Doctoral Dissertation

EVOLUTION OF MITOCHONDRIA AND MITOCHONDRION-RELATED ORGANELLES WITH SPECIAL REFERENCE TO THE FREE-LIVING ANAEROBIC STRAMENOPILE *Cantina marsupialis*

March 2016

Graduate School of Marine Science and Technology
Tokyo University of Marine Science and Technology
Doctoral Course of Applied Marine Bioscience

Fumiya NOGUCHI

Doctoral Dissertation

EVOLUTION OF MITOCHONDRIA AND MITOCHONDRION-RELATED ORGANELLES WITH SPECIAL REFERENCE TO THE FREE-LIVING ANAEROBIC STRAMENOPILE *Cantina marsupialis*

March 2016

Graduate School of Marine Science and Technology
Tokyo University of Marine Science and Technology
Doctoral Course of Applied Marine Bioscience

Fumiya NOGUCHI

Table of contents

Chapter 1

General introduction

The rise of eukaryotes through endosymbiosis was one of critical evolutionary events of life on earth. However, there are many questions surrounding the precise origins of the first eukaryotic common ancestor and of mitochondria.

Mitochondria were established through a single endosymbiotic event at an early stage of eukaryotic evolution, and therefore all known eukaryotes have mitochondria (or their secondary degenerated organelles). This organelle in eukaryotes certainly originated from an alpha-proteobacterium, but its precise origin remains unclear. It was originally proposed that the obligate intracellular parasitic bacterial group Rickettsiales (one order of alpha-proteobacteria) is closely related to the mitochondrial origin (Andersson et al. 1998). However, Thrash et al. (2011) have recently argued that the SAR11 clade (including *Pelagibacter*) other than Rickettsiales shares a common ancestor with mitochondria, but this hypothesis remains controversial (Rodríguez-Ezpeleta and Embley 2012). Furthermore, a host for an endosymbiontic origin of mitochondria (namely, a primitive eukaryotic cell) is also enigmatic, although there are several hypotheses regarding the host cell based on the comparative biochemistry of energy metabolism, such as the hydrogen hypothesis and syntrophy hypothesis (Martin and Müller 1998; López-García and Moreira 1999).

Recent phylogenetic analyses support the proposal that eukaryotes emerged from an archaeal group comprising Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota (the TACK superphylum) (Guy and Ettema 2011). In addition, using metagenomic approaches, Spang et al. (2015) found that the novel archaeal phylum Lokiarchaeota was related to the TACK superphylum, which robustly branches with eukaryotes through phylogenomic analyses and potentially has an expanded repertoire of

eukaryotic signature proteins including actin, small GTPases, and endosomal sorting complexes for transport (ESCRT) complexes. These findings strongly suggest that Lokiarchaeota was tightly associated with the emergence of the eukaryotic cell, but it is not clear whether the organism prior to the acquisition of mitochondria was a Lokiarchaeote itself or a "chimera" of a Lokiarchaeote and an unknown bacterium.

After the acquisition of mitochondria, eukaryotes vastly diversified and adapted to various environments (including hypoxic/anoxic habitats). According to the five-kingdom classification system proposed by Whittaker (1969), which appears in biology textbooks, such diverse eukaryotes are composed of three kingdoms, Animalia, Plantae, and Fungi, all of which are multicellular, along with another single-celled kingdom Protista (Fig. 1-1). However, this classification system does not consider the organismal phylogeny. To solve this problem, eukaryotes are currently classified into five major clades (so-called "supergroup"), Amoebozoa, Opisthokonta, SAR (comprising Rhizaria, alveolates, and stramenopiles), Excavata, and Archaeplastida (including red algae, green algae, land plants, and glaucophytes) mainly based on phylogenetic (or phylogenomic) analyses, although a significant fraction of eukaryotes (such as Cryptophyta, Haptophyta, Centrohelida, Apusozoa, and Breviata) cannot be assigned to any of these major clades (Adl et al. 2012) (Fig. 1-2). Remarkably, in this new classification system, two kingdoms of Whittaker's classification, Animalia and Fungi, are subgroups of Opisthokonta, while Whittaker's Plantae kingdom is part of Archaeplastida. Therefore, it is apparent that the diversity of unicellular eukaryotes (protists) is much greater than that of multicellular eukaryotes, and biochemical, physiological, ecological, and evolutionary approaches to various lineages of protists are important to understand the

early evolution and diversification of eukaryotes.

In this doctoral dissertation, I focused on mitochondria in the major eukaryotic (protistan) group stramenopiles. Stramenopiles (also referred to as heterokonts) are a large lineage of SAR as mentioned above and characterized by the presence of an anterior cilium with tripartite flagellar hairs (mastigonemes) in two opposing rows and mitochondria with tubular cristae (e.g., Adl et al. 2012). This lineage with more than 100,000 known species contains an ecologically diverse assemblage, such as primary producers (e.g., brown algae and diatoms), consumers (e.g., bicosoecids), decomposers (e.g., labyrinthulids), and parasites (e.g., oomycetes and *Blastocystis*). Stramenopiles are also known to include several hypoxic/anoxic specialists with both biochemically and morphologically degenerated mitochondria, so-called mitochondrion-related organelles (MROs), instead of canonical mitochondria. MRO-bearing eukaryotes belonging to Amoebozoa, Excavata, and Microsporidia (a fungal lineage) were previously regarded as mitochondrion-lacking organisms (often referred to as Archezoa), which diverged before the acquisition of mitochondria (Cavalier-Smith 1989). In the late 1990s, mitochondrion-related genes were found in the genome of archezoan organisms (e.g., Clark and Roger 1995; Germot et al. 1996; Hirt et al. 1997), and therefore this hypothesis is not accepted in the field of evolutionary biology at present.

MROs are known to lack the proteins responsible for oxidative phosphorylation (i.e., aerobic respiration) generally found in eukaryotes adapted to hypoxic or anoxic environments, implying that studies of MROs are key to clarifying the mechanisms of environmental adaptation and diversification of eukaryotes. Such “noncanonical” mitochondria are traditionally classified into two forms, i.e., hydrogenosomes and

mitosomes (e.g., Embley and Martin 2006; van der Giezen et al. 2005). However, MROs have been recently discovered from anaerobic organisms across a broad range of eukaryotic lineages, and it has become evident that MROs are extremely diverse in the context of both metabolism and structure, meaning that the traditional simple classification system of MROs mentioned above must be revised. Nevertheless, only MROs in the medically important parasite *Blastocystis* have been intensively studied among stramenopiles.

In this situation, my collaborators successfully isolated the free-living anaerobic species *Cantina marsupialis* (formerly described as *Cafeteria marsupialis*), which represents an independent lineage within the radiation of stramenopiles and possesses organelles with only a few tubular cristae (Yubuki et al. 2015). To elucidate some of the evolutionary mechanisms involved in the adaptation to hypoxic/anoxic environments and diversification associated with such adaptation, not only of stramenopiles but also of eukaryotes as a whole, I investigated the metabolic capacity of MROs in *Cantina* using transcriptome data and compared it with those of previously reported MROs in other anaerobic eukaryotes.

The mitochondrion is a powerful clue to understand the origin and early evolution of eukaryotes, because this organelle must have been tightly associated with the emergence of eukaryotes. The mitochondrial phospholipid cardiolipin is an especially intriguing example. It is known that cardiolipin is synthesized by either of two distinct enzymes, cardiolipin synthase (CLS) with two phospholipase D domains, called CLS_pld, or the other with one CDP-alcohol phosphatidyltransferase domain, called CLS_cap. Recently, Tian et al. (2012) have reported that CLS_cap and CLS_pld are patchily and

complementarily distributed at higher taxonomic levels of eukaryotes. Furthermore, in their study, the eukaryotic CLS_cap homologues were closely related to alpha-proteobacterial homologues, while the eukaryotic CLS_pld homologues did not have a phylogenetic affiliation with any specific bacterial homologues. Based on those findings, Tian et al. (2012) proposed that a primitive eukaryote prior to the acquisition of mitochondria inherited CLS_pld from its ancient bacterial ancestor; and then when the primitive eukaryote evolved into the last common ancester (LECA), endosymbiotic event of an alpha proteobacterium brought in CLS_cap, suggesting that the LECA harbored both CLS_cap and CLS_pld. Eventually, either of the two types of cardiolipin synthase was differentially lost in various eukaryotic lineages.

Considering these scenarios, it is reasonable to assume that the cell prior to the acquisition of mitochondria was a chimera of an archaeon and a bacterium rather than an archaeon alone, although Martin and Müller (1998) argued that an archaeon (methanogen) was a host for the endosymbiotic origin of mitochondria. Tian et al. (2012) showed that stramenopiles have only CLS_cap, but lacked CLS_pld. However, I found that *Cantina* "exceptionally" possesses CLS_pld through the transcriptome analyses mentioned above. Because Tian et al. (2012) surveyed a very limited number of stramenopiles, I thoroughly examined the presence or absence of CLS_cap and CLS_pld in a broad range of stramenopile taxa and revisited the evolutionary pattern of these enzymes in the stramenopile lineage as well as the hypothesis regarding the eukaryotic evolution proposed by Tian et al. (2012).

References

Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EA, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick RS, Schoch CL, Smirnov A, Spiegel FW (2012) The revised classification of eukaryotes. *J. Eukaryot. Microbiol* **59**:429–493

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

Burki F (2014) The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb Perspect Biol* **6**:a016147

Cavalier-Smith T (1989) Archaebacteria and archezoa. *Nature* **339**:100–101

Clark CG, Roger AJ (1995) Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica, Proc. Natl. Acad. Sci. USA* **92**:6518–6521

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

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

Guy L & Ettema TJ (2011) Thearchaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol* **19**:580–587

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

Martin W, Muller M (1998) The hydrogen hypothesis for the first eukaryote. *Nature* **392**:37–41

Rodriguez-Ezpeleta N, Embley TM (2012) The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria. *PLos ONE* **7**:e30520

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

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

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

van der Giezen M, Tovar J, Clark CG (2005) Mitochondrion-derived organelles in protists and fungi. *Int Rev Cytol* **244**:175–225

Whittaker RH (1969) New concepts of kingdoms of organisms. *Science* **163**:150–160

Yubuki N, Pánek T, Yabuki A, Čepička I, Takishita K, Inagaki Y, Leander BS（2015) Morphological identities of two different marine stramenopile environmental

sequence clades: *Bicosoeca kenaiensis* (Hilliard, 1971) and *Cantina marsupialis* (Larsen and Patterson, 1990) gen. nov., comb. nov. *J. Eukaryot. Microbiol* **62**:532–542

