Investigations on the Mitochondrial Protein Composition of Microbial Eukaryotes from Diverse Physical Environments.

A thesis submitted for the degree of
Master of Philosophy (MPhil)

Luke Williams

School of Biosciences, Stacey Building,
University of Kent,
Canterbury, Kent,
CT2 7NJ

2017

Supervisors:
Dr. Anastasios Tsaousis
Dr. Mark Wass

Word count: 14,117

Abstract

Ciliates are a group of microbial eukaryotes that are found in diverse environments, which have been identified to contain different classes of mitochondria. It is understood that mitochondria have modified their functions overtime depending on the environment of their host; these differences in adaptations allow mitochondria to be classified into five different groups/classes. Four of these classes generate adenosine triphosphate (ATP), while the fifth has undergone evolutionary reduction having no apparent role in ATP synthesis. The aim of this project was to evaluate the diversity of these organelles in ciliates. This was carried out using computational methods, where we analysed and translated genomic and transcriptomic data collected from 27 ciliates found in diverse environments. The ciliate data were searched against a database of nine different mitochondrial proteomes and multiple anaerobically functioning proteins in order to assemble the potential mitochondrial proteome for each organism based on sequence similarity.

Our method allowed the prediction of mitochondrial class diversity within one 'group' of eukaryotes, identifying all five classes throughout the 27 ciliates, including the proposal of a potential mitosome present within the rumen-living ciliate *Entodinium caudatum,* as well as confirming the existence of hydrogenosomes present within *Polyplastron multivesiculatum.* By analysing functional proteins of the citric acid cycle and the electron transport chain, as well as identifying any proteins that function in anaerobic conditions; we have been able to make clear distinctions between each mitochondrial class and provide a strong likelihood that all five mitochondrial classes are present across the ciliate clade,

along with the possible presence of intermediated organelles that do not fit the already classified groups.

This diversity of mitochondrial class seen throughout the ciliate clade further supports the claim of anaerobically functioning mitochondria potentially being the ancestral mitochondria of ciliates.

Acknowledgements

I would like to thank Dr. Eleni Gentekaki at the Mae Fah Luang University and the late Prof. Denis H. Lynn at the University of British Columbia for providing the transcriptomic ciliate data and my supervisors Dr. Anastasios Tsaousis and Dr. Mark Wass for their support throughout.

Declaration

I certify that this is my own work, except where acknowledged by reference and citation.

Contents

List of Figures

List of Tables

Abbreviations

TCA	Tricarboxylic acid or citric acid cycle
ETC	Electron transport chain
Fe/S	Iron sulphur cluster
PNO	Pyruvate: NADP+ oxidoreductase
PFO	Pyruvate: ferredoxin oxidoreductase
AOX	Alternative oxidase
ATP	Adenosine triphosphate
ISC	Iron sulphur cluster
TIM	Translocase of the inner membrane
TOM	Translocase of the outer membrane
mtDNA	Mitochondrial DNA

Introduction

Mitochondria are viewed as vital organelles due to their characteristic as energy providers, and are considered to have a fundamental role in the evolution of eukaryotic cells for this reason. Until recently it was thought that mitochondria were present in all eukaryotes; however the recent discovery from Karnkowska et al 2016 altered this [1]. They identified that the microbial eukaryotic *Monocercomonoides sp.* lacks all hallmark mitochondrial proteins. It has been understood that mitochondria have been able to modify their function dependent on the environment of their host, and this adaptation in diverse eukaryotes remains somewhat of a mystery. Recent studies have observed these different anaerobic functional mitochondria including hydrogen producing hydrogenosomes as well as remnant mitochondria such as mitosomes that are found in several parasites; all share the distinctive double membrane [2, 3].

Mitochondrial proteins are involved in many pathways, especially those relating to energy production for cells. Proteins targeted to these pathways and the mitochondria as an organelle are encoded by both the mitochondrial and nuclear genome. Mitochondrial genomes studied to date encode no more than 70 of the proteins that mitochondria utilise to function, which include subunits of complex I, III, IV and V of the electron transport chain [4]. Since a typical aerobic mitochondrial proteome compromises 1200 proteins, most mitochondrial proteins are encoded by the nuclear genome and are targeted to the mitochondrion using a protein import machinery specific to this organelle [5].

The ancestral origin of mitochondria appears likely to have arisen from primary endosymbiosis between a host cell related to Asgard Archaea and intracellular

Rickettsiales α-proteobacteria [6–9]. The 'engulfed' α-proteobacterium became reliant on the protective environment of the host, and conversely the Achaea host came to rely on the energy provided by the α-proteobacteria, which over time via many evolutionary steps developed into the mitochondria we know today.

There are a few hypothesis surrounding the origin of mitochondria, whether mitochondria evolved as sister to, or within (as outlined above) *Rickettsiales*, where the ancestor was an energy parasite [10]. While others propose that mitochondria emerged from a deeper position within α-proteobacterium, which would suggest that the pre-mitochondrial α-proteobacterium was free-living and a facultative anoxygenic photosynthesiser – a physiology that may have likely been ancestral to all α-proteobacterium [11, 12, 13].

Mitochondria have been categorised into five different classes by Müller et al [14] depending on their function/size/morphology; classes 1-5, which are aerobic mitochondria (textbook type), anaerobic mitochondria, hydrogen-producing mitochondria, hydrogenosome and mitosome respectively. It was first believed that hydrogenosomes did not share a common ancestor with mitochondria until the discovery of organellar chaperone proteins common to both mitochondria and hydrogenosomes, suggesting a possible shared ancestral origin [11–14]. Following the discoveries of shared chaperone proteins, other common shared proteins and functions between both have been discovered, including the discovery of the further reduced form of mitochondria called mitosomes. The presence of mitochondria being throughout most eukaryotes indicates the organelles' importance [3, 15, 16].

Mitochondrial Functions

The following sections provide an overview of the functions and metabolic pathways of mitochondria, followed by the diversity between each mitochondrial class. The variations seen with mitochondrial functions and the pathways they contain help distinguish the type of mitochondria present within a species.

-The Mitochondrial Genome

A unique feature of these organelles is that they possess their own DNA, where it was derived from the circular genome of the engulfed ancestral α-proteobacteria. In humans, mitochondrial DNA (mtDNA) is a circular negatively supercoiled double stranded molecule that contains 16,569 base pairs, which encode for 13 polypeptides, 22 transfer RNA (tRNAs) and two ribosomal RNAs (rRNA). All 13 proteins encoded by mtDNA are important subunits of the electron transport chain and ATP synthase. mtDNA encodes for seven of the 46 subunits of complex I, one of the 11 subunits of complex III, three of the 13 subunits of complex IV and two of the 16 subunits of complex V (ATP synthase). mtDNA has been identified in aerobic and anaerobic functioning mitochondria, along with the ciliate of *Nyctotherus ovalis* and *Blastocystis sp.* that contain a reduced class of mitochondria - class III Hydrogen-producing mitochondria.

-Iron-sulphur Cluster

Iron-sulphur cluster (Fe/S) proteins perform critical roles in many metabolic reactions within cells; their functions include involvement in electron transfer within complex I, II and III of electron transport chain, enzyme catalysis like aconitase in the TCA, as well as nuclear gene expression. The biosynthesis of these inorganic Fe/S centres and their insertion into apoproteins require dedicated and complex cellular machinery that is found within the mitochondria as well as within the cytosol.

The Fe/S cluster assembly machinery within the mitochondrion is known as the ISC (iron sulphur cluster) and is vital for the formation of Fe/S proteins (Figure 1). The assembly pathway begins when a ferrous form of iron is imported into the mitochondria via the mitoferrin transporter protein. Frataxin (Fxn) then facilitates the delivery of the reduced iron molecule to the IscU scaffold protein and directly binds to IscU [21]. In conjunction to this, an aminotransferase known as cysteine deslufurase (IscS) catalyses the conversion of cysteine to alanine and sulfane sulphur. The sulfur produced is then transferred to the IscU proteins and along with the reduced iron form transient Fe/S proteins on the IscU scaffold [7, 18, 19]. Isd11 is bound to cysteine desulfurase and is believed to be a stabilising component preventing aggregation of cysteine desulfurase [24]. Ferredoxin and ferredoxin reductase (Fdx1 and FdxR) are important in the maturation of Fe/S proteins and apparently function early in Fe/S cluster formations, serving as an electron transfer chain providing electrons (NADH preferred) for the reduction of sulfur to sulphide formed in the cysteine desulfurase [21, 22].

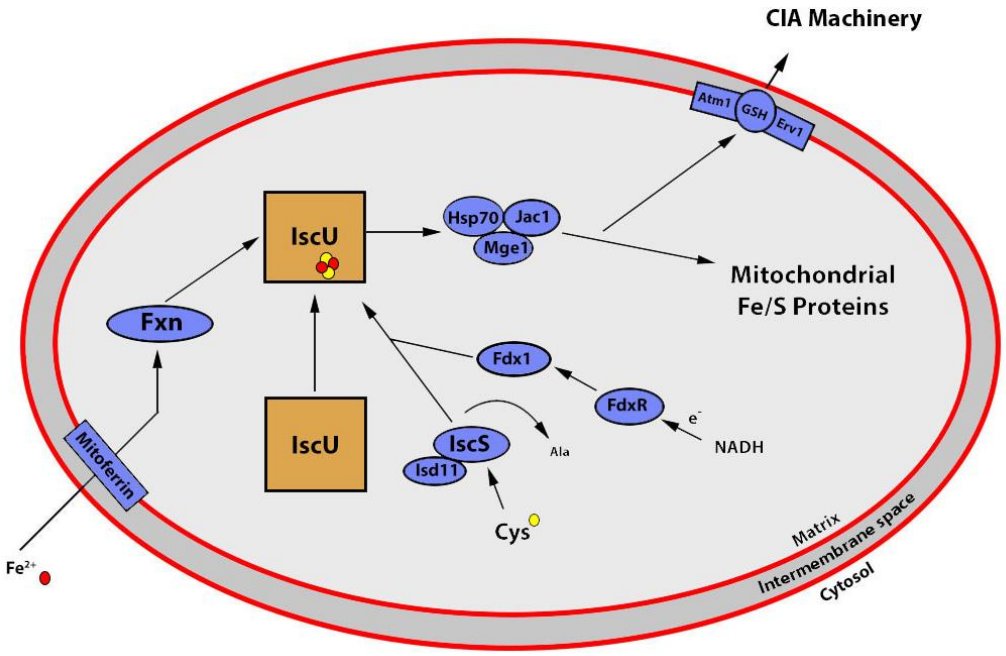


Figure 1 - Overview of the Fe/S protein maturation pathways and functional proteins involved. Abbreviations: Fxn – frataxin, IscS – cysteine desulfurase, FdxR – ferredoxin reductase, Fdx1 – ferredoxin.

The Fe/S cluster assembly machinery encompasses a dedicated chaperone complex system that transfers the Fe/S cluster to apoproteins to be further synthesised within the mitochondria or to be exported to the cytosol via the transporter, ATM1. This complex is formed by a specialised Hsp70, a DnaJ-type cochaperone Jac1 and a nucleotide exchange factor Mge1, which all play a role in the biogenesis of Fe/S proteins [27]. The specialised Hsp70 requires the ADP/ATP nucleotide factor Mge1 for function, but like with other Hsp70s it also demonstrates weak intrinsic ATPase activity, and that is where Jac1 assists, stimulating the ATP-dependent binding between the chaperone complex and IscU [28].

-β-Oxidation Pathway

Fatty acid oxidation was first identified by Franz Knoop, where his experiments showed that fatty acids must be degraded by oxidation at the β-Carbon, followed by cleavage of the C_α-C_β bond.

β-oxidation is an important metabolic process, which primarily takes place within mitochondria and is a vital energy source for eukaryotic organisms by utilising the energy provided by fatty acids. However, for this process to start the fatty acids firstly need to be activated to be able to enter through the mitochondria membranes. This is done by acyl-CoA synthetase to yield fatty acyl-CoA. At this stage the fatty acid is able to enter into the mitochondria with the aid of carnitine acyltransferase carrier proteins [29].

β-oxidation is a catabolic process by which fatty acids are degraded, it drives ATP synthesis by oxidative phosphorylation as well as providing acetyl-CoA for further catabolism in the TCA cycle, β-oxidation also leads to the production of NADH and $FADH_2$, which are co-enzymes used in the electron transport chain.

There are four different steps required for the degradation of fatty acids (Figure 2), each of which are catalysed by a specific enzyme with specificity for short, medium and long chain acyl-CoA intermediates. This process is known as the β-oxidation cycle. The four steps are:

Oxidation – conversion of acyl-CoA to Δ2-3-trans-enoyl-CoA in the presence of FAD. This conversion is catalysed by acyl-CoA dehydrogenase.

Hydration – Enoyl-CoA hydratase adds water across the double bond to produce L-3-hydroxyacyl-CoA.

Oxidation – L-3-hydroxyacyl-CoA dehydrogenase catalyses L-3-hydroxyacyl-CoA to L-3-ketoacyl-CoA, with the involvement of NAD^+ as a cofactor.

Cleavage – The splitting of L-3-ketoacyl-CoA at the 2,3 position by 3-ketoacyl-CoA thiolase, in the presence of CoA to produce one acetyl-CoA and one acyl-CoA. The acetyl-CoA goes onto the TCA cycle, whereas the acyl-CoA feeds back to the acyl-CoA dehydrogenase to begin another cycle of β-oxidation [29].

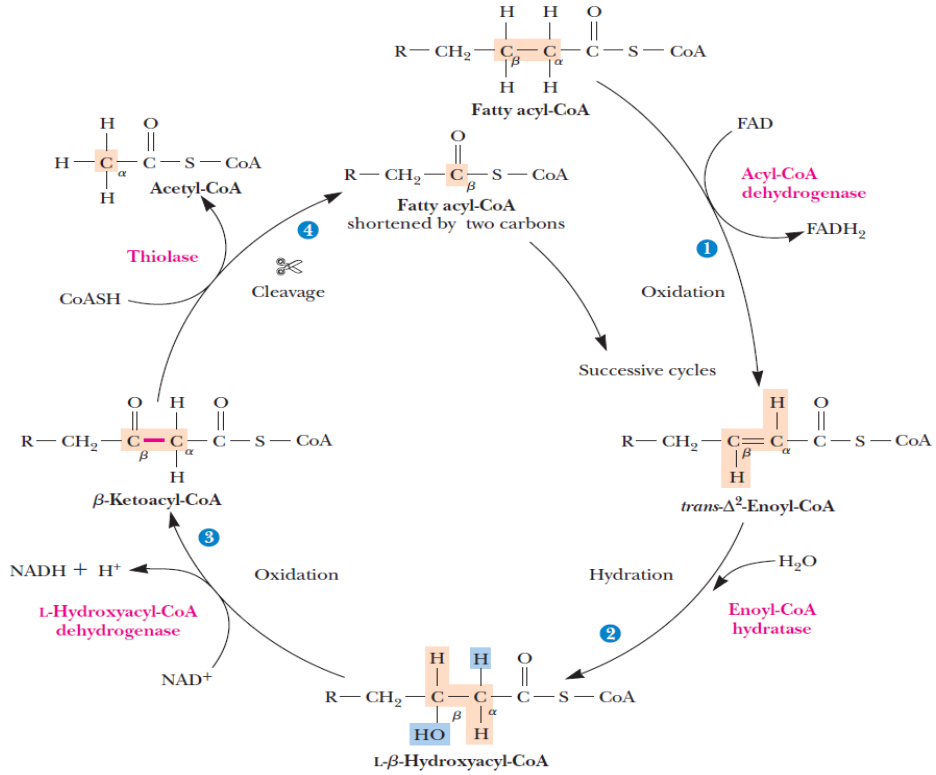


Figure 2 - The β-oxidation of saturated fatty acids. It is a cycle containing 4 steps catalysed by specific enzymes, with each cycle producing a single molecule of NADH, $FADH_2$ and acetyl-CoA. Biochem 4th ed

-Protein Import

The molecular machines that drive protein import to the mitochondria were established in the last common ancestor for all eukaryotes [30]. Most proteins destined for the mitochondria are encoded by the nucleus, so an efficient protein transport system is required. Proteins targeted for the mitochondria carry either an N-terminal signal sequence or internal signal sequences, which are recognised by the TOM (Translocase of the Outer Membrane) complex and the TIM (Translocase of the Inner Membrane) prior to being sorted ensuring proper delivery and assignment of the protein to function within the organelle. The majority of the proteins destined for the mitochondria contain a N-terminal signal sequence comprised of 10-80 residues, which are usually hydrophobic and form an amphipathic α-helix [31].

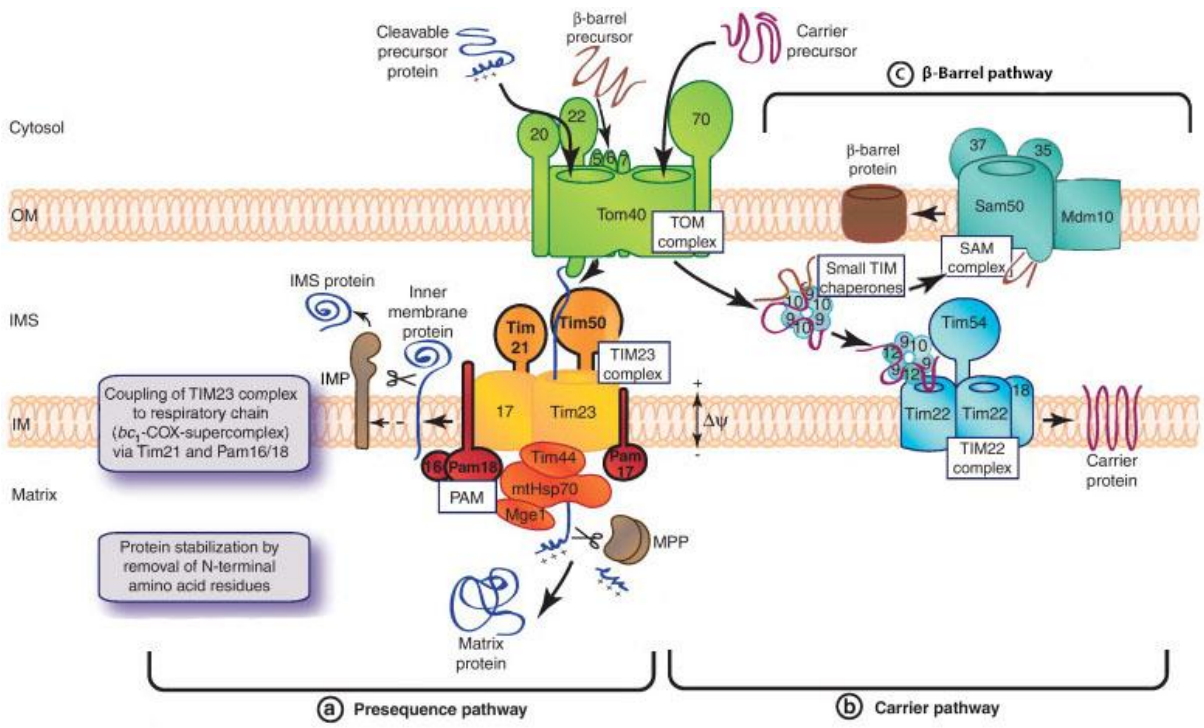


Figure 3 – Import protein pathways. Thomas Becker et al 2012 (modified). a) Import proteins involved in presequences protein import pathway. b) Import proteins involved in proteins destined to be carrier proteins within the membrane. c) Shows the proteins involved in β-barrel protein pathway.

Proteins can also contain internal signal sequences which are less defined and can be found at any point throughout the protein. However, these proteins do not enter the matrix but rather targeted to either, the inner membrane, the intermembrane space or the outer membrane [32]. The same N-terminal signal patterns also appears on proteins targeted to hydrogenosomes and mitosomes, however the N-terminal targeting signal sequences is less defined for mitosomes [26, 28]. There is also an example of mitosomes, which includes *Entamoeba histolytica,* potentially disposing the use of N-terminal targeted proteins, relying on internal signals for protein trafficking [5, 29].

The protein translocases found on the mitochondrial membranes are formed with a core translocation unit that is enhanced by one or more subunit modules that have a distinct function, aiding the translocation of targeted proteins. These are known as TOM complexes which are composed of core translocases TOM40, 22 and 7, alongside other subunits like TOM5 and TOM6 that assist in protein transfer, and receptor subunits TOM70 and TOM20 that promote protein binding [30–32]. Mitochondrial targeted proteins with the aid of cytosolic HSP70 chaperone proteins bind to TOM70, where TOM20/22 recognise the presequence target signal, transferring the protein towards TOM40 channel likely via TOM5 where the N-terminal signal binds to the *trans* site within the TOM40 complex.

TOM complexes are the main entry gate for mitochondrial proteins; however other import machineries are required for protein insertion in the outer membrane. This pathway is called the β-barrel pathway, where β–barrel proteins are inserted into the outer membrane via the SAM (sorting assembly machinery) complex (Figure 3c). Precursor proteins for β-barrels consist of the β-strand carboxy terminal which

once translocated into the intermembrane space, a complex known as ‘small TIMs or tiny TIMs’ bind and direct the proteins to the SAM complex.

There are two major TIM complexes embedded on the inner membrane, TIM23 and TIM22. TIM22 only binds to proteins that are designated for integration into the inner membrane and is composed of four membrane embedded subunits, TIM22, 12, 54 and 18, where they assist in the docking of proteins and their assembly. TIM22 also has interactions with the small TIM complex which shuttles to and from TOM40 and TIM22 collecting and delivering the targeted proteins [30, 32, 33].

The TIM23 complex translocates presequence proteins into the mitochondrial matrix where the positive charges on the presequences of the translocating protein are drawn through the TIM23 channel of the inner membrane. This is aided by the membrane potential already present between the mitochondrial matrix and intermembrane space. Associated with TIM23 complex is TIM50, a receptor that guides proteins to bind to the translocase channel regulating the opening and closing of this channel [32, 34]. TIM21 interacts with the second core translocase TIM17, both determining the destination of the protein and whether it should be translocated into the mitochondrial matrix [35, 36].

Hsp70 is present outside the channel opening, anchored to the inner membrane by PAM (Presequence translocase-Associated Motor) complex proteins and TIM44, where Hsp70 activity is harnessed to aid proteins to be released into the matrix, this complex of proteins act like a motor to drive the translocation. Once they have entered the mitochondrial matrix, the positively charged presequence cleavage motif present on the protein is identified by mitochondrial processing peptidase

(MPP) which proceeds to cleaving the proteins N-terminal presequence [37, 38]. MPP typically consists of two subunits, a α-subunit that is involved with binding to the protein and the β-subunit that possesses the catalytic activity to cleave off the presequence [44].

Aerobic Mitochondrial Metabolism

Aerobic mitochondria are considered to be text-book mitochondria; these membrane bound organelles are essential in aerobic eukaryotes and are known as the energy suppliers of the cells due to their production of ATP. The significance of both the TCA cycle and oxidative phosphorylation carried out via the electron transport chain are major factors for why these organelles have been retained among diverse aerobic eukaryotes [45]. They are also recognised to be involved in other cellular processes such as cellular proliferation, apoptosis, steroid and haem synthesis, Fe/S cluster assembly, lipid metabolism, amino acid metabolism and fusion/fission [10, 41].

Three major pathways are commonly associated with aerobic mitochondria; these are the citric acid cycle (TCA), oxidative phosphorylation or electron transport chain (ETC) and β-oxidation. Figure 4 shows the major metabolic pathways seen within mitochondria.

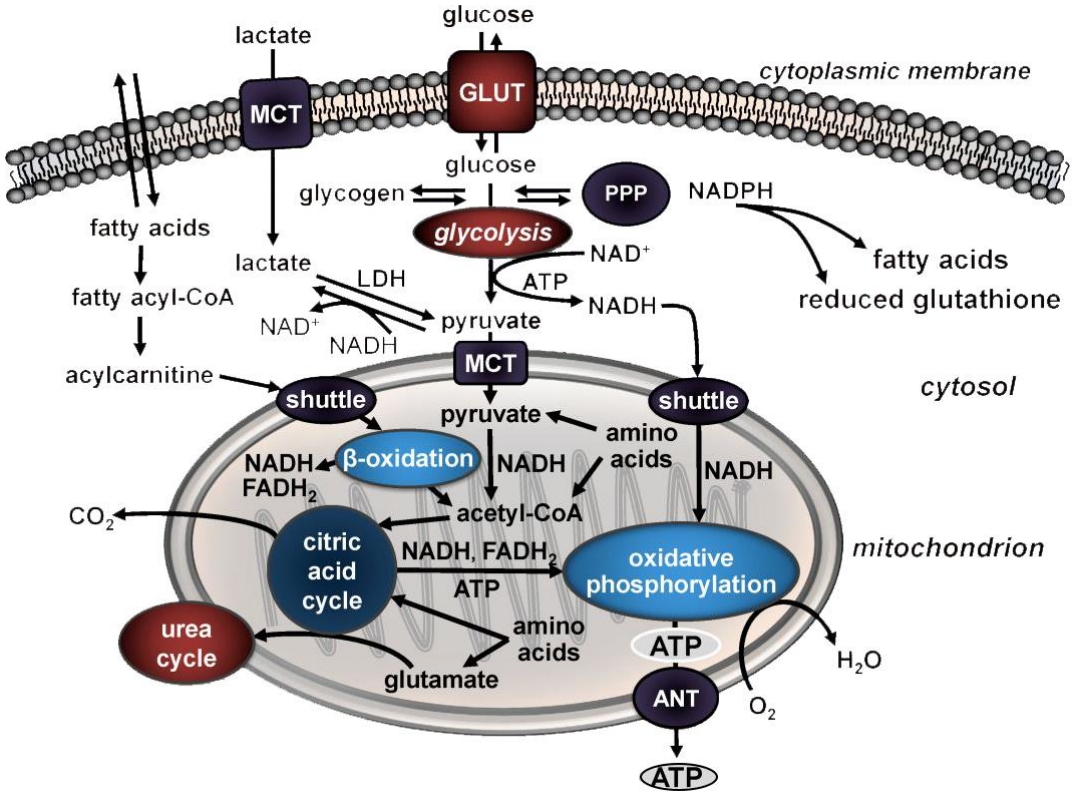


Figure 4 – Mitochondrial metabolic pathway overview, showing the three major pathways that mitochondria are involved in, oxidative phosphorylation, β-oxidation and citric acid cycle.

-The Citric Acid Cycle

The TCA cycle was proposed by Hans Adolf Krebs in 1937 and is the most central metabolic pathway for all the aerobic processes required by the organism, with input from carbohydrate, fatty acid and amino acid oxidation pathways. The catabolism of these three pathways each produces acetyl-CoA. Glycolysis is the most recognised, which is the breakdown of glucose yielding pyruvate where the enzyme pyruvate dehydrogenase decarboxylates pyruvate to produce acetyl-CoA, which then enters the TCA cycle to undergo oxidation [47]. This is carried out by eight enzymes within the cycle (Figure 5), these enzymes catalyse a series of organic reactions that cumulatively oxidase one acetyl-CoA to produce two CO_2 molecules along with three NADHs, one $FADH_2$ and one GTP. The production of NADH and FADH is vital, as these are used via the oxidative phosphorylation pathway by being re-oxidised by O_2 completing the breakdown of this metabolic fuel driving ATP synthesis, ultimately producing the energy needed for the cells to function.

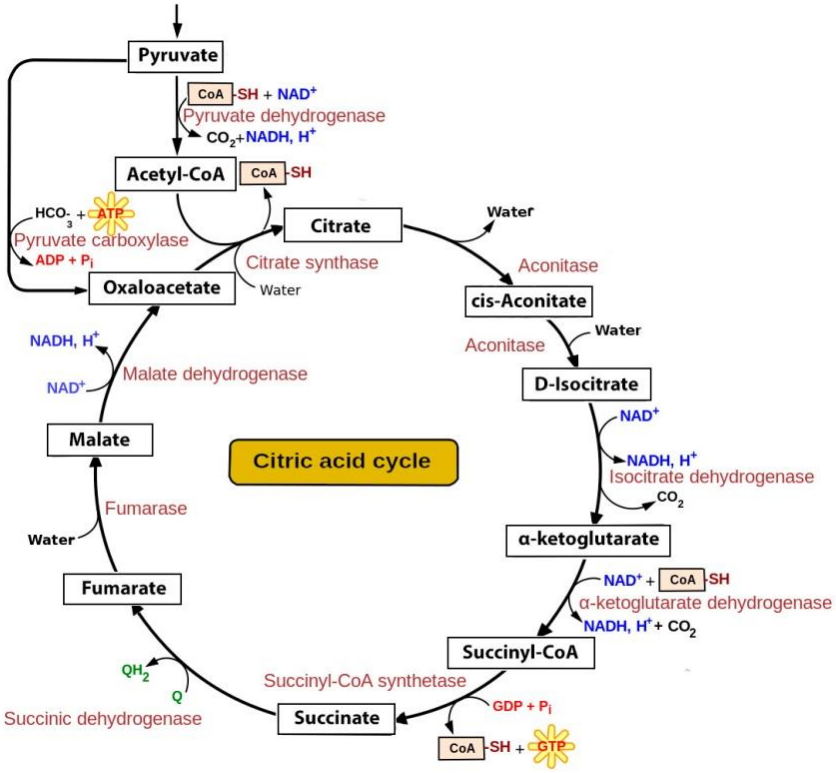


Figure 5 – The citric acid cycle (TCA), showing each enzyme involved.

-The Electron Transport Chain

Mitochondria are distinctively recognisable by their double membrane, where the inner membrane contains a series of protein complexes that form the electron transport chain. The electron transport chain is comprised of complex I, II, III, IV and the ATP synthase. It uses oxygen as a terminal electron acceptor and carries out electron transfer via redox reactions, creating a proton gradient across the membrane, driving ATP synthesis. A basic overview of the electron transport chain is shown in Figure 6.

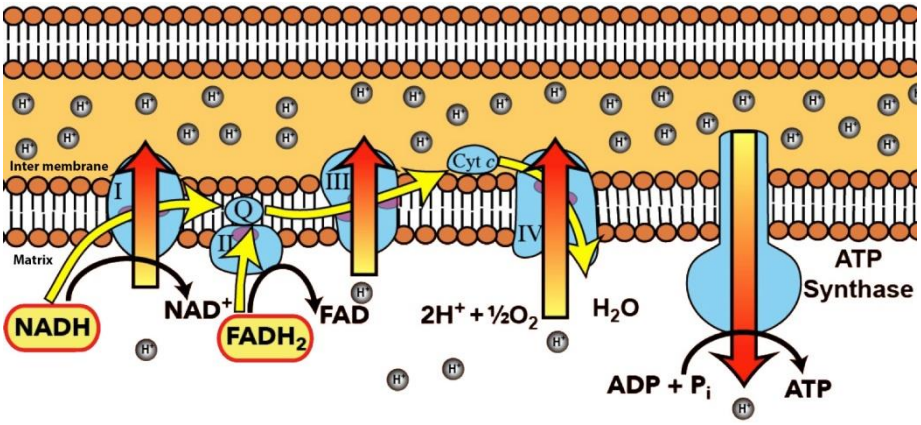


Figure 6 – Basic overview of the steps involved in oxidative phosphorylation along the electron transport chain. Image modified from science music videos
Source: (*https://www.sciencemusicvideos.com)*

-Complex I NADH:ubiquinone oxidoreductase

This is the first enzyme complex of the electron transport chain, catalysing the initial reaction of the electron transport chain by the reduction of ubiquinone by NADH. It is a complex enzyme and among the largest membrane-bound protein assemblies, comprised of up to 46 dissimilar subunits, with 14 of these subunits forming the functional core of the enzyme - seven hydrophobic subunits are encoded by the mitochondrial genome and the other seven are encoded via the nuclear genome. When one molecule of NADH binds to the hydrophilic domain of complex I, NADH undergoes oxidation and ultimately ends up transferring two

electrons to ubiquinone causing the translocation of four protons across the inner membrane. This helps create a proton gradient differential between the matrix and the intermembrane space which is required for ATP synthesis.

-Complex II Succinate Dehydrogenase

Complex II differs from the other complexes within the electron transport chain as it is not capable of pumping protons across the inner membrane; however it is unique as it participates in both the electron transport chain and the TCA cycle. In the TCA cycle it catalyses the oxidation of succinate to fumarate. Complex II is comprised of four subunits; two subunits are hydrophobic b-type cytochromes, which anchor the complex to the inner membrane containing a ubiquinone binding site [48]. The two further subunits are hydrophilic and comprise of a flavoprotein that contains a covalently bound flavin-adenine dinucleotide (FAD) prosthetic group and a succinate binding site, the other subunit contains three iron-sulphur centres [49].

-Complex III Cytochrome c Oxidoreductase

Complex III is composed of 11-12 dissimilar subunits, with the cytochrome c subunit being coded by the mitochondrial genome. Complex III procedures a proton gradient via a series of reactions known as the Q-cycle which was originally proposed by Peter Mitchell in 1975 [50]. The Q-cycle (shown in Figure 7) results in the transfer of four protons from the matrix to the intermembrane space via two cycles of ubiquinol oxidation.

Step 1 of the cycle begins when ubiquinol (QH_2) binds to a site on complex III and pumps its two protons into the intermembrane space. Ubiquinone (Q) also binds to a separate binding site. Ubiquinol is oxidised releasing its two electrons - one electron is donated to the Rieske iron-sulphur protein, once the Rieske iron-sulphur protein receives the electron, it unbinds from its position and transfers over to the haem group of cytochrome c_1.The electron is then further transferred to the Cytochrome c externally bound to the complex, where it then transports to complex IV, and is replaced by a new reduced Cytochrome c.

The second electron transfers to cytochrome b which contains two haem groups; the b_L haem group and the b_H haem group. After transferring to each haem group the electron binds to ubiquinone (Q) causing ubiquinone to become a semiquinone radical.

