1295.
Shepard EM, Mus F, Betz JN, Byer AS, Duffus BR, Peters JW, Broderick JB (2014) [FeFe]-hydrogenase maturation. Biochemistry **53**: 4090-4104.
Shima S, Pilak O, Vogt S, Schick M, Stagni MS, Meyer-Klaucke W, Warkentin E, Thauer RK, Ermler U (2008) The crystal structure of [Fe]-hydrogenase reveals the geometry of the active site. Science **321**: 572-575.
Shinzato N, Watanabe I, Meng X-Y, Sekiguchi Y, Tamaki H, Matsui T, Kamagata Y (2007) Phylogenetic analysis and fluorescence in situ hybridization detection of archaeal and bacterial endosymbionts in the anaerobic ciliate *Trimyema compressum*. Microbial Ecology **54**: 627-636.
Siqueira-Castro ICV, Greinert-Goulart JA, Bonatti TR, Yamashiro S, Franco RMB (2016) First report of predation of *Giardia* sp. cysts by ciliated protozoa and confirmation of predation of *Cryptosporidium* spp. oocysts by ciliate species. Environmental Science and Pollution Research: 1-6.
Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics **4**: 1581-1590.
Smith DG, Gawryluk RM, Spencer DF, Pearlman RE, Siu KM, Gray MW (2007) Exploring the mitochondrial proteome of the ciliate protozoon *Tetrahymena thermophila*: direct analysis by tandem mass spectrometry. Journal of Molecular Biology **374**: 837-863.
Sogin ML (1989) Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. American Zoologist **29**: 487-499.
Söllner T, Griffiths G, Pfaller R, Pfanner N, Neupert W (1989) MOM19, an import receptor for mitochondrial precursor proteins. Cell **59**: 1061-1070.
Soslau G, Nass MM (1971) Effects of ethidium bromide on the cytochrome content and ultrastructure of L cell mitochondria. The Journal of cell biology **51**: 514.
Soucy SM, Huang J, Gogarten JP (2015) Horizontal gene transfer: building the web of life. Nature Reviews Genetics **16**: 472-482.
Sousa FL, Neukirchen S, Allen JF, Lane N, Martin WF (2016) Lokiarchaeon is hydrogen dependent. Nature Microbiology **1**: 16034.
Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJ (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature **521**: 173-179.
Stahl AD, Amann R (1991) Development and Application of Nucleic Acid Probes, in: Goodfellow M, Stackebrandt E (eds) Nucleic acid techniques in bacterial systematics. John Wiley & Sons, pp. 205-248.

Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, van der Giezen M, Roger AJ (2014) A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist *Pygsuia*. Current Biology **24**: 1176-1186.
Stairs CW, Leger MM, Roger AJ (2015) Diversity and origins of anaerobic metabolism in mitochondria and related organelles. Philosophical Transactions of the Royal Society B **370**: 20140326.
Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**: 1312-1313.
Stechmann A, Hamblin K, Pérez-Brocal V, Gaston D, Richmond GS, Van der Giezen M, Clark CG, Roger AJ (2008) Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. Current Biology **18**: 580-585.
Steinbüchel A, Müller M (1986) Anaerobic pyruvate metabolism of Tritrichomonas foetus and Trichomonas vaginalis hydrogenosomes. Molecular and biochemical parasitology **20**: 57-65.
Stover NA, Rice JD (2011) Distinct cyclin genes define each stage of ciliate conjugation. Cell Cycle **10**: 1699-1701.
Strauss M, Hofhaus G, Schröder RR, Kühlbrandt W (2008) Dimer ribbons of ATP synthase shape the inner mitochondrial membrane. The EMBO Journal **27**: 1154-1160.
Sun F, Huo X, Zhai Y, Wang A, Xu J, Su D, Bartlam M, Rao Z (2005) Crystal structure of mitochondrial respiratory membrane protein complex II. Cell **121**: 1043-1057.