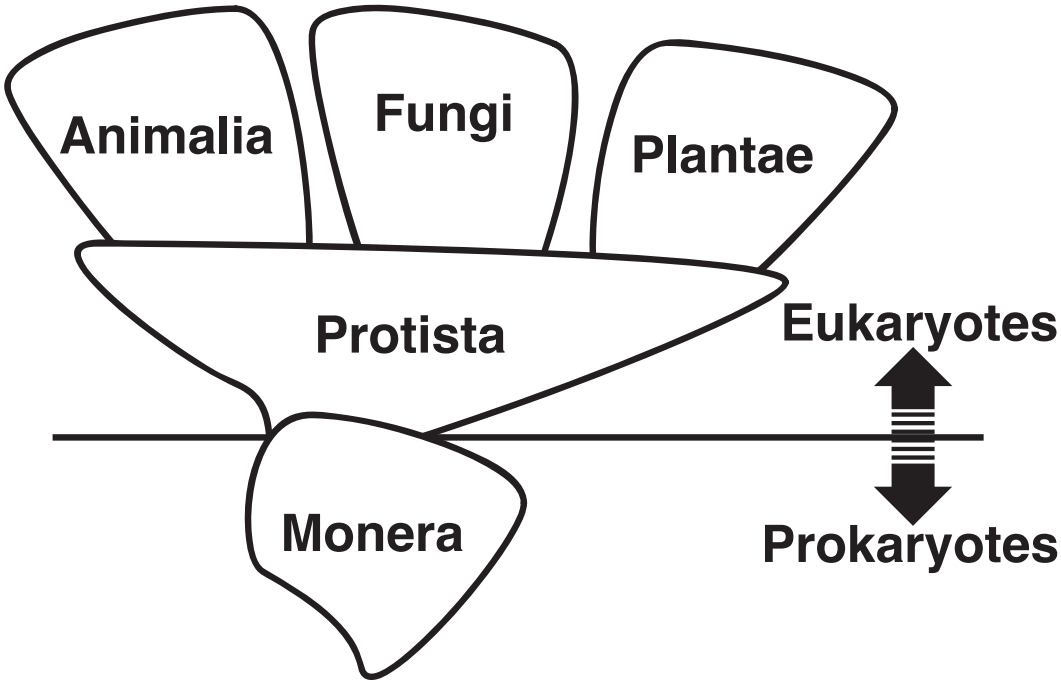

Fig.1-1. The five-kingdom classification proposed by Whittaker (1969). All living creatures are divided into five kingdoms based on nutrient systems: Animalia; Fungi; Plantae; Protista; and Monera. All members of the kingdom Animalia are multicellular heterotrophic eukaryotes (mammals, reptiles, fish, etc.). The kingdom Plantae is composed of multicellular, photosynthetic (autrophic) terrestrial organisms. The kingdom Fungi is characterized by heterotrophic eukaryotes that obtain their nutrients by absorption. The kingdom Protista consists mainly of unicellular eukaryotic organisms that cannot be classified as Plantae, Animalia, or Fungi. Prokaryotic organisms (eubacteria and archaebacteria) belong to the kingdom Monera.

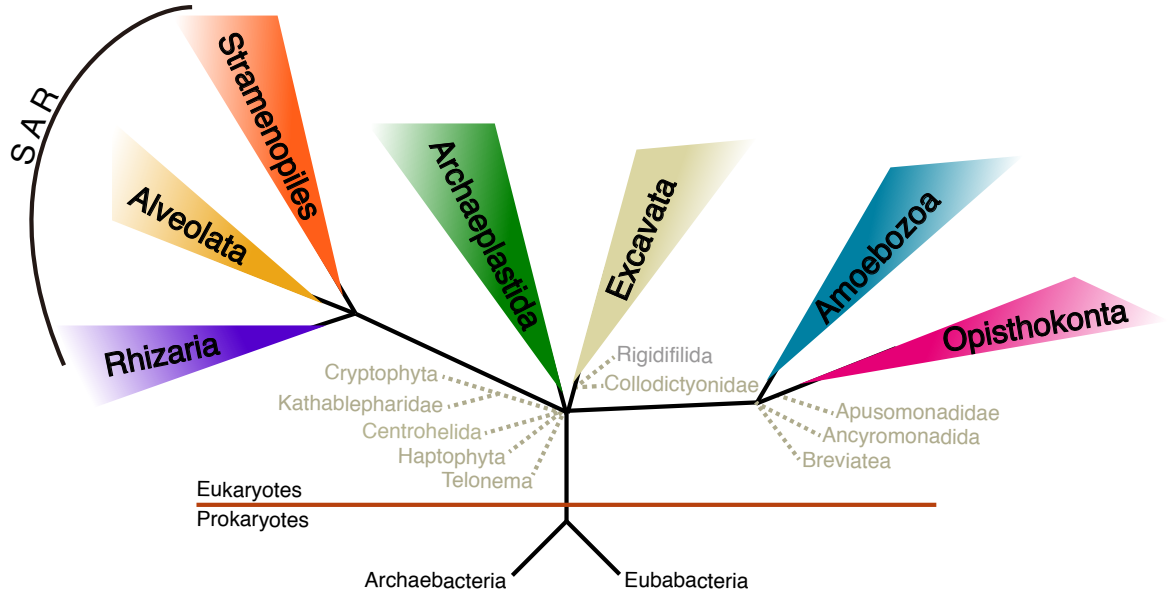

Fig.1-2. The eukaryotic tree inferred by Adl et al. (2012) and Burki (2014). All eukaryotes are classified into the 5 supergroups Opisthokonta, Amoebozoa, Excavata, Archaeplastida, and SAR (comprised of stramenopiles, Alveolata, and Rhizaria) as well as several other lineages (e.g., Cryptophyta, Breviatea, Collodictyonidae). Dotted lines indicate uncertain relationships, including conflicting research findings.

Chapter 2

Metabolic capacity of mitochondrion-related organelles in the free-living anaerobic stramenopile *Cantina marsupialis*

Abstract

Functionally and morphologically degenerate mitochondria, so-called mitochondrion-related organelles (MROs), are frequently found in eukaryotes inhabiting hypoxic or anoxic environments. In the last decade, MROs have been discovered from a phylogenetically broad range of eukaryotic lineages and these organelles have been revealed to possess diverse metabolic capacities. In this study, the biochemical characteristics of an MRO in the free-living anaerobic protist *Cantina marsupialis*, which represents an independent lineage in stramenopiles, were inferred based on RNA-seq data. I found transcripts for proteins known to function in one form of MROs, the hydrogenosome, such as pyruvate:ferredoxin oxidoreductase, iron-hydrogenase, acetate:succinate CoA-transferase, and succinyl-CoA synthase, along with transcripts for acetyl-CoA synthetase (ADP-forming). These proteins possess putative mitochondrial targeting signals at their N-termini, suggesting dual ATP generation systems through anaerobic pyruvate metabolism in *Cantina* MROs. In addition, MROs in *Cantina* were also shown to share several features with canonical mitochondria, including amino acid metabolism and an "incomplete" tricarboxylic acid cycle. Transcripts for all four subunits of complex II (CII) of the electron transport chain were detected, while there was no evidence for the presence of complexes I, III, IV, or F1Fo-ATPase. *Cantina* MRO biochemistry challenges the categories of mitochondrial organelles recently proposed.

2-1. Introduction

Mitochondria are organelles that arose from an α-proteobacterium through endosymbiosis and are ubiquitously found in eukaryotic cells (e.g., Andersson et al. 1998; Embley and Martin 2006; Gray et al. 1999). It is widely held that mitochondria efficiently yield adenosine triphosphate (ATP) via oxidative phosphorylation under aerobic conditions. However, eukaryotes, most of which are protists, adapting to hypoxic oranoxic environments often harbor biochemically reduced mitochondria lacking proteins responsible for oxidative phosphorylation, along with metabolic pathways of the tricarboxylic acid (TCA) cycle, amino acid metabolism, and/or fatty acid oxidation. In general, these organelles are also morphologically degenerate. Such "non-canonical" organelles of mitochondrial origin are collectively called mitochondrion-related organelles (MROs) and are commonly classified into two forms: i.e., hydrogenosomes and mitosomes (e.g., Embley and Martin 2006; van der Giezen et al. 2005). MROs traditionally called "hydrogenosomes" in trichomonads are double membrane-bound, cristae-lacking organelles with no genome. Using enzymes involved in pyruvate metabolism, some of which are not present in canonical mitochondria, such as pyruvate:ferredoxin oxidoreductase (PFO) andiron-only hydrogenase ([Fe]-Hyd), this type of organelle produces ATP via substrate-level phosphorylation with molecular hydrogen as one of the waste products (Lindmark and Müller 1973). Mitosomes are smaller and more biochemically reduced than hydrogenosomes: it is unlikely that mitosomes participate in ATP synthesis or produce molecular hydrogen, and their function is not yet fully understood. So far, iron-sulfur cluster assembly systems have been discovered in

mitosomes of the diplomonad *Giardia intestinalis* and the microsporidian *Encephalitozoon cuniculi* (Goldberg et al. 2008; Tovar et al. 1999, 2003). On the other hand, mitosomes of the parasitic amoeba *Entamoeba histolytica* are primarily involved in sulfate activation (Mi-ichi et al. 2009, 2011).

During the past 10 years, MROs have been discovered from anaerobic organisms across a phylogenetically broad range of eukaryotic lineages. It has become evident that MROs are unexpectedly diverse in the context of both biochemistry and structure. For example, the commensal ciliate in the cockroach hindgut *Nyctotherus ovalis* and the parasitic stramenopiles *Blastocystis* spp. were shown to possess MROs with hydrogenosome-like characteristics (i.e., anaerobicmetabolism producing ATP and molecular hydrogen), along with canonical mitochondrion-like characteristics, including the presence of an "incomplete" electron transport chain (ETC), an "incomplete" TCA cycle, an organelle genome, and cristae structure (Boxma et al. 2005; de Graaf et al. 2011; Denoeud et al. 2011; Pérez-Brocal and Clark 2008; Stechmann et al. 2008). Furthermore, it was reported that the free-living anaerobic protists *Mastigamoeba balamuthi* and *Pygsuia biforma* also harbor MROs with both hydrogenosome and mitochondrion-like characteristics, although their organelles likely lost their own genome DNA and obvious cristae structure (Gill et al. 2007; Nývltová et al. 2015; Stairs et al. 2014). Considering such diversity of MROs, it is apparently difficult to simply classify MROs into the two types of hydrogenosomes and mitosomes. Accordingly, Müller et al. (2012) proposed five classes of the mitochondrial family of organelles based on functional lines. Class 1 is a canonical mitochondrion using oxygen as a terminal electron acceptor. Class 2 is an anaerobically functioning organelle capable of using an endogenously produced electron

acceptor (fumarate in many cases), instead of oxygen. Class 3 is a hydrogen-producing mitochondrion with [Fe]-Hyd that allows it to use protons as terminal electron acceptors. This class of organelle possesses a proton-pumping ETC but lacks F1Fo-ATPase. MROs in *Nyctotherus* and *Blastocystis* typically represent this class. Class 4 is a hydrogenosome, which also generates ATP via hydrogen-producing fermentation, as in the case of class 3, but completely lacks a membrane-associated ETC and a genome. Class 5 is a mitosome not involved in ATP synthesis, as mentioned above.