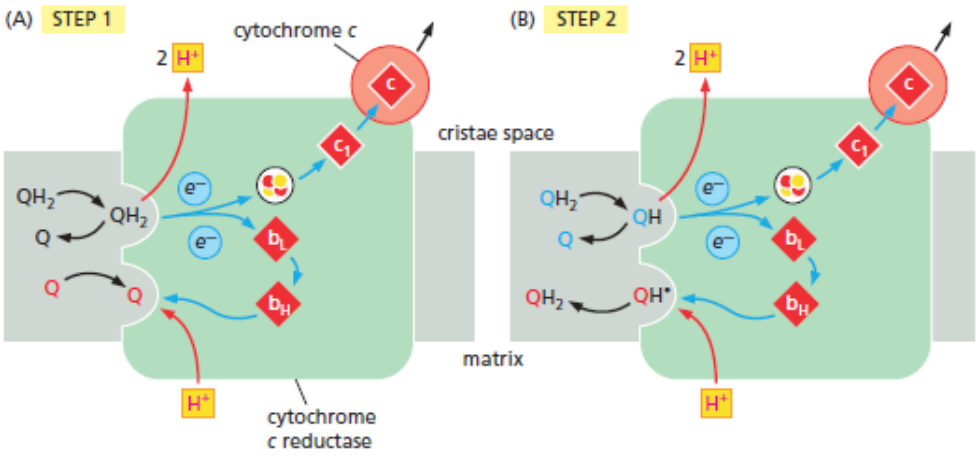


Figure 7 – Overview of the Q-cycle within complex III, that results in the transfer of four protons from the matrix into the intermembrane space via two cycles of ubiquinol oxidation.

Step 2 of the cycle follows the same pathway as previously mentioned in *step 1*; a new ubiquinol binds to complex III releasing the two electrons and causing the two protons to be pumped into the intermembrane space. One electron is transferred to cyctochrome c, whereas the second electron transfers through cytochrome b

and to the semiquinone radical causing two protons to enter from the matrix to form ubiquinol (QH_2) to be recycled and used by complex III again, completing the cycle [51].

-Complex IV Cytochrome c Oxidase

Complex IV is comprised of up to 13 dissimilar subunits, three of which are encoded by the mitochondrial genome. Complex IV like complex III creates a proton gradient, pumping four protons into the intermembrane space during the movement of the electrons from cytochrome c and the formation of two final water molecules. This happens when two reduced cytochrome c molecules bind to complex IV releasing the two electrons; one electron reduces the copper atom group (Cu_B) and the other reduces the haem group (Haem a_3). Once both are in their reduced form, oxygen is able to bind to the two electrons forming a peroxide bridge. This is followed by two more reduced cytochrome c molecules transferring their two electrons, combining this with two protons that are obtained from the matrix to help break the peroxide bridge, establishing two hydroxide groups. Once the two hydroxide groups have formed, another two protons are extracted from the matrix breaking the hydroxide group bonds, forming two water molecules (H_2O) and returning the copper and haem group to their original oxidised state [51].

-ATP Synthase Complex V

ATP synthase generates ATP, which provides the cell with the energy needed to function. ATP is the most used source of energy and is formed by the addition of inorganic phosphate (P_i) to adenosine diphosphate (ADP). The energy required for

ATP synthase to produce ATP molecules is gained from the translocation of protons across the mitochondrial inner membrane creating an electrochemical or proton gradient. The ATP synthase has three major parts; a membrane-embedded F_0 domain, a catalytic F_1 domain and a joining stalk (Figure 8) [51].

The F_1 domain is composed of five different types of polypeptide chains, alpha, beta, gamma, epsilon and delta. Three α-subunits and three β-subunits combine to form a hexameric ring, forming into three pairs of α β subunits; they form three catalytic sites which all adopt three different conformations for different partial reactions of the synthesis of ATP. The γ and ε subunits organise to form a central stalk that is anchored in the F_0 domain but extends through the inner cavity of the hexamer ring and as it rotates, it stimulates the synthesis and release of ATP molecules. Finally the δ subunit holds the αβ hexamer ring in place to prevent it from rotating and connects to the *b* subunit that acts as an arm connection to the stationary region of F_0 domain.

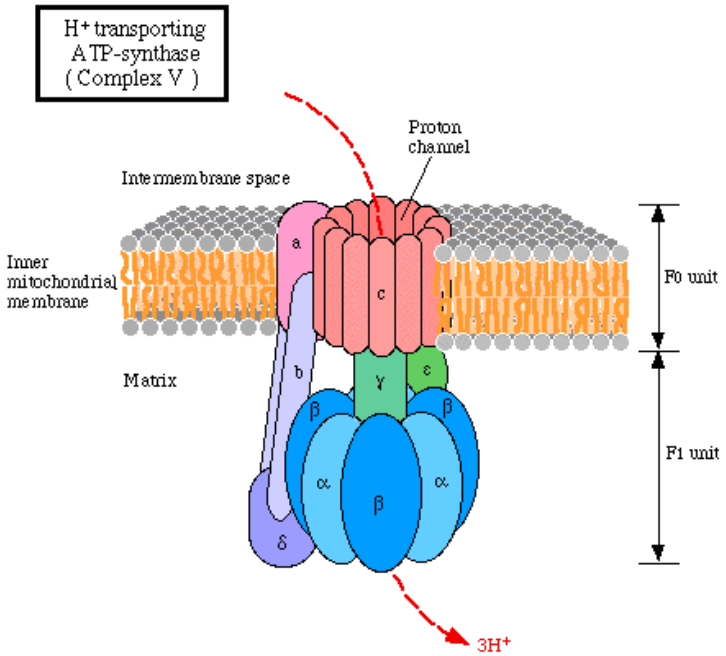


Figure 8 - ATP Synthase has three major parts; a membrane-embedded F_0 domain, a catalytic F_1 domain and a joining stalk. The F_1 domain is composed of five different types of polypeptide chains, alpha, beta, gamma, epsilon and delta. F_0 is a region found within the inner membrane of the mitochondria and consists of 9 – 12 c subunits known as the rotor.

F_0 is a region found within the inner membrane of the mitochondria and consists of 9 – 12 c subunits which is known as the ‘rotor’ and acts like a proton channel allowing protons to flow into the mitochondrial matrix. There is also a single *α* subunit that binds to the outer side of the *c* subunit ring which helps connect both the F_0 and F_1 domains and assists to channel protons through to the rotor.

Due to the electrochemical gradient between the intermembrane space and the matrix, protons travel through the channel down the gradient causing the rotor to rotate, which in turn causes the γε subunits to rotate. The γ subunit has a small ‘bent end’ found central within the αβ hexamer ring, and this causes the αβ subunits shape to deform slightly, triggering conformational changes to the F_1 catalytic sites which leads to the formation of ATP molecules.

Three ATP molecules are synthesised from one rotation of the *c* subunit rotor and γ subunit. The amount of protons required is dependent on the number of *c* subunits e.g. 12 *c* subunits requires 10 protons which yields a ratio of H+/ATP of four [51].

Mitochondria classification and Mitochondria-related organelles

As previously mentioned, mitochondria have been categorised into five different classes by Müller et al as shown in Figure 9 [14]. The classification is dependent on their function/size/morphology; classes 1-4 (ATP producing) or class 5 (no ATP production). Classes 1-4 are then further defined to produce a clear differentiation between the mitochondria and mitochondria-related organelles (MROs). Class 2-4 are anaerobic type mitochondria that use endogenous or environmental compounds as an electron acceptor. Class 3-5 mitochondria are often referred to as MROs, with class 5 being the most reduced and referred to as mitosomes.

Most hydrogenosomes and mitosomes do not contain a genome, although there are a few exceptions in the microbial world (e.g. the anaerobic Stramenopile *Blastocystis)*;

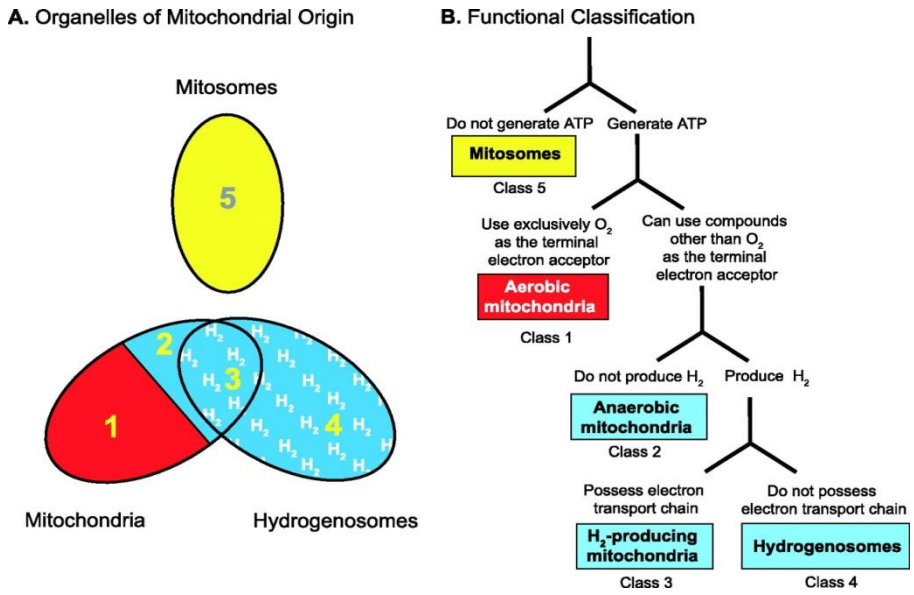


Figure 9 - Functional classification of mitochondria origin adapted from Müller et al [14]. Mitochondria have been divided into five classes, class 1-5. Class 1 is the canonical type, which uses oxygen as the terminal acceptor. Class 2 is an anaerobic functioning mitochondrion, which uses endogenously produced electron acceptor. Class 3 is a hydrogen producing mitochondrion that possesses a hydrogenase and also has an electron transport chain. Class 4 are hydrogenosomes, which are anaerobically functioning and produce ATP by using protons as electron acceptors. Class 5 are mitosomes, which do not generate ATP. Red shows oxygen use, blue shows ATP is produced without oxygen, and yellow shows no ATP production.

Class II: Anaerobic mitochondria

Anaerobic mitochondria are almost functionally identical to that of textbook-type aerobic mitochondria, with the difference being they function anaerobically, using compounds other than oxygen as the final electron acceptor e.g. acetyl-CoA. These types of mitochondria are found in fungi such as, *Fusarium oxysporum* and the parasitic flat worm *Fasciola hepatica.* Most of the organisms that have these anaerobic mitochondria use an endogenously produced electron acceptor, such as fumarate which is converted to succinate by fumarate reductase as an end product that is excreted. The use of environmentally available acceptors like nitrate can also be used as a final electron acceptor, where the fungal plant pathogen *Fusarium oxysporum* is shown to utilise nitrate as its electron acceptor, also referred to as denitrification [52].

There are three major distinctions between class II mitochondria and class I; 1) the enzyme catalysing the conversion between fumarate and succinate (fumarate reductase), 2) the quinone that connects the electron transfer to the electron transport chain (rhodoquinone), and 3) the presence of acetate:succinate CoA-transferase which transfers acetyl-CoA to succinate yielding acetate and succinyl-CoA, an enzyme that is usually typical of hydrogenosomes [53]. The quinone used is rhodoquinone (RQ) and has a lower redox potential than ubiquinone so functions as a sufficient electron donor to fumarate in eukaryotes [47, 48]. Figure 10 shows the energy metabolism in anaerobic mitochondria of *Fasciola hepatica.*

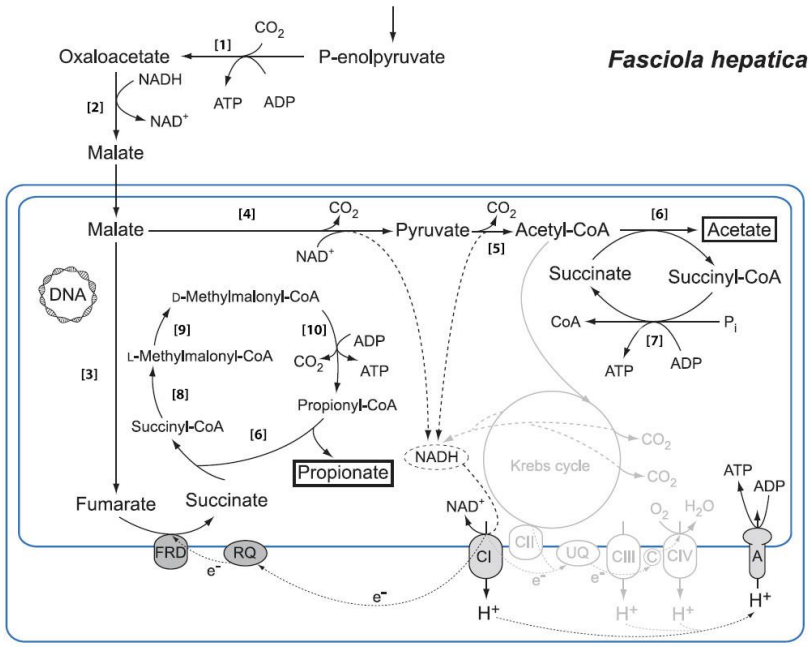


Figure 10 - Major pathways of energy metabolism in anaerobic mitochondria of *Fasciola hepatica.* The main end products of the adult parasites are acetate and propionate, with minor amounts of lactate and succinate. Aerobic mitochondrial metabolism occurs in free-living and juvenile parasitic stages of *F. hepatica,* leading to the reduction of oxygen and water production, Aerobic respiration is shaded in grey and anaerobic respiration and fermentation pathways are shaded in black.

Abbreviations: CI to CIV, respiratory complexes I to IV; UQ, ubiquinone; RQ, rhodoquinone; C, cytochrome *c*; A, ATPase; FRD, fumarate reductase; (6), acetate:succinate CoA-transferase (subfamily 1B); (7), succinyl-CoA synthetase; (4), malic enzyme; (5), pyruvate dehydrogenase complex; (3), fumarase; (1), phosphoenolpyruvate carboxykinase (ATP dependent); (2), malate dehydrogenase; (8), methylmalonyl-CoA mutase; (9), methylmalonyl-CoA epimerase; (10), propionyl-CoA carboxylase. (Müller et al [14])

Class III: Hydrogen-producing (H_2)

Hydrogen-producing mitochondria do not contain a complete electron transport chain; they typically have a truncated version containing complex I and II. They also contain an iron-only hydrogenase [56]. Although they have complexes of the electron transport chain, they do not use oxygen as a terminal acceptor, but instead use endogenously produced fumurate. Class III mitochondria also have conservation of Fe/S cluster protein biogenesis and protein import, and contain a mitochondrial genome encoding for subunits of complex I and II of the electron transport chain. *Nyctotherus ovalis* that live in the hindgut of cockroaches and the

parasite *Blastocystis sp.* that thrives in the gastrointestinal tract of various species, both contain class III mitochondria, encompassing a mitochondrial genome and complex I and II of the electron transport chain. *Nyctotherus ovalis* is commonly referred to as the 'missing link' in the evolution of mitochondria and hydrogenosomes, suggesting it as an intermediate between the two [2]. Although *Blastocystis sp.* contain hydrogenase, no hydrogen production has been demonstrated [50, 51].

H_2-producing mitochondria use pyruvate to produce acetate and succinate end products, where succinate is part of the TCA cycle; however, pyruvate undergoes the TCA cycle in a reductive direction. It is converted to malate via decarboxylation which enters the hydrogen-producing mitochondria via the conserved protein import complexes found throughout all mitochondrial classes, where it then undergoes part of the TCA cycle: Malate → Fumarate → Succinate.

The end product acetate is also produced from pyruvate within its matrix where pyruvate is converted to acetyl-CoA via pyruvate dehydrogenase which is then consequently converted to acetate by acetate:succinate CoA-transferase. Acetate:succinate CoA-transferase transfers acetyl-CoA to succinate yielding acetate and succinyl-CoA, where via succinyl-CoA synthetase, substrate level phosphorylated ATP can be produced.

Metabolism via NADH oxidation is carried out by members of the electron transport chain where complex I transport electrons from NADH via rhodoquinone (RQ) to complex II, which then uses fumurate as an electron acceptor producing succinate (Figure 11). NADH is also potentially reduced by the iron hydrogenase as it contains two proteins; 51-kDa and 24-kDa derived from complex I of the electron

transport chain that form a protein complex that are C-terminally fused to the iron hydrogenase catalytic subunit. This forms a trimeric structure making it possible to release molecular hydrogen (step 7 in Figure 11), which produces a proton gradient. Hydrogenosomes however, lack an ATP synthase, so it is unlikely this gradient is used for the synthesis of ATP [2, 41, 52, 53]. The concentration of NADH can vary dependent on the metabolic state within the cell, and the iron hydrogenase could fulfil this assignment as it allows direct re-oxidation of NADH maintaining metabolic homeostasis under anaerobic conditions.

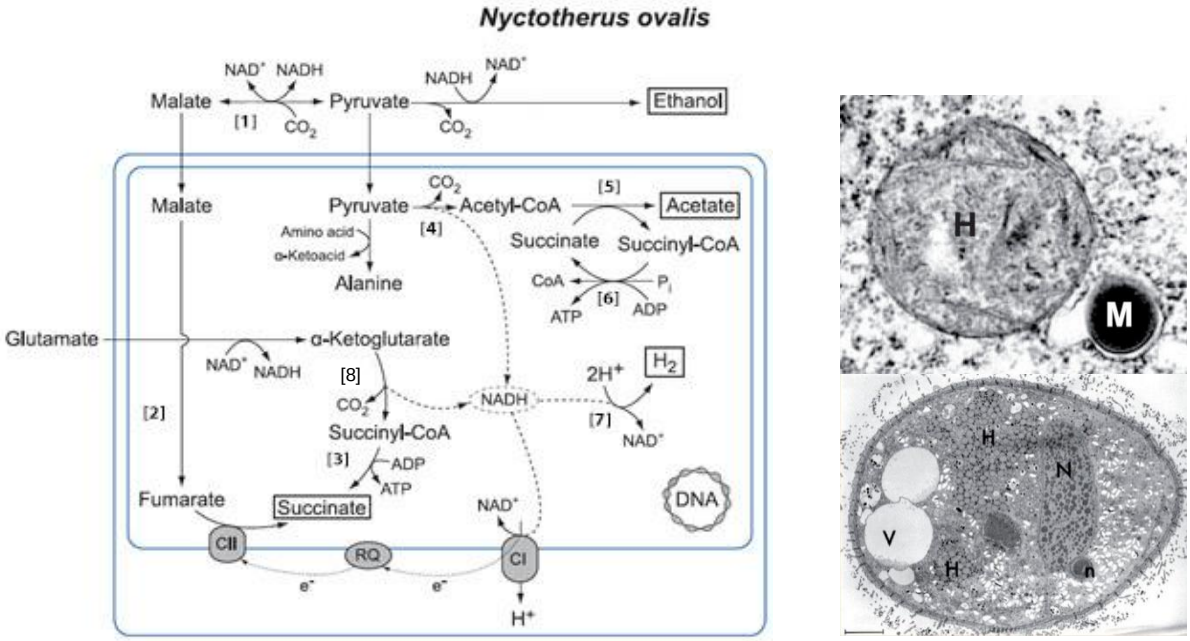


Figure 11 - Major pathways of energy metabolism in hydrogen-producing mitochondria of Nyctotherus ovalis. The TCA cycle is incomplete and is likely used in the reductive direction. ATP can be synthesized by substrate-level phosphorylation, producing acetate

Abbreviations: CI, respiratory complex I; RQ, rhodoquinone; CII, fumarate reductase/succinate dehydrogenase; (1), malic enzyme; (2), fumarase (predicted); (3), succinyl-CoA synthetase (4), pyruvate dehydrogenase complex; (5), acetate:succinate CoA-transferase subfamily 1A ; (6), succinyl-CoA synthetase; (7), hydrogenase; (8), α-ketoglutarate dehydrogenase (predicted).

The photograph (rightside) on the bottom shows *Nyctotherus ovalis*, with a length of ca. 80 um. The hydrogenosomes (H) are surrounded by endosymbiotic methane-producing archaebacteria (dark spots). N, macronucleus; n, micronucleus; V, vacuole. The top photograph shows a close-up view of a *Nyctotherus* hydrogenosome (H) and an associated methanogen (M). (Müller et al [14])

The hydrogenases like the one found in *Nyctotherus ovalis* are able to provide many advantages to an organism adapting to an anaerobic environment; until recently it was suggested that hydrogenases were not present in aerobic mitochondria, suggesting the evolution of these classes of mitochondria was via lateral gene transfer and fusion of functional domains [59]. However, recent studies in *Naegleria gruberi* have revealed non-mitochondrial (found in the cytosol) hydrogenases, allowing *Naegleria gruberi* to efficiently adapt to aerobic/anaerobic conditions [61].

Class IV: Hydrogenosomes

Heat shock chaperone proteins common to those in aerobic mitochondira which originated with the mitochondrial endosymbiosis have been preserved in hydrogenosomes, showing that hydrogenosomes have the same common ancestor as mitochondria. The hydrogenosomes within the sexually transmitted parasites *Trichomonas vaginalis* and the bovine parasite *Tritrichomonas foetus* are the best studied and more recently *Stygiella incarcerate* and *Mastigamoeba balamuthi* have been identified containing hydrogenosomes [55, 56]. Hydrogenosomes share common protein import pathways, conserved Fe/S cluster assembly mechanisms, NAD+ regeneration and contain the ATP producing enzyme succinyl-CoA synthetase. However, they lack a membrane associated electron transport chain, but rather decarboxylate pyruvate to H_2, CO_2 and acetate via hydrogenase, pyruvate:ferredoxin oxidoreductase (PFO) and acetate:succinate CoA-transferase respectively. Acetate:succinate CoA-transferase transfers acetyl-

CoA to succinate yielding acetate and succinyl-CoA, where via succinyl-CoA synthetase, like in class III H_2-producing mitochondria, substrate level ATP can be produced (step 3 in Figure 12b), which is exported into the cytosol by a mitochondrial carrier family (MCF) [57, 58]. It is vital to highlight that although hydrogenosomes decarboxylate pyruvate, they do this via pyruvate:ferredoxin oxidoreductase (PFO), which is not related to the pyruvate dehydrogenase complex found in mitochondria.

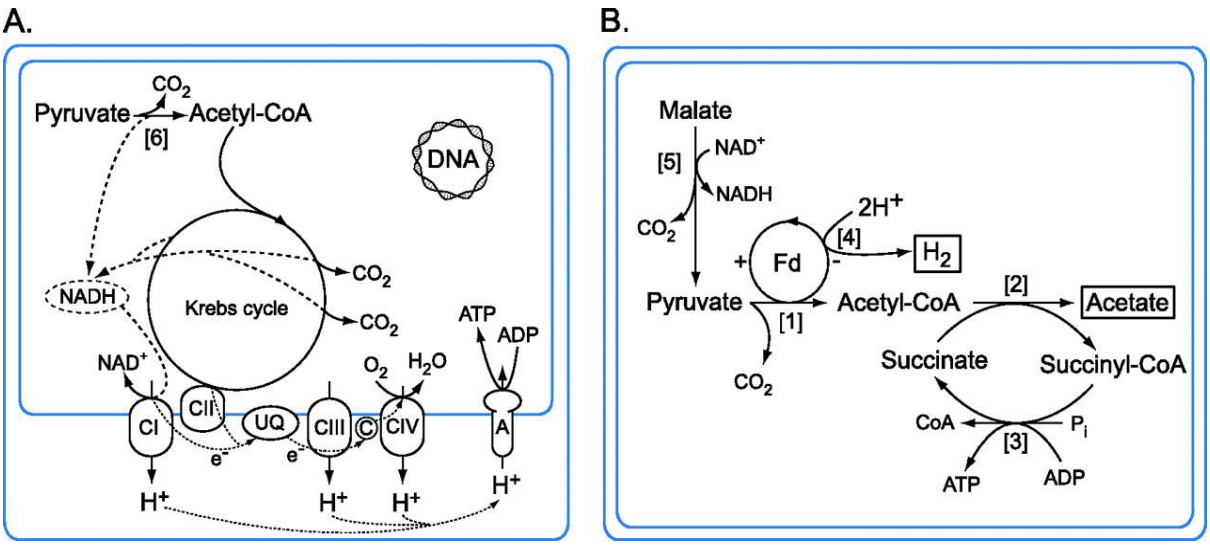


Figure 12 - A basic metabolic comparison between aerobic mitochondria - class 1 and hydrogenosomes - class 4. (A) Metabolic pathway of the pyruvate oxidation and oxidation phosphorylation in standard oxygen respiring mitochondria. (B) Metabolic pathway of fermentative pyruvate oxidation in Trichomonad hydrogenosomes. End products have been boxed (Acetate and H_2).

Abbreviations: CI to CIV, respiratory complexes I to IV; UQ, ubiquinone; C, cytochrome *c*; A, ATPase; Fd, ferredoxin; (1), pyruvate:ferredoxin oxidoreductase; (2), acetate:succinate CoA-transferase; (3), succinyl-CoA synthetase; (4), hydrogenase; (5), malic enzyme; (6), pyruvate dehydrogenase complex. (Müller et al [14])

Hydrogenosomes lack a genome, so all proteins targeted to the organelle are encoded via the nuclear genome and therefore must be transported through the double membrane. This is carried out by the conserved import protein complexes TOM and SAM present on the outer membrane as well as TIM and PAM complexes found on the inner membrane, as found in typical mitochondria.

The hydrogenosomes of *Trichomonas vaginalis* contain the two proteins 51-kDa and 24-kDa derived from complex I of the electron transport chain, that form a protein complex which help maintain redox balance by the reoxidisation of NADH from the malic enzyme reaction (step 6 in Figure 13) reducing [2Fe-2S]-ferredoxin using the NADH and ferredoxin as likely electron donors [60]. These complex I proteins form the β and γ subunits of the trimeric iron hydrogenase.

This links to the potential reduction or variation from class III type mitochondria to class IV, where the class III *Nyctotherus ovalis* contains mitochondrial complex I and II dependent respiratory chain activity, a genome and a chimeric iron hydrogenase.

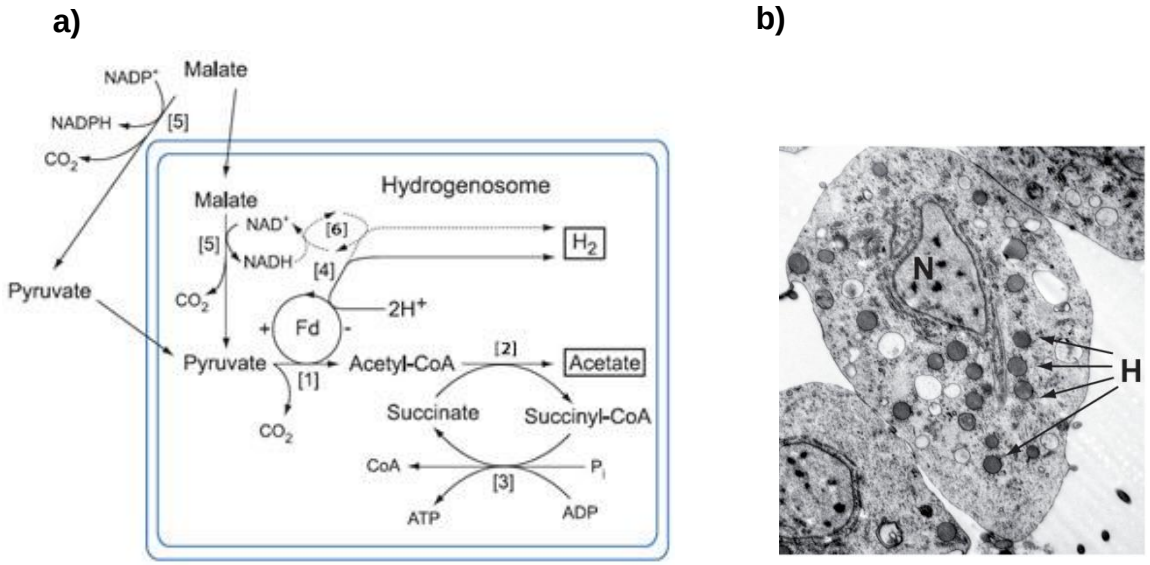


Figure 13 - a) Major pathways of hydrogenosome within the parasite *Trichomonas vaginalis*. Hydrogenosomal pyruvate breakdown involves pyruvate:ferredoxin oxidoreductase and functional 51-kDa and 24-kDa subunits of the NADH dehydrogenase module in complex I, which reoxidize NADH and associates with iron hydrogenase in *Trichomonas.*

Abbreviations: (4), hydrogenase; (5), malic enzyme; (6), 51-kDa and 24-kDa subunits of the NADH dehydrogenase module of complex I. (Müller et al [14])
b) The transmission electron micrograph shows hydrogenosomes (H) and the nucleus (N). (Photograph from Müller [14])

Class V: Mitosomes

Mitosomes have undergone evolutionary reduction in perspective of both physical size and their biochemical complexity and have been found in microbial parasites including *Entamoeba, Giardia, Cryptosporidium* and the Microsporidia*.* Characteristically they appear to have no role in ATP synthesis, as mitosomes appear to be found in eukaryotes that synthesize ATP in their cytosol (Figure 14), nor do they produce hydrogen [3, 16]. They also lack a genome and electron transport chain on their inner membranes.

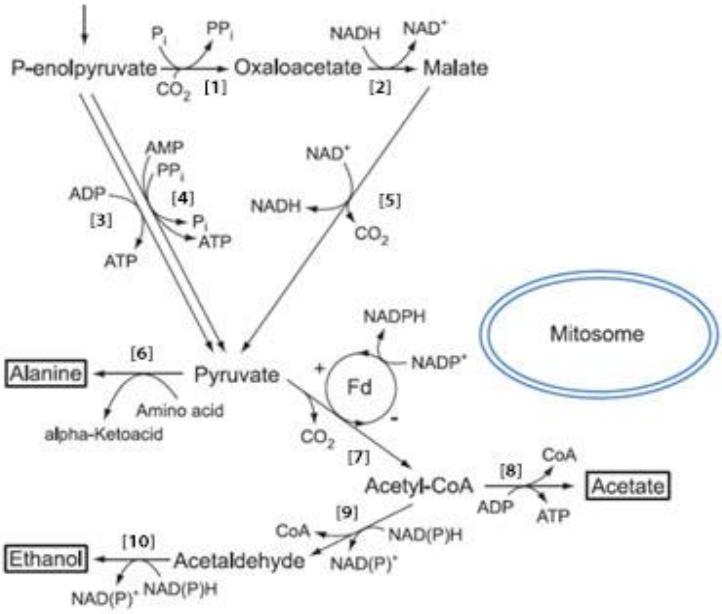


Figure 14 - Shows major metabolic pathways for parasite *Entamoeba histolytica*; its energy metabolic pathways are localized in the cytosol, which also harbours mitosomes.

Abbreviations: Fd, ferredoxin; (1), phosphoenolpyruvate carboxytransferase (PPi dependent); (2), malate dehydrogenase; (3), pyruvate kinase; (4), pyruvate:orthophosphate dikinase; (5), malic enzyme; (6), alanine aminotransferase; (7), pyruvate:ferredoxin oxidoreductase ;(8), acetyl-CoA synthetase (ADP forming ; (9/10), alcohol dehydrogenase E. (Müller et al [14])

Most mitosomes are now thought to import proteins into their matrix similar to that of mitochondria and also contain mitochondrial proteins of Fe/S cluster assembly. The Fe/S cluster assembly is thought to be a common function for mitochondrial

homologues and studies have revealed the presence of all key components of the Fe/S cluster assembly machinery, including cysteine desulfurase, scaffold proteins (IscU), ferredoxin which provides reducing agents for the formation of Fe/S clusters and a set of chaperones which are involved in the transfer of Fe/S clusters to apoproteins. This pathway is the only conserved biosynthetic pathway found in these organelles [16, 59].

Entamoeba histolytica may represent an exception to the rule however, as Fe/S-cluster biogenesis appears to localise outside of its mitosome, so gives further questions to the function of the mitosome [60, 61]. The identification of enzymes that activate sulfate localised to the mitosome suggests other possible functions [62, 63].

The protein import process is also conserved; mitosomes do not contain a mitochondrial genome, so any function they possess will require the ability to import nuclear encoded proteins into their matrix. The TOM and SAM complexes of the outer membrane and TIM complexes of the inner membrane have been identified, however TIM proteins have not been identified in *Giardia*, but PAM16, 18 and Hsp70 proteins that interact with TIM23 complex are present. Small TIMs can also be seen, though they appear to be more 'simple', as seen in *Cryptosporidium* where they possess a single small TIM protein rather than a complex [71].

Although mitosome function still remains somewhat mysterious, it is clear that their involvement in core energy metabolism is marginal at best, if even any involvement at all. However, recent findings of alternative oxidase (AOX) genes within two phylogenetic clades of microsporidia suggest that mitosomes may retain an alternative respiratory pathway [72].

Ciliates

-What are they? Where are they found?

Ciliates are single celled eukaryotes that were first observed microscopically by van Leeuwenhoek in 1674. Ciliates have been separated from other protists for a vast amount of time and are among the top five groups of protists in terms of species numbers; there are, upwards of 8,000 different ciliate species [73].

Ciliates are a distinct group of protists distinguished by three major features, 1) the presence of cilia that are derived from kinetosomes which are used for movement and to help direct food to the cytostome (cell mouth) to undergo phagocytosis. The high number of cilia led to the name 'Ciliates'. 2) The containment of two kinds of nuclei – a macronucleus that controls physiological and biochemical functions and a micronucleus that is a germ-line reserve. 3) The process of conjugation, a sexual process where partners fuse to exchange genetic material [74].

Ciliates are important components of the 'microbial loop' as they are often responsible for the consumption of bacteria and primary production (organic compounds) in certain habitats or even be consumed by many aquatic animals [74]. They are diverse in shape and size and may have evolved over two billion years ago based on the rate of small subunit rRNA evolution [75].

Ciliates have a cytoskeleton that is an elaborate complex compiled from microtubular and microfiliments [74]. They are found in diverse environments and are mainly free-living, but they can also be found parasitising other organisms. Free-living ciliates are found in any sort of habitat that contains water, be it in soil, hot springs or the freezing temperatures of the Antarctic sea ice and feed on bacteria, algae or even other ciliates (e.g.*Didinium*). They are the top predators in

microbial food webs and were likely the major predatory group before the evolution of animals [74]. Symbiotic species live as commensals in sea urchins as well as parasites of fish and lobsters like *Anophryoides sp.* [76]. Some ciliates like *Entodinium caudatum* are present in the rumen of many grazing animals where they stabilise the large populations of bacteria that serve to break down the cellulose of the animals' diet [67, 70–72].