Sutak R, Dolezal P, Fiumera HL, Hrdy I, Dancis A, Delgadillo-Correa M, Johnson PJ, Müller M, Tachezy J (2004) Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote *Trichomonas vaginalis*. Proceedings of the National Academy of Sciences of the United States of America **101**: 10368-10373.
Suyama Y, Miura K (1968) Size and structural variations of mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of America **60**: 235.
Swart EC, Bracht JR, Magrini V, Minx P, Chen X, Zhou Y, Khurana JS, Goldman AD, Nowacki M, Schotanus K (2013) The *Oxytricha trifallax* Macronuclear Genome: A Complex Eukaryotic Genome with 16,000 Tiny Chromosomes. PLoS Biology **11**: e1001473.
Swart EC, Nowacki M, Shum J, Stiles H, Higgins BP, Doak TG, Schotanus K, Magrini VJ, Minx P, Mardis ER (2012) The *Oxytricha trifallax* mitochondrial genome. Genome Biology and Evolution **4**: 136-154.
Szostak JW, Blackburn EH (1982) Cloning yeast telomeres on linear plasmid vectors. Cell **29**: 245-255.
Taanman J-W (1999) The mitochondrial genome: structure, transcription, translation and replication. Biochimica et Biophysica Acta (BBA)-Bioenergetics **1410**: 103-123.
Tard C, Pickett CJ (2009) Structural and functional analogues of the active sites of the [Fe]-,[NiFe]-, and [FeFe]-hydrogenases. Chemical Reviews **109**: 2245-2274.
Thauer RK, Kaster A-K, Seedorf H, Buckel W, Hedderich R (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. Nature Reviews Microbiology **6**: 579-591.
Tielens A (1994) Energy generation in parasitic helminths. Parasitology Today **10**: 346-352.
Tielens AG, Rotte C, van Hellemond JJ, Martin W (2002) Mitochondria as we don't know them. Trends in Biochemical Sciences **27**: 564-572.
Tielens AG, van Grinsven KW, Henze K, van Hellemond JJ, Martin W (2010) Acetate formation in the energy metabolism of parasitic helminths and protists. International Journal for Parasitology **40**: 387-397.
Tielens AG, Van Hellemond JJ (1998) The electron transport chain in anaerobically functioning eukaryotes. Biochimica et Biophysica Acta (BBA)-Bioenergetics **1365**: 71-78.
Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature Reviews Genetics **5**: 123-135.
Tovar J, Fischer A, Clark CG (1999) The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. Molecular Microbiology **32**: 1013-1021.

Tran-Betcke A, Warnecke U, Böcker C, Zaborosch C, Friedrich B (1990) Cloning and nucleotide sequences of the genes for the subunits of NAD-reducing hydrogenase of *Alcaligenes eutrophus* H16. Journal of Bacteriology **172**: 2920-2929.
Uyeda K, Rabinowitz JC (1971) Pyruvate-ferredoxin oxidoreductase III. Purification and properties of the enzyme. Journal of Biological Chemistry **246**: 3111-3119.
Uzarska MA, Nasta V, Weiler BD, Spantgar F, Ciofi-Baffoni S, Saviello MR, Gonnelli L, Mühlenhoff U, Banci L, Lill R (2016) Mitochondrial Bol1 and Bol3 function as assembly factors for specific iron-sulfur proteins. eLife **5**: e16673.
van Bruggen JJ, Stumm CK, Vogels GD (1983) Symbiosis of methanogenic bacteria and sapropelic protozoa. Archives of Microbiology **136**: 89-95.
van der Giezen M, Birdsey GM, Horner DS, Lucocq J, Dyal PL, Benchimol M, Danpure CJ, Embley TM (2003) Fungal hydrogenosomes contain mitochondrial heat-shock proteins. Molecular Biology and Evolution **20**: 1051-1061.