Stramenopiles are a large monophyletic group of eukaryotes comprising an ecologically diverse assemblage, such as primary producers (e.g., brown algae and diatoms), consumers (e.g., bicosoecids), decomposers (e.g., labyrinthulids), and parasites (e.g., *Blastocystis)* (Adl et al. 2012). This group is also known to include hypoxic/anoxic specialists with MROs instead of canonical mitochondria, such as the commensal *Proteromonas lacerate* and the free-living flagellate *Rictus lutensis* (Leipe et al. 1996; Yubuki et al. 2010), in addition to *Blastocystis,* although MROs in the two former organisms have not been intensively studied. On the other hand, environmental (culture-independent) PCR surveys of 18S ribosomal RNA (rRNA) gene sequences from various environments revealed numerous uncultured stramenopile lineages without phylogenetic affinity to any reported stramenopile lineages (e.g., Massana et al. 2014), indicating that my knowledge on the diversity of this eukaryotic group remains insufficient. Very recently, one of the sequences recovered in environmental PCR surveys of anoxic sediments (designated as NAMAKO-31 in Takishita et al. 2007), which represents an independent lineage within the radiation of stramenopiles, was shown to be derived from the free-living anaerobic species *Cantina marsupialis* (formerly described as *Cafeteria*

marsupialis), through cell isolation/cultivation (Yubuki et al. 2015). This anaerobic protist has organelles with only a few tubular cristae, suggesting that these organelles (MROs) are biochemically degenerate to some extent. However, there is no information on MROs in *C. marsupialis* at the molecular/biochemical levels.

In this chapter, I characterized the metabolic capacity of *Cantina* MROs based on RNA-seq data. As a result, *Cantina* MROs were shown to harbor both hydrogenosome- and mitochondrion-like characteristics, as those of the anaerobic protists *Nyctotherus*, *Blastocystis*, *Mastigamoeba*, and *Pygsuia*. It was demonstrated or suggested that MROs in several eukaryotes tolerant or adaptive to oxygen-depleted environments possess rhodoquinone (RQ) with low redox potential and reduce fumarate to succinate with complex II (CII) of ETC as fumarate reductase (FRD) using electrons transferred via RQ (Müller et al. 2012; Stairs et al. 2014). Markedly, unlike in the case of such RQ-bearing eukaryotes, the enzyme tightly associated with fumarate respiration (reduction), fumarate hydratase, along with RQ, was not found in *Cantina*. *Cantina* MROs possess CII based on RNA-seq data, and therefore this ETC complex may function as succinate dehydrogenase rather than as FRD. On the basis of the unique metabolism of *Cantina* MROs, I also reconsider the criteria for the functional classification of organelles of mitochondrial origin proposed by Müller et al. (2012) and discuss evolutionary events associated with mitochondria that could have occurred during the course of adaptation to oxygen-depleted environments.

2-2. Materials and methods

2-2-1. Cultures

Cells of *Cantina marsupialis* strain YPF1205 were an aerobically maintained in 3.0% Lysogeny Broth (LB) mediumin artificial seawater at 20◦C with bacterial prey. For analyzing quinone composition, the bacterial prey cells were isolated from the original culture of *Cantina* with a capillary pipette, and a bacterial culture without *Cantina* was established using the same medium. The absence of the *Cantina* cells in the bacterial prey culture was confirmed by PCR using universal primers of eukaryotic 18S rRNA gene, 18S-42F and 18S-1520R (López-García et al. 2003) (Table 2-3).

2-2-2. Analyses of RNA-seq data

Cells of the original *Cantina* culture were harvested by centrifugation at 4,170 g for 60 minutes at 4◦C. Total RNA was isolated from the harvested cells using TRIzol reagent (Thermo Fisher Scientific USA). Construction of cDNA libraries and paired-end sequencing with illumina HiSeq2000 (100 bp per read) were performed by Eurofins Genomics. Raw sequencing reads were deposited in the Sequence Read Archive under the accession number DRA003063. 2.27 billion raw sequence read data were filtered using TRIMMOMATIC software version 0.30 (Lohse et al. 2012) to remove adapter sequences and low-quality bases. Filtered sequences were then assembled into 80,091 transcript contigs using the TRINITY package (release 2013-02-25) (Grabherr et al. 2011). The 80,091 contigs were subjected to blastx against the non-redundant (nr) database at NCBI and CBOrg (containing a database of mitochondrial/hydrogenosomal proteomes) (Gaston et al. 2009). Contigs that had hits to the entries in both the nr (with an e-value of less than 0.001) and the CBOrg (with no threshold) databases were retrieved. Among these retrieved contigs, those for which the top hit in the nr database was a eukaryotic homologue were

considered as genes encoding proteins putatively operating in mitochondrion-related organelles (MROs) of *Cantina*. Two *in silico* methods were used for predicting N-terminal mitochondrial targeting signals (MTSs) of these putative MRO proteins, MitoProt (Claros and Vincens 1996) and TargetP (Emanuelsson et al. 2000). I here provisionally defined the N-terminal peptides with Mitoprot or TargetP prediction scores (>0.5) as MTSs. In addition, several known mitochondrial/MRO proteins, such as the mitochondrial pyruvate carrier (Bricker et al. 2012), putative methyltransferase RQUA (Stairs et al. 2014), ADP-forming acetyl-CoA synthetase (ACS) (Fritz-Laylin et al. 2010; Nývltová et al. 2015), pyruvate:NADP+ oxidoreductase (Rotte et al. 2001; Stechmann et al. 2008) and alternative oxidase (McDonald and Vanlerberghe 2006) were independently searched with tblastn against the *Cantina* RNA-seq data using reported eukaryotic homologues as queries.

2-2-3. Phylogenetic analyses

I conducted phylogenetic analyses of five enzymes, ACS (ADP-forming), two types of acetate:succinate CoA-transferase (ASCT1B and ASCT1C), iron-only hydrogenase ([Fe]-Hyd), and pyruvate:ferredoxin oxidoreductase (PFO), which are involved in anaerobic pyruvate metabolism and are frequently found in MROs. The deduced amino acid sequences of these five enzymes from *Cantina* obtained in this study were separately aligned with the corresponding sequences from phylogenetically diverse organisms using MAFFT v7.037b (Katoh and Standley 2013). The alignments were inspected by eye and manually edited. I then excluded ambiguously aligned sites from the datasets prior to phylogenetic analyses. The analyzed datasets had the following

dimensions: ACS (ADP forming): 63 taxa, 658 sites; ASCT1B: 76 taxa, 417 sites; ASCT1C: 32 taxa, 504 sites; [Fe]-Hyd: 140 taxa, 293 sites; and PFO: 82 taxa, 948 sites. The alignment data are available on request to the corresponding author. For each single-gene dataset, the maximum-likelihood (ML) phylogenetic tree and corresponding bootstrap support values (100 replicates) were calculated using RAxML ver. 7.2.6 (Stamatakis 2006). The ML tree was selected from 20 heuristic tree search initiated from randomized parsimony starting trees. In ML bootstrap analyses, a single tree search per replicate was performed. Bayesian analyses were also performed using MrBayes Version 3.2.3. (Ronquist and Huelsenbeck 2003). Six parallel Metropolis coupled Markov chain MonteCarlo (MCMCMC) runs, each consisting of three heated and one cold chains with default chain temperatures, were run for 1.0–3.0 × 106 generations. Log-likelihood scores and trees with branch lengths were sampled every 1,000 generations. The first 2.5–7.5 × 105 generations were excluded as burn-in, and the remaining trees were summarized to obtain Bayesian posterior probabilities. Convergence of parallel MCMCMC runs was judged by average standard deviation of split frequencies (ASDSF). For both ML and Bayesian analyses of all five alignments, the LG model with four categories of among-site rate variation (LG + Γ model), which was selected as the most appropriate model with Aminosan (Tanabe 2011), was applied.

2-2-4. Quinone composition analyses

Genomic DNA from the cells of the bacterial prey alone (*Cantina*-free) culture and the original culture with *Cantina* (harvested by centrifugation at 4,170 g for 60 minutes at 4°C) was extracted with using a DNeasy Blood & Tissue Kit (QIAGEN, Japan).

Using these two genomic DNA samples as templates, bacterial 16S rRNA genes were independently PCR-amplified with the primers B27f and U1492r (Jiang et al. 2006) (Table2-3) using a HotStarTaq Plus Master Mix Kit (QIAGEN). HotStarTaq DNA Polymerase was activated by 5-min incubation at 95°C prior to thermal cycling. Thermal cycling consisted of 30 cycles of 1 min at 94°C, 1 min at 55°C, and 2 min at 72°C, followed by a final elongation step of 10 min at 72°C. After the PCR-amplified DNA fragments were cloned into the pCR21-TOPO (Thermo Fisher Scientific USA), 96 clones of each library were sequenced on both strands and subsequently analyzed using blastn against the nr database for identification.

For determining the quinone composition in *Cantina*, the cells of the bacterial prey (*Cantina*-free) culture and original culture with *Cantina* were harvested as stated above. Lipid extraction from the harvested cells and subsequent high performance liquid chromatography (HPLC) analyses followed the methods of Nishijima et al. (1997), although mass spectrometry analyses were not conducted. The experiments were performed on a Waters 600 series HPLC system equipped with a photodiode array detector. An Inertsil ODS-3 (ϕ 4.6 $\times$ 150 mm, 5 μm, GL Science) column was used. Methanol-isopropanol (3:1, v/v) was used as the solvent system at a flow rate of 1 ml min^{-1}. A 20-μL volume of the sample solution was injected with an autosampler. Three kinds of standard mixture were used as references: ubiquinone, Q-6, Q-7, Q-8, Q-9, and Q-10; menaquinone, MK-6, MK-7, MK-8, MK-9, and MK-10; and hydrogenated menaquinone, MK-8(H_2), MK-8(H_4), MK-9(H_2), MK-9(H_4), MK-10(H_2), and MK-10(H_4). Quinone peaks were identified by their retention times and patterns of UV spectra, which were detected with the photodiode-array detector. Ubiquinone and rhodoquinone have UV

spectra with absorption maxima (λ_{max}) at 275 nm and 280 nm, respectively, while menaquinone has a UV spectrum with λ_{max} at 245, 270, and 330 nm (Nishijima et al. 1997). These characteristic UV spectra were compared with those of peaks from the culture samples, which were detected on HPLC chromatograms.

2-3. Result

2-3-1. Overview of the mitochondrion-related organelle (MRO) proteome of *Cantina marsupialis* ased on RNA-seq data

Through homology-based analyses, 126 genes encoding putative mitochondrial (MRO) proteins were identified (Table 2-1). Among 121 annotated sequences (except for transcripts for 5 hypothetical proteins), 67, 42, 4, and 8 were predicted to be localized to the MRO matrix, inner membrane, intermembrane space, and outer membrane, respectively, based on mitochondrial proteomes of other eukaryotes, such as human, yeast, *Blastocystis*, *Nyctotherus*, and *Pygsuia*. Mitochondrial targeting signals (MTSs) were identified in 83 sequences. Twelve sequences were N-terminally truncated. In 31 sequences (including transcripts for 27 mitochondrial membrane-associated proteins), the MTSs were not identified, although their N-terminal ends were obtained. I failed to detect transcripts for several proteins that were expected to be present, suggesting that RNA-seq data does not cover all transcripts of *Cantina*. Since no transcripts for proteins involved in transcription or translation in MROs, such as ribosomal proteins, transcription initiation factors, and translation elongation factors, were detected, it is strongly suggested that an organelle genome is missing in *Cantina*. The metabolic capacity of *Cantina* MROs based on the putative protein composition is shown in Figure 2-1 and in Table 2-1. A comparison

of the metabolic capacities of mitochondria and several types of MROs (including *Cantina* MROs) is shown in Table 2-2.