-Reproductive Survival

Ciliates reproduce asexually by division, where the micronucleus undergoes mitosis, while macronuclei divide in two ways; either by splitting apart using the extramacronuclear microtubules or by dividing using the intramacronuclear microtubules [80]. However, ciliates can also reproduce sexually through a process known as conjugation. This is when two ciliates of opposite mating types come together and form a cytoplasmic bridge between them. The macronuclei disintegrate and the micronuclei divide by meiosis, these conjugating cells then exchange haploid micronuclei over the formed cytoplasmic bridge. After exchanging, they separate with both containing their newly exchanged genetic material and form a new macronuclei from their micronuclei and divide to give rise to progeny with the new combination of genes. This process is crucial for the survival of ciliate lineages [67, 73].

The ciliate *Tetrahymena thermophila* actually expresses seven different mating types. [81] After mating, each new exconjugent has a recombinant micronucleus from the two parental *Tetrahymena* species. The new exconjugent does not necessarily express one of the two parental mating types, but however they can randomly choose up to seven mating types [74].

-Classification Identification

Ciliophora is the phylum, which is commonly referred to as ciliates. This phylum can be divided into two sub phyla Postciliodesmatophora and Intramacronucleata, and 11 classes based on features of nuclear division and fibrillar pattern in their somatic kinetids [74]. The defining characteristic between the two sub phyla is, Intramacronucleata species have the ability to divide its macronucleus, whereas Postciliodesmatophora classified species do not [74]. Phylogenetic relationships among the ciliates can then be further understood via their basal body framework. The cell cortex of a ciliate is supported by a very complex framework formed from kinetosomes, microtubules and microfilaments. The kinetosomes form the central unit of an organellar structure which is known as the kinetid. The cell cortex and the patterns of the fibres and microtubules within the kinetid provide the primary characteristic used to distinguish ciliate taxa [74].

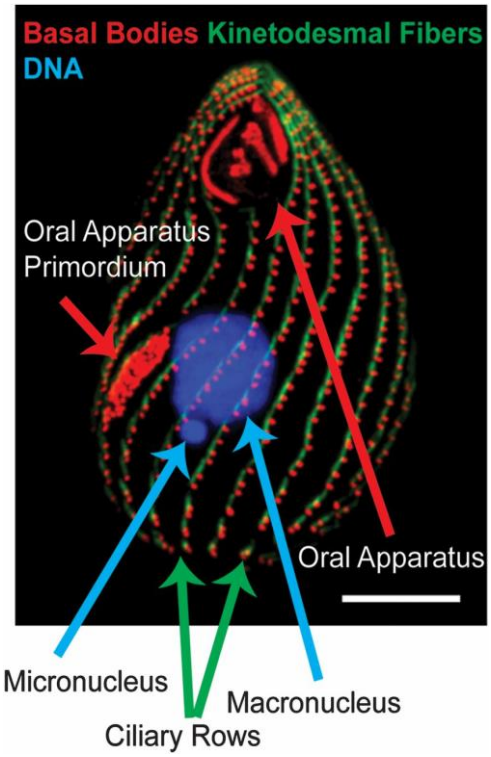


Figure 15 - *Tetrahymena* cell image from Marisa D. Ruehle 2016, illustrating its crystal-like organization of ciliary units. A single *Tetrahymena* cell labeled for basal bodies (a-centrin, red), kinetodesmal fibers (5D8, green), and DNA (Hoerscht-33342, blue). Bar, 10 mm.

Hypothesis

With knowledge of the diverse mitochondrial classes already outlined and how each of them functions differently, the aim of this project is to assign the mitochondrial class to 27 different ciliates gathered from diverse environments and to assess the diversity of mitochondrial classes present in the ciliates. This will be completed using computational methods, analysing newly available transcriptomic and genomic data to identify mitochondrial proteins.

Methods

-Flow chart overview of method

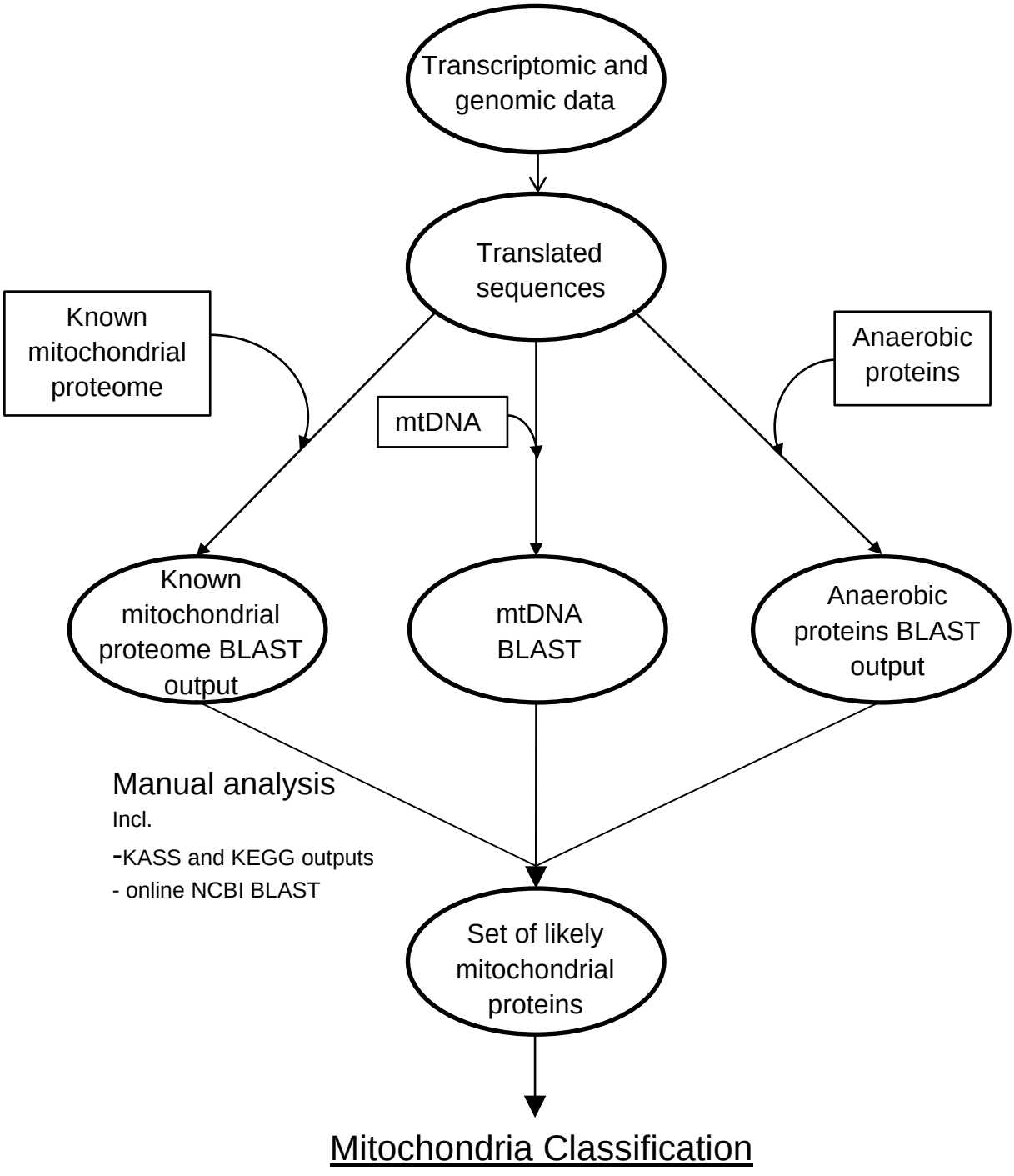


Figure 16 - The flow chart shows a basic overview of the steps taken to carry out our analysis to identify mitochondrial diversity from newly generated ciliate transcriptomes along with four known ciliate genomes. The original data was translated into protein sequences which would then be compared against a formed database to produce outputs of mitochondrial proteins, mtDNA encoded proteins and anaerobically functioning proteins. With these outputs, manual analysis was then possible and mitochondria classification could be achieved. Identifying and classifying the type of mitochondria present in each ciliate based on their content of mitochondrial proteins present in the transcriptomic and genomic data.

-Mitochondrial database and query ciliate data preparation

Nine different known mitochondrial proteomes (*Acanthamoeba castellanii, Arabidopsis thaliana, Entamoeba histolytica, Giardia intestinalis*, *Human*, *Mouse, Tetrahymena thermophila, Trichomonas vaginalis, Trypanosoma brucei)* were collected from various sources (MitoMiner, NCBI and the publications - see Table 1), which included ciliate species as well as the human mitochondrial proteome. These proteomes were combined into a local BLAST database [82] where it was used as a reference for identifying mitochondrial proteins present in the query ciliate transcriptomic and genomic data.

11 newly generated ciliate transcriptomic data and 12 other ciliate transcriptomic data was received from Dr. Eleni Gentekaki at the Mae Fah Luang University and the late Prof. Denis H. Lynn from the University of Guelph. These generated sequences were in mRNA fasta format and were initially checked for contamination against a local NCBI 'nt' database, removing any sequences with a percentage identity exceeding 95% match to probable contamination sources e.g. bacterial food source. After the removal of contamination, the sequences were translated from the mRNA triplet codons into the corresponding amino acid protein sequences, considering all six reading frames. To check if contamination filtering was successful, Phylogenetic trees were created (Figure 23 and 24). Predicted iron hydrogenase and PFO/PNO sequences from the ciliates were aligned to existing iron hydrogenase and PFO/PNO multiple sequence alignment using MAFFT [83]. IQ-TREE was then used to create phylogenetic trees, using a bootstrap value of 1000 and the substitution model LG [84]. The multiple sequence alignments used were kindly provided by Michelle M Leger from Dalhousie University [62].

Genetic code six (ciliate code) was used along with the species specific non-standard codon variations of *Euplotes sp., Mesodinium sp. and Condylostoma magnum* for the translation of the query ciliate sequences. Each newly translated sequence was compared against the newly formed local mitochondrial BLAST database producing sequences with similarity to mitochondrial proteins in the database within a cut off e-value of 1e-05. Fragments shorter than 10 amino acids were discarded.

Recent observations show that *Condylostoma magnum* is able to code the three standard stop codons as glutamine or tryptophan [69, 70]. Therefore this newly discovered genetic code (code 28) was used, similarly with *Mesodinium sp.* coding tyrosine rather than a stop codon [86]. Code 6 has similar codon translations with two stop codons being coded as Glutamine, so it was used as the genetic code for the translation of the query ciliates for the mitochondrial protein identification analysis.

There are four ciliates (*Ichthyophthirius multifiliis, Nyctotherus ovalis, Paramecium tetraurelia* and *Tetrahymena thermophila*) with genomic data that were also analysed alongside the 23 transcriptomic data also provided by Dr. Eleni Gentekaki and Prof. Denis H. Lynn (see table 4). However, prior to undergoing the same methods previously discussed, the open reading frames (ORF) were identified using online application SMS ORF Finder [87].

-Ciliate mitochondrial data analysis

TargetP [88] and TMHMM [89] were incorporated into the pipeline providing further information to aid in analysis. TargetP was used to predict if the protein is targeted to the mitochondria by locating any N-terminal presequences. TMHMM was used to predict any transmembrane helices in the proteins.

On completion of the translated ciliate sequences being searched against the local mitochondrial BLAST database, a table output was produced containing each predicted protein and the outputs for TargetP and TMHMM. Each entry was then further analysed allowing additional protein analysis to be carried out, which included running further searches using online NCBI BLASTp nr database [82] if required to confirm the predictions of certain mitochondrial proteins are correct e.g. iron hydrogenase.

KAAS *(KEGG Automatic Annotation Server)* [90] and KEGG *(Kyoto Encylopedia of Genes and Genomes)* [91] was used to aid protein function prediction and provides figures and pathways to visualise what proteins are present.

-Anaerobic functioning and mtDNA encoded protein analysis

The identification of proteins that function under anaerobic conditions *iron hydrogenase and pyruvate:NADP+ oxidoreductase (PNO), pyruvate:ferredoxin oxidoreductase (PFO))* and rhodoquinone (RQ) required further analysis. To perform this task a new database was created containing the translated ciliate sequences as reference; the query anaerobic proteins are then searched against the newly created local ciliate BLAST database to identify anaerobic functioning proteins within the ciliates. For these searches the e-value cut off was set at 1e-05. The identified sequences were then further analysed via the online NCBI BLASTp nr database [82] to support the identification, allowing anaerobic proteins to be confirmed in the query ciliate.

mtDNA encoded proteins were identified using the same method as anaerobic functioning proteins. After gathering the mitochondrial genomes for *Tetrahymena thermophila, Paramecium tetraurelia and Nyctotherus ovalis* from Uniprot [92], they were searched against the local ciliate BLAST database. These three genomes were selected due to them containing the minimum complete set of mitochondrial-genome encoded proteins that could be used for identifying mtDNA-encoded proteins in other ciliates.

Prior to searching the mitochondrial genomes against the ciliate BLAST database, the open reading frame were read using SMS ORF Finder [93] to search for the 'part' of the protein that has the potential to be translated, to be searched against the database.

-Mitochondria classification criteria for each ciliate

Classification of mitochondria type followed the following criteria:

- Five or more TCA enzymes identified
- Subunits of two or more ETC complexes.
- The presence of anaerobic functioning proteins.

The combination of these criteria leads to the mitochondria classification prediction.

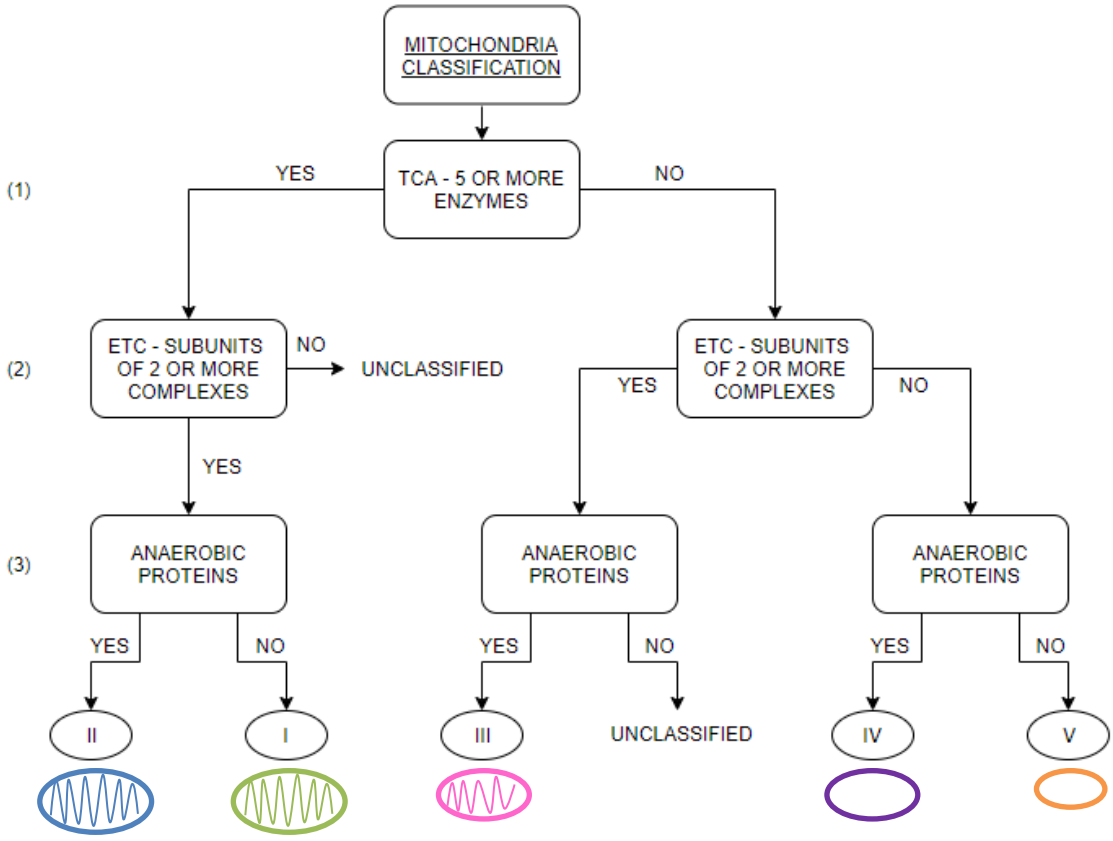


Figure 17 - Basic overview flowchart showing the criteria used to categorise the class of mitochondria predicted in each ciliate. Each step asks the following:

1. Does the ciliate contain five or more enzyme steps of the TCA?
2. Does the ciliate contain subunits from two or more ETC complexes?
3. Does the ciliate contain anaerobic proteins?

Mitochondrial proteome BLAST database data source table

Table 1 - sources for mitochondrial proteome BLAST database.

Species	Source
Acanthamoeba castellanii	Gawryluk RMR, Chisholm KA, Pinto DM, Gray MW. Data showing the compositional complexity of the mitochondrial proteome of a unicellular eukaryote (*Acanthamoeba castellanii*, supergroup Amoebozoa). Data Br. 2014;1:12–4.
Arabidopsis thaliana	Kruft V, Eubel H, Jänsch L, Werhahn W, Braun HP. Proteomic approach to identify novel mitochondrial proteins in *Arabidopsis*. Plant Physiol. 2001 Dec;127(4):1694–710. Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH. Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. Plant Cell. 2004 Jan;16(1):241–56.
Entamoeba histolytica.	Mi-ichi F, Abu Yousuf M, Nakada-Tsukui K, Nozaki T. Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway. Proc Natl Acad Sci U S A. 2009 Dec 22;106(51):21731–6.
Giardia intestinalis.	Jedelský PL, Doležal P, Rada P, Pyrih J, Smíd O, Hrdý I, et al. The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. PLoS One. Public Library of Science; 2011;6(2):e17285.
Human	Smith AC, Robinson AJ. MitoMiner v3.1, an update on the mitochondrial proteomics database. Nucleic Acids Res. 2016 Jan 4;44(D1):D1258-61.
Mouse	Smith AC, Robinson AJ. MitoMiner v3.1, an update on the mitochondrial proteomics database. Nucleic Acids Res. 2016 Jan 4;44(D1):D1258-61.
Tetrahymena thermophila	Smith DGS, Gawryluk RMR, Spencer DF, Pearlman RE, Siu KWM, Gray MW. Exploring the Mitochondrial Proteome of the Ciliate Protozoon *Tetrahymena thermophila*: Direct Analysis by Tandem Mass Spectrometry. J Mol Biol. 2007 Nov;374(3):837–63.
Trichomonas vaginalis	Schneider RE, Brown MT, Shiflett AM, Dyall SD, Hayes RD, Xie Y, et al. The *Trichomonas vaginalis* hydrogenosome proteome is highly reduced relative to mitochondria, yet complex compared with mitosomes. Int J Parasitol. 2011;41(13):1421–34.
Trypanosome brucei	Panigrahi AK, Ogata Y, Zíková A, Anupama A, Dalley RA, Acestor N, et al. A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. Proteomics. 2009 Jan;9(2):434–50. Acestor N, Panigrahi AK, Ogata Y, Anupama A, Stuart KD. Protein composition of *Trypanosoma brucei* mitochondrial membranes. Proteomics. 2009 Dec;9(24):5497–508.

Results

Mitochondria are classified dependent on their protein composition and the pathways in which they encompass [14]. Class I and II contain a TCA and electron transport chain as well as β-oxidation, Fe/S cluster and protein import. Class III and IV contain reduced or no proteins that function within the TCA or electron transport chain pathways, while retaining Fe/S and protein import function. Class V is the most reduced form and contains no TCA or electron transport chain pathways.

27 different ciliates from diverse environments were investigated to predict the class of mitochondria present within each organism from their protein composition. The study shows that between them, the investigated ciliates contain all five mitochondrial classes, and potentially some further not yet defined intermediates, due to presence of certain anaerobic proteins and complexes of the electron transport chain. An overview chart of their class and protein content can be seen in Tables 2 and 3, where the number of identified proteins and their location are shown.

Before the ciliates could be investigated, contamination was removed from all the transcriptomic data, including the new genomic data of *Metopus sp.*, removing the majority of bacterial source and/or their known food source which lowered the number of contigs for each ciliate, for example with *Metopus sp.* where around 400,000 individual contigs were removed due to contamination. The final number of contigs for each ciliate varied, with *Metopus sp.* having the most at 92,542 individual contigs and having mitochondrial protein identification percentage of 7%; and the smallest being *Nyctotherus ovalis* at 241 contigs with a mitochondrial

protein identification percentage of 54%. The number of mitochondrial functioning proteins identified for all the ciliates from the number of contigs varied from 7% up to 73%, with the majority being around 10-20% (Table 5).

The environment of each ciliate was also considered to see if the type of environment the ciliate exists within correlates with the class of mitochondria it has (Table 4). No association was present apart from ciliates found within the rumen or hindgut as they all appear to contain mitochondrial-related organelles, with *Nyctotherus ovalis, Polyplastron multivesiculatum* and *Entodinium caudatum* containing class III, IV and V respectively from our methods of class prediction.

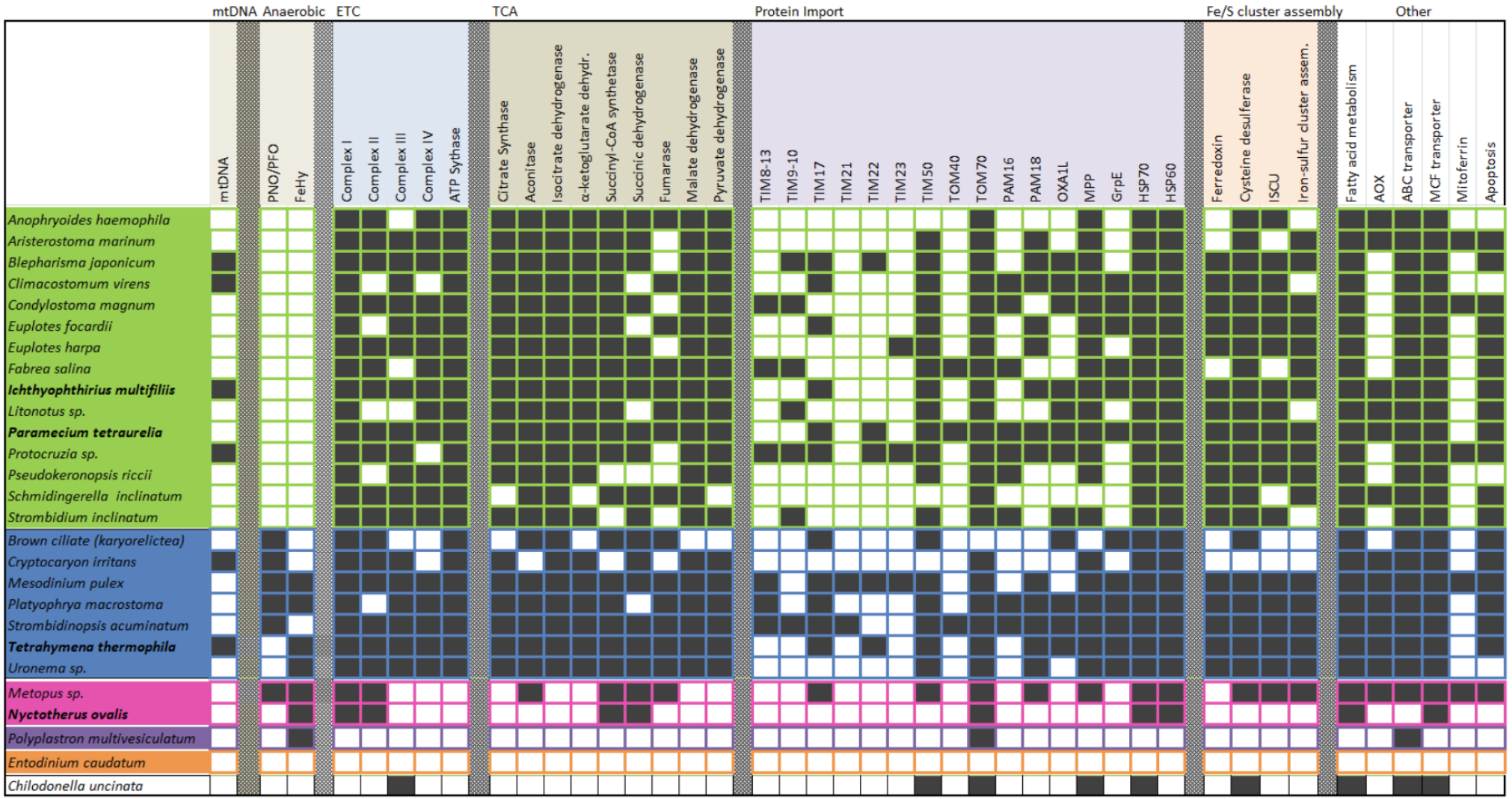
Classifying the type of mitochondria present in each ciliate followed a main method of prediction. This method was identifying proteins that function within the TCA cycle and electron transport chain pathways as well as any anaerobically functioning proteins. Other pathways like β-oxidation, Fe/S cluster assembly and protein import were investigated to see what proteins were present in each ciliate and whether any presented itself as interesting or different to expectation.

Ciliate	Contigs	#Hits		TCA KEGG	ETC	Fe/S cluster assembly proteins	Fatty Acid beta oxidation	Import	AOX	Apop	Transporters ABC	Transporters MCF		mtDNA		Anaerobic Hits PNO/ PFO	Anaerobic Hits FeHy
Anophryoides haemophila	7689	5637		12	8	21	10	18	5	-	4	138		-		-	-
Aristerostoma marinum	25813	3194		9	11	7	10	34	2	2	13	32		-		-	-
Blepharisma japonicum	18751	2578		9	13	10	11	33	-	3	9	21		1		-	-
Climacostomum virens	15019	1981		9	7	8	7	19	-	-	9	13		1		-	-
Condylostoma magnum	26217	4691		8	14	25	11	51	-	5	10	55		-		-	-
Euplotes focardii	25584	2152		8	16	7	9	15	-	2	7	73		-		-	-
Euplotes harpa	25118	2546		9	13	6	10	16	-	2	12	51		-		-	-
Fabrea salina	10034	2135		11	9	5	12	31	-	1	10	27		-		-	-
Ichthyophthirius multifiliis	8097	3970		11	22	13	11	52	2	3	11	57		1		-	-
Litonotus sp.	22501	1941		9	7	5	7	19	1	2	8	32		-		-	-
Paramecium tetraurelia	81474	25664		11	28	37	16	433	6	10	12	72		1		-	-
Protocruzia sp.	49352	4999		8	13	10	10	101	-	1	16	99		1		-	-
Pseudokeronopsis riccii	24119	2487		7	8	6	13	29	-	-	8	20		-		-	-
Schmidingerella inclinatum	19527	2089		7	14	4	12	17	2	5	6	31		-		-	-
Strombidium inclinatum	23215	2211		7	10	7	8	24	-	4	11	30		-		-	-
Brown Ciliate	11646	1103		8	6	8	11	6	-	1	4	2		-		1	-
Cryptocaryon irritans	2658	797		7	10	1	7	10	1	2	3	7		3		2	-
Mesodinium pulex	69163	6668		13	31	21	23	57	2	8	12	61		-		1	1
Platyophrya macrostoma	44091	6606		10	17	16	23	82	7	8	12	44		-		2	1
Strombidinopsis acuminatum	51161	6204		13	33	18	26	66	2	10	17	70		-		2	-
Tetrahymena thermophila	36993	11587		11	27	13	12	138	2	9	28	136		1		-	1
Uronema sp.	13437	1735		12	6	7	-	27	1	-	10	-		-		-	1
Metopus sp.	92542	6591		6	9	19	16	77	1	3	18	30		-		3	4
Nyctotherus ovalis	241	131		4	12	1	1	1	-	-	-	6		-		-	6
Polyplastron multivesiculatum	624	232		-	-	-	-	1	-	-	1	-		-		-	1
Entodinium caudatum	1061	288		-	-	-	-	-	-	-	-	-		-		-	-
Chilodonella uncinata	9553	770		-	2	2	5	5	-	-	4	6		-		-	-

Table 2 - each ciliate is separated into a predicted class where each column shows number of identified proteins.

Class I, **Class II**, **Class III**, **Class IV**, **Class V**, (white = no classification).

Note: Contigs – number of sequences in original data. #Hits – number of proteins identified in mitochondrial BLAST database. TCA (KEGG) - Number of identified steps in cycle (not number of proteins identified). ETC - Number of different subunits identified. Fe/S – number of protein hits from BLAST search, however if hsp70 is present it is counted as 1. Fatty Acid β-oxidation – number of identified β-oxidation proteins via KEGG. Import – number of protein hits from blast search, does not contain hsp70 or hsp60. Apop – number of apoptotic proteins identified. ABC – number identified in KEGG. MCF – number of protein hits from blast search. mtDNA – number of proteins identified as encoded by mitochondrial genome. Anaerobic hits – number of anaerobic functioning proteins identified (FeHy = iron Hydrogenase). Ciliates shown in **BOLD** represent genomic data.



Subunits hits identified
No identification

Class 1
Class 2
Class 3
Class 4
Class 5
Unclassified

Table 3 - Each ciliate separated in relation to predicted mitochondrial class, showing specific proteins present within each. Each dark coloured in square shows an identified protein.

Class I, **Class II**, **Class III**, **Class IV**, **Class V**

Ciliates shown in **BOLD** represent genomic data.

Ciliate	Clade	Environment	Data Type Collected	Extraction Conditions
Anophryoides haemophila	Oligohymenophorea	Lobster Parasite	Transcriptomic	-
Aristerostoma marinum	Colpodea	Marine brackish plankton	Transcriptomic (ESTs)	TRIzol® Reagent [3]
Blepharisma japonicum	Heterotrichea	Stagnant pond water, or soil zooplankton	Transcriptomic	-
Climacostomum virens	Heterotrichea	Zooplankton Brackish water/fresh	Transcriptomic	-
Condylostoma magnum	Heterotrichea	Marine ciliate brackish zooplankton	Transcriptomic (ESTs)	Mass cultures [1]
Euplotes focardii	Spirotrichea	Sand sediments Antarctica Psychrophilic	Transcriptomic (ESTs)	TRIzol® Reagent [3]
Euplotes harpa	Spirotrichea	Marine zooplankton brackish	Transcriptomic (ESTs)	Mass cultures [1]
Fabrea salina	Heterotrichea	Marine brackish	Transcriptomic	-
Ichthyophthirius multifiliis	Oligohymenophorea	Fish Parasite / mosquito larvae (Ich)	Genomic	-
Litonotus sp.	Litostomatae	Marine, Brackish, Predatory	Transcriptomic (ESTs)	Mass cultures [1]
Paramecium tetraurelia	Oligohymenophorea	Freshwater	Genomic	-
Protocruzia sp.	Incertae sedis	Marine, Brackish, sp. dependent	Transcriptomic (ESTs)	TRIzol® Reagent [3]
Pseudokeronopsis riccii	Spirotrichea	Marine, Brackish, sp. dependent	Transcriptomic (ESTs)	Mass cultures [1]
Schmidingerella inclinatum	Spirotrichea	Marine zooplankton Predatory ciliate	Transcriptomic (ESTs)	TRIzol® Reagent [3]
Strombidium inclinatum	Spirotrichea	Marine Brackish zooplankton	Transcriptomic (ESTs)	Mass cultures [1]
Brown Ciliate	*Karyorelictea*	Marine	Transcriptomic	-
Cryptocaryon irritans	Prostomatea	Fish Parasite - white spot disease	Transcriptomic	-
Mesodinium pulex	Litostomatae	Marine Brackish	Transcriptomic	-
Platyophrya macrostoma	Colpodea	Fresh water pond/lakes marine Plankton	Transcriptomic (ESTs)	RNeasy® Mini Kit [2]
Strombidinopsis acuminatum	Spirotrichea	Marine Brackish plankton	Transcriptomic (ESTs)	TRIzol® Reagent [3]
Tetrahymena thermophila	Oligohymenophorea	Freshwater	Genomic	-
Uronema sp.	Oligohymenophorea	Marine, Brackish sp. Dependent	Transcriptomic	-
Metopus sp.	Armophorea	Fresh water marine/soil/brackish water	Transcriptomic	-
Nyctotherus ovalis	Armophorea	Cockroach hindgut symbionts	Genomic	-
Polyplastron multivesiculatum	Litostomatae	Rumen symbionts	Transcriptomic	-
Entodinium caudatum	Litostomatae	Rumen	Transcriptomic	-
Chilodonella uncinata	Phyllopharyngea	Fish Parasite	Transcriptomic	-

Table 4 - Each ciliate is separated into their predicted class, the phylogenetic clade and the environment in which the ciliate exists. All ciliate data was provided by Dr. Eleni Gentekaki and the late Prof. Denis H. Lynn.
**Note: ' - ' shows where extraction conditions are not known.*

RNA extraction – *Procedure provided by Dr. Eleni Gentekaki*

1. *Condylostoma magnum*, *Litonotus* sp., *Euplotes harpa*, *Pseudokeronopsis riccii* and *Strombidium inclinatum*

   Mass cultures (~10^4-10^6 cells in ~0.5-4.0 L of medium) of each strain were filtered twice with a sieve of the appropriate pore size and then serially centrifuged 3 times (90-180x g, 10-12 min). The pellet was recovered and resuspended in sterile medium each time. After the final centrifugation, the pellet was resuspended in 50-60 mL of sterile medium + ampicillin (0.1 mg/mL). *S. inclinatum* cells were too fragile to be washed and concentrated through centrifugation; instead, the culture medium was poured on paper filters, and the ciliates recovered from the filter and brought to a sterile medium. Ciliates were checked at the stereomicroscope to evaluate their viability and the presence of eukaryotic contaminants (especially food organisms). After 6-20 hours in the medium with antibiotic the cells were centrifuged again several times (or recovered from a paper filter in the case of *S. inclinatum*) and the pellet washed with sterile medium. After one last centrifugation at 1000xg for 10 min, the supernatant was removed and total RNA was extracted from the cell pellet using the NucleoSpin® RNA II kit (Macherey-Nagel) according to manufacturer's specifications.