Van Der Giezen M, Rechinger KB, Svendsen I, Durand R, Hirt RP, Fevre M, Embley TM, Prins RA (1997a) A mitochondrial-like targeting signal on the hydrogenosomal malic enzyme from the anaerobic fungus *Neocallimastix frontalis:* support for the hypothesis that hydrogenosomes are modified mitochondria. Molecular Microbiology **23**: 11-21.
van der Giezen M, Sjollema KA, Artz RR, Alkema W, Prins RA (1997b) Hydrogenosomes in the anaerobic fungus *Neocallimastix frontalis* have a double membrane but lack an associated organelle genome. FEBS Letters **408**: 147-150.
van der Giezen M, Slotboom DJ, Horner DS, Dyal PL, Harding M, Xue G-P, Embley TM, Kunji ERS (2002) Conserved properties of hydrogenosomal and mitochondrial ADP/ATP carriers: a common origin for both organelles. The EMBO Journal **21**: 572-579.
van der Giezen M, Tovar J, Clark CG (2005) Mitochondrion-Derived Organelles in Protists and Fungi. International Review of Cytology **244**: 175-225.
Van Hellemond JJ, Klockiewicz M, Gaasenbeek CP, Roos MH, Tielens AG (1995) Rhodoquinone and complex II of the electron transport chain in anaerobically functioning eukaryotes. Journal of Biological Chemistry **270**: 31065-31070.
Van Hellemond JJ, Tielens A (1994) Expression and functional properties of fumarate reductase. Biochemical Journal **304**: 321.
Van Hoek A, van Alen TA, Sprakel V, Hackstein J, Vogels GD (1998) Evolution of anaerobic ciliates from the gastrointestinal tract: phylogenetic analysis of the ribosomal repeat from *Nyctotherus ovalis* and its relatives. Molecular Biology and Evolution **15**: 1195-1206.
van Hoek AH, Akhmanova AS, Huynen MA, Hackstein JH (2000a) A mitochondrial ancestry of the hydrogenosomes of *Nyctotherus ovalis*. Molecular Biology and Evolution **17**: 202-206.
van Hoek AH, van Alen TA, Sprakel VS, Leunissen JA, Brigge T, Vogels GD, Hackstein JH (2000b) Multiple acquisition of methanogenic archaeal symbionts by anaerobic ciliates. Molecular Biology and Evolution **17**: 251-258.
Velours J, Arselin G (2000) The *Saccharomyces cerevisiae* ATP synthase. Journal of Bioenergetics and Biomembranes **32**: 383-390.
Verni F, Gualtieri P (1997) Feeding behaviour in ciliated protists. Micron **28**: 487-504.
Verni F, Rosati G (2011) Resting cysts: A survival strategy in Protozoa Ciliophora. Italian Journal of Zoology **78**: 134-145.
Vignais PM, Billoud B, Meyer J (2001) Classification and phylogeny of hydrogenases. FEMS Microbiology Reviews **25**: 455-501.
Voncken FG, Boxma B, van Hoek AH, Akhmanova AS, Vogels GD, Huynen M, Veenhuis M, Hackstein JH (2002) A hydrogenosomal [Fe]-hydrogenase from the anaerobic chytrid *Neocallimastix* sp. L2. Gene **284**: 103-112.
Voordouw G, Brenner S (1985) Nucleotide sequence of the gene encoding the hydrogenase from *Desulfovibrio vulgaris* (Hildenborough). The FEBS Journal **148**: 515-520.
Voordouw GT, Strang JD, Wilson FR (1989) Organization of the genes encoding [Fe] hydrogenase in *Desulfovibrio vulgaris* subsp. *oxamicus* Monticello. Journal of Bacteriology **171**: 3881-3889.

Vossbrinck C, Maddox J, Friedman S, Debrunner-Vossbrinck B, Woese C (1987) Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. Nature.
Wagener S, Bardele C, Pfennig N (1990) Functional integration of *Methanobacterium formicicum* into the anaerobic ciliate *Trimyema compressum*. Archives of Microbiology **153**: 496-501.
Watanabe KI, Bessho Y, Kawasaki M, Hori H (1999) Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal *Dicyema*. Journal of Molecular Biology **286**: 645-650.
Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics **25**: 1189-1191.
Wawrzyniak I, Roussel M, Diogon M, Couloux A, Texier C, Tan KS, Vivarès CP, Delbac F, Wincker P, El Alaoui H (2008) Complete circular DNA in the mitochondria-like organelles of *Blastocystis hominis*. International Journal for Parasitology **38**: 1377-1382.
Wesolowski M, Fukuhara H (1981) Linear mitochondrial deoxyribonucleic acid from the yeast *Hansenula mrakii*. Molecular and Cellular Biology **1**: 387-393.
Whatley JM, John P, Whatley FR (1979) From Extracellular to Intracellular: The Establishment of Mitochondria and Chloroplasts. Proceedings of the Royal Society of London. Series B, Biological Sciences **204**: 165-187.
Williams BA, Hirt RP, Lucocq JM, Embley TM (2002) A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. Nature **418**: 865-869.
Williams TA, Foster PG, Cox CJ, Embley TM (2013) An archaeal origin of eukaryotes supports only two primary domains of life. Nature **504**: 231-236.
Williams TA, Foster PG, Nye TM, Cox CJ, Embley TM (2012) A congruent phylogenomic signal places eukaryotes within the Archaea. Proceedings of the Royal Society of London B: Biological Sciences **279**: 4870-4879.
Woese CR (1987) Bacterial evolution. Microbiological Reviews **51**: 221.
Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences **87**: 4576-4579.
Wurm CA, Jakobs S (2006) Differential protein distributions define two sub-compartments of the mitochondrial inner membrane in yeast. FEBS Letters **580**: 5628-5634.
Xia D, Yu C-A, Kim H, Xia J-Z, Kachurin AM, Zhang L, Yu L, Deisenhofer J (1997) Crystal structure of the cytochrome bc1 complex from bovine heart mitochondria. Science **277**: 60-66.
Yagi T, Dinh TM (1990) Identification of the NADH-binding subunit of NADH-ubiquinone oxidoreductase of *Paracoccus denitrificans*. Biochemistry **29**: 5515-5520.
Yang X, Trumpower B (1986) Purification of a three-subunit ubiquinol-cytochrome c oxidoreductase complex from *Paracoccus denitrificans*. Journal of Biological Chemistry **261**: 12282-12289.
Yankovskaya V, Horsefield R, Törnroth S, Luna-Chavez C, Miyoshi H, Léger C, Byrne B, Cecchini G, Iwata S (2003) Architecture of succinate dehydrogenase and reactive oxygen species generation. Science **299**: 700-704.
Yarlett N, Hann AC, Lloyd D, Williams A (1981) Hydrogenosomes in the rumen protozoon *Dasytricha ruminantium* Schuberg. Biochemical Journal **200**: 365-372.
Yarlett N, Orpin C, Munn E, Yarlett N, Greenwood C (1986) Hydrogenosomes in the rumen fungus *Neocallimastix patriciarum*. Biochemical Journal **236**: 729-739.
Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature.
Zaug AJ, Cech TR (1980) In vitro splicing of the ribosomal RNA precursor in nuclei of *Tetrahymena*. Cell **19**: 331-338.
Zerbes RM, van der Klei IJ, Veenhuis M, Pfanner N, van der Laan M, Bohnert M (2012) Mitofilin complexes: conserved organizers of mitochondrial membrane architecture. Biological Chemistry.
Zick M, Rabl R, Reichert AS (2009) Cristae formation—linking ultrastructure and function of mitochondria. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1793**: 5-19.

Zierdt C, Donnolley C, Muller J, Constantopoulos G (1988) Biochemical and ultrastructural study of *Blastocystis hominis*. Journal of Clinical Microbiology **26**: 965-970.
Zufall RA, McGrath CL, Muse SV, Katz LA (2006) Genome architecture drives protein evolution in ciliates. Molecular Biology and Evolution **23**: 1681-1687.