2-3-2. Phylogenies of proteins associated with anaerobic pyruvate metabolism

As previously reported (e.g., Jerlström-Hultqvist et al. 2013; Nývltová et al. 2015; Stairs et al. 2014), the eukaryotic homologues of ADP-forming acetyl-CoA synthetase (ACS), two types of acetate:succinate CoA-transferase (ASCT1B andASCT1C), and iron-only hydrogenase ([Fe]-Hyd) were not monophyletic in each phylogenetic tree. In the phylogeny of ACS (ADP-forming) (Fig. 2-2), homologues of *Cantina* and *Blastocystis* grouped together with 100% bootstrap percentage (BP) and1.00 Bayesian posterior probability (PP). Likewise, the clade composed of *Cantina* and *Blastocystis* was also recovered in both phylogenies of ASCT1B (Fig. 2-3) and ASCT1C (Fig. 2-4) with 86% BP/1.00 PP and 87% BP/1.00 PP, respectively. *Cantina* has two homologues of [Fe]-Hyd with the MTS, and these two homologues were branched together with 99% BP and 1.00 PP. This *Cantina* [Fe]-Hyd clade was nested with a clade com-posed of the *Blastocystis* homologues with 59%BP and 0.99 PP (Fig. 2-5). In the pyruvate:ferredoxin oxidoreductase (PFO) phylogeny, the monophyly of eukaryotic homologues (except for the homologue of *Bombus impatiens*) was recovered with 75% BP and 0.99 PP (Fig. 2-6). In this eukaryotic radiation of PFO, the phylogenetic positions of two closely related homologues of *Cantina* could not be resolved.

2-3-3. Quinone composition

The bacterial communities of the *Cantina*-free culture and original culture with

Cantina were identical based on clone sequencing of 16S rRNA gene: Only three bacterial species, *Vibrio* sp. (99% sequence homology with *Vibrio chagasii* strain LMG21353), *Arcobacter* sp. (98% sequence homology with *Arcobacter bivalviorum* strain F4), and *Marinifilum* sp. (96% sequence homology with *Marinifilum* sp. strain THREE-2) were identified in both cultures. The 16S rRNA gene sequences from these three prey bacteria were deposited in the DDBJ/EMBL/GenBank databases as Accession Numbers LC028386–LC028388.

With HPLC analyses, menaquinone-6 (MK-6) (15.83%), menaquinone-7 (MK-7) (8.62%), menaquinone-8 (MK-8) (14.94%), and ubiquinone-8 (UQ-8) (60.61%) were detected in the *Cantina*-free culture, while MK-6 (16.33%), MK-7 (7.80%), MK-8 (9.93%), UQ-8 (63.59%), and ubiquinone-7 (UQ-7) (2.35%) were detected in the original culture with *Cantina* (Table. 2-4). According to previous studies (Collins and Jones 1981; Vandamme et al. 1991; Na et al. 2009; Sasi Jyothsna et al. 2013), it is suggested that MK-8 and UQ-8 were derived from the bacterial prey of *Cantina*, *Vibrio* sp., and that MK-6 and MK-7 originated from the other bacterial prey, *Arcobacter* sp. and *Mirinifilum* sp., respectively. Thus, UQ-7 is possibly derived from *Cantina*. The peak corresponding to rhodoquinone was not found in either culture sample.

2-4. Discussion

2-4-1. Dual ATP generation via anaerobic metabolism of pyruvate

A mitochondrial pyruvate carrier involved in the transport of pyruvate from the cytosol into mitochondria has recently been identified in humans and yeast (Bricker et al. 2012), and furthermore, Stairs et al. (2014) also found its transcripts in a transcriptome of the

anaerobic protist *Pygsuia*. In contrast, no transcript encoding this carrier was found in *Cantina*. However, because transcripts for a mitochondrial tricarboxylate carrier, a mitochondrial 2-oxoglutarate/malate carrier, and malic enzyme were detected in this analyses, it is likely that *Cantina* does not directly incorporate pyruvate from the cytosol into MROs but transports malate from the cytosol into MROs and decarboxylate malate into pyruvate in MROs, as in the case of trichomonad hydrogenosomes (Lindmark and Müller 1973). Based on RNA-seq data, *Cantina* MROs have pyruvate:ferredoxin oxidoreductase (PFO), iron-only hydrogenase ([Fe]-Hyd), two types of acetate:succinate CoA-transferase (ASCT1B and ASCT1C) and [2Fe-2S] ferredoxin, which are fundamentally found in classes 3 and 4 mitochondria, were detected, while pyruvate dehydrogenase ubiquitously present in aerobic mitochondria (i.e., class 1 mitochondria) and pyruvate:NADP+ oxidoreductase present in mitochondria or MROs of a limited number of eukaryotes including *Blastocystis* were likely lacking. The *Cantina* [Fe]-Hydseems to be a "single-domain" enzyme, unlike the case of *Blastocystis* and *Nyctotherus* in which homologues were fused with flavodoxin and two subunits of a mitochondrial complex I, respectively (Boxma et al. 2007; Stechmann et al. 2008). In the trichomonad hydrogenosomes, electrons and acetyl-CoA are generated via oxidative decarboxylation of pyruvate with PFO. The resulting electrons are transferred to [2Fe-2S] ferredoxin and then to [Fe]-Hyd, which reduces protons to produce molecular hydrogen. The coenzyme A moiety of the resulting acetyl-CoA is transferred to succinate with ASCT, generating acetate and succinyl-CoA. Succinate is subsequently regenerated by the TCA cycle enzyme succinyl-CoAsynthetase (SCS), coupled with ATP generation through substrate-level phosphorylation (Lindmark and Müller 1973). Thus, it is possible that the

MROs analyzed in this study generate ATP via hydrogen-producing fermentation. Intriguingly, ADP-forming acetyl-CoA synthetase (ACS) seems to operate in *Cantina* MROs, because this enzyme has the mitochondrial targeting signal (MTS). *Giardia* and *Entamoeba* also produce electrons and acetyl-CoA via oxidative decarboxylation of pyruvate with PFO and generate molecular hydrogen with [Fe]-Hyd (reviewed in Müller et al. 2012). However, their ATP generation through substrate-level phosphorylationis conducted with ACS (ADP-forming) instead of ASCT and SCS, using acetyl-CoA as a substrate, and all of these biochemical reactions are localized in the cytosol (not in mitosomes) (reviewed in Müller et al. 2012). It seems that *Mastigamoeba* also has ACS (ADP-forming) in its MROs like *Cantina*, but lacks ASCT and SCS (Nývltová et al. 2015). Based on these findings, *Cantina* plausibly has two types (i.e., trichomonad hydrogenosome and *Giardia*/*Entamoeba* cytosolic types) of ATP generation systems through pyruvate metabolism in its MROs. Such dual ATP synthesis systems were suggested to occur in mitochondria of the heterolobosean amoebae *Naegleria gruberi* based on its genome analyses (Fritz-Laylin et al. 2010). In addition, I found that the anaerobic stramenopile parasite *Blastocystis* has ACS (ADP-forming) with the MTS through BLAST searches. Because *Blastocystis* MROs are known to possess the trichomonad hydrogenosome-type ATP generation system using two types of ASCT (ASCT1B and ASCT1C), along with PFO, [Fe]-Hyd, and SCS, there would also be two ways of ATP synthesis via anaerobic pyruvate metabolism in this parasite. However, I cannot rule out the possibility that either ASCT or ACS in *Cantina*, *Naegleria*, and *Blastocystis* is involved in acetate formation with concomitant ATP production and that the other is associated with an unknown metabolism. In the molecular phylogenies, the

respective homologues of ASCT1B, ASCT1C, and ACS, all of which are limited to *Cantina* and *Blastocystis* in the stramenopile clade (except for ACS of a single species of diatom), each forms a robust monophyly (Figs. 2-2–4). Considering these findings, together with the fact that *Cantina* and *Blastocystis* are distantly related in the stramenopile lineage (Yubukiet al. 2015), the following three evolutionary scenarios could be proposed: 1) each enzyme was present in a common ancestor of stramenopiles, and many stramenopile lineages other than *Cantina* and *Blastocystis* lost the enzyme; 2) the genes encoding these enzymes were laterally transferred between *Cantina* and *Blastocystis*; and 3) *Cantina* and *Blastocystis* independently acquired the genes in question from two closely related, unknown donors. On the other hand, the phylogenetic positions of [Fe]-Hyd and PFO from *Cantina* could not be fully resolved, and so the evolutionary scenarios of these enzymes from *Cantina* remain unclear.

2-4-2. An incomplete TCA cycle

In aerobic (class 1) mitochondria, complexes I (CI) and II (CII) of the electron transfer chain (ETC) oxidize NADH and FADH2, respectively, and transfer the resulting electrons to the electron carrier ubiquinone (UQ). These electrons are utilized for the reduction of molecular oxygen at complex IV (CIV) of the ETC. In this case, CII functions as succinate dehydrogenase (SDH), converting succinate to fumarate. In contrast, it is known that rhodoquinone (RQ), which has a redox potential lower than that of UQ, functions as an electron carrier in anaerobic (class 2) mitochondria of an aerotolerant invertebrates and *Euglena* and hydrogen-producing (class 3) mitochondria of *Nyctotherus* under hypoxic/anoxic conditions (while class 2 mitochondria can also carry out aerobic

respiration using UQ, as do class 1 mitochondria). This means that non-canonical mitochondria in such facultatively and obligately anaerobic eukaryotes can reduce fumarate to succinate with CII as fumarate reductase (FRD) using electrons, which are generated by the oxidation of NADH at CI and are transferred to CII via RQ (Müller et al. 2012). Recently, the gene encoding the putative methyltransferase RQUA, an enzyme responsible for the RQ biosynthetic pathway, was identified from the RQ-bearing purple non-sulfur bacterium *Rhodospirillum rubrum* (Lonjers et al. 2012), and this gene was also found in the anaerobic protist *Pygsuia*, along with *Euglena*, *Blastocystis*, and a few other eukaryotes (Stairs et al. 2014). The RQUA gene was not found in RNA-seq data of *Cantina*, and RQ was not detected with high performance liquid chromatography (Table 2-4). In addition, the enzyme fumarate hydratase, which is consistently found in organelles capable of performing the reductive TCA cycle (Denoeud et al. 2011; Müller et al. 2012; Stairs et al. 2014; Stechmann et al. 2008), seems to be absent in *Cantina* MROs based on RNA-seq data. The absence of the gene encoding this enzyme was also suggested by the results of PCR experiments with primers specific to the stramenopile fumarate hydratasegenes designed in this study (Table 2-3). Although PCR-amplified products were obtained, all sequenced 96 clones encoded fumarate hydratase probably from bacterial prey. These findings provide insight into the unique metabolic characteristics of *Cantina* MROs. Transcripts for a subset of the TCA cycle enzymes, all three subunits (E1, E2, and E3) of 2-oxoglutarate dehydrogenase, both the α and β subunits of succinyl-CoA synthase, all four subunits (A, B, C, and D) of succinate dehydrogenase, and malate dehydrogenase were identified, while transcripts for fumarate hydratase (as mentioned above), citrate synthase, aconitate hydratase, and isocitrate dehydrogenase were not detected. Thus,