2. *Platyophrya macrostoma*

   RNA was isolated from 10^6 *P. macrostoma* cells using the RNeasy® Mini Kit (Qiagen) according to the manufacturer's instructions. Total RNA was treated with DNase (Qiagen) and concentration was determined using the Quant-iT™ RiboGreen® RNA Assay Kit (Life Technologies).

3. *Protocruzia adherens, Aristerostoma* sp.*, Schmidingerella inclinatum*, *Strombidinopsis acuminatum* and *Euplotes focardii*

   Total RNA was extracted with TRIzol® Reagent (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. For *Strombidinopsis acuminatum*, because the concentration was too low, the cells were concentrated on a filter using gentle suction before transferring them to Trizol in a Petri dish. After flushing the dish several times with a pipette, the solution was transferred to a 1.5 ml microfuge tube and the remaining steps were followed according to the manufacturer's specifications.

Ciliate	Contigs	#Hits	% of hits (hits/contigs)
Anophryoides haemophila	7689	5637	73%
Aristerostoma marinum	25813	3194	12%
Blepharisma japonicum	18751	2578	14%
Climacostomum virens	15019	1981	13%
Condylostoma magnum	26217	4691	18%
Euplotes focardii	25584	2152	8%
Euplotes harpa	25118	2546	10%
Fabrea salina	10034	2135	21%
Ichthyophthirius multifiliis	8097	3970	49%
Litonotus sp.	22501	1941	9%
Paramecium tetraurelia	81474	25664	31%
Protocruzia sp.	49352	4999	10%
Pseudokeronopsis riccii	24119	2487	10%
Schmidingerella inclinatum	19527	2089	11%
Strombidium inclinatum	23215	2211	10%
Brown Ciliate	11646	1103	9%
Cryptocaryon irritans	2658	797	30%
Mesodinium pulex	69163	6668	10%
Platyophrya macrostoma	44091	6606	15%
Strombidinopsis acuminatum	51161	6204	12%
Tetrahymena thermophila	36993	11587	31%
Uronema sp.	13437	1735	13%
Metopus sp.	92542	6591	7%
Nyctotherus ovalis	241	131	54%
Polyplastron multivesiculatum	624	232	37%
Entodinium caudatum	1061	288	27%
Chilodonella uncinata	9553	770	8%

Table 5 - A table showing the number of original contig files after contamination removal and number of identified protein hits after analysis. The last column shows the percentage of identified hits against original contig number. Each ciliate is separated into their predicted class as in previous tables.

The Citric Acid Cycle

The TCA cycle is a central metabolic pathway for aerobic processes. 24 of the 27 investigated ciliates contained TCA functioning proteins, however, the number of identified proteins of the TCA varied from four to 13 (shown in Tables 2 and 3) for the pathway.

The first step in classification was from the identification of TCA functioning proteins. Class I and II are defined by a high amount of TCA functioning proteins, where identification of five or more components of the TCA cycle allowed mitochondrial classification – separating them from the other three reduced classes that contain less TCA functioning proteins. The differentiation between class II compared with class I was determined from investigations to any anaerobically functioning proteins. If the ciliate was seen to contain anaerobic functioning proteins, including PNO/PFO or iron hydrogenase, this would be reason to classify them as class II anaerobic mitochondria, due to the probability of them utilising these proteins. Seven of the ciliates contain a complete or almost complete TCA, along with anaerobic functioning proteins shown in 'blue' in Table 6b. Table 6a and b show class I and class II classified ciliates in relation to TCA alone (*shown in green and blue respectively)*.

One of the ciliates, *Nyctotherus ovalis,* contains TCA functioning proteins, however, it is not classed within I or II due to lack of TCA components. *N. ovalis* is classified as Class III, when combined with the prevalence of iron hydrogenase along with predictions of complex I and II of the electron transport chain (Table 3). *N. ovalis* contains a reduced form of the TCA cycle and is also likely used in a

reductive direction, differing from that of class I and II. This is similar to that of *Metopus sp.,* which also contained a reduced TCA cycle.

Table 6 shows the class of mitochondria present in each ciliate dependent on their TCA protein composition, along with other contributing factors like anaerobic functioning proteins to allow a better separation. It is clear to see that without the identification of anaerobic functioning proteins it would be difficult to predict between class I and II due to them having a majority of TCA proteins, as seen in Figures 18a and b, where *Anophryoides haemophila* and *Mesodinium pulex* both have a complete TCA, however, due to the identification of anaerobic functioning proteins within *Mesodinium pulex* they can be classified separately.

a)

Ciliate	TCA
Anophryoides haemophila	12
Aristerostoma marinum	9
Blepharisma japonicum	9
Climacostomum virens	9
Condylostoma magnum	8
Euplotes focardii	8
Euplotes harpa	9
Fabrea salina	11
Ichthyophthirius multifiliis	11
Litonotus sp.	9
Paramecium tetraurelia	11
Protocruzia sp.	8
Pseudokeronopsis riccii	7
Schmidingerella inclinatum	7
Strombidium inclinatum	7

b)

Ciliate	TCA
Brown Ciliate	8
Cryptocaryon irritans	7
Mesodinium pulex	13
Platyophrya macrostoma	10
Strombidinopsis acuminatum	13
Tetrahymena thermophila	11
Uronema sp.	12

c)

Metopus sp.	6
Nyctotherus ovalis	4
Polyplastron multivesiculatum	-
Entodinium caudatum	-
Chilodonella uncinata	-

Table 6 - Shows the containment of TCA proteins within each ciliate.
Colour code is as follows: **Class I, Class II, Class III, Class IV and Class V.**

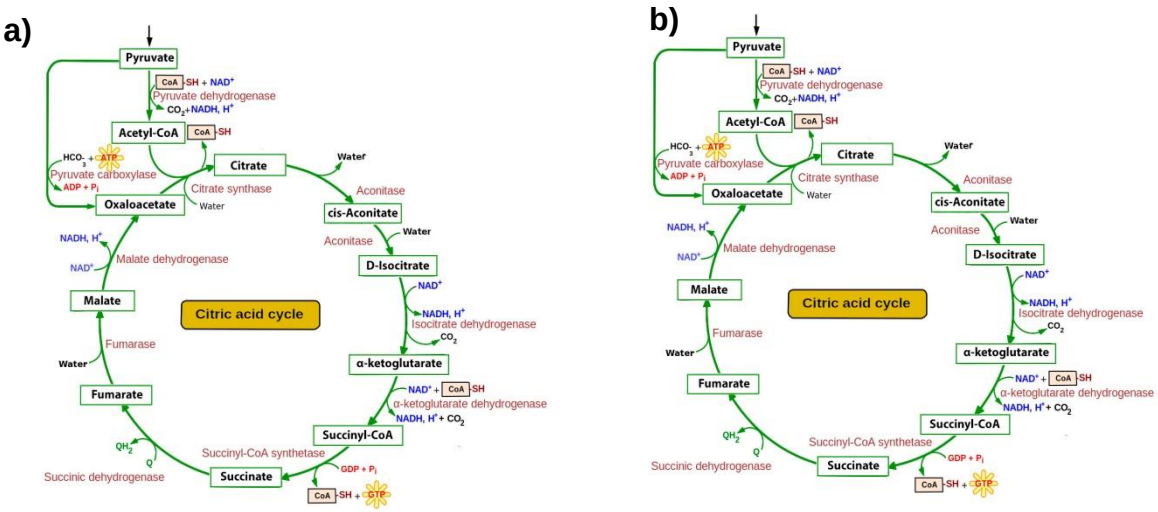
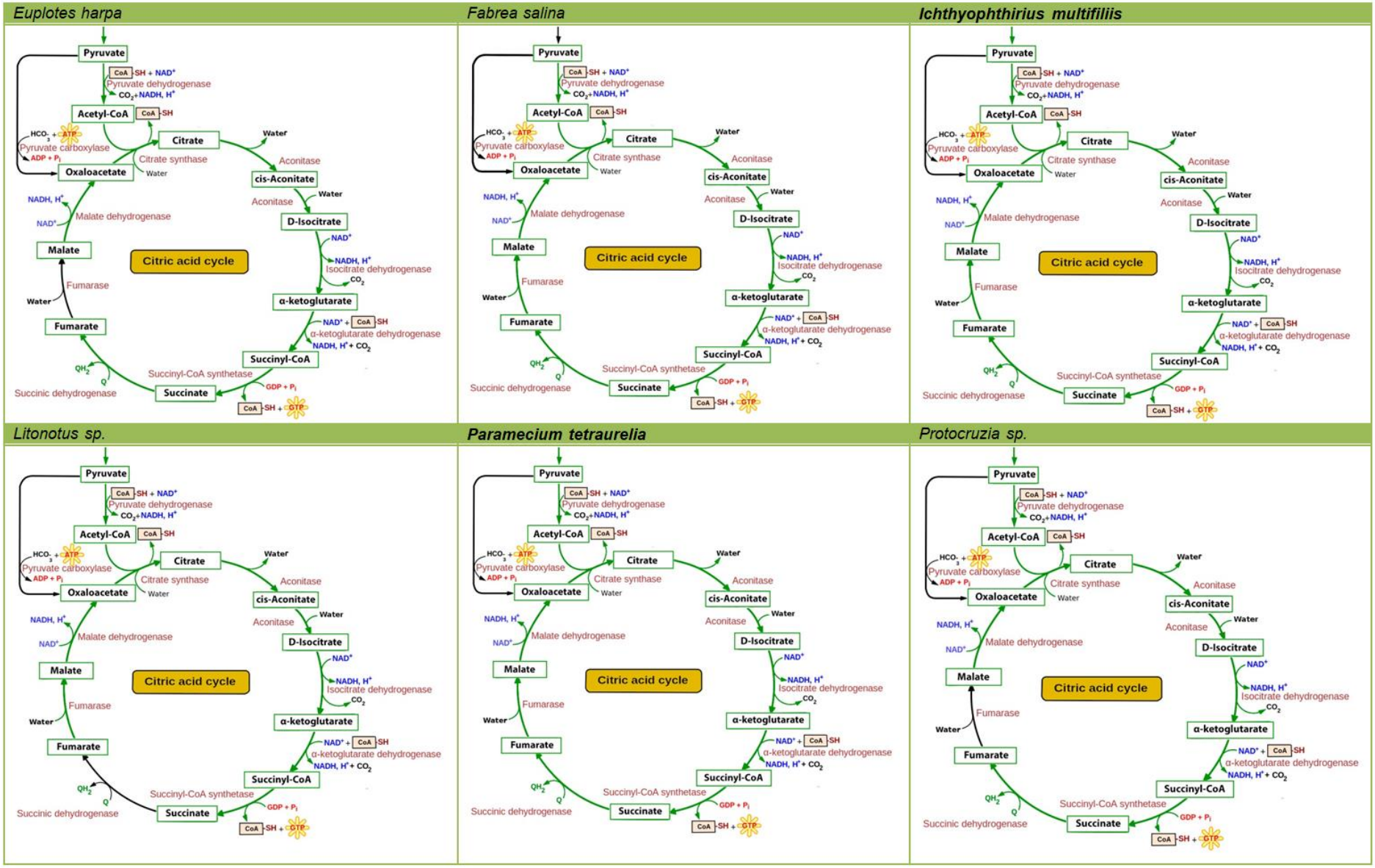


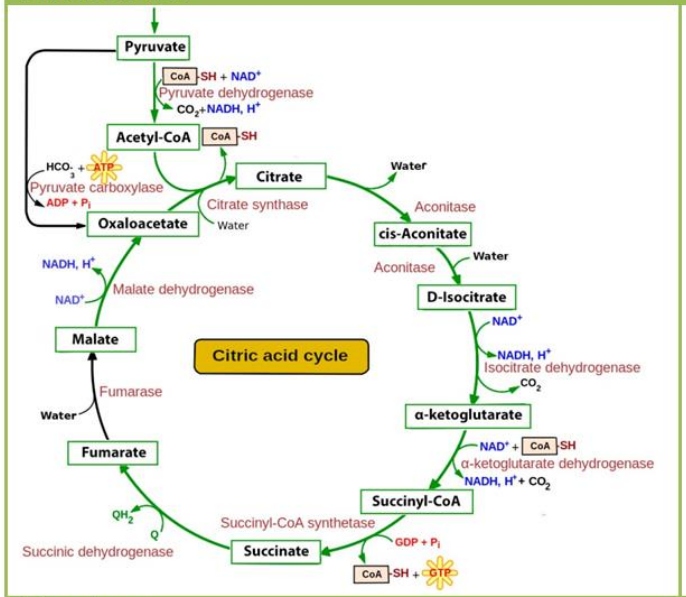
Figure 18 **-** a) Class I *Anophryoides haemophila* TCA proteins. b) Class II *Mesodinium pulex* TCA proteins.

Three ciliates had no TCA proteins identified; *Polyplastron multivesiculatum, Entodinium caudatum* and *Chilodonella uncinata* (seen in Figure 19). So it is not possible to effectively class them, however, like with differentiating between class I and II, the same could be done for *Polyplastron multivesiculatum* as an iron hydrogenase was identified which suggests it likely contains an anaerobically functioning type of mitochondria, and due to the lack of TCA proteins, likely a reduced class. All 27 ciliates in relation to their TCA protein composition can be seen in Figure 19 and Tables 2 and 3.

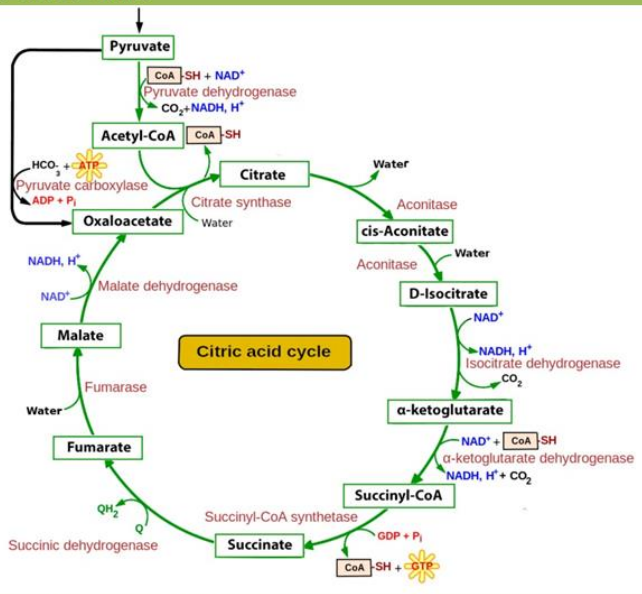
TCA Citric Acid Cycle Figure 19



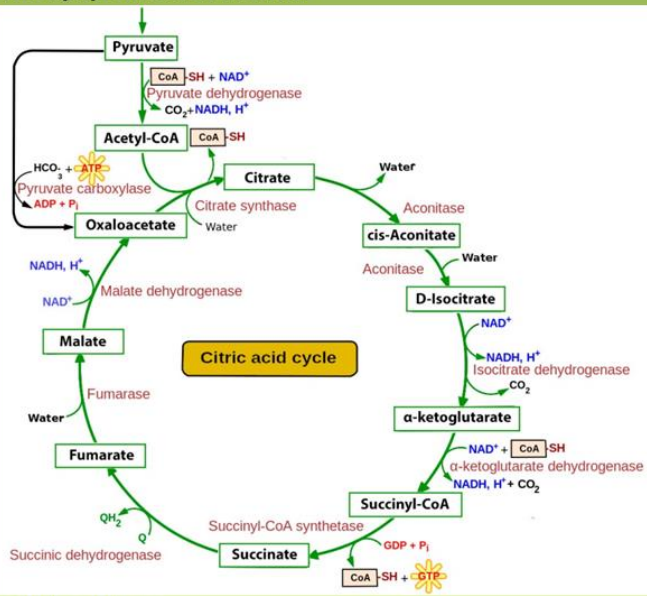
Euplotes harpa



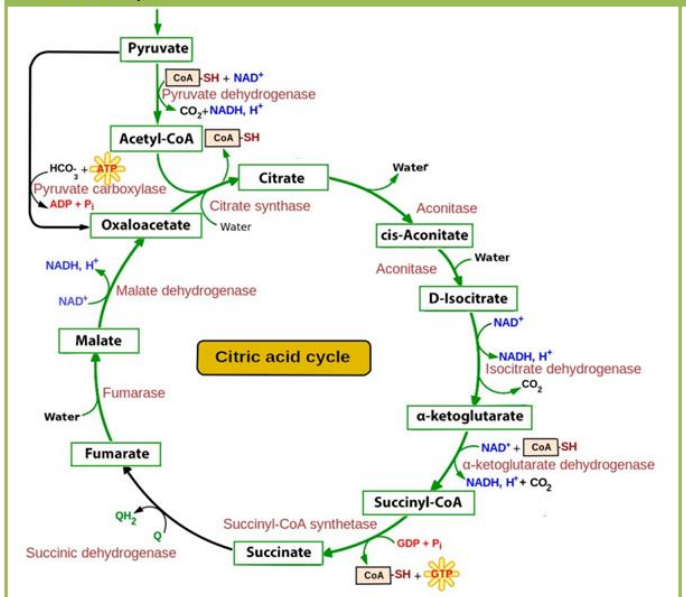
Fabrea salina



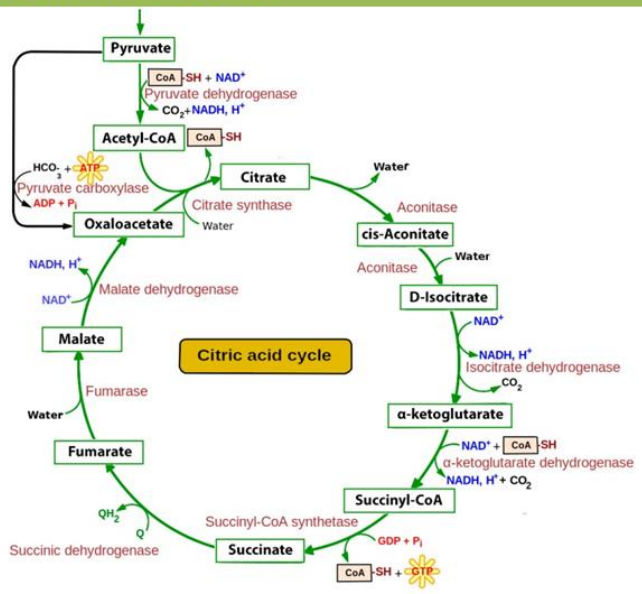
Ichthyophthirius multifiliis



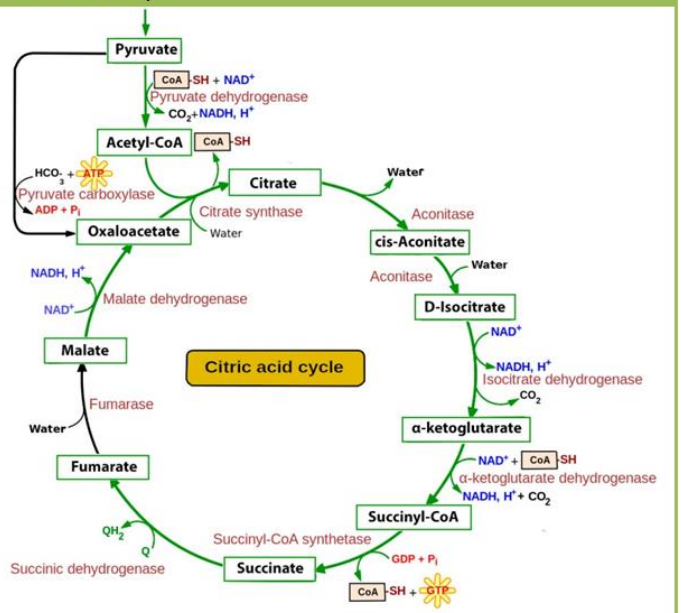
Litonotus sp.

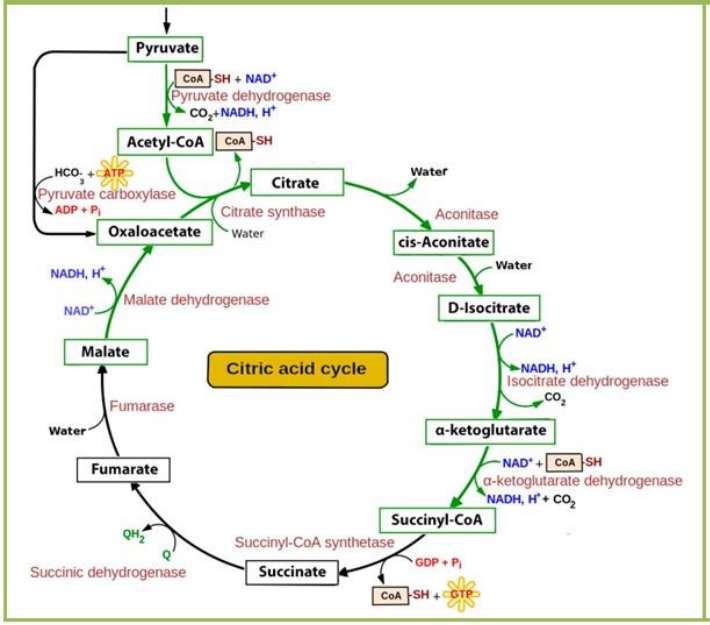
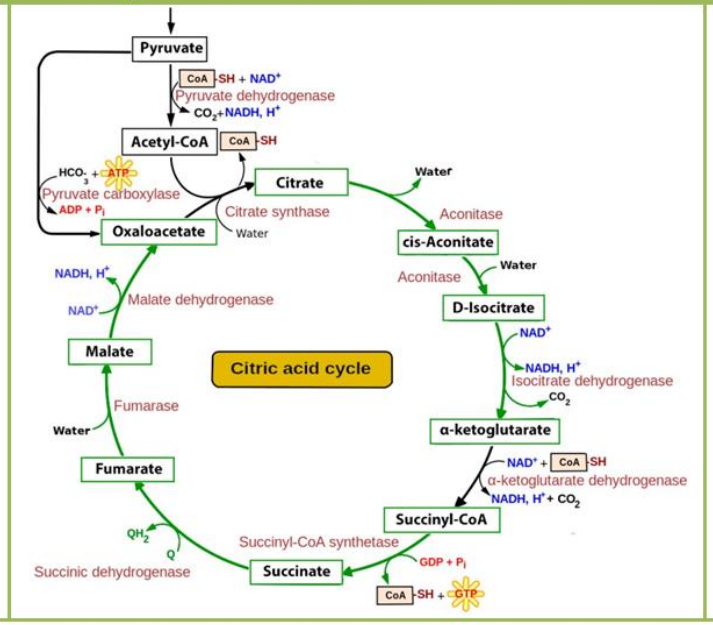
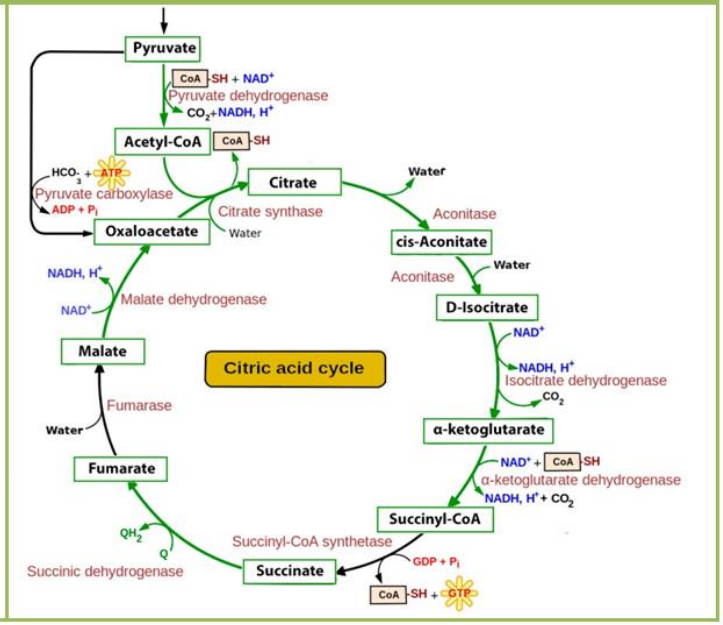


Paramecium tetraurelia

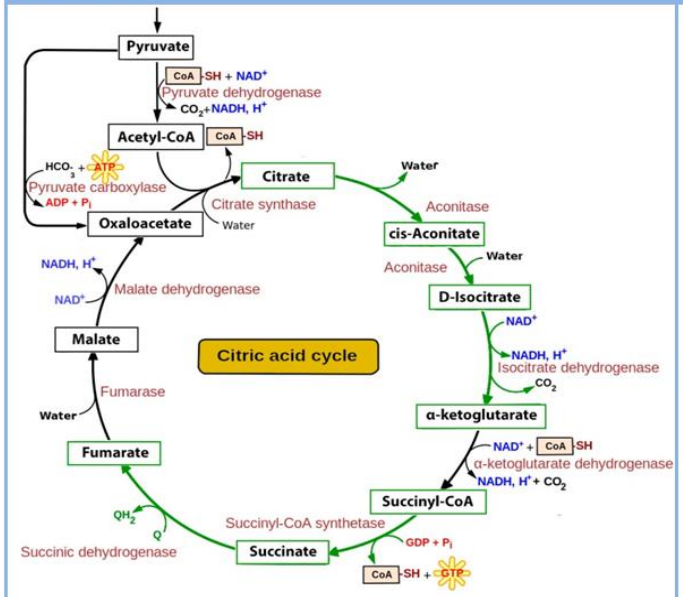


Protocruzia sp.

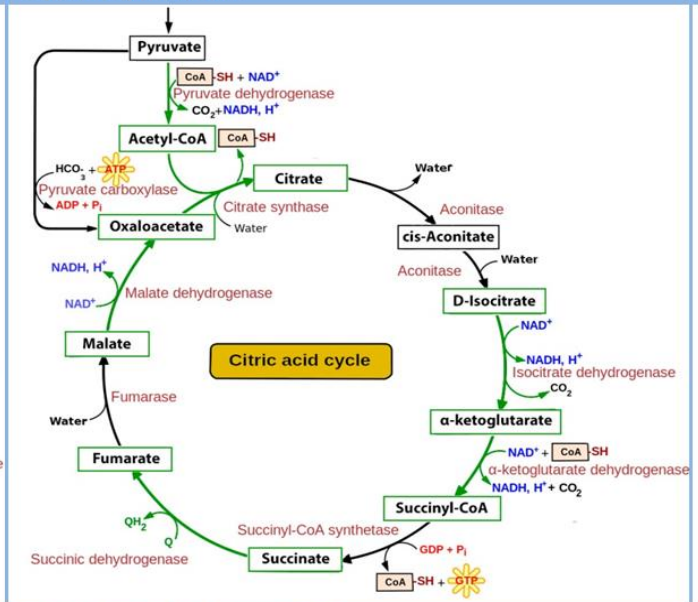


Pseudokeronopsis riccii	*Schmidingerella inclinatum*	*Strombidium inclinatum*
		

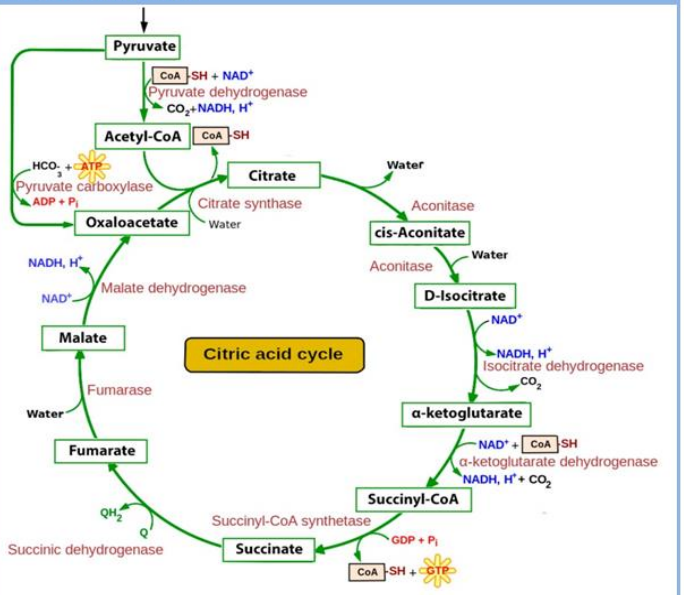
Brown Ciliate



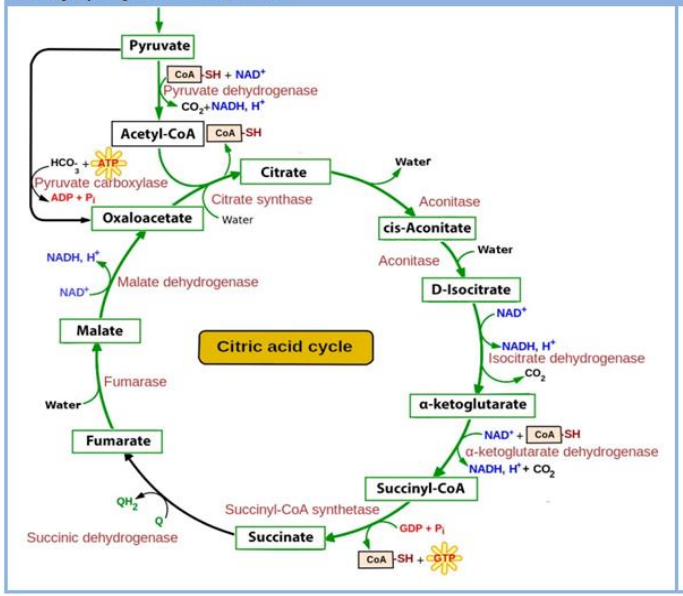
Cryptocaryon irritans



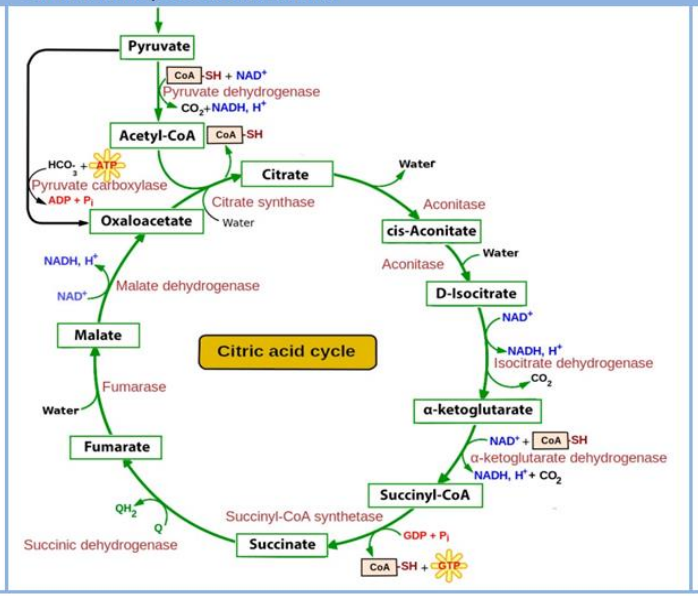
Mesodinium pulex



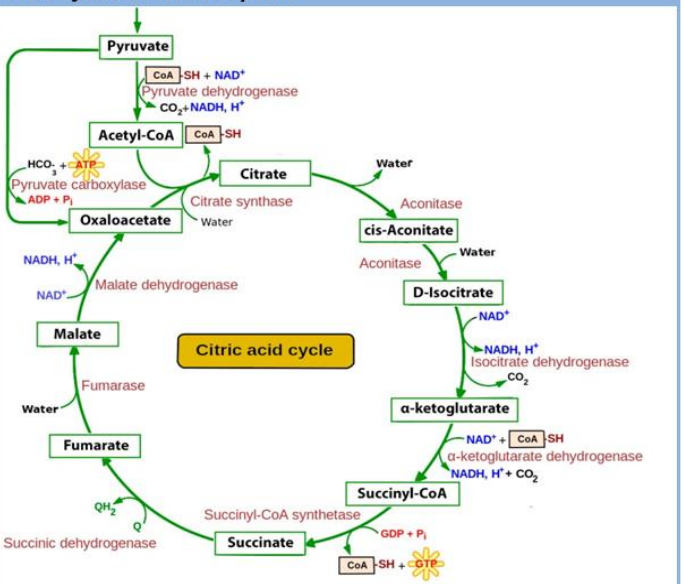
Platyophrya macrostoma



Strombidinopsis acuminatum



Tetrahymena thermophila



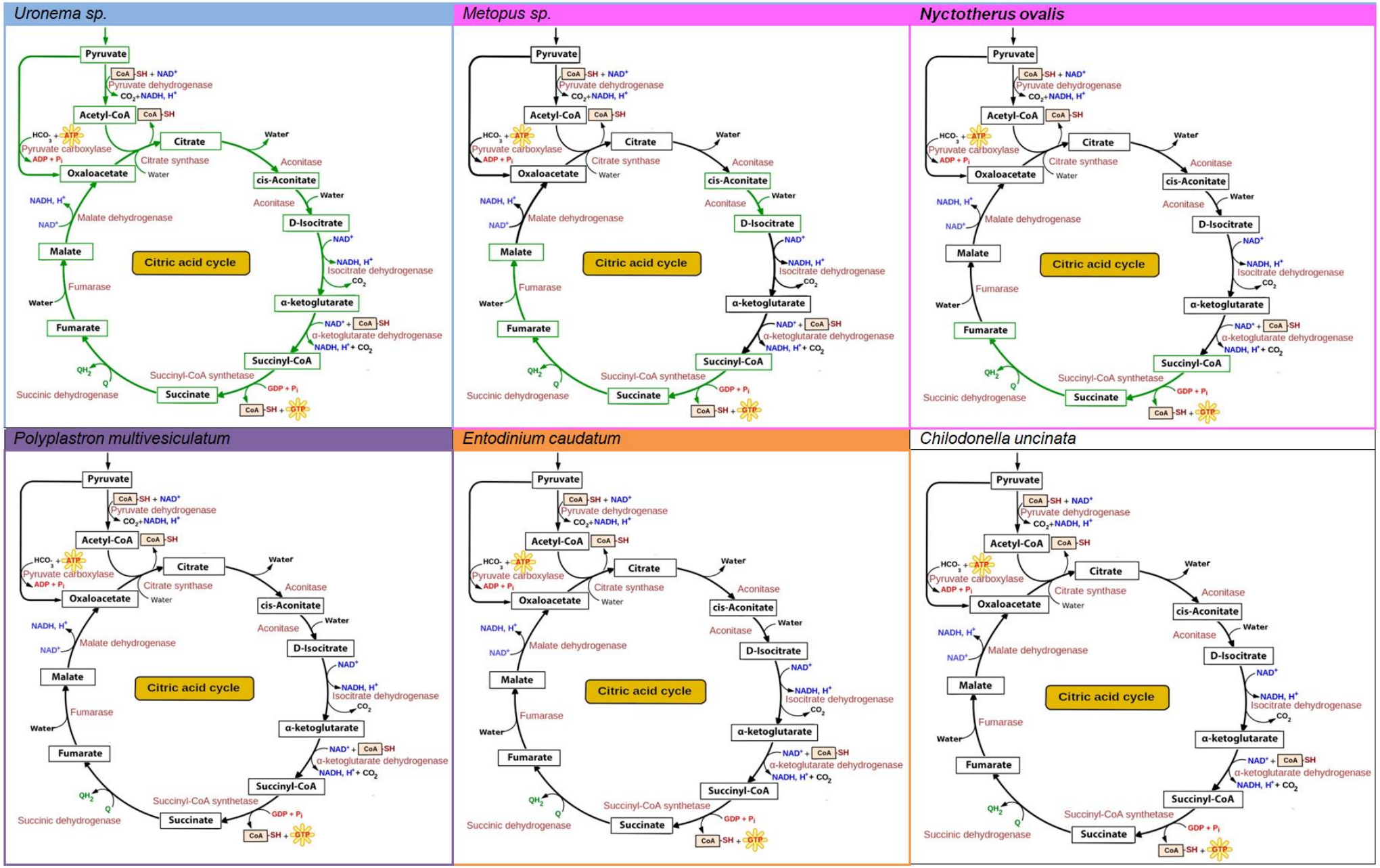


Figure 19 - The TCA content for each ciliate, with steps coloured in green showing when the protein has been identified.

The Electron Transport Chain

Mitochondria are distinctively recognisable by their double membrane, where the inner membrane contains a series of protein complexes that form the electron transport chain; however, these complexes are not always present on the inner membrane due to alternate methods of energy production utilised by the organism. 25 of the 27 ciliates contained proteins of the electron transport chain, varying from as little as two subunits all the way to 33 subunits being identified as seen in Table 2 and Figure 21.

All five complexes have been identified between all the ciliates, with many having a complete electron transport chain (Table 3). *Mesodinium pulex* is an example where subunits for all complexes were identified shown in Figure 20a; leading to a likelihood that the mitochondria present is either class I or II. *Nyctotherus ovalis* contains 12 identified subunits, all of which form components of complex I and II, shown in Figure 20b. This further supports the classification of *N. ovalis* being class III H_2-production mitochondria according to Müller [10] outline, due to its reduced electron transport chain.

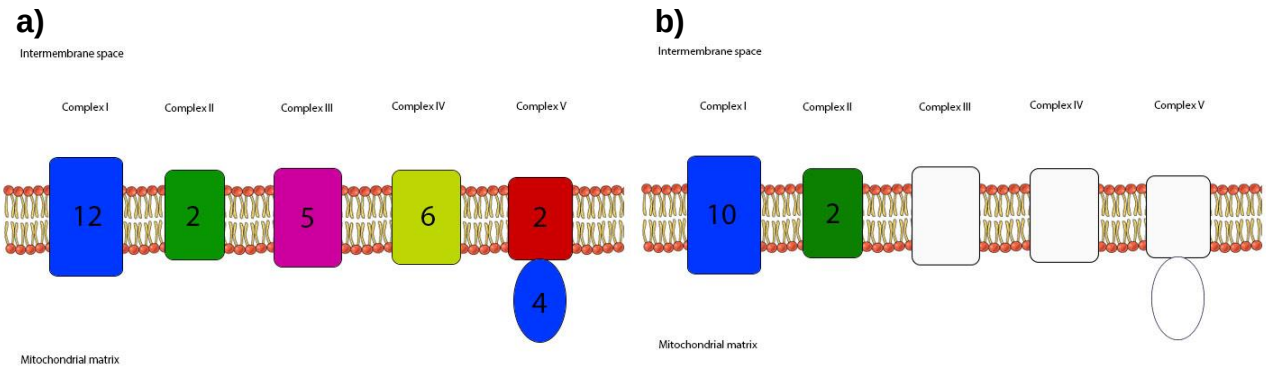
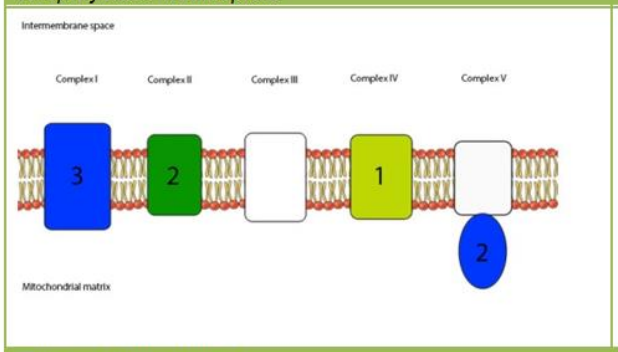
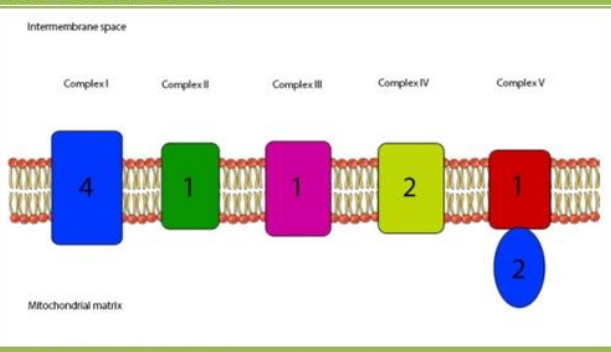
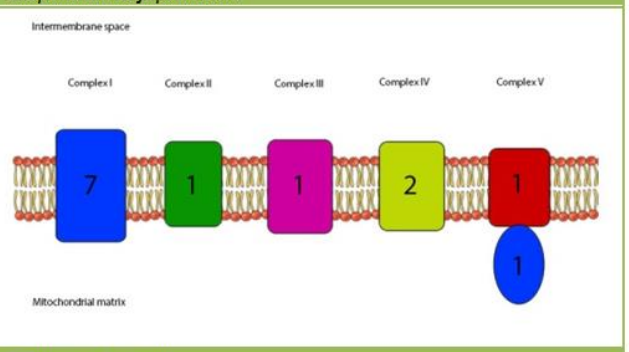
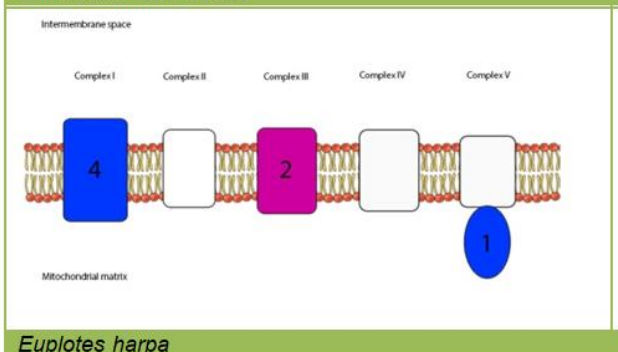
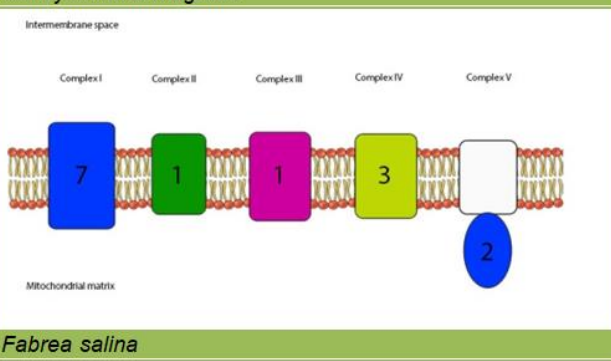
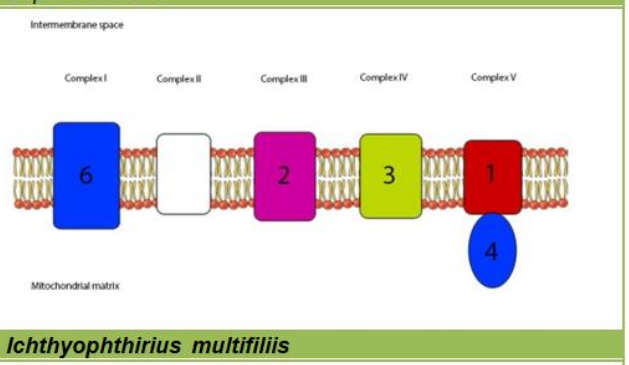
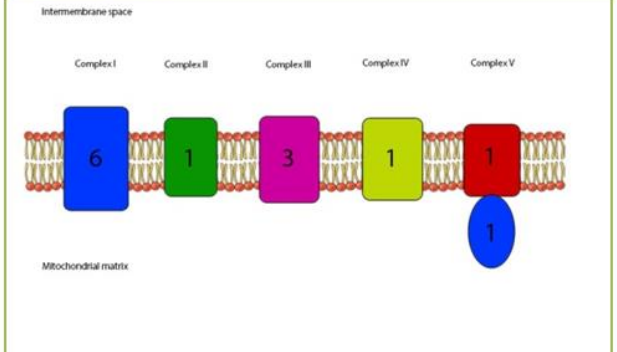
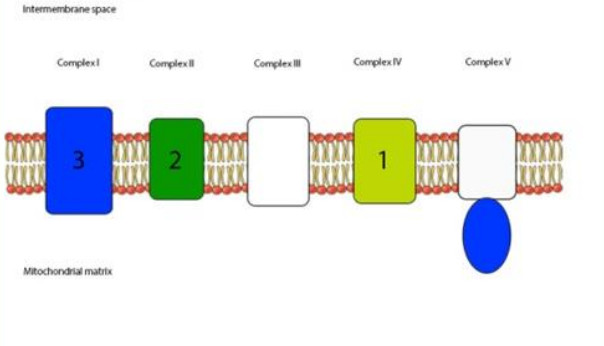
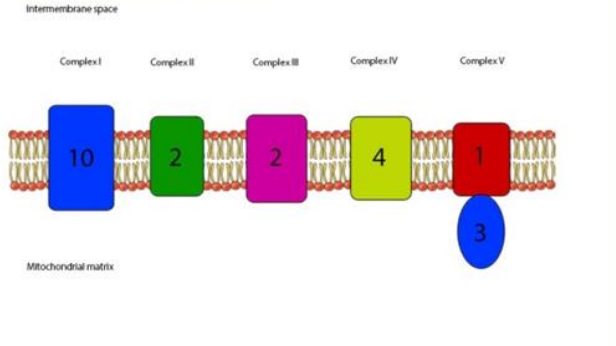


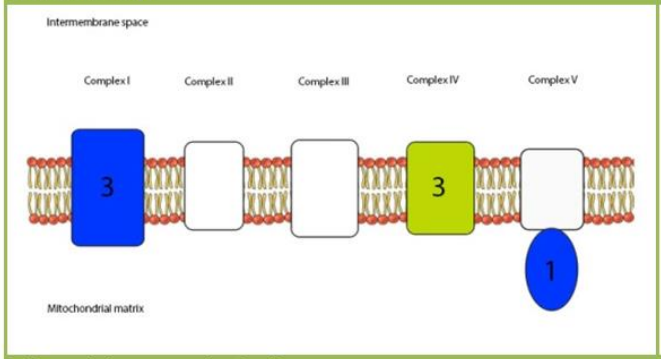
Figure 20 - a) *Mesodinium pulex* electron transport chain. Coloured in complex shows subunits identified. Number equals number of individual subunits identified. b) *Nyctotherus ovalis* electron transport chain. Coloured in complex shows subunits identified. Number equals the number of individual subunits identified.

For the two ciliates *Polyplastron multivesiculatum* and *Entodinium caudatum,* no electron transport chain proteins were identified, which leads to the potential of them being either class IV or V. Then in conjunction with TCA proteins identification and anaerobic proteins, a specific class could then be predicted. All 27 ciliates in relation to their electron transport chain protein composition can be seen in Figure 21, as well as Table 3.

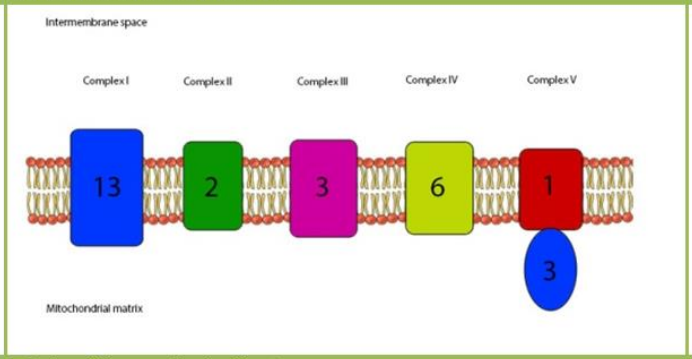
Electron Transport Chain Figure 21

Anophryoides haemophila	*Aristerostoma marinum*	*Blepharisma japonicum*
		
Climacostomum virens	*Condylostoma magnum*	*Euplotes focardii*
		
Euplotes harpa	*Fabrea salina*	***Ichthyophthirius multifiliis***
		

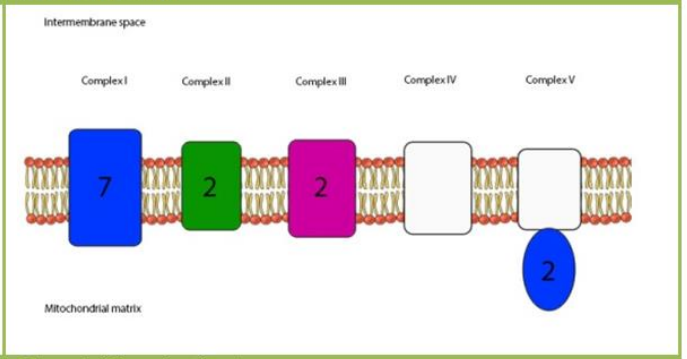
Litonotus sp.



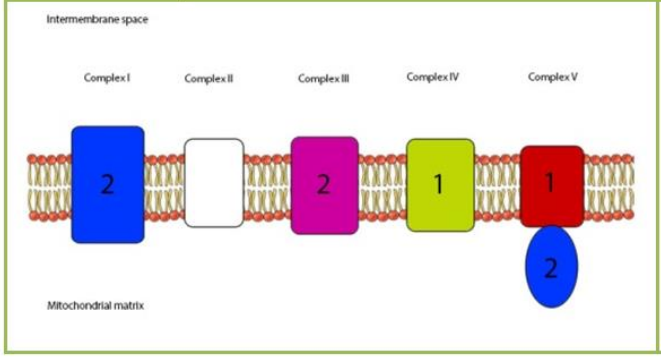
Paramecium tetraurelia



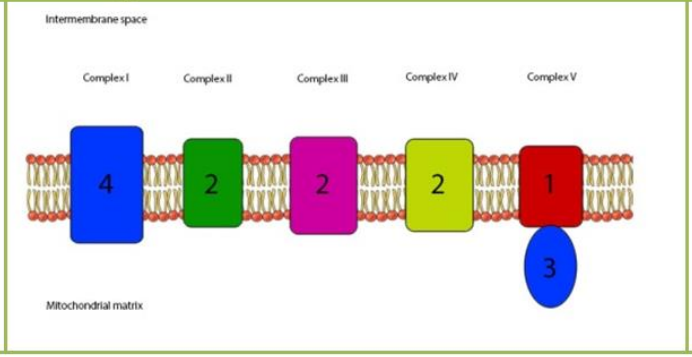
Protocruzia sp.



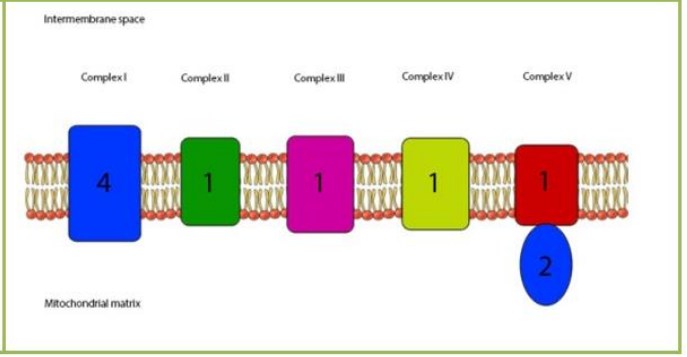
Pseudokeronopsis riccii



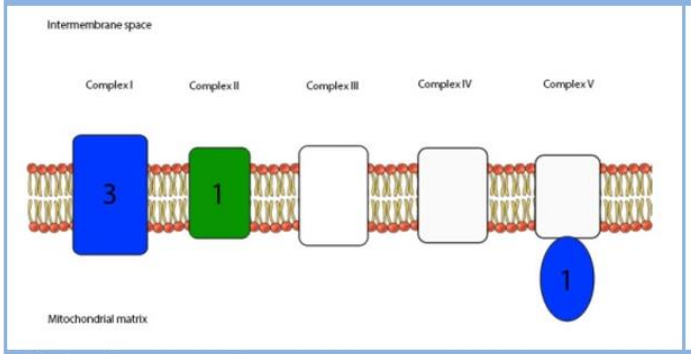
Schmidingerella inclinatum



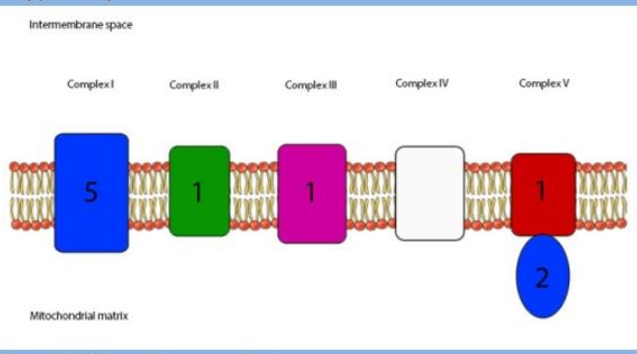
Strombidium inclinatum



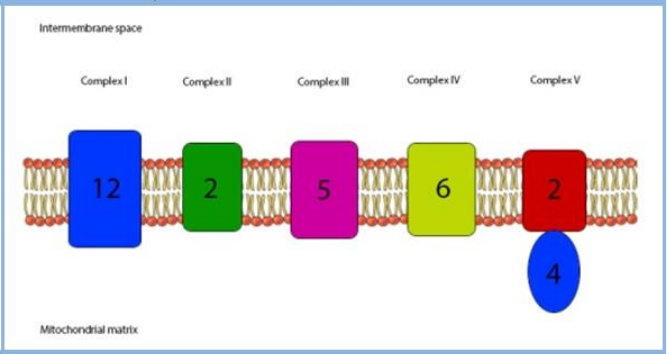
Brown Ciliate



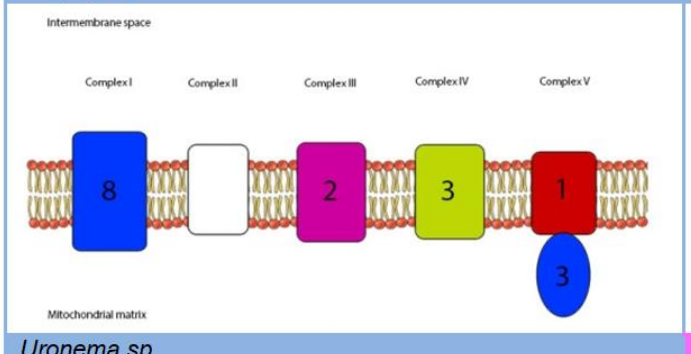
Cryptocaryon irritans



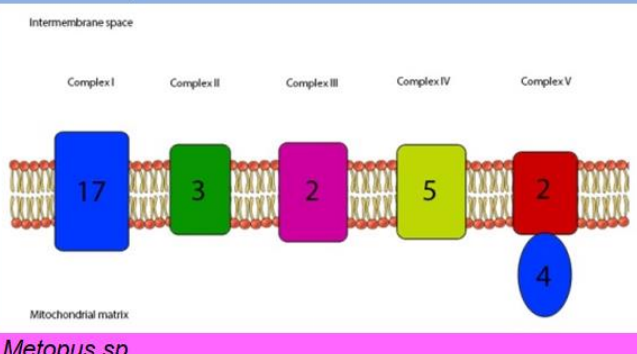
Mesodinium pulex



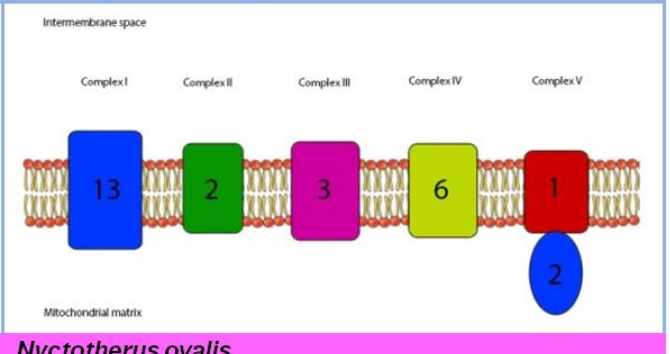
Platyophrya macrostoma



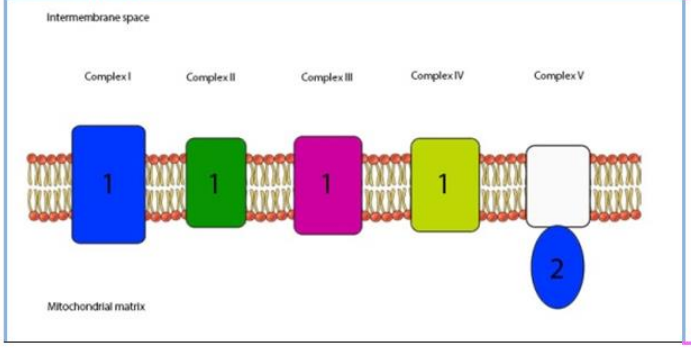
Strombidinopsis acuminatum



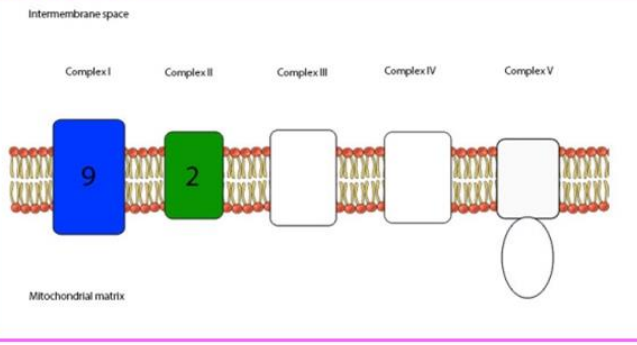
Tetrahymena thermophila



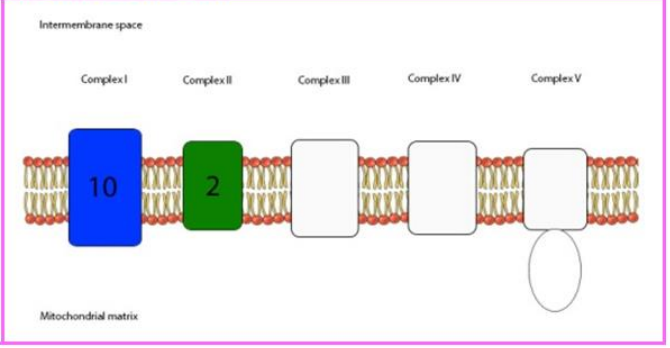
Uronema sp.



Metopus sp.



Nyctotherus ovalis



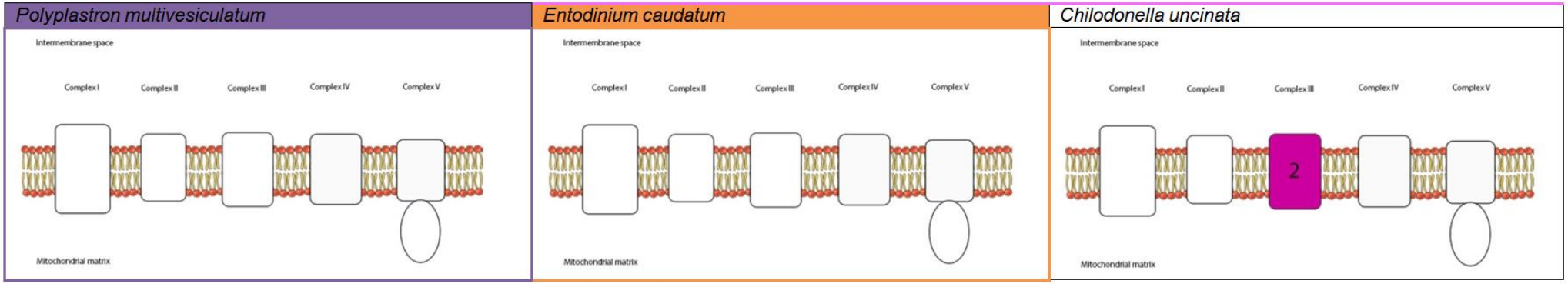


Figure 21 - The ETC content for each ciliate, with each complex coloured showing when the proteins forming the complex are identified and the number showing the number of protein subunits identified.

Anaerobic proteins

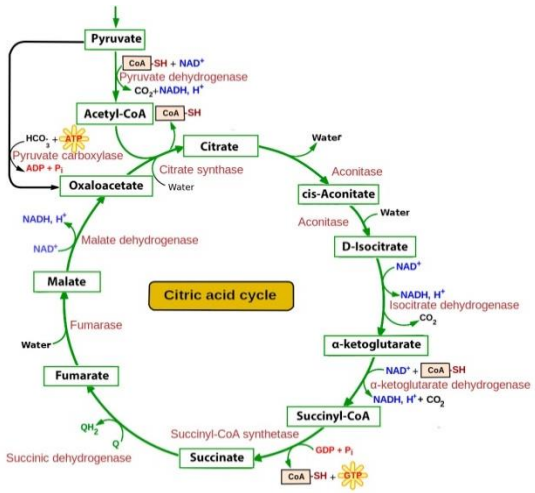
Anaerobic functioning proteins (PNO/PFO and iron hydrogenase) were important to help distinguish the class of mitochondria, especially between class I and II. When analysing the mitochondrial protein composition of each ciliate, if they contain a substantial TCA and electron transport chain protein composition, anaerobic proteins were required to separate the classification between class I and II. It was also beneficial for ciliates containing limited or no TCA or electron transport chain proteins, helping predict whether it was likely to be class III and IV.

PNO/PFO's were identified in six of the ciliates, and seven ciliates contained an iron hydrogenase, both are shown in Table 7.

Ciliate	PNO/PFO	FeHy
Brown Ciliate	1	-
Cryptocaryon irritans	2	-
Mesodinium pulex	1	1
Platyophrya macrostoma	2	1
Strombidinopsis acuminatum	2	-
Tetrahymena thermophila	-	1
Uronema sp.	-	1
Metopus sp.	3	4
Nyctotherus ovalis	-	6
Polyplastron multivesiculatum	-	1
Entodinium caudatum	-	-
Chilodonella uncinata	-	-

Table 7 - Ciliates containing proteins that function in anaerobic conditions, along with their predicted mitochondrial class.

a)



b)

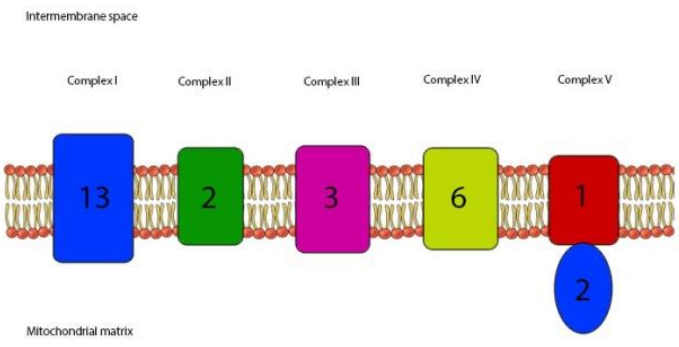


Figure 22 – a) and b) show protein composition of TCA cycle and electron transport chain for *Tetrahymena thermophila.*

Those shown in blue are class II mitochondria, due to the abundance of TCA and electron transport chain proteins, which due to data available from this analysis; it is difficult to investigate further if they belong to a further reduced class. Good examples of this are the ciliates that contain identification hits for iron hydrogenase; a protein that is usually associated with hydrogenosomes. However, our methods show, in the example of *Tetrahymena thermophila,* which has an abundance of TCA and electron transport chain proteins (Figure 22a and b), suggests that it may not be a reduced mitochondrial class, like that of *N. ovalis* or *Metopus sp.*. The combination of TCA and electron transport chain protein composition along with anaerobic proteins allows the type of mitochondria present to be predicted in relation to classification traits outlined by Müller *et al* [10].

Following the construction of the phylogenetic trees, it demonstrated that there was some contamination present. This was clearly seen especially for *Metopus sp.,* in regards to the predicted PFO sequences (Figure 24). However, it also confirmed the presence of iron hydrogenases, as many have a close similarity to those of *Nyctotherus ovalis* (Figure 23).

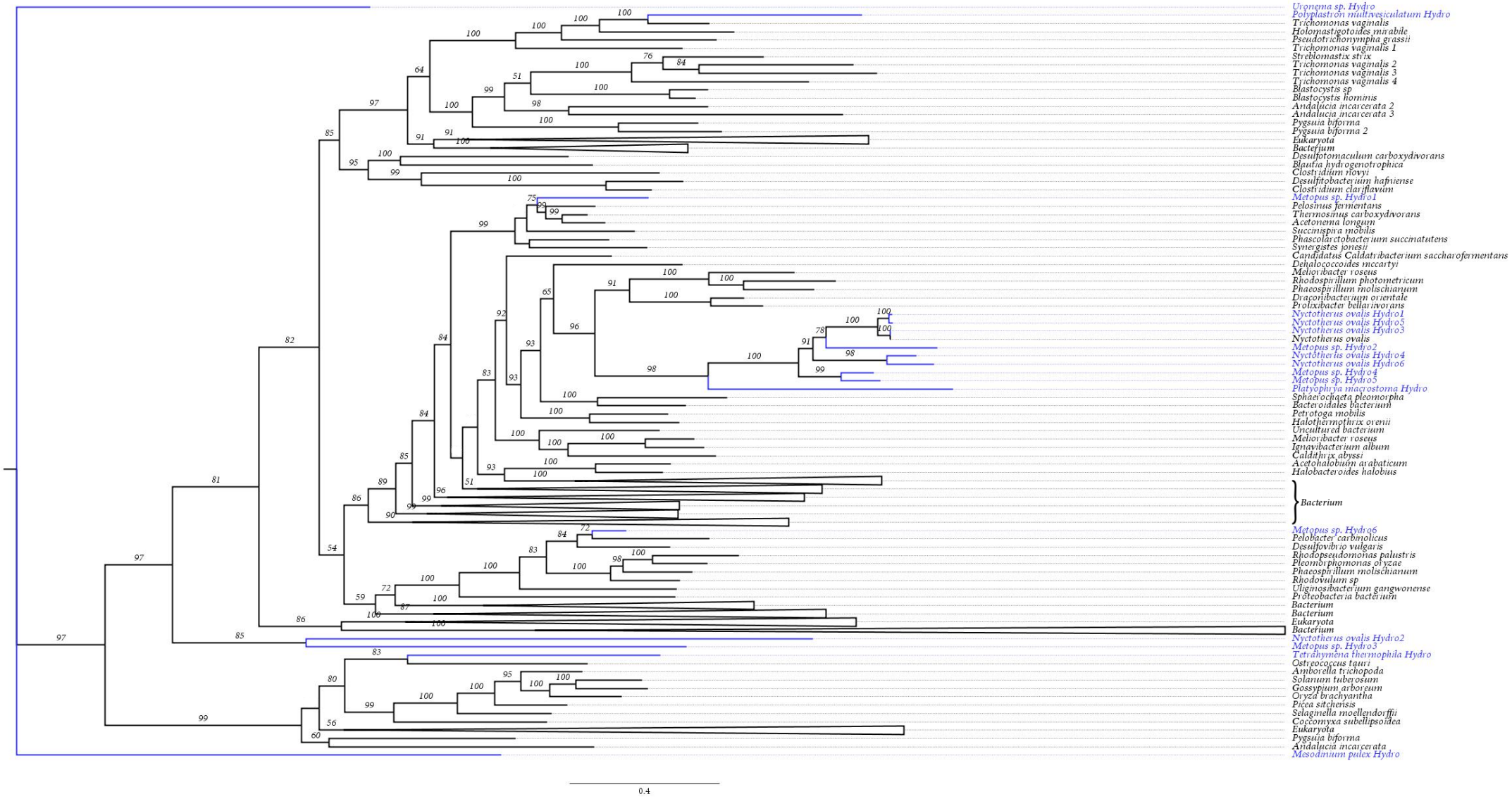


Figure 23 – Unrooted Maximum Likelihood phylogenetic tree of the predicted Iron Hydrogenase sequences aligned to other known iron hydrogenase sequences [62] to identify for contamination. Bootstrap 1000.