Cantina MROs seem to possess an incomplete TCA cycle, similar to the anaerobic protists *Nyctotherus*, *Blastocystis*, *Mastigamoeba*, and *Pygsuia*. In MROs at least in *Nyctotherus*, *Blastocystis*, and *Pygsuia*, the incomplete TCA cycle is suggested to run in the reductive (reverse) direction using RQ. These sequential reactions include the reduction of fumarate using CII (FRD), as mentioned above. In contrast, it is possible that the incomplete TCA cycle of *Cantina* MROs runs in the oxidative direction and that CII functions as SDH, using UQ rather than RQ. Indeed, preliminary quinone composition analyses suggest that *Cantina* has UQ-7 (Table 2-4), although I cannot completely exclude the possibility that UQ-7 was detected as a minor component of quinone from bacterial prey. Considering the subset of TCA cycle enzymes found in *Cantina* and the above-mentioned possible oxidative direction of TCA cycle, 2-oxoglutarate and fumarate are suggested to be the starter and end product of its incomplete TCA cycle, respectively. There could be two possible systems generating 2-oxoglutarate in *Cantina* MROs: the reactions catalyzed by glutamate dehydrogenase and aspartate aminotransferase. In the former reaction, glutamate is used as a substrate, and this substrate would be imported from the cytosol into MROs via a glutamate-aspartate transporter and/or a mitochondrial 2-oxodicarboxylate carrier. In the latter reaction, oxaloacetic acid, along with glutamate, is used as a substrate, and oxaloacetic acid is assumed to be converted from malate, which would be transferred from the cytosol via a mitochondrial tricarboxylate carrier and/or a mitochondrial 2-oxoglutarate/malate carrier. Succinyl-CoA synthesized from 2-oxoglutarate with 2-oxoglutarate dehydrogenase is likely involved in the incomplete TCA cycle, in which the SCS reaction produces one molecule of ATP per molecule of 2-oxoglutarate via substrate-level phosphorylation. Furthermore, succinyl-CoA may also be linked to a

reverse pathway of β-oxidation of odd-numbered chain fatty acids (rather than to a ATP-consuming pathway in the regular direction), because transcripts for propionyl-CoA carboxylase and methylmalonyl-CoA mutase were detected in this study. It has been shown that this reverse β-oxidation, in which succinyl-CoA is metabolized to propionyl-CoA, runs in anaerobic mitochondriain several invertebrates and MROs (hydrogen-producing mitochondria) in *Blastocystis* (Denoeudet al. 2011; Müller et al. 2012; Stechmann et al. 2008). In this pathway, the propionyl-CoA carboxylase reaction generates propionyl-CoA and produces ATP via substrate-level phosphorylation. Again, note that ACS (ADP-forming) seems to operate in *Cantina* MROs. It was demonstrated that ACS (ADP-forming) from *Giardia*, *Entamoeba*, and the hyperthermophilic archaeon *Pyrococcus furiosus* can utilize not only acetyl-CoA but also propionyl-CoA as a substrate in vitro (Glasemacher et al. 1997; Jones and Ingram-Smith 2014; Sánchezet al. 2000). Therefore, if the *Cantina* ACS were to function similarly to the ACS present in *Giardia*, *Entamoeba*, and *P. furiosus*, then it may be possible for *Cantina* MROs to generate propionate from the end-product of reverse β-oxidation, propionyl-CoA, resulting in the synthesis of ATP via substrate-level phosphorylation. In consequence, two molecules of ATP per molecule of succinyl-CoA are assumed to be yielded in the consecutive reactions from succinyl-CoA to propionate.

2-4-3. Membrane-associated electron transport system without proton pumping

It is common knowledge that the ETC of canonical mitochondria comprises CI, CII, CIII, and CIV. CI, CIII, and CIV actively pump protons into the intermembrane space and contribute a proton gradient across the mitochondrial inner membrane. The resulting

gradient is harnessed by F1Fo-ATPase to efficiently produce ATP via oxidative phosphorylation. On the other hand, CII, which is the only enzyme participating in both the ETC and the TCA cycle, does not perform proton pumping. Some eukaryotes adapting to oxygen-depleted environments completely lack the ETC, but others retain a "truncated" ETC. For example, only CI and CII are considered to operate in *Nyctotherus* and *Blastocystis.* CI in their organelles potentially pumps protons from NADH and passes the electrons from NADH through RQ to CII (FRD) (Boxma et al. 2005; de Graaf et al. 2011; Stechmann et al. 2008). In addition, it was reported that *Pygsuia* and *Mastigamoeba* retain only CII in the ETC in their MROs (Gill et al. 2007; Nývltová et al. 2015; Stairset al. 2014). It was also suggested that MROs in *Blastocystis* and *Pygsuia* have alternative oxidase (AOX), electron-transferring flavoprotein (ETF), electron-transferring flavoprotein dehydrogenase (ETF-DH), and glycerol-3-phosphate dehydrogenase (G3PDH), along with RQ, as components of the membrane-associated electron transport system (Stairs et al. 2014; Stechmann et al. 2008). In this study, no transcripts for subunits of CI, CIII, and CIV of the ETC, or of F1Fo ATPase were retrieved, but transcripts for all four subunits of CII (i.e., succinate dehydrogenase) were detected. Thus, *Cantina* seems to possess only CII in the ETC in its organelles, similar to *Pygsuia* and *Mastigamoeba*. I failed to find transcripts even for the two soluble subunits of NADH:ubquinone oxidore-ductase (NuoE and NuoF) from CI often found in MROs of anaerobic protists. Because transcripts for AOX, ETF, ETF-DH, G3PDH, and alternative NAD(P)H dehydrogenase (ANDH) were also found in this study, it is reasonable to assume that ANDH and ETF-DH/G3PDH in *Cantina* MROs donate the electrons derived from NAD(P)H and FADH2 to AOX through UQ, respectively. These reactions are not accompanied by proton pumping, as in the case of the

CII-mediated reaction. AOX is known to reduce the final electron acceptor, molecular oxygen, to water, and so it is uncertain how *Cantina* inhabiting oxygen-depleted environments provides its MROs with molecular oxygen. This enzyme is limited within eukaryotes, and the evolutionary history of eukaryotic homologues remains unclear (e.g., Suzuki et al. 2005). Similarly, the evolutionary origin of AOX of *Cantina* could not be resolved based on phylogeny (Fig. 2-7). In *Cantina* MROs, a few tubular cristae have been observed with electron microscopy (Yubuki et al. 2015). Such structure may be retained to pack the above-mentioned components of the electron transport system.

2-4-4. Possible "simple" protein import machineries

The components of six major complexes: translocase of the outer membrane of mitochondria (TOM), translocase of the inner membrane of mitochondria (TIM), small-TIM, sorting and assembly machinery (SAM), presequence translocase-associated motor (PAM), and mitochondrial processing peptidase (MPP) involved in mitochondrial protein import were compared among *Cantina* and other anaerobic microbial eukaryotes. The TOM complex plays a crucial role in the first step of the translocation of nuclear-encoded mitochondrial proteins through the mitochondrial outer membrane. The number and repertoire of subunits of the TOM complex vary considerably among different lineages of eukaryotes. Nevertheless, the subunit Tom40 occurs over a broad range of eukaryotes, because this subunit is a core translocation channel (Wiedemann et al. 2004). Among subunits of the TOM complex, transcripts only for Tom40 were found in *Cantina*, as in the case of many other anaerobic protists (Jerlström-Hultqvist et al. 2013; Makiuchi and Nozaki 2014; Stairs et al. 2014; Zubácová et al. 2013). Although I cannot exclude the

possibility that other TOM subunits failed to be detected through homology (BLAST)-based analyses due to their highly divergent sequences, the TOM machinery of *Cantina* MROs may fulfill its role with very limited number of subunits. The Tim9-Tim10 and Tim8-Tim13 chaperone complexes (small TIM complexes) in the intermembrane space are significant for the translocation of β-barrel membrane proteins and Tim23, respectively (Wiedemann et al. 2004). Among subunits of the small TIM complex, transcripts for Tim8, Tim13, and Tim9 were identified, although transcripts for Tim10 were not detected. Therefore, *Cantina* likely has these two complexes. In *Blastocystis*, *Pygsuia*, and *Trichomonas*, either one or both of the two complexes were confirmed to be present, while mitosome-containing eukaryotes, such as *Giardia*, *Entamoeba*, and *Encephalitozoon* seem to lack them (Makiuchi and Nozaki 2014). The Tim8-Tim13 complex may guide β-barrel outer membrane proteins to the SAM machinery, which sorts and assembles these proteins (Wiedemann et al. 2004). As in the case of many anaerobic microbial eukaryotes investigated (reviewed in Makiuchi and Nozaki 2014), transcripts only for Sam50 among subunits of the SAM complex were found in *Cantina*. One outer membrane protein processed by the SAM machinery, Tom40, is highly likely present in *Cantina* MROs based on RNA-seq data, as mentioned above. These findings support the hypothesis that the SAM system could operate in *Cantina* MROs, although it remains unknown whether Sam50 is a sole component of the SAM machinery. It is known that the TIM22 and TIM23 complexes are involved in the assembly of inner membrane proteins and matrix proteins, respectively (Wiedemann et al. 2004). In *Cantina*, transcripts for Tim22, Tim23, Tim17, and Tim50 were found, although transcripts for other TIM subunits, such as Tim54, Tim18, Tim12, and Tim21 identified in yeasts, were not retrieved. Thus,

Cantina is thought to possess the machineries corresponding to both complexes. Transcripts only for Tim22 among subunits of the TIM22 complex were identified in *Cantina*, and several anaerobic protists, such as *Giardia*, *Entamoeba*, and *Mastigamoeba*, along with the anaerobic free-living heterolobose amoeba *Sawyeria marylandensis* with hydrogenosomes, are in the same situation, having no subunits of the TIM22 complex other than Tim22 in their organelles (reviewed in Makiuchi and Nozaki 2014). Transcripts for subunits of the PAM complex, Hsp70, Mge1, Mdj2, Tim (Pam) 16, and Tim44 were detected in *Cantina*, although transcripts for other PAM subunits, such as Pam17, Tim15, and Tam41identified in yeasts were not found. In addition, both the α and β subunits of MPP were identified. Therefore, the possible presence of the PAM complex and mitochondrial processing peptidase in *Cantina* MROs, both of which contribute to the maturation of proteins imported into the matrix, supports the hypothesis that the TIM23 machinery functions in this organism.

2-4-5. Comparison of other metabolism between MROs in *Cantina* and other anaerobic eukaryotes

Cantina MROs are likely responsible for the metabolism of seven amino acids, threonine, glycine, valine, cysteine, alanine, glutamate, and aspartic acid, while canonical mitochondria are involved in the synthesis and catabolism of all 20 essential amino acids. For example, transcripts for threonine 3-dehydrogenase, glycine C-acetyltransferase, and serine hydroxymethyltransferase were detected, and thus threonine could be converted into serine via glycine. Transcripts for branched-chain aminotransferase, 3-hydroxyisobutyryl-CoA hydrolase, and 3-hydroxyisobutyrate dehydrogenase were

detected, suggesting that valine is metabolized into propionyl-CoA. Transcripts for four proteins associated with the glycine cleavage system (T-, L-, H-, and P-proteins) were identified. Transcripts for alanine aminotransferase, which converts alanine into pyruvate and is linked with pyruvate metabolism and iron-sulfur assembly machinery, were detected. A subset of metabolism of these seven amino acids was often found in other biochemically degenerate organelles other than mitosomes (e.g., Carltonet al. 2007; de Graaf et al. 2011; Gill et al. 2007; Jerlström-Hultqvist et al. 2013; Nývltová et al. 2015; Stairs et al. 2014; Stechmann et al. 2008; Zubácová et al. 2013). Transcripts for enzymes involved in the β-oxidation of even-numbered chain fatty acids, short-chain acyl-CoA dehydrogenase, hydroxyacyl-CoA dehydrogenase, and acetyl-CoA acetyltransferase were detected, although transcripts for enoyl-CoA hydratase were not found. As mentioned above, the pathway of β-oxidation of odd-numbered chain fatty acids may run in the reverse direction, generating propionyl-CoA along with ATP in *Cantina* MROs. On the other hand, *Cantina* likely possesses the "normally running" pathway of β-oxidation of even-numbered chain fatty acids, in which fatty acyl-CoA is finally broken down into acetyl-CoA. One step in this pathway, the acyl-CoA dehydrogenase (short-chain) reaction, which catalyzes the oxidation of acyl-CoA into 2-enoyl-CoA, may be tightly associated with the electron transport system in *Cantina*: the resulting FADH2 in acyl-CoA dehydrogenase could be oxidized by ETF, which gives the electrons to ETF-DH. This pathway has been found in the organelles of *Sawyeria*, *Blastocystis*, and *Pygsuia* among MRO-bearing microbial eukaryotes known to date (Barberà et al. 2010; Denoeud et al. 2011; Stairs et al. 2014). Almost all eukaryotes employ the iron-sulfur cluster assembly (ISC) system as an iron-sulfur cluster assembly machinery in mitochondria. It is thought that this system was

inherited from an α-proteobacterial ancestor of mitochondria (Lill and Mühlenhoff et al. 2006). In contrast, several exceptions were reported among MRO-bearing anaerobic protists. *Mastigamoeba* MROs lack the ISC system, and instead adopt a nitrogen-fixation (NIF)-related iron-sulfur cluster biogenesis system, which was acquired by lateral gene transfer from an ε-proteobacterium (Nývltová et al.2013). Furthermore, *Pygsuia* MROs utilize a sulfur mobilization factor (SUF) system laterally acquired from a Methanomicrobiales archaeon instead of the ISC system (Stairs et al. 2014). Nevertheless, transcripts for components of the iron-sulfur assembly machinery, IscS catalyzing the removal of elemental sulfur from cysteine, the scaffold proteins IscU and Isa2, the chaperone proteins frataxin and HscB, the transporter mitoferrin, the redox protein glutaredoxin, the P-loop NTPase involved in Fe-S protein biogenesis Ind1, and three hydrogenase maturation proteins (HydE, HydF, and HydG) were identified in *Cantina*. Thus, *Cantina* seems to have the canonical ISC system rather than such non-canonical iron-sulfur cluster assembly systems in its MROs. Intriguingly, transcripts for a novel ornithine carbamoyltransferase (OTC)-carbamate kinase (CK) fusion protein with the MTS were identified in this study. Therefore, the arginine dihydrolase pathway or urea cycle, in both of which OCT and CK are involved, possibly runs in this organism. Among MRO-bearing anaerobic microbial eukaryotes, the presence of the former pathway was demonstrated or suggested in *Giardia*, *Trichomonas*, and the rumen anaerobic fungus *Neocallimastix frontalis* (Gelius-Dietrich et al. 2007; Morada et al. 2010; Touz et al. 2008). However, as in the case of *Blastocystis*, *Sawyeria*, and the free-living anaerobic excavate *Trimastix pyriformis* (Barberà et al. 2010; Stechmann et al. 2008; Zubácová et al. 2013), I could not conclude which metabolic system occurs in *Cantina*, because transcripts for

enzymes involved in either of these two metabolic systems other than OTC and CK were not detected. Also, the presence of an unknown metabolic pathway, in which OTC and CK are utilized, cannot be ruled out.

References

Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EA, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick RS, Schoch CL, Smirnov A, Spiegel FW (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**:429–493

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

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

Boxma B, de Graaf RM, van der Staay GW, van Alen TA, Ricard G, Gabaldón T, van Hoek AH, Moon-van der Staay SY, Koopman WJ, van Hellemond JJ, Tielens AG, Friedrich T, Veenhuis M, Huynen MA, Hackstein JH (2005) An anaerobic mitochondrion that produces hydrogen. *Nature* **434**:74–79

Boxma B, Ricard G, van Hoek AH, Severing E, Moon-van der Staay SY, van der Staay GW, van Alen TA, de Graaf RM, Cremers G, Kwantes M, McEwan NR, Newbold CJ, Jouany JP, Michalowski T, Pristas P, Huynen MA, Hackstein JH

(2007) The [FeFe] hydrogenase of *Nyctotherus ovalis* has a chimeric origin. *BMC Evol Biol* **7**:230

Bricker DK, Taylor EB, Schell JC, Orsak T, Boutron A, Chen YC, Cox JE, Cardon CM, Van Vranken JG, Dephoure N, Redin C, Boudina S, Gygi SP, Brivet M, Thummel CS, Rutter J (2012) A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, *Drosophila*, and humans. *Science* **337**:96–100

Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, Wortman JR, Bidwell SL, Alsmark UC, Besteiro S, Sicheritz-Ponten T, Noel CJ, Dacks JB, Foster PG, Simillion C, Van de Peer Y, Miranda-Saavedra D, Barton GJ, Westrop GD, Müller S, Dessi D, Fiori PL, Ren Q, Paulsen I, Zhang H, Bastida-Corcuera FD, Simoes-Barbosa A, Brown MT, Hayes RD, Mukherjee M, Okumura CY, Schneider R, Smith AJ, Vanacova S, Villalvazo M, Haas BJ, Pertea M, Feldblyum TV, Utterback TR, Shu CL, Osoegawa K, de Jong PJ, Hrdy I, Horvathova L, Zubacova Z, Dolezal P, Malik SB, Logsdon JM Jr, Henze K, Gupta A, Wang CC, Dunne RL, Upcroft JA, Upcroft P, White O, Salzberg SL, Tang P, Chiu CH, Lee YS, Embley TM, Coombs GH, Mottram JC, Tachezy J, Fraser-Liggett CM, Johnson PJ (2007) Draft genome sequence of thesexually transmitted pathogen *Trichomonas vaginalis*. *Science* **315**:207–212

Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* **241**:779–786

Collins MD, Jones D (1981) Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. *Microbiol Rev* **45**:316–354

de Graaf RM, Ricard G, van Alen TA, Duarte I, Dutilh BE, Burgtorf C, Kuiper JW, van der Staay GW, Tielens AG, Huynen MA, Hackstein JH (2011) The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. *Mol Biol Evol.* **28**:2379–2391

Denoeud F, Roussel M, Noel B, Wawrzyniak I, Da Silva C, Diogon M, Viscogliosi E, Brochier-Armanet C, Couloux A, Poulain J, Segurens B, Anthouard V, Texier C, Blot N, Poirier P, Ng GC, Tan KS, Artiguenave F, Jaillon O, Aury JM, Delbac F, Wincker P, Vivarès CP, El Alaoui H (2011) Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. *Genome Biol.* **12**:R29

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

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

Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, Shu S, Neupane R, Cipriano M, Mancuso J, Tu H, Salamov A, Lindquist E, Shapiro H, Lucas S, Grigoriev IV, Cande WZ, Fulton C, Rokhsar DS, Dawson SC (2010) The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* **140**:631–642

Gaston D, Tsaousis AD, Roger AJ (2009) Predicting proteomes of mitochondria and related organelles from genomic and expressed sequence tag data. *Methods Enzymol.* **457**:21–47

Gelius-Dietrich G, Ter Braak M, Henze K (2007) Mitochondrial steps of arginine biosynthesis are conserved in the hydrogenosomes of the chytridiomycete *Neocallimastix frontalis*. *J Eukaryot Microbiol* **54**:42–44

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

Glasemacher J, Bock AK, Schmid R, Schönheit P (1997) Purification and properties of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate formation and ATP synthesis, from the hyperthermophile *Pyrococcus furiosus*. *Eur J Biochem* **244**:561–567

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

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**:644–652

Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. *Science* **283**:1476–1481

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

Jiang H, Dong H, Zhang G, Yu B, Chapman LR, Fields MW (2006) Microbial diversity in water and sediment of Lake Chaka, an athalassohaline lake in northwestern China. *Appl Environ Microbiol* **72**:3832–3845

Jones CP, Ingram-Smith C (2014) Biochemical and kinetic characterization of the recombinant ADP-forming acetyl coenzyme A synthetase from the amitochondriate protozoan *Entamoeba histolytica. Eukaryot Cell* **13**:1530–1537

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

Leipe DD, Tong SM, Goggin CL, Slemenda SB, Pieniazek NJ, Sogin ML (1996) 16S-like rDNA sequences from *Developayella elegans*, *Labyrinthuloides haliotidis*, and *Proteromonas lacertae* confirm that the stramenopiles are a primarily heterotrophic group. *Eur J Protistol* **32**:449–458

Lill R, Mühlenhoff U (2006) Iron-sulfur protein biogenesis in eukaryotes: components and mechanisms. *Annu Rev Cell Dev Biol* **22**:457–486

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

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res* 40:W622–627

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

López-García P, Philippe H, Gail F, Moreira D (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc Natl Acad Sci USA* **100**:697–702

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

Massana R, del Campo J, Sieracki ME, Audic S, Logares R (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J* **8**:854–866

McDonald AE, Vanlerberghe GC (2006) Origins, evolutionary history, and taxonomic distribution of alternative oxidase and plastoquinol terminal oxidase. *Comp Biochem Physiol Part D Genomics Proteomics* **1**:357–364

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

Mi-ichi F, Makiuchi T, Furukawa A, Sato D, Nozaki T (2011) Sulfate activation in mitosomes plays an important role in the proliferation of *Entamoeba histolytica*. *PLoS Negl Trop Dis* **5**:e1263