*Only Bootstrap values above 50 shown

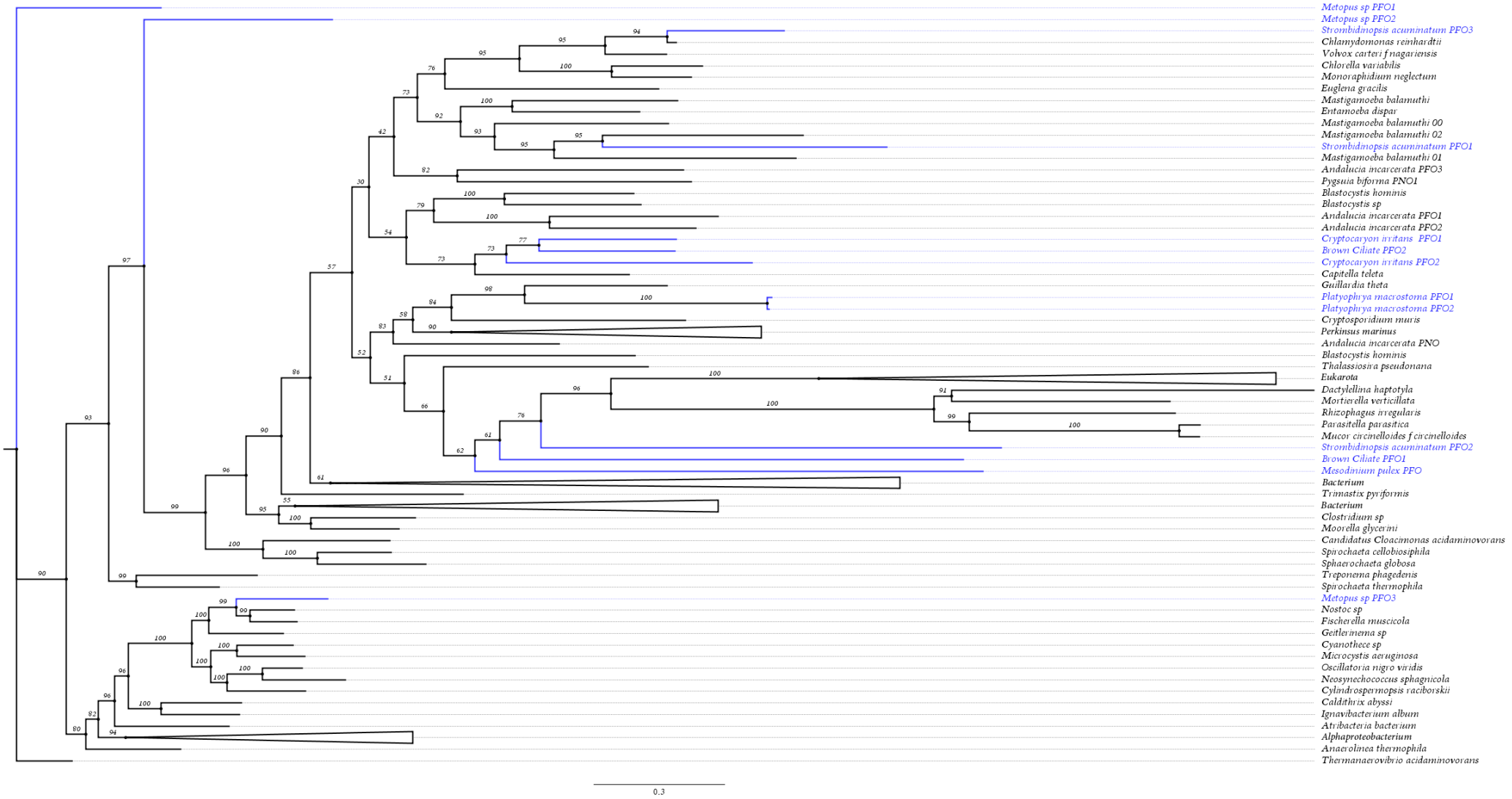


Figure 24 – Unrooted Maximum Likelihood phylogenetic tree of the predicted PFO/PNO sequences aligned to other known PFO/PNO sequences [62] to identify for contamination. Bootstrap 1000.

*Only Bootstrap values above 50 shown

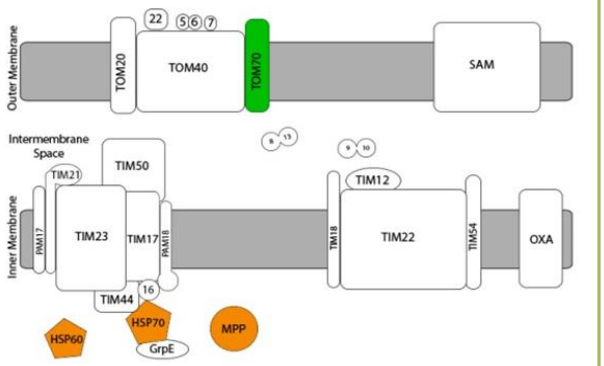
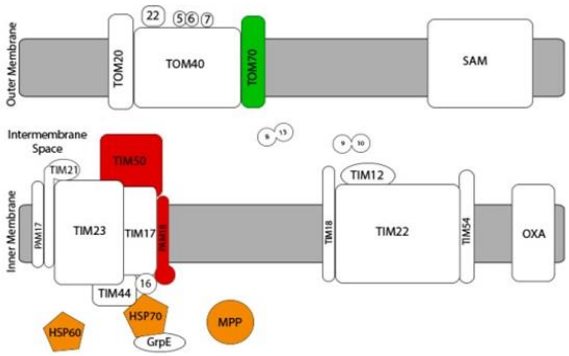
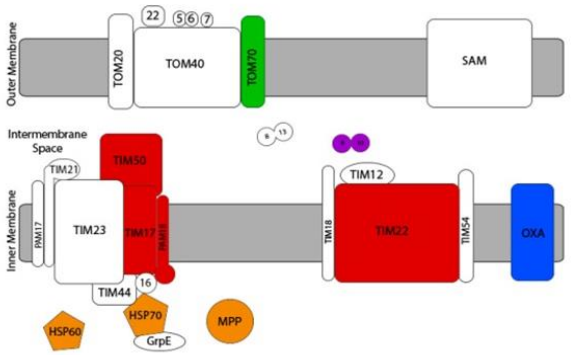
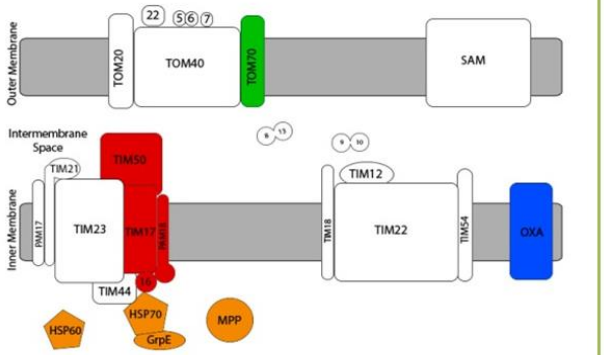
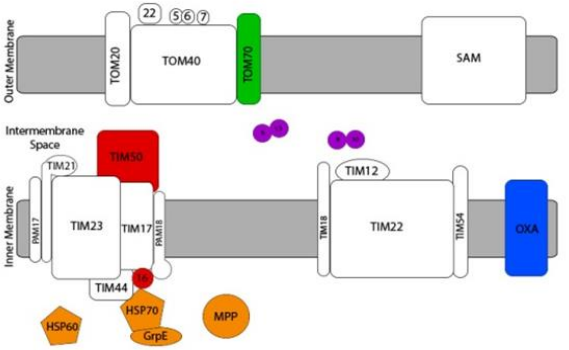
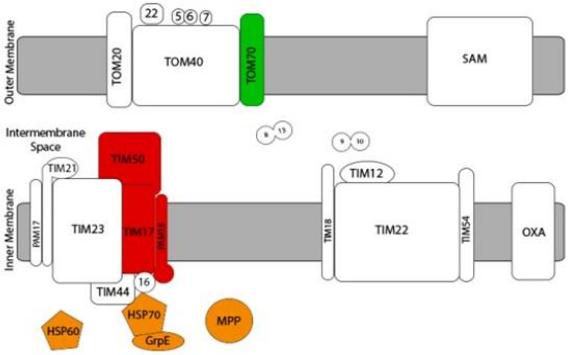
Protein Import, β-Oxidation, Fe/S pathways and Mitochondrial Class

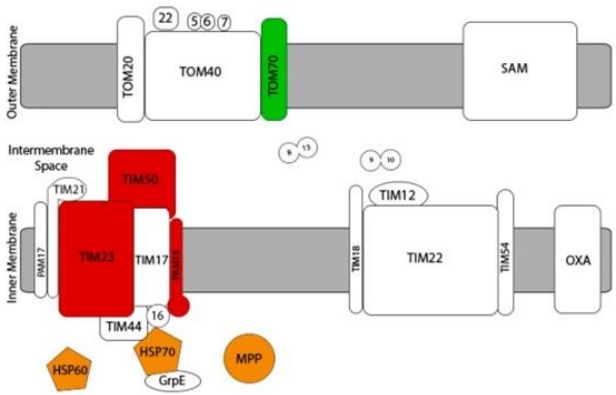
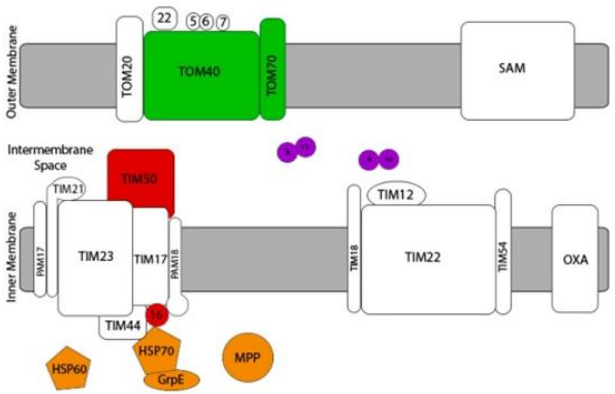
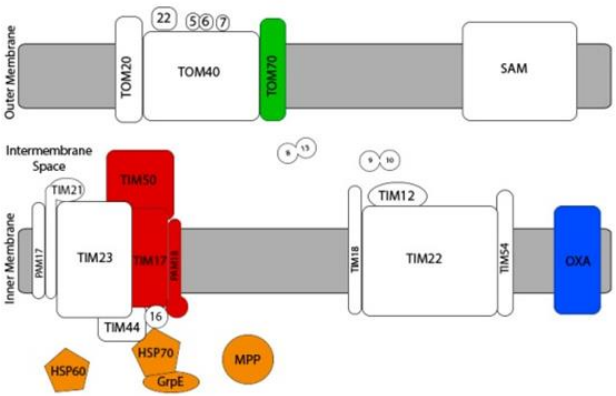
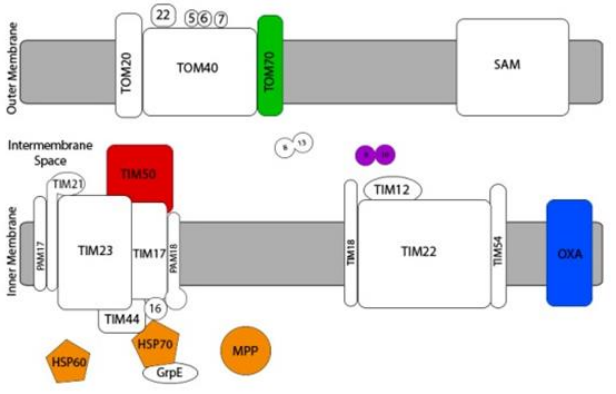
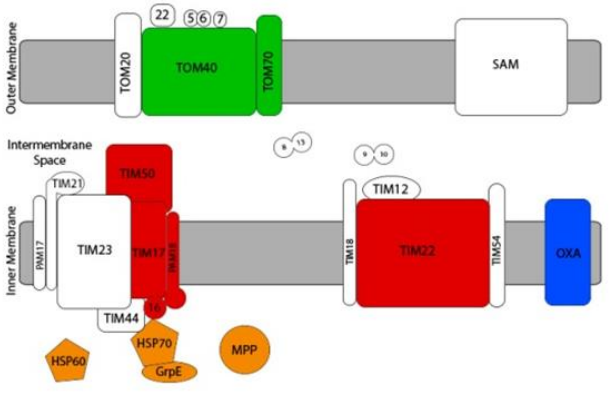
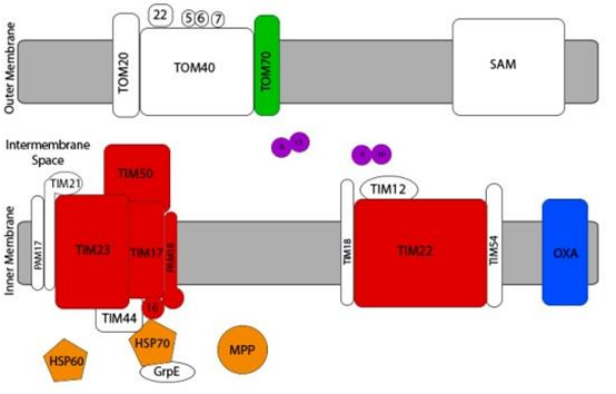
Protein import, β-oxidation and the Fe/S cluster assembly pathways were also investigated to help build a strong mitochondrial overview. These three pathways were present for all the ciliates predicted containing class I, II or III mitochondria, which is unsurprising due to the conservation already seen in these vital pathways. A mitochondrial TOM70 has also been identified within *P. multivesiculatum (class IV),* suggesting the presence of protein import. The number of contigs for the *P. multivesiculatum* transcriptome was very low causing protein identification to be more difficult from a possible lack of coverage (Table 2 column 2 and Table 4).

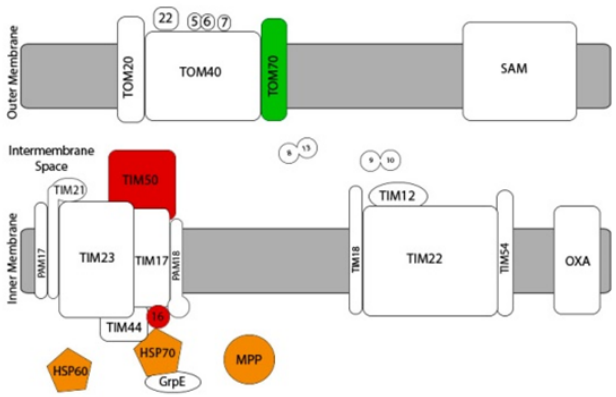
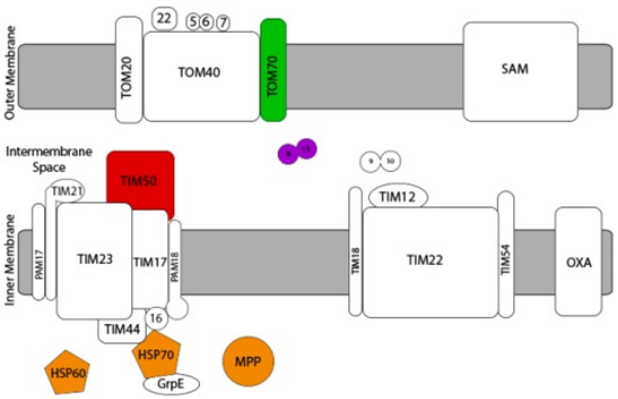
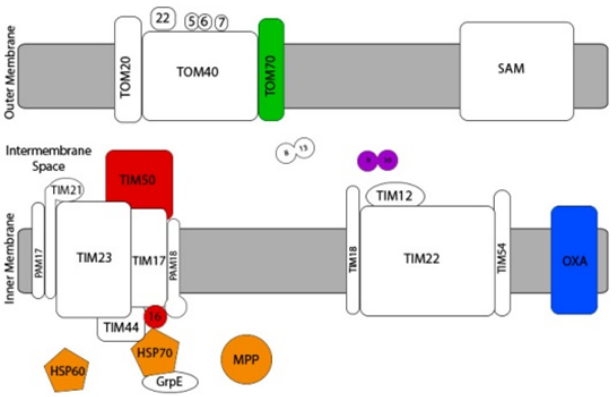
The protein composition for protein import and Fe/S cluster for each ciliate is shown in Figures 25 and 26. It is clear that protein import is widespread throughout all classes, with TIM50 being present in almost all ciliates along with mitochondrial Hsp70 (Table 3). β-oxidation was identified in all but three ciliates (Table 3 under 'fatty acid metabolism'), which included both class III containing ciliates *N. ovalis* and *Metopus sp.*, using the fatty acid degradation as a form of energy source. Fe/S cluster assembly was likewise shown to be conserved throughout the mitochondrial classes, as all but two ciliates had Fe/S cluster assembly proteins present (Figure 26 and Table 3).

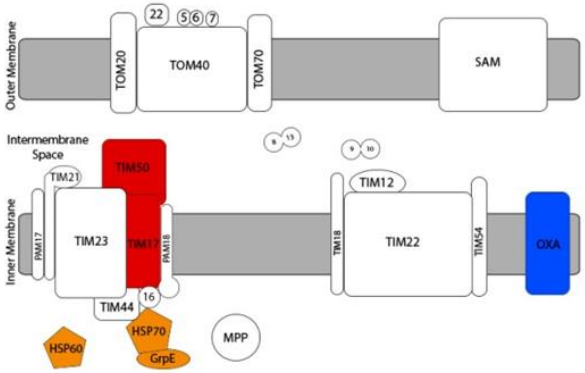
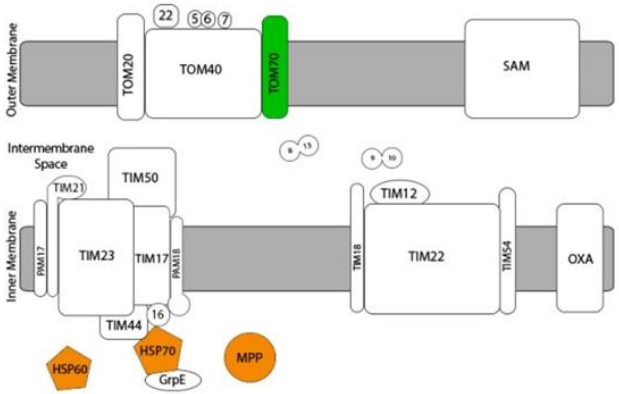
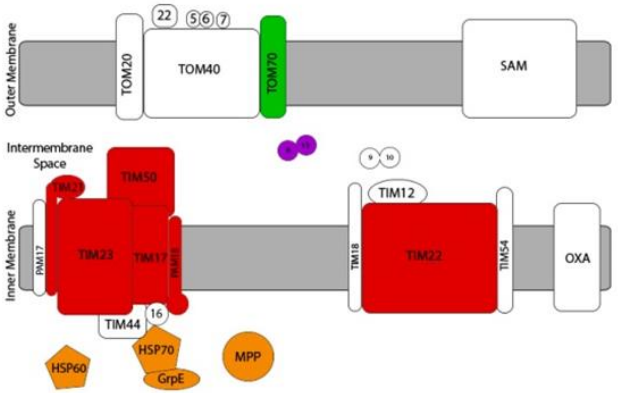
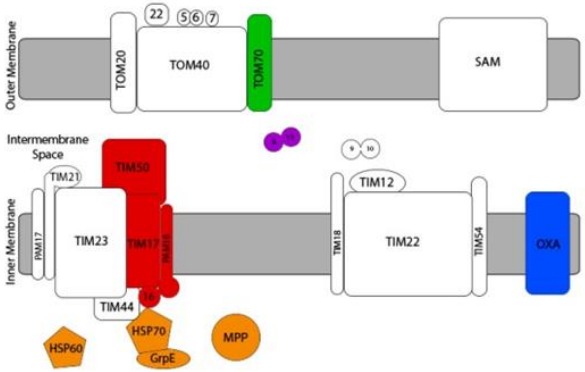
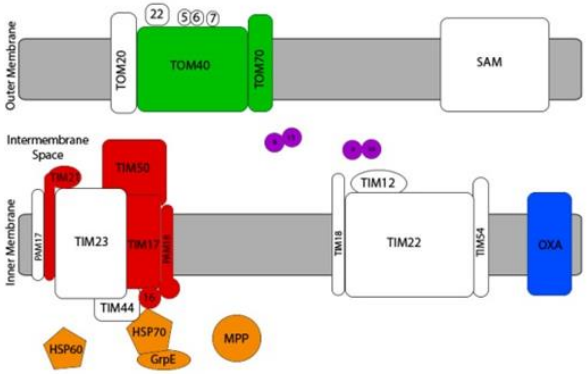
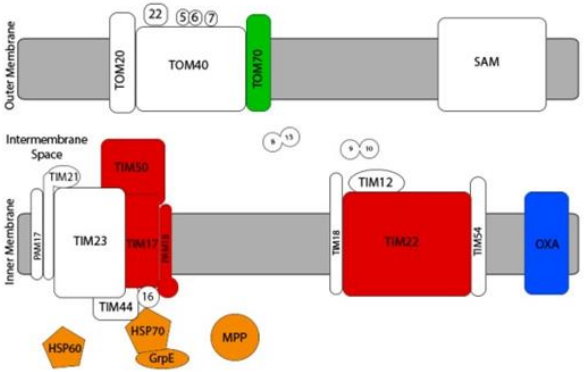
Complete protein compositions in relation to their pathways are shown in the 'Mitochondrial pathway overview figures' (Figure 27). They show all the pathways and specific proteins that were identified from the methods used within each ciliate.

Protein Import Figure 25

Anophryoides haemophila	*Aristerostoma marinum*	*Blepharisma japonicum*
		
Climacostomum virens	*Condylostoma magnum*	*Euplotes focardii*
		

Euplotes harpa	*Fabrea salina*	***Ichthyophthirius multifiliis***
		
Litonotus sp.	***Paramecium tetraurelia***	***Protocruzia sp.***
		

Pseudokeronopsis riccii	*Schmidingerella inclinatum*	*Strombidium inclinatum*
		

Brown Ciliate	*Cryptocaryon irritans*	*Mesodinium pulex*
		
Platyophrya macrostoma	*Strombidinopsis acuminatum*	***Tetrahymena thermophila***
		

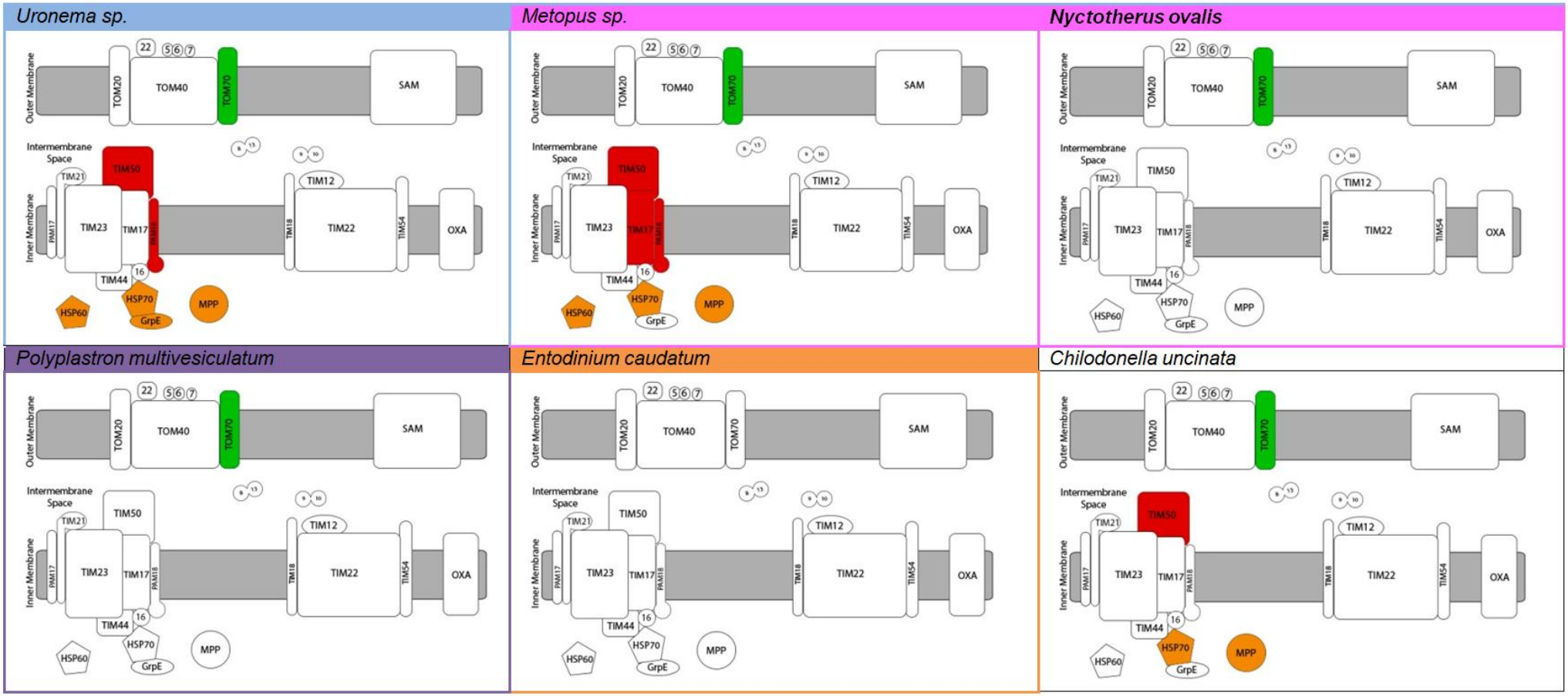
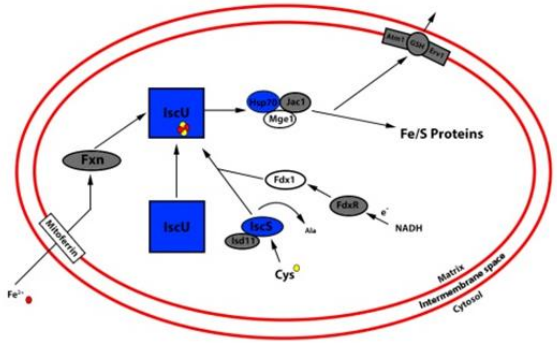
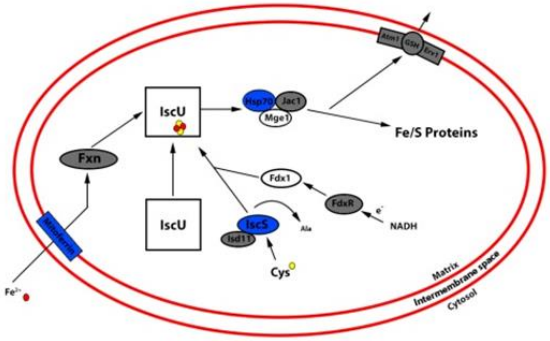
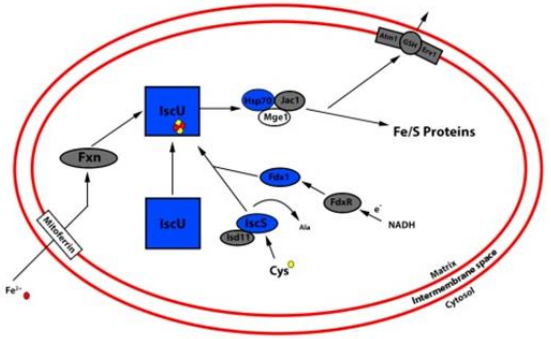
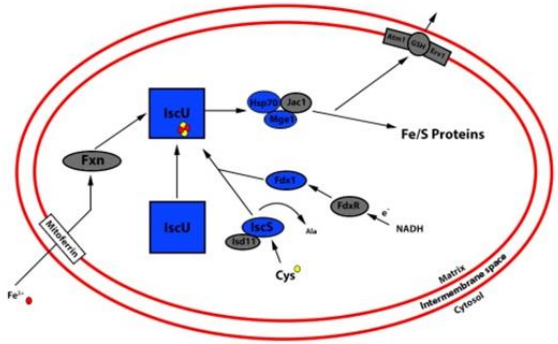
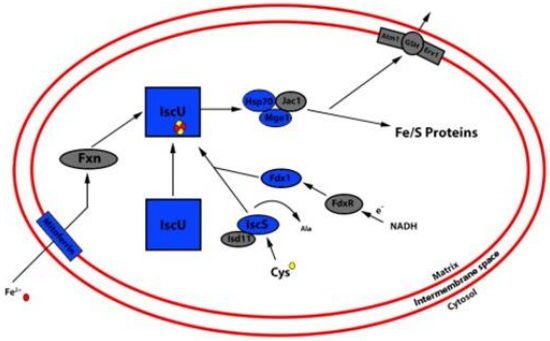
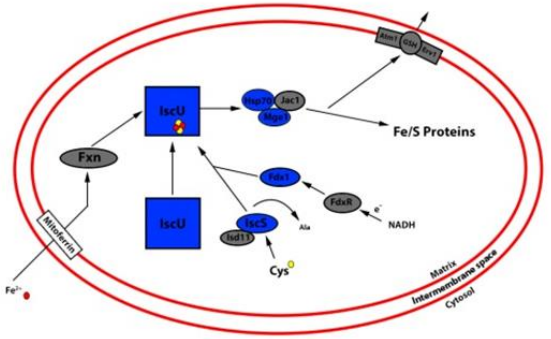
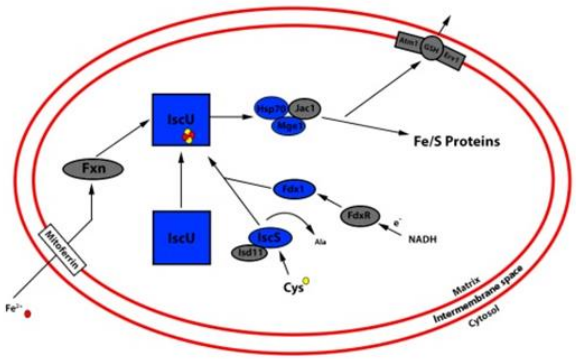
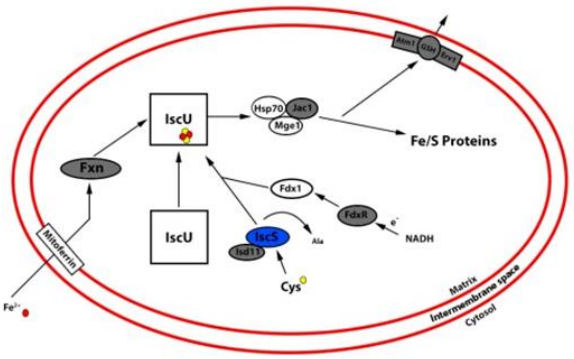
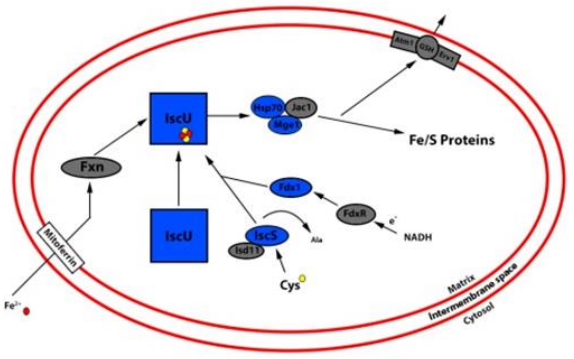
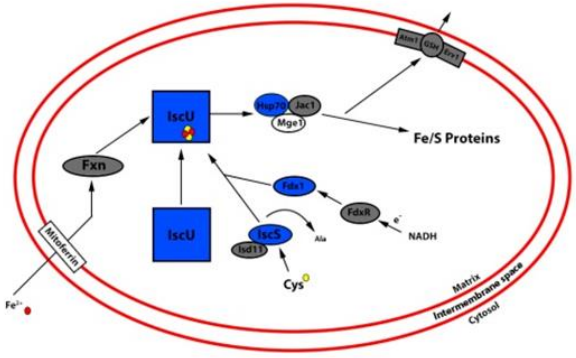
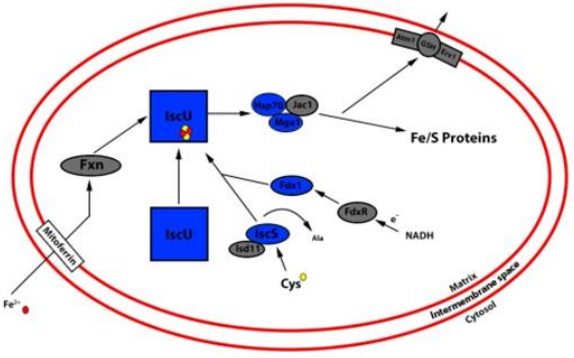
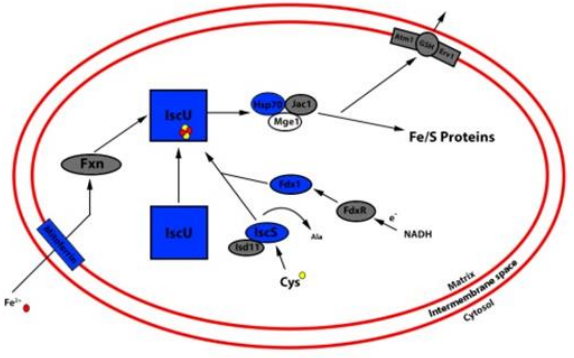
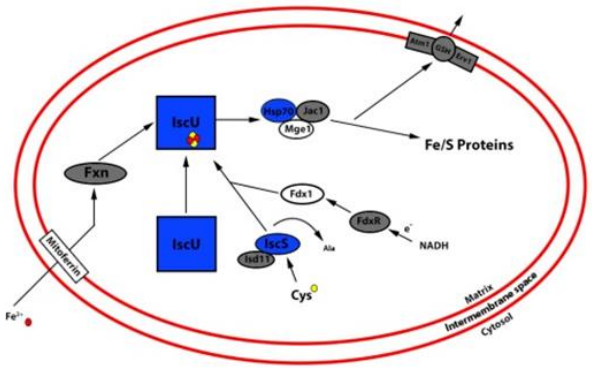
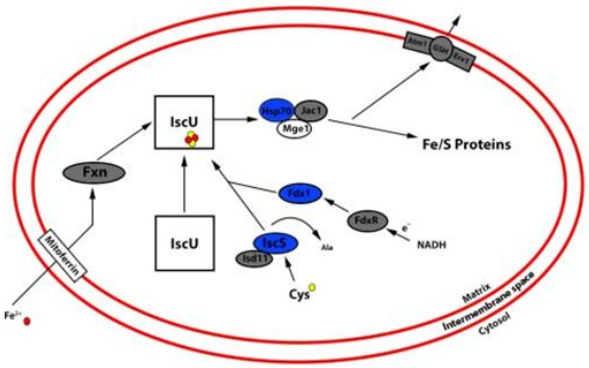
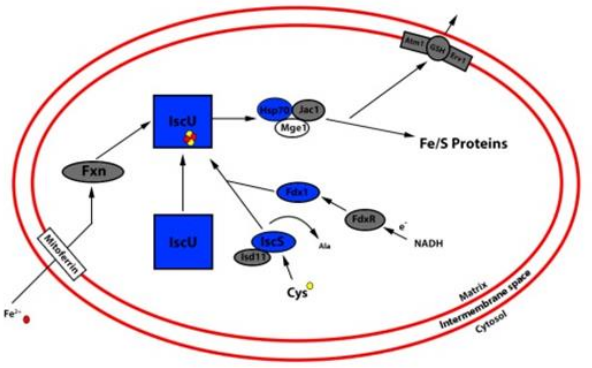


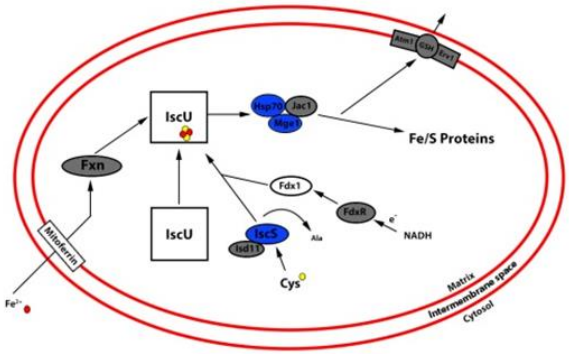
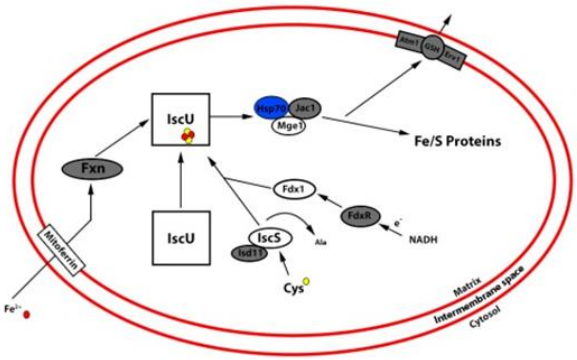
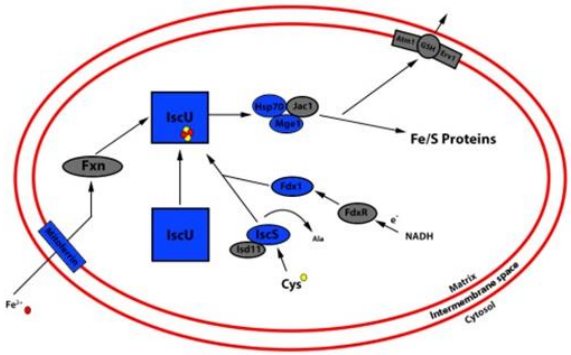
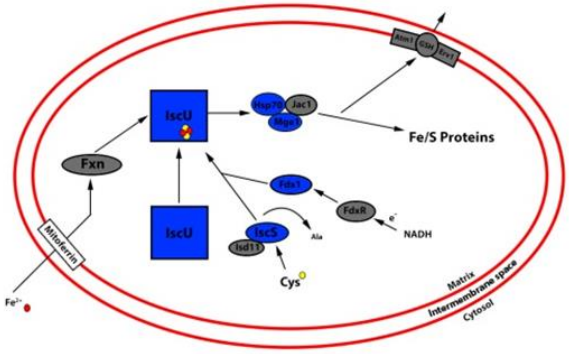
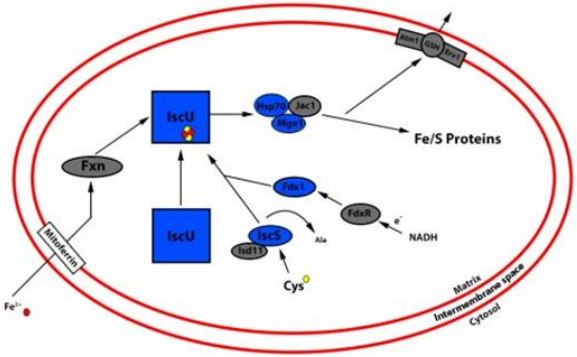
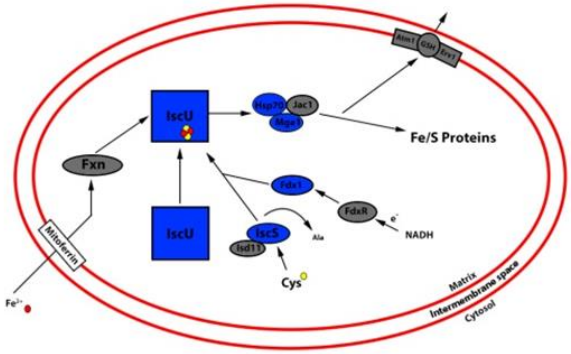
Figure 25 - Protein import proteins identified in each ciliate shown coloured in relation to the import protein complex.

Fe/S Cluster Assembly Figure 26

Anophryoides haemophila	*Aristerostoma marinum*	*Blepharisma japonicum*
		
Climacostomum virens	*Condylostoma magnum*	*Euplotes focardii*
		

Euplotes harpa	*Fabrea salina*	***Ichthyophthirius multifiliis***
		
Litonotus sp.	***Paramecium tetraurelia***	*Protocruzia sp.*
		

Pseudokeronopsis riccii	*Schmidingerella inclinatum*	*Strombidium inclinatum*
		

Brown Ciliate	*Cryptocaryon irritans*	*Mesodinium pulex*
		
Platyophrya macrostoma	*Strombidinopsis acuminatum*	***Tetrahymena thermophila***
		

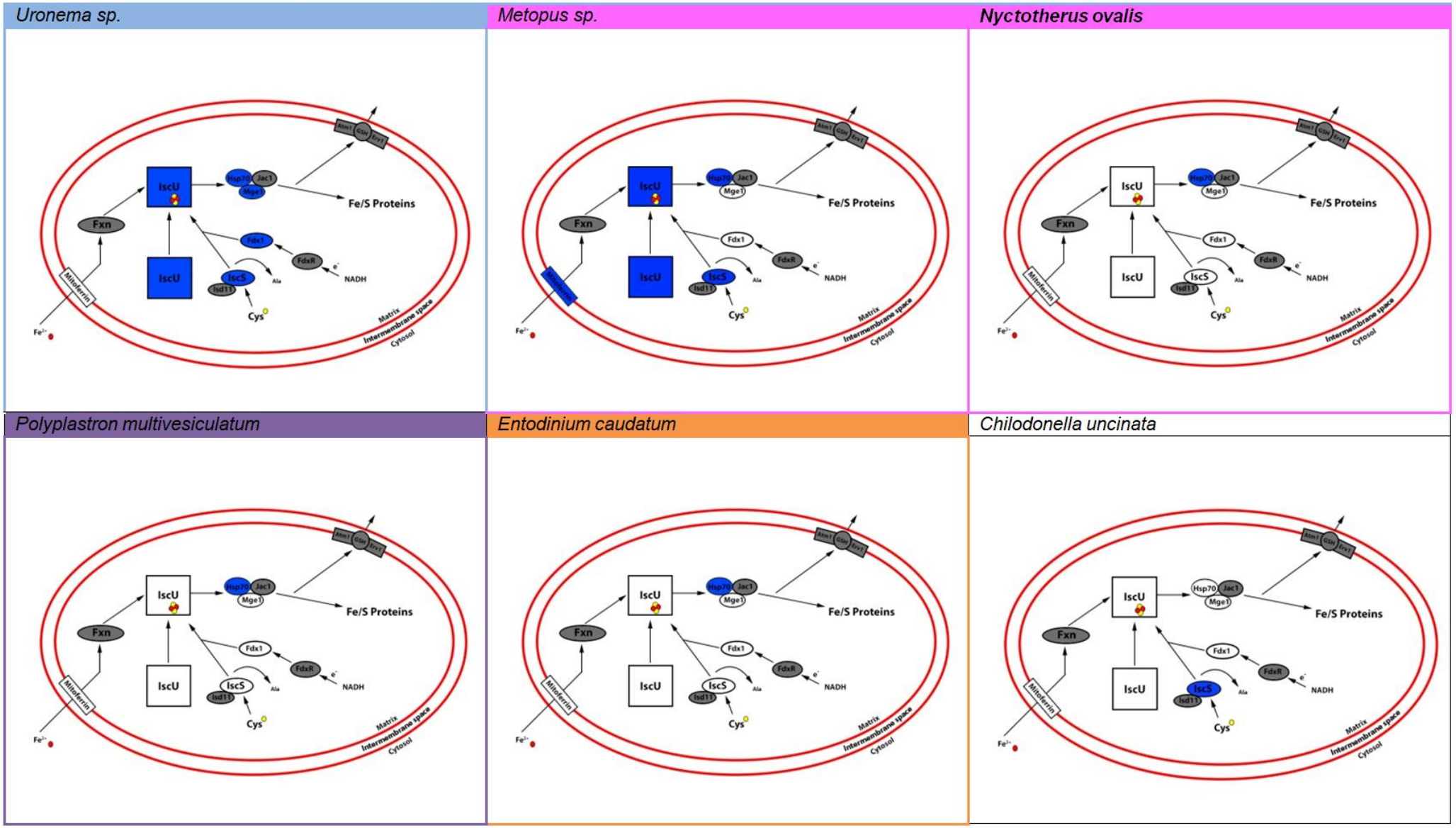
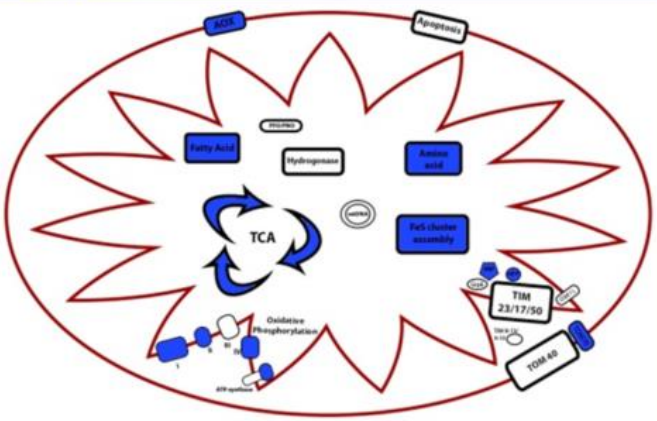
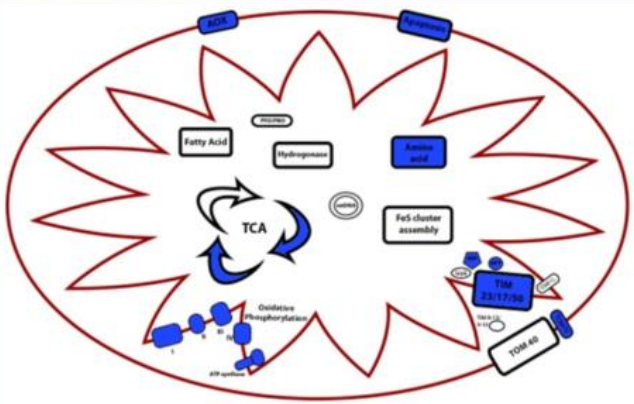
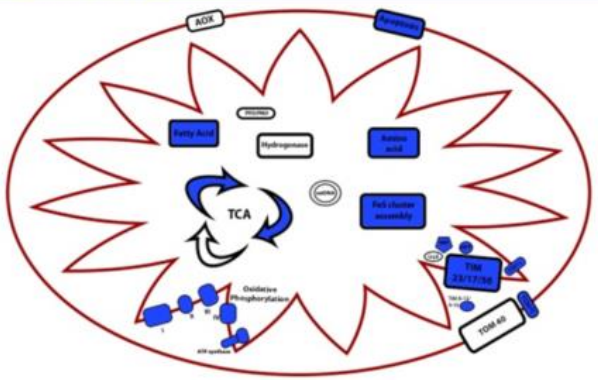
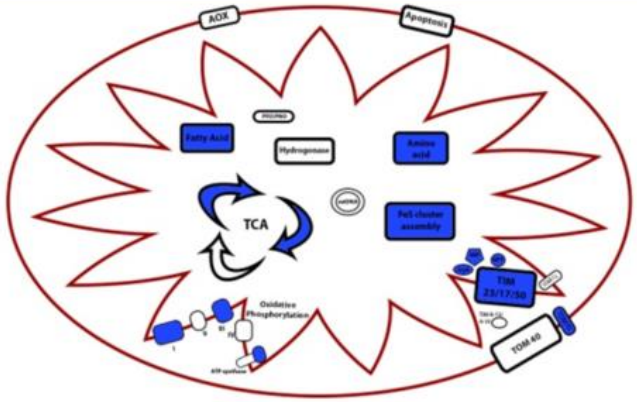
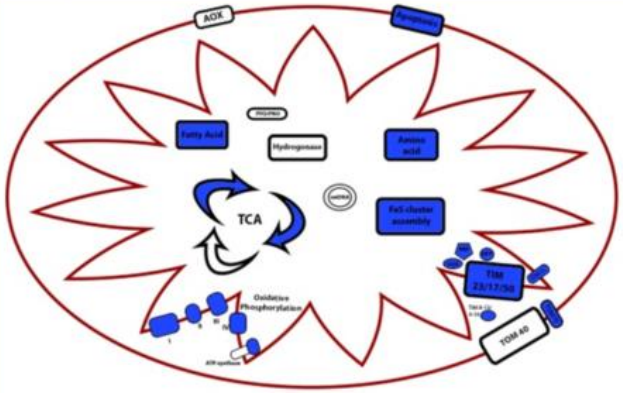
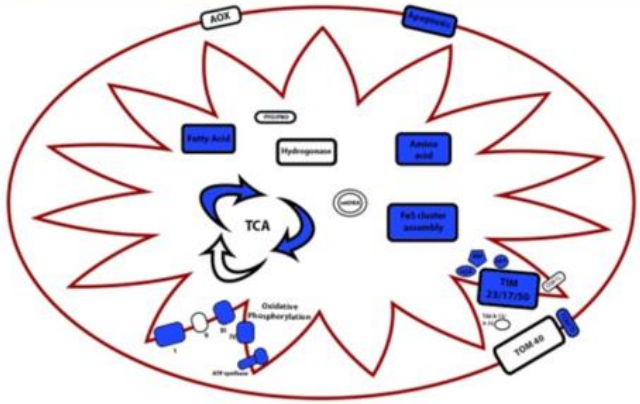
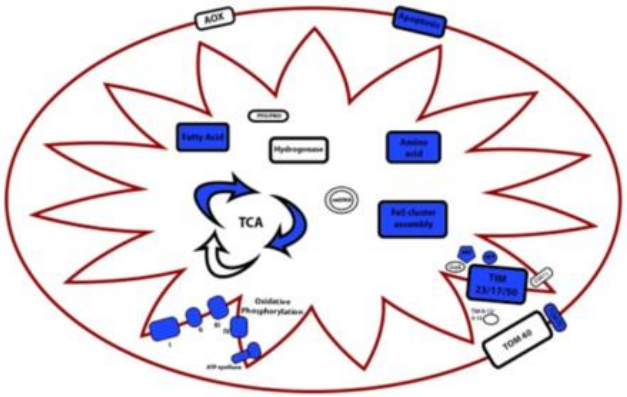


Figure 26 - Fe/S cluster assembly proteins identified in each ciliate shown in blue.

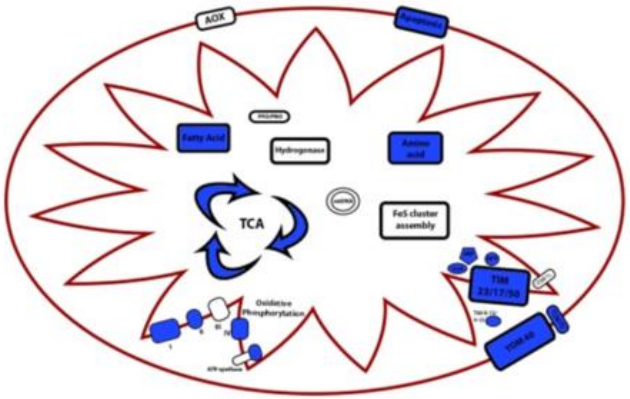
Mitochondrial Pathways Overview Figure 27

Anophryoides haemophila	*Aristerostoma marinum*	*Blepharisma japonicum*
		
Climacostomum virens	*Condylostoma magnum*	*Euplotes focardii*
		

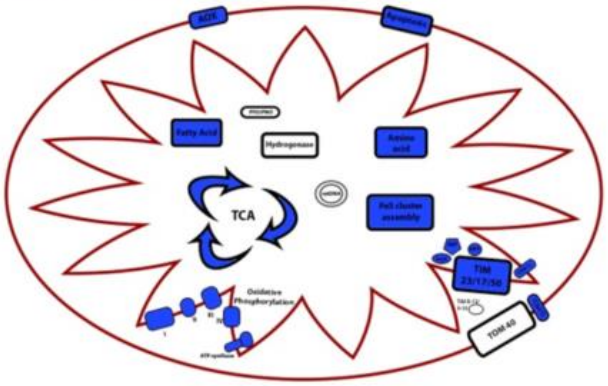
Euplotes harpa



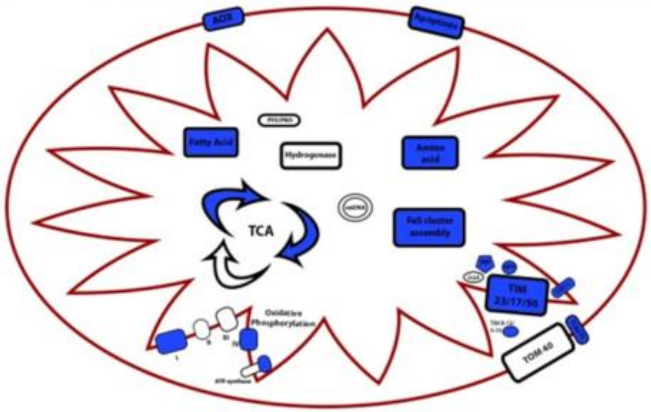
Fabrea salina



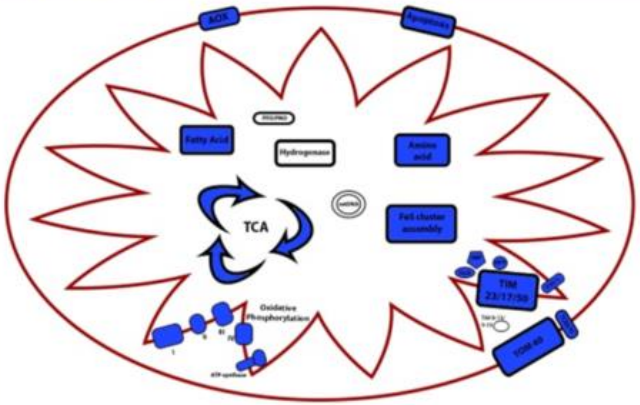
Ichthyophthirius multifiliis



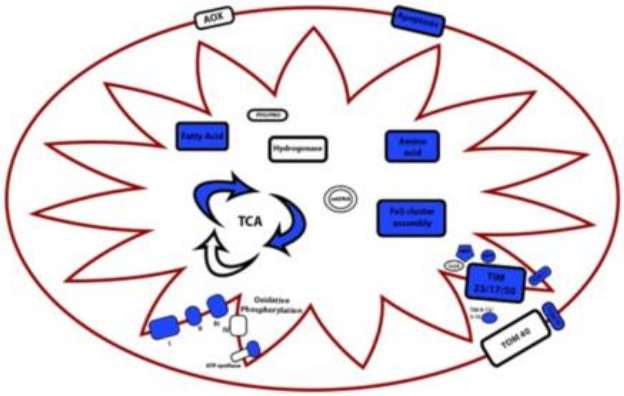
Litonotus sp.



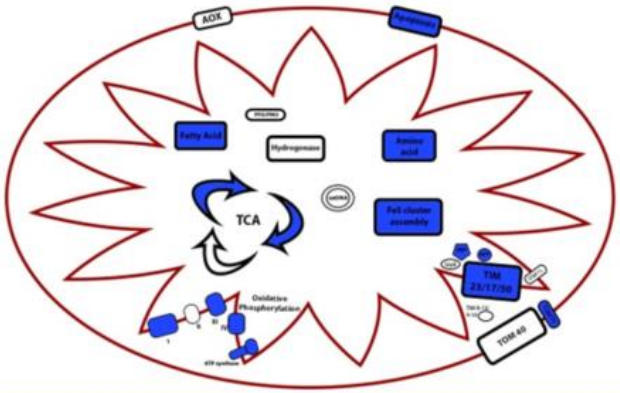
Paramecium tetraurelia



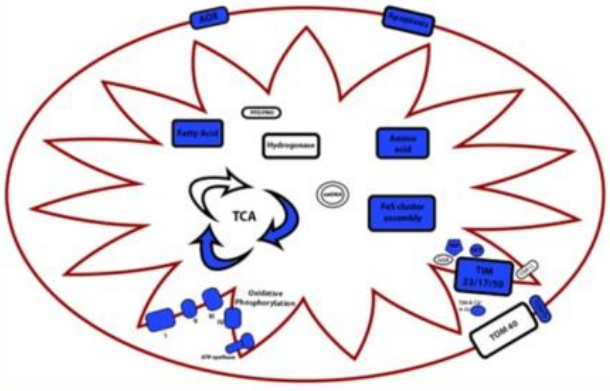
Protocruzia sp.



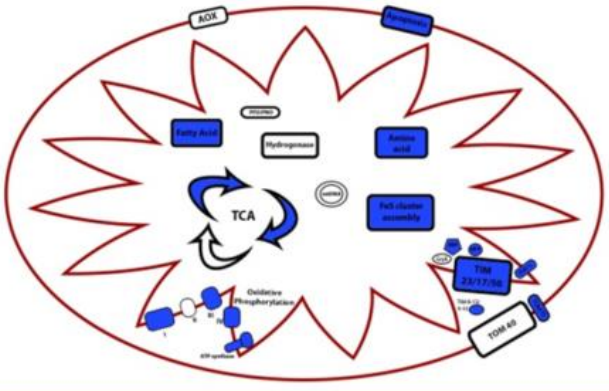
Pseudokeronopsis riccii



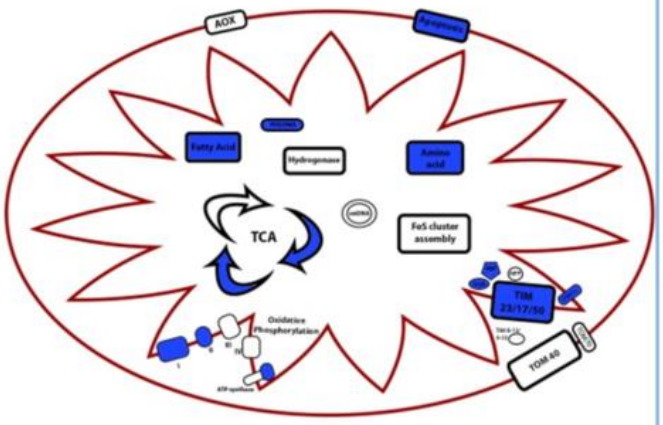
Schmidingerella inclinatum



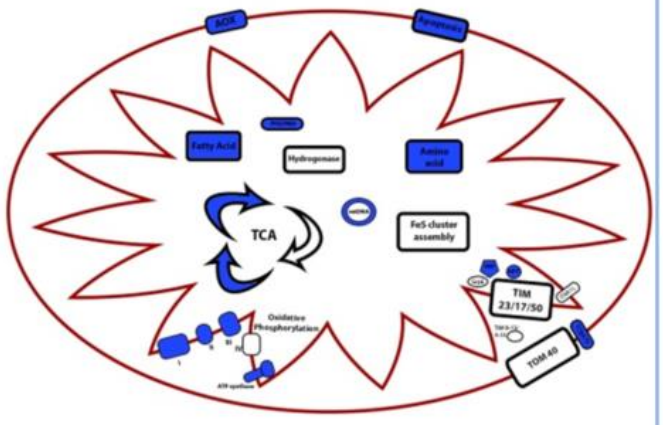
Strombidium inclinatum



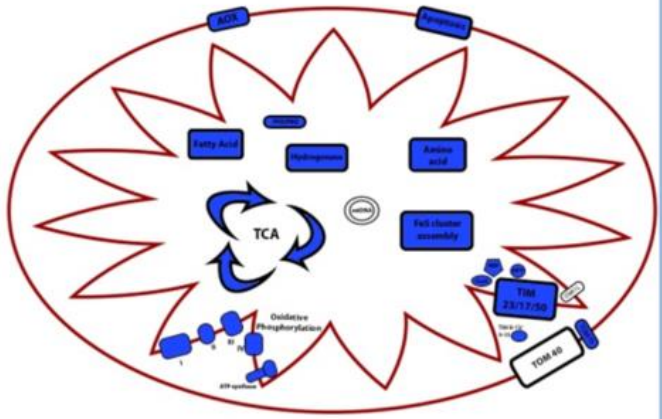
Brown Ciliate



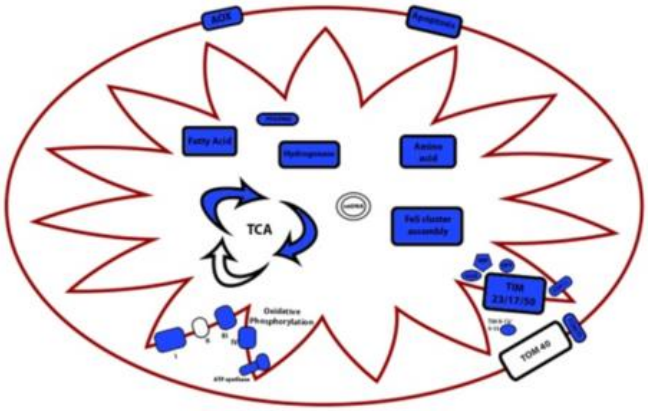
Cryptocaryon irritans



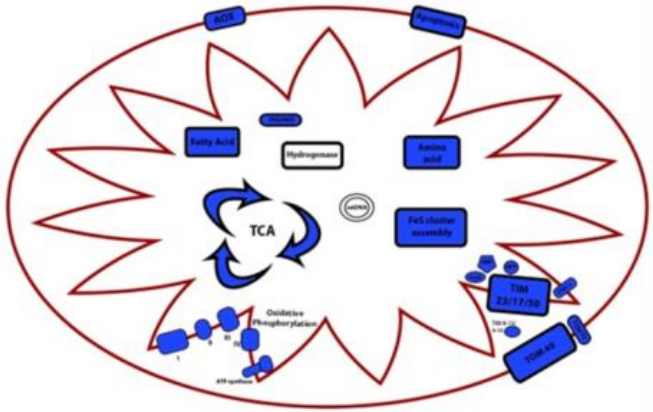
Mesodinium pulex



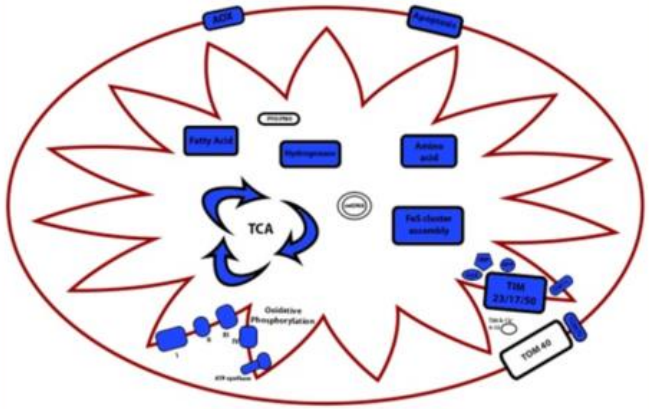
Platyophrya macrostoma



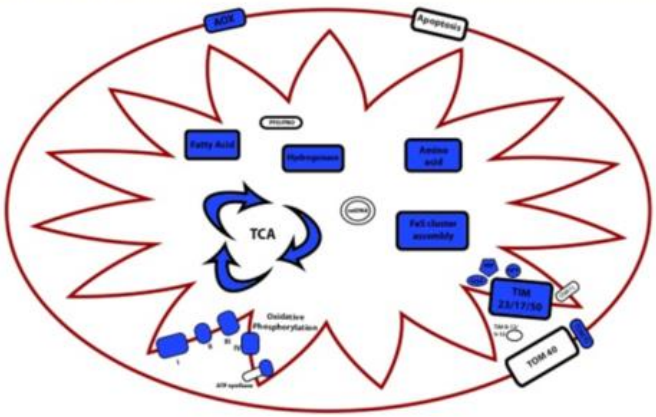
Strombidinopsis acuminatum



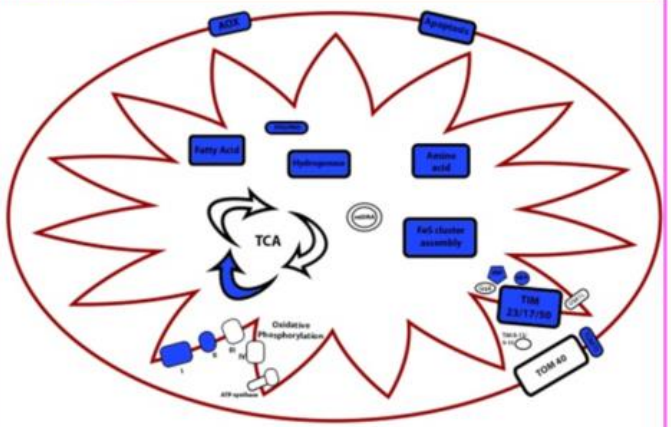
Tetrahymena thermophila



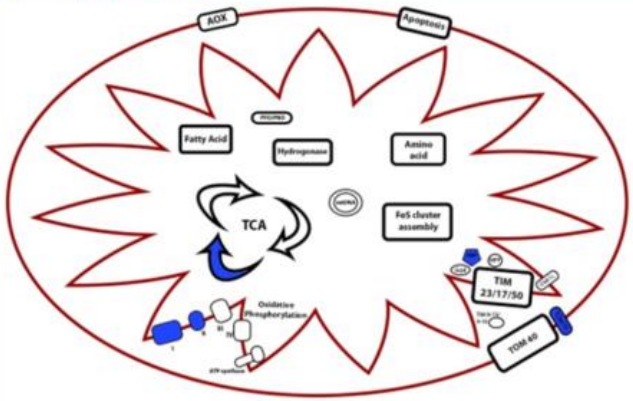
Uronema sp.

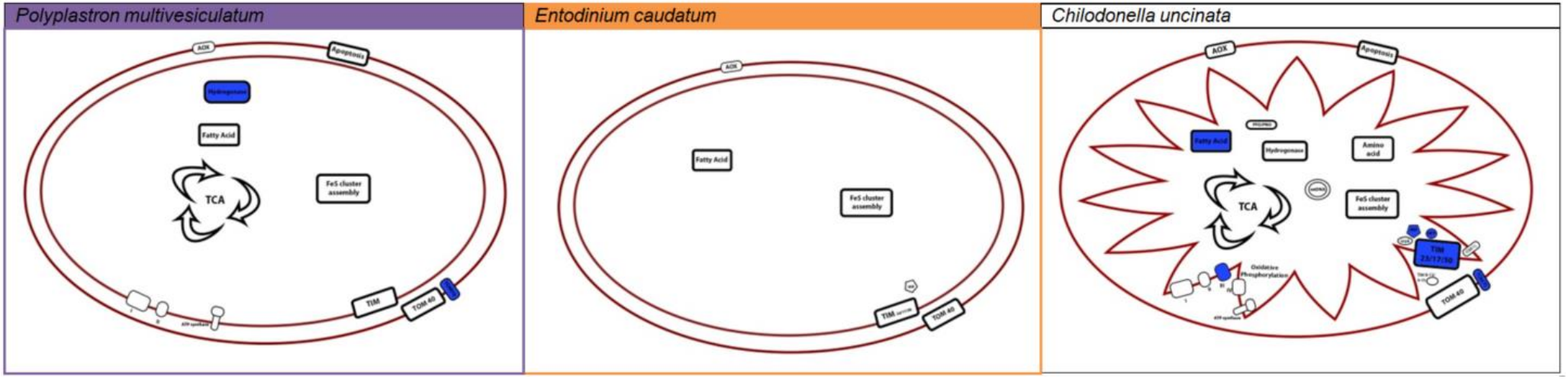


Metopus sp.



Nyctotherus ovalis





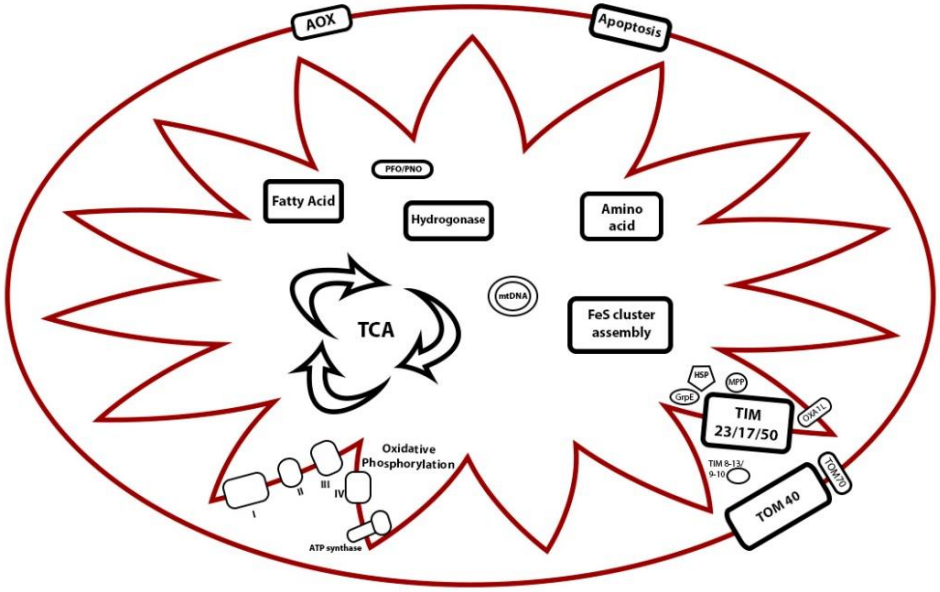


Figure 27 - A basic overview of mitochondrial pathways/proteins identified in each ciliate. Pathways and proteins present are shown in blue.