Morada M, Manzur M, Lam B, Tan C, Tachezy J, Rappelli P, Dessì D, Fiori PL, Yarlett N (2010) Arginine metabolism in *Trichomonas vaginalis* infected with *Mycoplasma hominis*. *Microbiology* **156**:3734–3743

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

Na H, Kim S, Moon EY, Chun J (2009) *Marinifilum fragile* gen. nov., sp. nov., isolated from tidal flat sediment. *Int J Syst Evol Microbiol* **59**:2241–2246

Nishijima M, Araki-Sakai M, Sano H (1997) Identification of isoprenoid quinones by frit-FAB liquid chromatography-mass spectrometry for the chemotaxonomy of microorganisms. *J Microbiol Methods* **28**:113–122

Nývltová E, Šuták R, Harant K, Šedinová M, Hrdy I, Paces J, Vlček Č, Tachezy J (2013) NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. *Proc Natl Acad Sci USA* **110**:7371–7376

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

Pérez-Brocal V, Clark CG (2008) Analysis of two genomes from the mitochondrion-like organelle of the intestinal parasite *Blastocystis*: complete sequences, gene content, and genome organization. *Mol Biol Evol* **25**:2475–2482

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574

Rotte C, Stejskal F, Zhu G, Keithly JS, Martin W (2001) Pyruvate: NADP+oxidoreductase from the mitochondrion of *Euglena gracilis* and from the apicomplexan *Cryptosporidium parvum*: a biochemical relic linking pyruvate metabolismin mitochondriate and amitochondriate protists. *Mol Biol Evol* **18**:710–720

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

Sasi Jyothsna TS, Rahul K, Ramaprasad EV, Sasikala Ch, Ramana ChV (2013) *Arcobacter anaerophilus* sp. nov., isolated from an estuarine sediment and emended description of the genus *Arcobacter*. *Int J Syst Evol Microbiol* **63**:4619–4625

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

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

Suzuki T, Hashimoto T, Yabu Y, Majiwa PA, Ohshima S, Suzuki M, Lu S, Hato M, Kido Y, Sakamoto K, Nakamura K, Kita K, Ohta N (2005) Alternative oxidase (AOX) genes of African trypanosomes: phylogeny and evolution of AOX and plastid terminal oxidase families. *J Eukaryot Microbiol* **52**:374–381

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

Takishita K, Tsuchiya M, Kawato M, Oguri K, Kitazato H, Maruyama T (2007) Genetic diversity of microbial eukaryotes in anoxic sediment of the saline

meromictic lake Namako-ike (Japan): on the detection of anaerobic or anoxic-tolerant lineages of eukaryotes. *Protist* **158**:51–64

Tanabe AS (2011) Kakusan4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Mol Ecol Resour* **11**:914–921

Touz MC, Rópolo AS, Rivero MR, Vranych CV, Conrad JT, Svard SG, Nash TE (2008) Arginine deiminase has multiple regulatory roles in the biology of *Giardia lamblia*. *J Cell Sci* **121**:2930–2938

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

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

van der Giezen M, Tovar J, Clark CG (2005) Mitochondrion-derived organelles in protists and fungi. *Int Rev Cyto*. **244**:175–225

Vandamme P, Falsen E, Rossau R, Hoste B, Segers P, Tytgat R, De Ley J (1991) Revision of *Campylobacter*, *Helicobacter*, and *Wolinella* taxonomy: emendation of generic descriptions and proposal of *Arcobacter* gen. nov. *Int J Syst Bacteriol* **41**:88–103

Wiedemann N, Frazier AE, Pfanner N (2004) The protein import machinery of mitochondria. *J Biol Chem* **279**:14473–14476

Yubuki N, Leander BS, Silberman JD (2010) Ultrastructure and molecular phylogenetic position of a novel phagotrophic stramenopile from low oxygen environments *Rictus lutensis* gen. et sp. nov. (Bicosoecida, incertae sedis). *Protist* **161**:264–278

Yubuki N, Pánek T, Yabuki A, Čepička I, Takishita K, Inagaki Y, Leander BS (2015) Morphological identities of two different marine stramenopile environmental sequence clades: *Bicosoeca kenaiensis* (Hilliard 1971) and *Cantina marsupialis* (Larsen and Patterson, 1990) gen. nov., comb. nov. *J Eukaryot Microbiol* **62**:532–542

Zubácová Z, Novák L, Bublíková J, Vacek V, Fousek J, Rídl J, Tachezy J, Dolezal P, Vlcek C, Hampl V (2013) The mitochondrion-like organelle of *Trimastix pyriformis* contains the complete glycine cleavage system. *PLoS One* **8**:e55417

Table 2-1. List of the proteins putatively localized in mitochondrion-related organelles of *Cantina marsupialis*

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
succinate dehydrogenase (ubiquinone) flavoprotein subunit	ETC	Complete	SDHA	MM	27	0.9889	19	0.894
succinate dehydrogenase (ubiquinone) iron-sulfur subunit	ETC	Complete	SDHB	MM	25	0.9126	17	0.788
succinate dehydrogenase (ubiquinone) cytochrome b560 subunit	ETC	Complete	SDHC	IM	※	0.9997	23	0.958
succinate dehydrogenase (ubiquinone) membrane anchor subunit	ETC	Complete	SDHD	IM	13	0.4496	5	0.845
electron transfer flavoprotein alpha subunit	ETC	Complete	ETFα	MM	23	0.8124	18	0.857
electron transfer flavoprotein beta subunit	ETC	Complete	ETFβ	MM	※	0.5086	※	0.343
electron transfer flavoprotein dehydrogenase	ETC	Complete	ETF-DH	IM	28	0.9624	19	0.869
alternative NADH dehydrogenase	ETC	Complete	ADH	IM	25	0.93	22	0.594

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
alternative oxidase *	ETC	Complete	AOX	IM	34	0.9501	33	0.48
glycerol-3-phosphate dehydrogenase	ETC	Complete	G3PDH	IM	23	0.5889	5	0.337
heat shock protein 70 (Hsp70)	FeS cluster	Complete	Hsp70_1	MM	24	0.9481	※	0.815
heat shock protein 70 (Hsp70)	FeS cluster	3' end missing	Hsp70_2	MM	30	0.7247	29	0.751
Fe-S protein assembly co-chaperone HscB protein	FeS cluster	Complete	HscB	MM	28	0.8913	27	0.712
frataxin	FeS cluster	Complete	Fd	MM	25	0.9749	24	0.668
ferredoxin	FeS cluster	Complete	Frx	MM	30	0.9936	22	0.746
scaffold protein Isa2	FeS cluster	Complete	Isa2	MM	43	0.9227	42	0.672

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
hydrogenase maturase HydE	FeS cluster	Complete	HydE	MM	23	0.8004	15	0.897
hydrogenase maturase HydF	FeS cluster	Complete	HydF	MM	20	0.9927	19	0.84
hydrogenase maturase HydG	FeS cluster	Complete	HydG	MM	32	0.9208	24	0.429
mitoferrin (Mrs3/Mrs4)	FeS cluster	Complete	Mfrn	OM	※	0.0218	※	0.059
cysteine desulfurase IscS	FeS cluster	Complete	IscS	MM	25	0.968	24	0.801
iron-sulfur clusters assembley enzyme IscU	FeS cluster	Complete	IscU	MM	23	0.9944	15	0.852
glutaredoxin	FeS cluster	Complete	Grx	MM	22	0.9938	34	0.884
P-loop NTPase Ind1	FeS cluster	Complete	Ind1	MM	27	0.9858	19	0.527

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
pyruvate carboxylase	Pyruvate metabolism	Complete	PYC	MM	25	0.865	※	0.938
iron hydrogenase	Pyruvate metabolism	Complete	FeHyd_1	MM	36	0.9787	28	0.9
iron hydrogenase	Pyruvate metabolism	Complete	FeHyd_2	MM	32	0.9965	31	0.929
pyruvate:ferredoxin oxidoreductase	Pyruvate metabolism	Complete	PFO_1	MM	31	0.9907	30	0.566
pyruvate:ferredoxin oxidoreductase	Pyruvate metabolism	Complete	PFO_2	MM	17	0.9795	34	0.835
acetyl-CoA synthase *	Pyruvate metabolism	Complete	ACS	MM	31	0.9658	30	0.814
acetate:succinate CoA-transferase 1B	Pyruvate metabolism	Complete	ASCT1B	MM	22	0.9058	※	0.468
acetate:succinate CoA-transferase 1C	Pyruvate metabolism	Complete	ASCT1C	MM	※	0.8359	20	0.803
succinyl-CoA synthetase alpha subunit	Pyruvate metabolism	Complete	SCS_α	MM	20	0.7704	19	0.772

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
succinyl-CoA synthetase beta subunit	Pyruvate metabolism	Complete	SCS_β	MM	17	0.9776	12	0.731
NADP-dependent malic enzyme	Pyruvate metabolism	Complete	ME	MM	※	0.7114	※	0.288
propionyl-CoA carboxylase alpha chain	β-oxidation	Complete	PCC_α	MM	※	0.9256	19	0.481
propionyl-CoA carboxylase beta chain	β-oxidation	Complete	PCC_β	MM	12	0.8692	16	0.788
methylmalonyl-CoA mutase	β-oxidation	Complete	MMM	MM	24	0.8803	16	0.6
acetyl-CoA C-acetyltransferase	β-oxidation	Complete	ACAT	MM	18	0.8414	10	0.521
hydroxyacyl-Coenzyme A dehydrogenase	β-oxidation	Complete	HADH	MM	20	0.8496	※	0.555
short-chain specific acyl-CoA dehydrogenase	β-oxidation	Complete	SCAD	MM	21	0.9873	41	0.802
long chain fatty acid CoA ligase	β-oxidation	Complete	LCAC_1	OM	※	0.0199	※	0.071

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
long chain fatty acid CoA ligase	β-oxidation	Complete	LCAC_2	OM	※	0.0687	※	0.065
long chain fatty acid CoA ligase	β-oxidation	3' end missing	LCAC_3	OM	※	0.007	※	0.267
long chain fatty acid CoA ligase	β-oxidation	3' end missing	LCAC_4	OM	※	0.0566	※	0.059
long chain fatty acid CoA ligase	β-oxidation	Complete	LCAC_5	OM	※	0.0411	※	0.054
aminomethyltransferase (glycine cleavage system_T_protein)	AA_metab-olism	Complete	GCS_T	MM	24	0.9333	15	0.66
glycine cleavage system protein H	AA_metab-olism	Complete	GCS_H	MM	30	0.9885	22	0.905
glycine dehydrogenase (decarboxylating)	AA_metab-olism	Complete	GCS_P	MM	20	0.996	23	0.85
serine hydroxymethyltransferase	AA_metab-olism	Complete	SHMT	MM	27	0.9374	26	0.789
glycine C-acetyltransferase	AA_metab-olism	Complete	GCAT	MM	※	0.9436	5	0.744