The final larger figure shows a closer view to allow recognition of the pathways and proteins within in each mitochondrial overview figure.

Pathways/proteins shown are: AOX, apoptosis, fatty acid metabolism, PNO/PFO, hydrogenase, amino acid metabolism, mtDNA, Fe/S cluster assembly, TCA cycle, complex I – V (ATPsynthase) of the electron transport chain, and proteins involved in protein import (TIM, TOM, tiny TIMs, OXA1L, HSP, MPP, GrpE).

Discussion

In this investigation 27 different ciliates were analysed from diverse environments to predict a mitochondrial class dependent on their mitochondrial protein composition. Mitochondria have always been considered vital to organisms that encompass them due to their involvement in energy production and have been seen to adapt and modify their function depending on the environment of their host. It has been observed that there are different functioning mitochondria, including anaerobically hydrogen producing mitochondria like hydrogenosomes and remnant mitochondria known as mitosomes, which have limited involvement in energy production. By analysing and predicting the likely mitochondrial class present within each ciliate, it allowed the proposal of the different possible approaches in which the organism utilises energy and if there are any evolutionary trends, where the predicted class relates to the phylogenetic classification of the ciliates.

To carry out the study, 27 transcriptomic and genomic ciliate data gathered from various ciliates were analysed using computational methods to identify mitochondrial functioning proteins to allow class prediction (Tables 2 and 3). Through doing so, we have been able to predict the class diversity of mitochondria within one ‘group’ of eukaryotes. These predictions have shown clear distinctions between the class as well as potentially suggesting a presence of intermediate organelles that do not fit the already classified groups.

Figure 28 shows a phylogenetic tree generated by Dr. Eleni Gentekaki, of the 27 ciliates, along with the predicted mitochondrial class present in each. The phylogenetic tree could suggest that all classes of mitochondria within ciliates potentially arose from a common ancestor where the ancestral ciliate contained

anaerobically functioning mitochondrion, due to the diverse appearance of anaerobic class II mitochondria seen throughout the phylogenetic tree. The figure also shows diversity of all mitochondria classes; different mitochondria classes are mixed across the tree, with individual branches having a mixture of both class I and II types: for example in the Oligohymenophorea branch and also Litostomatea where *Litonotus sp.*, *Polyplastron multivesiculatum* and *Entodinium caudatum* contain class I, IV and V predicted mitochondria or mitochondrial related organelles (MRO's) respectively. One branch, the Armophorea class, contains *Nyctotherus ovalis* and *Metopus sp*., where both of these ciliates contain anaerobic functioning class III mitochondria. This is the only branch that does not contain any class I aerobic mitochondria, which is expected as Lynn 2004 [94] classified this group 'as ciliates that live in anoxic environments and no longer contain class I mitochondria'.

Entodinium caudatum is predicted containing class V mitosomes due to the lack of mitochondrial functioning proteins identified from the small transcriptomic data. It is possible however, that *Entodinium* contains no mitochondrial organelle whatsoever, similar to *Monocercomonoides*. However, we are unable to prove absence without further investigation and acquiring cultures of *Entodinium* to provide a visual aid under electron microscopy, identifying whether a mitochondrial organelle is present or not. *Entodinium caudatum* are found in the rumen of grazing animals, so it is very possible it could contain an anaerobically function remnant mitochondria similar to *Polyplastron multivesiculatum,* or a mitosome, or even no mitochondrial organelle, as *Monocercomonoides* is found in the intestines of chinchillas [1]. However, *Entodinium* has been shown not to express hydrogenase activity [79, 80], and previous electron microscopy of certain species of rumen ciliates revealed no mitochondrial-shaped organelles [97]. So it is likely

to be either mitosome or no mitochondrial organelle; mitosome has been suggested elsewhere [98].

The environment in which the ciliate resides does not seem to show a trend in relation to the type of mitochondria they encompass, apart from ciliates found within rumen or hindgut environments, where they appear to contain class III, IV or V mitochondria. Ciliates are found in diverse environments and it appears their class of mitochondria is just as diverse as seen in Table 4.

It is possible that there could be further intermediate classes of mitochondria, although further investigation would be required. For the seven (blue) anaerobic functioning mitochondria containing ciliates, their composition of mitochondria functioning proteins varies for electron transport chain, TCA and the significant anaerobic proteins. One of these examples is the Brown Ciliate where it contains a PNO/PFO protein, a reduced electron transport chain and TCA cycle, not too dissimilar to those seen in *Metopus sp.* and *Nyctotherus ovalis*. However, it contains an ATP synthase subunit which would not be expected from class III or IV mitochondria. More so, the lack of iron hydrogenase prevents it being grouped as a class III or IV based on our categorising criteria. A possible alternate is where some ciliates contain both aerobic mitochondria as well as hydrogenosomes, utilising both classes of mitochondria for varied types of respiration to changing environments or organism stresses. Many species have been seen to contain alternative oxidase (AOX) (apart from the Brown Ciliate), providing an alternative reduced route and adaptation to oxygen stress.

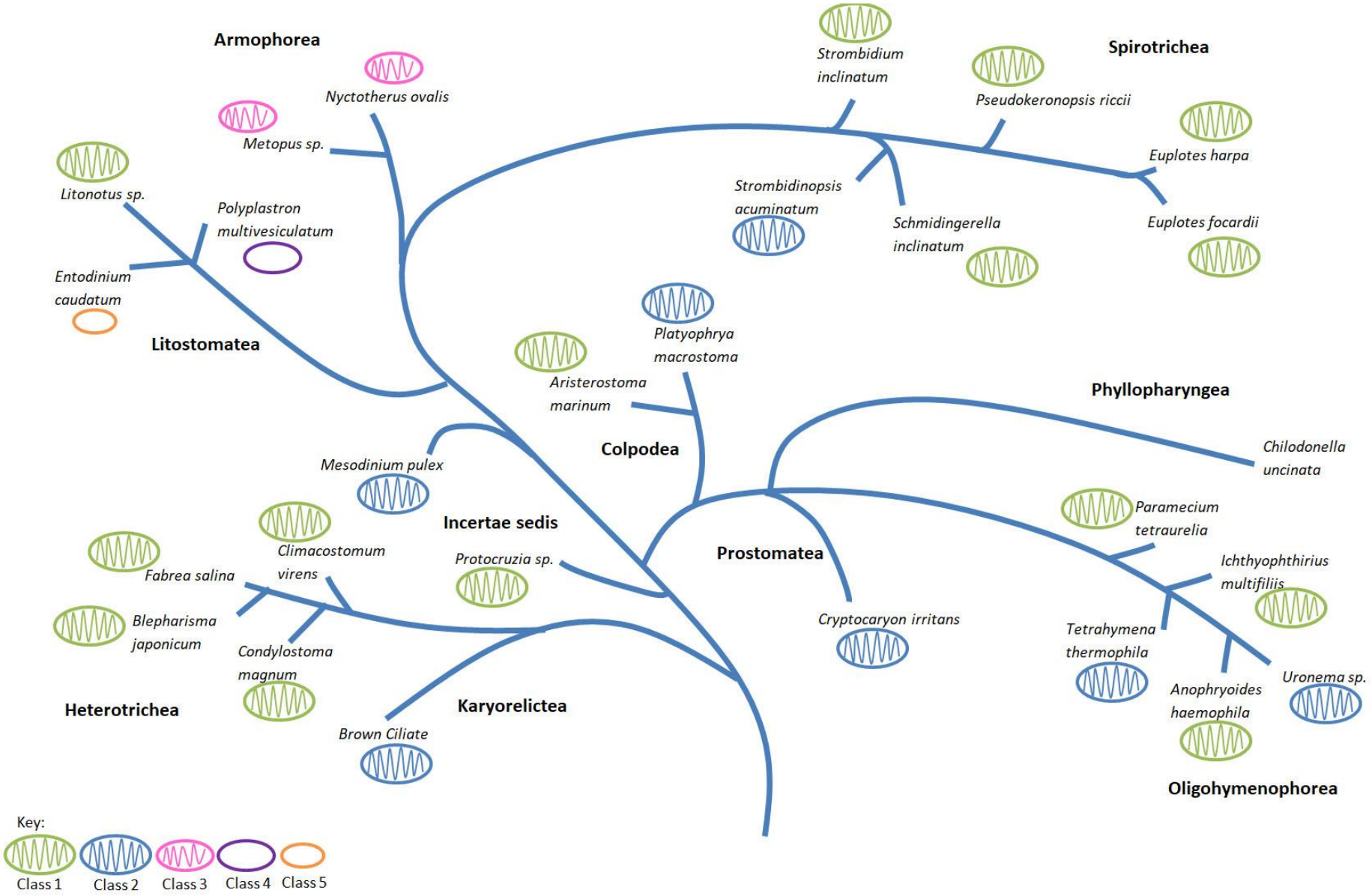


Figure 28 - Phylogenetic tree for the 27 ciliates investigated combined with their predicted class of mitochondria. Assisted by Dr. Eleni Gentekaki at the Mae Fah Luang University.

Alongside identifying five different classes within the ciliate clade, our results also confirmed the discoveries from previous literature where *Polyplastron multivesiculatum* was shown to contain class IV hydrogenosomes, as we were able to identify iron hydrogenase from the transcriptomic data [99]. Complex I and II of the electron transport chain and two TCA cycle steps was also clearly identified within *Nyctotherus ovalis* matching previous discoveries allowing easy conformation. This aided the prediction of *Metopus sp*., containing the same class III mitochondria due to the reduced electron transport chain and TCA cycle (Table 3). *Metopus sp*. did appear to contain contamination, however there are three predicted iron hydrogenase sequences that have similarity to *Nyctotherus ovalis*, which would keep the classification of Class III in relation to our criteria (Figure 23).

Unfortunately the contamination check was carried out as an afterthought, which led to contamination being revealed for some of the anaerobic functioning proteins, which could alter the results slightly. Some sequences even show little similarity to other 'similar proteins of interest', for example *Metopus sp.* PFO1 (Figure 24).

Our computational methodology does come with some limitations that could be prevented with access to ciliate cultures to view under electron magnetic microscope (EM). Also, due to these results being 'protein sequences predictions that resemble the query sequences' as well as only having transcriptomic data, and with some species sequence data being very small in size (lack of sequences), it is not possible to know if a complete mitochondrial species analysis is being viewed, as specific proteins or complexes may not be in the original data. These are some of the disadvantages with using transcriptomic data, as it contains messenger RNA sequences from the whole cell/organism, reflecting the genes that are actively expressed at any given moment, so it is not possible to know exactly if the

transcriptomic data contains likely proteins of interest, as it is possible these proteins were not expressed. The results from the analysis could also vary depending on genetic code and downstream/upstream generation. However, we tried to manage this as effectively as possible by considering all six reading frames and using the correct genetic code for the species. Once genomic data is more readily available for these species, these methods would provide a more consistent prediction for identifying mitochondrial proteins present within the ciliate.

As mentioned, there is currently a lack of cultures for the ciliates we investigated, and having access to cultures to view under EM would allow a visual aid to confirm if mitochondria are present. A prime example of this is for *Chilodonella* uncinata, where due to the lack of data, a clear class prediction is unable to be suggested. However, this method of predicting mitochondrial functioning protein present within sequence data can be very useful for future studies when combined with laboratory work to further support findings. Further investigation should be carried out on the species with low amount of mitochondrial proteins identified, especially *Entodinium caudatum* as no proteins have been identified and it would be interesting to see if this is a remnant mitosome or in fact contains no mitochondrial organelle. However, with our methods we cannot prove absence.

Our study successfully predicted the mitochondria classification for 27 ciliates, identifying all five classes and possible further unclassified intermediates within a single eukaryotic group.

References

[1] A. Karnkowska *et al.*, “A Eukaryote without a Mitochondrial Organelle,” *Curr. Biol.*, vol. 26, no. 10, pp. 1274–1284, May 2016.

[2] B. Boxma *et al.*, “An anaerobic mitochondrion that produces hydrogen.,” *Nature*, vol. 434, no. 7029, pp. 74–9, Mar. 2005.

[3] J. Tovar, A. Fischer, and C. G. Clark, “The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite Entamoeba histolytica.,” *Mol. Microbiol.*, vol. 32, no. 5, pp. 1013–21, Jun. 1999.

[4] M. W. Gray, B. F. Lang, and G. Burger, “Mitochondria of protists.,” *Annu. Rev. Genet.*, vol. 38, pp. 477–524, Jan. 2004.

[5] N. Pfanner and A. Geissler, “Versatility of the mitochondrial protein import machinery.,” *Nat. Rev. Mol. Cell Biol.*, vol. 2, no. 5, pp. 339–49, May 2001.

[6] B. F. Lang, M. W. Gray, and G. Burger, “Mitochondrial Genome Evolution and the Origin of Eukaryotes,” *Annu. Rev. Genet.*, vol. 33, no. 1, pp. 351–397, Dec. 1999.

[7] Z. Wang and M. Wu, “An integrated phylogenomic approach toward pinpointing the origin of mitochondria.,” *Sci. Rep.*, vol. 5, p. 7949, Jan. 2015.

[8] D. Yang, Y. Oyaizu, H. Oyaizu, G. J. Olsen, and C. R. Woese, “Mitochondrial origins.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 82, no. 13, pp. 4443–7, Jul. 1985.

[9] W. Martin and M. Müller, “The hydrogen hypothesis for the first eukaryote.,” *Nature*, vol. 392, no. 6671, pp. 37–41, Mar. 1998.

[10] Z. Wang and M. Wu, “Phylogenomic Reconstruction Indicates Mitochondrial Ancestor Was an Energy Parasite,” *PLoS One*, vol. 9, no. 10, p. e110685, Oct. 2014.

[11] N. Rodríguez-Ezpeleta and T. M. Embley, “The SAR11 Group of Alpha-Proteobacteria Is Not Related to the Origin of Mitochondria,” *PLoS One*, vol. 7, no. 1, p. e30520, Jan. 2012.

[12] T. Cavalier-Smith, “Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium.,” *Proceedings. Biol. Sci.*, vol. 273, no. 1596, pp. 1943–52, Aug. 2006.

[13] S. A. Muñoz-Gómez, J. G. Wideman, A. J. Roger, and C. H. Slamovits, “The origin of mitochondrial cristae from alphaproteobacteria,” *Mol. Biol. Evol.*, vol. 34, no. 4, p. msw298, Jan. 2017.

[14] M. Muller *et al.*, “Biochemistry and evolution of anaerobic energy metabolism in eukaryotes.,” *Microbiol. Mol. Biol. Rev.*, vol. 76, no. 2, pp. 444–95, Jun. 2012.

[15] E. T. Bui, P. J. Bradley, and P. J. Johnson, “A common evolutionary origin for mitochondria and hydrogenosomes.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93,

no. 18, pp. 9651–6, Sep. 1996.

[16] A. Germot, H. Philippe, and H. Le Guyader, “Presence of a mitochondrial-type 70-kDa heat shock protein in Trichomonas vaginalis suggests a very early mitochondrial endosymbiosis in eukaryotes,” *Evolution (N. Y).*, vol. 93, pp. 14614–14617, 1996.

[17] D. S. Horner, R. P. Hirt, S. Kilvington, D. Lloyd, and T. M. Embley, “Molecular Data Suggest an Early Acquisition of the Mitochondrion Endosymbiont,” *Proc. R. Soc. B Biol. Sci.*, vol. 263, no. 1373, pp. 1053–1059, Aug. 1996.

[18] A. J. Roger, C. G. Clark, and W. F. Doolittle, “A possible mitochondrial gene in the early-branching amitochondriate protist Trichomonas vaginalis.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 25, pp. 14618–22, Dec. 1996.

[19] A. V. Goldberg *et al.*, “Localization and functionality of microsporidian iron–sulphur cluster assembly proteins,” *Nature*, vol. 452, no. 7187, pp. 624–628, Apr. 2008.

[20] J. Tovar *et al.*, “Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation.,” *Nature*, vol. 426, no. 6963, pp. 172–6, Nov. 2003.

[21] J. Gerber, U. Mühlenhoff, and R. Lill, “An interaction between frataxin and Isu1/Nfs1 that is crucial for Fe/S cluster synthesis on Isu1,” *EMBO Rep.*, vol. 4, no. 9, pp. 906–911, Aug. 2003.

[22] D. Kessler, “Enzymatic activation of sulfur for incorporation into biomolecules in prokaryotes,” *FEMS Microbiol. Rev.*, vol. 30, no. 6, pp. 825–840, Nov. 2006.

[23] K. Nishio, M. Nakai, K. Nishio, and M. Nakai, “Transfer of Iron-Sulfur Cluster from NifU to Apo-Ferredoxin*,” 2000.

[24] A. C. Adam, C. Bornhövd, H. Prokisch, W. Neupert, and K. Hell, “The Nfs1 interacting protein Isd11 has an essential role in Fe/S cluster biogenesis in mitochondria,” *EMBO J.*, vol. 25, no. 1, pp. 174–183, Jan. 2006.

[25] Y. Takahashi2 and M. Nakamura, “Functional Assignment of the ORF2-iseS-iscU-iscA-hscB-hscA-fdx- ORF3 Gene Cluster Involved in the Assembly of Fe-S Clusters in Escherichia coli1,” *J. Biochem*, vol. 126, pp. 917–926, 1999.

[26] J. Li, S. Saxena, D. Pain, and A. Dancis, “Adrenodoxin Reductase Homolog (Arh1p) of Yeast Mitochondria Required for Iron Homeostasis,” *J. Biol. Chem.*, vol. 276, no. 2, pp. 1503–1509, Jan. 2001.

[27] R. Lill and G. Kispal, “Maturation of cellular Fe-S proteins: an essential function of mitochondria.,” *Trends Biochem. Sci.*, vol. 25, no. 8, pp. 352–6, Aug. 2000.

[28] R. Dutkiewicz, B. Schilke, S. Cheng, H. Knieszner, E. A. Craig, and J. Marszalek, “Sequence-specific Interaction between Mitochondrial Fe-S Scaffold Protein Isu and Hsp70 Ssq1 Is Essential for Their *in Vivo* Function,” *J. Biol. Chem.*, vol. 279, no. 28, pp. 29167–29174, Jul. 2004.

[29] H. Schulz, “Beta oxidation of fatty acids.,” *Biochim. Biophys. Acta*, vol. 1081, no. 2, pp. 109–20, Jan. 1991.

[30] P. Dolezal, V. Likic, J. Tachezy, and T. Lithgow, “Evolution of the Molecular Machines for Protein Import into Mitochondria,” *Science (80-.).*, vol. 313, no. 5785, pp. 314–318, Jul. 2006.

[31] P. J. Bradley, C. J. Lahti, E. Plümper, and P. J. Johnson, “Targeting and translocation of proteins into the hydrogenosome of the protist Trichomonas: similarities with mitochondrial protein import.,” *EMBO J.*, vol. 16, no. 12, pp. 3484–93, Jun. 1997.

[32] J. Brix, S. Rüdiger, B. Bukau, J. Schneider-Mergener, and N. Pfanner, “Distribution of binding sequences for the mitochondrial import receptors Tom20, Tom22, and Tom70 in a presequence-carrying preprotein and a non-cleavable preprotein.,” *J. Biol. Chem.*, vol. 274, no. 23, pp. 16522–30, Jun. 1999.

[33] S. D. Dyall and P. J. Johnson, “Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis.,” *Curr. Opin. Microbiol.*, vol. 3, no. 4, pp. 404–11, Aug. 2000.

[34] M. Van Der Giezen and J. Tovar, “Degenerate mitochondria,” *EMBO Rep.*, vol. 6, no. 6, pp. 525–530, 2005.

[35] P. Rehling, K. Brandner, and N. Pfanner, “Mitochondrial import and the twin-pore translocase,” *Nat. Rev. Mol. Cell Biol.*, vol. 5, no. 7, pp. 519–530, Jul. 2004.

[36] C. Meisinger *et al.*, “Protein Import Channel of the Outer Mitochondrial Membrane: a Highly Stable Tom40-Tom22 Core Structure Differentially Interacts with Preproteins, Small Tom Proteins, and Import Receptors,” *Mol. Cell. Biol.*, vol. 21, no. 7, pp. 2337–2348, Apr. 2001.

[37] C. M. Koehler, “NEW DEVELOPMENTS IN MITOCHONDRIAL ASSEMBLY,” *Annu. Rev. Cell Dev. Biol.*, vol. 20, no. 1, pp. 309–335, Nov. 2004.

[38] W. Neupert and J. M. Herrmann, “Translocation of proteins into mitochondria.,” *Annu. Rev. Biochem.*, vol. 76, pp. 723–49, Jan. 2007.

[39] M. Meinecke *et al.*, “Tim50 Maintains the Permeability Barrier of the Mitochondrial Inner Membrane,” *Science (80-.).*, vol. 312, no. 5779, pp. 1523–1526, Jun. 2006.

[40] A. Chacinska *et al.*, “Mitochondrial Presequence Translocase: Switching between TOM Tethering and Motor Recruitment Involves Tim21 and Tim17,” *Cell*, vol. 120, no. 6, pp. 817–829, Mar. 2005.

[41] D. Mokranjac, D. Popov-[Cbreve]eleketić, K. Hell, and W. Neupert, “Role of Tim21 in Mitochondrial Translocation Contact Sites,” *J. Biol. Chem.*, vol. 280, no. 25, pp. 23437–23440, Jun. 2005.

[42] G. Hawlitschek *et al.*, “Mitochondrial protein import: identification of processing peptidase and of PEP, a processing enhancing protein.,” *Cell*,

vol. 53, no. 5, pp. 795–806, Jun. 1988.

[43] S. D. Dyall and P. Dolezal, “Protein Import into Hydrogenosomes and Mitosomes,” in *Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 21–73.

[44] O. Šmíd *et al.*, “Reductive Evolution of the Mitochondrial Processing Peptidases of the Unicellular Parasites Trichomonas vaginalis and Giardia intestinalis,” *PLoS Pathog.*, vol. 4, no. 12, p. e1000243, Dec. 2008.

[45] T. Makiuchi and T. Nozaki, “Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa.,” *Biochimie*, vol. 100, pp. 3–17, May 2014.

[46] R. M. De Graaf *et al.*, “The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of nyctotherus ovalis,” *Mol. Biol. Evol.*, vol. 28, no. 8, pp. 2379–2391, Aug. 2011.

[47] J. G. Jones, A. D. Sherry, F. M. Jeffrey, C. J. Storey, and C. R. Malloy, “Sources of acetyl-CoA entering the tricarboxylic acid cycle as determined by analysis of succinate 13C isotopomers.,” *Biochemistry*, vol. 32, no. 45, pp. 12240–4, Nov. 1993.

[48] V. Yankovskaya, “Architecture of Succinate Dehydrogenase and Reactive Oxygen Species Generation,” *Science (80-.).*, vol. 299, no. 5607, pp. 700–704, Jan. 2003.

[49] I. Hassinen, “Regulation of Mitochondrial Respiration in Heart Muscle,” in *Mitochondria*, New York, NY: Springer New York, 2007, pp. 3–25.

[50] P. Mitchell, “Protonmotive redox mechanism of the cytochrome b-c1 complex in the respiratory chain: protonmotive ubiquinone cycle.,” *FEBS Lett.*, vol. 56, no. 1, pp. 1–6, Aug. 1975.

[51] D. L. (David L. Nelson, M. M. Cox, and A. L. Lehninger, *Lehninger principles of biochemistry*. .

[52] M. Kobayashi, Y. Matsuo, A. Takimoto, S. Suzuki, F. Maruo, and H. Shoun, “Denitrification, a Novel Type of Respiratory Metabolism in Fungal Mitochondrion*.”

[53] M. Müller, “The hydrogenosome.,” *J. Gen. Microbiol.*, vol. 139, no. 12, pp. 2879–89, Dec. 1993.

[54] A. G. M. Tielens and J. J. Van Hellemond, “The electron transport chain in anaerobically functioning eukaryotes,” *Biochim. Biophys. Acta - Bioenerg.*, vol. 1365, no. 1–2, pp. 71–78, Jun. 1998.

[55] J. J. Van Hellemond, M. Klockiewicz, C. P. Gaasenbeek, M. H. Roos, and A. G. Tielens, “Rhodoquinone and complex II of the electron transport chain in anaerobically functioning eukaryotes.,” *J. Biol. Chem.*, vol. 270, no. 52, pp. 31065–70, Dec. 1995.

[56] A. G. M. Tielens, C. Rotte, J. J. van Hellemond, and W. Martin, “Mitochondria

as we don’t know them.,” *Trends Biochem. Sci.*, vol. 27, no. 11, pp. 564–72, Nov. 2002.

[57] Y. Lantsman, K. S. W. Tan, M. Morada, and N. Yarlett, “Biochemical characterization of a mitochondrial-like organelle from Blastocystis sp. subtype 7,” *Microbiology*, vol. 154, no. 9, pp. 2757–2766, Sep. 2008.

[58] A. Stechmann *et al.*, “Organelles in Blastocystis that Blur the Distinction between Mitochondria and Hydrogenosomes,” *Curr. Biol.*, vol. 18, no. 8, pp. 580–585, Apr. 2008.

[59] B. Boxma *et al.*, “The [FeFe] hydrogenase of Nyctotherus ovalis has a chimeric origin,” *BMC Evol. Biol.*, vol. 7, no. 1, p. 230, 2007.

[60] G. J. Schut and M. W. W. Adams, “The iron-hydrogenase of Thermotoga maritima utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production.,” *J. Bacteriol.*, vol. 191, no. 13, pp. 4451–7, Jul. 2009.

[61] A. D. Tsaousis, E. Gentekaki, L. Eme, D. Gaston, and A. J. Roger, “Evolution of the Cytosolic Iron-Sulfur Cluster Assembly Machinery in Blastocystis Species and Other Microbial Eukaryotes,” *Eukaryot. Cell*, vol. 13, no. 1, pp. 143–153, Jan. 2014.

[62] M. M. Leger, L. Eme, L. A. Hug, and A. J. Roger, “Novel Hydrogenosomes in the Microaerophilic Jakobid Stygiella incarcerata.,” *Mol. Biol. Evol.*, vol. 33, no. 9, pp. 2318–36, Sep. 2016.

[63] E. Nyvltova *et al.*, “NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of Mastigamoeba balamuthi,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 18, pp. 7371–7376, Apr. 2013.

[64] J. Tjaden, I. Haferkamp, B. Boxma, A. G. M. Tielens, M. Huynen, and J. H. P. Hackstein, “A divergent ADP/ATP carrier in the hydrogenosomes of Trichomonas gallinae argues for an independent origin of these organelles.,” *Mol. Microbiol.*, vol. 51, no. 5, pp. 1439–46, Mar. 2004.

[65] M. van der Giezen *et al.*, “Conserved properties of hydrogenosomal and mitochondrial ADP/ATP carriers: a common origin for both organelles.,” *EMBO J.*, vol. 21, no. 4, pp. 572–9, Feb. 2002.

[66] P. L. Jedelský *et al.*, “The Minimal Proteome in the Reduced Mitochondrion of the Parasitic Protist Giardia intestinalis,” *PLoS One*, vol. 6, no. 2, p. e17285, Feb. 2011.

[67] V. Ali, Y. Shigeta, U. Tokumoto, Y. Takahashi, and T. Nozaki, “An Intestinal Parasitic Protist, *Entamoeba histolytica* , Possesses a Non-redundant Nitrogen Fixation-like System for Iron-Sulfur Cluster Assembly under Anaerobic Conditions,” *J. Biol. Chem.*, vol. 279, no. 16, pp. 16863–16874, Apr. 2004.

[68] M. van der Giezen, S. Cox, and J. Tovar, “The iron-sulfur cluster assembly genes iscS and iscU of Entamoeba histolytica were acquired by horizontal

gene transfer.," *BMC Evol. Biol.*, vol. 4, p. 7, Feb. 2004.

[69] F. Mi-ichi, M. Abu Yousuf, K. Nakada-Tsukui, and T. Nozaki, "Mitosomes in Entamoeba histolytica contain a sulfate activation pathway.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 51, pp. 21731–6, Dec. 2009.

[70] F. Mi-ichi, T. Makiuchi, A. Furukawa, D. Sato, and T. Nozaki, "Sulfate activation in mitosomes plays an important role in the proliferation of Entamoeba histolytica.," *PLoS Negl. Trop. Dis.*, vol. 5, no. 8, p. e1263, Aug. 2011.

[71] I. E. Gentle *et al.*, "Conserved Motifs Reveal Details of Ancestry and Structure in the Small TIM Chaperones of the Mitochondrial Intermembrane Space," *Mol. Biol. Evol.*, vol. 24, no. 5, pp. 1149–1160, Feb. 2007.

[72] B. A. P. Williams *et al.*, "A Broad Distribution of the Alternative Oxidase in Microsporidian Parasites," *PLoS Pathog.*, vol. 6, no. 2, p. e1000761, Feb. 2010.

[73] J. O. Corliss, "Why the world needs protists!," *J. Eukaryot. Microbiol.*, vol. 51, no. 1, pp. 8–22.

[74] D. H. Lynn, *The Ciliated Protozoa*. Dordrecht: Springer Netherlands, 2010.

[75] Andre-Denis G. Wright & Denis H. Lynn, "Maximum Ages of Ciliate Lineages Estimated Using a Small Subunit rRNA Molecular Clock: Crown Eukaryotes Date Back to the Paleoproterozoic," 1997.

[76] R. J. Cawthorn *et al.*, *Diseases of aquatic organisms.*, vol. 24, no. 2. Inter-Research Science Center, 1996.

[77] J. M. EADIE and P. N. HOBSON, "Effect of the Presence or Absence of Rumen Ciliate Protozoa on the Total Rumen Bacterial Count in Lambs," *Nature*, vol. 193, no. 4814, pp. 503–505, Feb. 1962.

[78] B. A. DEHORITY, "Rumen Ciliate Protozoa in Ohio White-tailed Deer (Odocoileus virginianus)," *J. Protozool.*, vol. 37, no. 6, pp. 473–475, Nov. 1990.

[79] D. E. Akin and H. E. Amos, "Mode of attack on orchardgrass leaf blades by rumen protozoa.," *Appl. Environ. Microbiol.*, vol. 37, no. 2, pp. 332–8, Feb. 1979.

[80] I. B. Raikov, *Research in Protozoology*. Elsevier, 1972.

[81] D. L. NANNEY, "NUCLEO-CYTOPLASMIC INTERACTION DURING CONJUGATION IN TETRAHYMENA," *Biol. Bull.*, vol. 105, no. 1, pp. 133–148, Aug. 1953.

[82] NCBI Resource Coordinators, "Database Resources of the National Center for Biotechnology Information," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D12–D17, Jan. 2017.

[83] T. Nakamura, K. D. Yamada, K. Tomii, and K. Katoh, "Parallelization of

MAFFT for large-scale multiple sequence alignments.," *Bioinformatics*, vol. 34, no. 14, pp. 2490–2492, Jul. 2018.

[84] O. Chernomor, A. von Haeseler, and B. Q. Minh, "Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices," *Syst. Biol.*, vol. 65, no. 6, pp. 997–1008, Nov. 2016.

[85] E. C. Swart, V. Serra, G. Petroni, and M. Nowacki, "Genetic Codes with No Dedicated Stop Codon: Context-Dependent Translation Termination," *Cell*, vol. 166, no. 3, pp. 691–702, Jul. 2016.

[86] S. M. Heaphy, M. Mariotti, V. N. Gladyshev, J. F. Atkins, and P. V. Baranov, "Novel Ciliate Genetic Code Variants Including the Reassignment of All Three Stop Codons to Sense Codons in *Condylostoma magnum*," *Mol. Biol. Evol.*, vol. 33, no. 11, pp. 2885–2889, Nov. 2016.

[87] P. Stothard, "The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.," *Biotechniques*, vol. 28, no. 6, pp. 1102, 1104, Jun. 2000.

[88] O. Emanuelsson, H. Nielsen, S. Brunak, and G. von Heijne, "Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence," *J. Mol. Biol.*, vol. 300, no. 4, pp. 1005–1016, 2000.

[89] A. Krogh *et al.*, "Predicting transmembrane protein topology with a hidden markov model: application to complete genomes," *Journal of molecular biology*, vol. 305, no. 3. pp. 567–580, 2001.

[90] Y. Moriya, M. Itoh, S. Okuda, A. C. Yoshizawa, and M. Kanehisa, "KAAS: an automatic genome annotation and pathway reconstruction server.," *Nucleic Acids Res.*, vol. 35, no. Web Server issue, pp. W182-5, Jul. 2007.

[91] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes.," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000.

[92] U. Consortium, "UniProt: the universal protein knowledgebase," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D158–D169, Jan. 2017.

[93] S. Matsuda, J.-P. Vert, H. Saigo, N. Ueda, H. Toh, and T. Akutsu, "A novel representation of protein sequences for prediction of subcellular location using support vector machines.," *Protein Sci.*, vol. 14, no. 11, pp. 2804–13, Nov. 2005.

[94] F.-M. Affa'a, D. A. Hickey, M. Struder-Kypke, and D. H. Lynn, "Phylogenetic Position of Species in the Genera Anoplophrya, Plagiotoma, and Nyctotheroides (Phylum Ciliophora), Endosymbiotic Ciliates of Annelids and Anurans," *J. Eukaryot. Microbiol.*, vol. 51, no. 3, pp. 301–306, May 2004.

[95] J. H. P. Hackstein, *Hydrogenosomes and mitosomes : mitochondria of anaerobic eukaryotes*. Springer, 2008.

[96] N. Yarlett, G. S. Coleman, A. G. Williams, and D. Lloyd, "Hydrogenosomes in known species of rumen entodiniomorphid protozoa," *FEMS Microbiol. Lett.*, vol. 21, no. 1, pp. 15–19, Jan. 1984.

[97] A. G. Williams and G. S. Coleman, *The Rumen Protozoa*. New York, NY: Springer New York, 1992.

[98] J. Tachezy, *Hydrogenosomes and mitosomes : mitochondria of anaerobic eukaryotes*. Springer, 2008.

[99] R. G. Paul, A. G. Williams^ ’, and R. D. Butler^, “Hydrogenosomes in the rumen entodiniomorphid ciliate Polyplastron multivesiculatum,” *J. Gen. Microbiol.*, vol. 136, 1990.