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
threonine dehydrogenase	AA_metab-olism	Complete	TDH	MM	30	0.9384	※	0.401
alalanine transferase	AA_metab-olism	Complete	AlaAT	MM	16	0.7418	※	0.31
branched-chain amino acid aminotransferase	AA_metab-olism	Complete	BCAT	MM	25	0.981	17	0.584
3-hydroxyisobutyryl-CoA hydrolase	AA_metab-olism	Complete	HIBCH	MM	※	0.6789	※	0.322
3-hydroxyisobutyrate dehydrogenase	AA_metab-olism	Complete	HIBADH	MM	37	0.7897	※	0.288
aspartate aminotransferase	AA_metab-olism	Complete	AspAT	MM	21	0.991	※	0.651
glutamate dehydrogenase	AA_metab-olism	Complete	GDH	MM	29	0.5113	21	0.587
ornitine aminotransferase	AA_metab-olism	Complete	OAT	MM	25	0.9812	※	0.656
translocase of the outer mitochondrial membrane 40 (Tom40)	TOMTIM	5' end missing	TOM40	OM	-	-	-	-

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
sorting assembly machinery 50 kDa subunit	TOMTIM	5' end missing	SAM50	OM	-	-	-	-
sulfydryl oxidase_Erv1	TOMTIM	Complete	Erv1	IMS	※	0.0578	23	0.03
mitochondrial import inner membrane translocase subunit Tim8	TOMTIM	5' end missing	Tim8	IMS	-	-	-	-
mitochondrial import inner membrane translocase subunit Tim13	TOMTIM	Complete	Tim13	IMS	※	0.0254	※	0.18
mitochondrial import inner membrane translocase subunit Tim44	TOMTIM	5' end missing	Tim44_1	IM	-	-	-	-
mitochondrial import inner membrane translocase subunit Tim44	TOMTIM	Complete	Tim44_2	IM	30	0.9523	8	0.864
mitochondrial import inner membrane translocase subunit Tim50	TOMTIM	5' end missing	Tim50	IM	-	-	-	-
mitochondrial import inner membrane translocase subunit Tim16	TOMTIM	Complete	Tim16	IM	32	0.4888	48	0.571
mitochondrial import inner membrane translocase subunit Tim9	TOMTIM	Complete	Tim9	IMS	※	0.0375	※	0.066

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
mitochondrial import inner membrane translocase subunit Tim17	TOMTIM	Complete	Tim17	IM	※	0.0958	※	0.095
mitochondrial import inner membrane translocase subunit Tim22	TOMTIM	Complete	Tim22_1	IM	※	0.165	※	0.12
mitochondrial import inner membrane translocase subunit Tim22	TOMTIM	Complete	Tim22_2	IM	※	0.0125	※	0.067
mitochondrial import inner membrane translocase subunit Tim23	TOMTIM	Complete	Tim23	IM	※	0.0336	※	0.042
mitochondrial GrpE-related protein 1 Mge1	TOMTIM	Complete	Mge1	MM	23	0.9927	22	0.884
mitochondrial DnaJ homolog 2 mdj2	TOMTIM	Complete	Mdj2	MM	※	0.6305	18	0.903
mitochondrial processing peptidase alpha subunit	TOMTIM	3' end missing	MPPα	MM	62	0.9912	15	0.761
mitochondrial processing peptidase beta subunit	TOMTIM	Complete	MPPβ	MM	26	0.9144	18	0.712
mitochondrial inner membrane protease	TOMTIM	Complete	IMP	MM	30	0.975	※	0.37

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
chaperonin 60	TOMTIM	Complete	Cpn60	MM	21	0.8093	12	0.748
10kDa chaperonin	TOMTIM	Complete	Cpn10	MM	29	0.9852	27	0.643
cell division protease ftsH	TOMTIM	5' end missing	ftsH_1	IM	-	-	-	-
cell division protease ftsH	TOMTIM	5' end missing	ftsH_2	IM	-	-	-	-
octapeptidyl aminopeptidase 1 oct1	TOMTIM	Complete	Oct	MM	22	0.9883	11	0.829
distribution and morphology protein 38	TOMTIM	Complete	Mdm38	MM	19	0.8841	18	0.878
2-oxoglutarate dehydrogenase E1 component	TCA cycle	5' end missing	αKDH_E1	MM	-	-	-	-
2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase)	TCA cycle	3' end missing	αKDH_E2	MM	29	0.9373	※	0.738
dihydrolipoamide dehydrogenase	TCA cycle	Complete	αKDH_E3 =L-protein	MM	26	0.8988	8	0.576

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
malate dehydrogenase	TCA cycle	Complete	MDH	MM	※	0.8833	6	0.526
thioredoxin	ROS	3' end missing	Trx	MM	※	0.0725	※	0.228
peroxiredoxin	ROS	Complete	Prx	MM	※	0.1334	※	0.078
Fe-superoxide dismutase	ROS	Complete	SOD_1	MM	※	0.3186	※	0.047
superoxide dismutase	ROS	Complete	SOD_2	MM	※	0.2288	※	0.074
ornithine carbamoyltransferase-carbamate kinase	Urea	Complete	OCT_CK _fusion	MM	23	0.965	29	0.771
mitochondrial carrier (MC) family	Other	Complete	MCF_1	IM	※	0.3274	※	0.049
mitochondrial carrier (MC) family	Other	Complete	MCF_2	IM	※	0.1581	※	0.091
mitochondrial carrier (MC) family	Other	Complete	MCF_3	IM	※	0.142	※	0.0883

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
mitochondrial carrier (MC) family	Other	Complete	MCF_4	IM	※	0.091	※	0.797
mitochondrial carrier (MC) family	Other	Complete	MCF_5	IM	※	0.0651	※	0.129
mitochondrial carrier (MC) family	Other	Complete	MCF_6	IM	※	0.7894	14	0.804
mitochondrial carrier (MC) family	Other	Complete	MCF_7	IM	※	0.0175	※	0.042
mitochondrial 2oxodicarboxylate carrier	Other	Complete	ODC	IM	※	0.6325	※	0.189
ABC transporter A family protein	Other	5' end missing	ABC_A	IM	-	-	-	-
ATP-binding cassette, subfamily B	Other	Complete	ABC_B	IM	26	0.8765	※	0.891
ABC transporter E family member	Other	Complete	ABC_E	IM	※	0.0866	※	0.12
ABC transporter F family member	Other	Complete	ABC_F	IM	※	0.566	※	0.209

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
tricarboxylate carrier	Other	Complete	TCC_1	IM	19	0.7502	26	0.705
tricarboxylate carrier	Other	Complete	TCC_2	IM	※	0.0796	※	0.059
tricarboxylate carrier	Other	Complete	TCC_3	IM	※	0.158	※	0.051
LAO/AO transport system ATPase	Other	Complete		IM	※	0.1994	※	0.369
glutmate-aspartate transporter	Other	5' end missing	GLAST	IM	-	-	-	-
sodium/potassium-transporting ATPase subunit alpha	Other	5' end missing		IM	-	-	-	-
carnitine/acyl carnitine carrier	Other	Complete	CAC	IM	※	0.0508	※	0.094
mitochondrial carrier protein	Other	Complete		IM	※	0.1323	※	0.145
mitochondrial 2-oxoglutarate malate carrier protein	Other	Complete	MOC_1	IM	※	0.1059	※	0.151

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
mitochondrial 2-oxoglutarate malate carrier proteinr	Other	Complete	MOC_2	IM	※	0.1881	※	0.098
solute carrier family 25, member 27	Other	Complete	SLC_25A 27	IM	※	0.097	※	0.138
solute carrier family 25 member 39	Other	Complete	SLC_25A 39_1	IM	※	0.2234	※	0.564
solute carrier family 25 member 39	Other	Complete	SLC_25A 39_2	IM	※	0.2233	※	0.31
pyridine nucleotide transhydrogenase	Other	5' and 3' missing	PNT	IM	-	-	-	-
aldehyde dehydrogenase	Other	Complete	ADH	MM	25	0.9903	17	0.791
nadph:adrenodoxin oxidoreductase	Other	Complete		MM	23	0.7219	15	0.664
hypothetical protein	Other	Complete		Unpredictable	20	0.941	※	0.387
hypothetical protein	Other	Complete		Unpredictable	26	0.7461	18	0.613

(continued next page)

Table 2-1. Continued

Protein Name	Category	Complete/ 5' end missing/ 3' end missing	Short Name	Predicted Localization	MTS length (MitoProt)	Score (MitoProt)	MTS length (TargetP)	Score (TargetP)
hypothetical protein	Other	Complete		Unpredictable	13	0.9432	※	0.216
hypothetical protein	Other	Complete		Unpredictable	18	0.9795	6	0.69
hypothetical protein	Other	Complete		Unpredictable	51	0.9959	31	0.92

Proteins detected with tblastn against the Cantina RNA-seq data are marked with an asterisk.

Table 2-2. Comparison of major metabolic properties of mitochondria and several types of mitochondrion-related organelles.

Property	*Homo*	*Fasciolola*	*Nyctotherus*	*Blastocystis*	*Pygsuia*	*Mastigamoeba*	*Cantina*	*Trichomonas*	*Giardia*
Class	I	II	III	III	?	?	?	IV	V
Organellar genome	+	+	+	+	-	-	-	-	-
Organellar genome maintenance proteins	+	+	+	+	-	-	-	-	-
Complex I (membrane-associated)	+	+	+	+	-	-	-	-	-
Complex II	+	+	+	+	+	+	+	-	-
Complex III/IV	+	+	-	-	-	-	-	-	-
F1Fo ATPase	+	+	-	-	-	-	-	-	-
Quinone	UQ	UQ/RQ	RQ	RQ	RQ	?	UQ	-	-
AOX	-	-	-	+	+	-	+	-	-
PFO	-	-	-	+	+	+	+	+	-
PDH	+	+	+	+	-	-	-	-	-
Iron-only hydrogenase	-	-	+	+	+	+	+	+	-
Fe-S assembly machinery	ISC	?	?	ISC	SUF	NIF	ISC	ISC	ISC
TCA cycle	+	+	incomplete	incomplete	incomplete	incomplete	incomplete	-	-
AA metabolism	+	+	+	+	+	+	+	+	-
Protein import machinery	+	+	+	+	+	+	+	+	+
Fatty acid metabolism	+	+	+	+	+	-	+	-	-

Abbreviations: UQ, ubiquinone; RQ, rhodoquinone; AOX, alternative oxidase; PFO, pyruvate:ferredoxin oxidoreductase; PDH, pyruvate dehydrogenase; TCA, tricarboxylic acid; AA, amino acid; +, presence as determined by RNA-seq/whole-genome sequence data and/or biochemical studies; -, absence as determined by RNA-seq/whole-genome sequence data and/or biochemical studies.

Table 2-3. The list and sequences of the primers.

Primer name	Sequences (5'→3')	References
Euk18S-42F	CTCAARGAYTAAGCCATGCA	López-García et al. 2003
Euk18S-1520R	CYGCAGGTTCACCTAC	López-García et al. 2003
Fumarase_F1	TGGGGNGCNCARAVNCARMG	This study
Fumarase_R1	ATNAGYTGNGCRCAVACCAT	This study
Fumarase_R2	GGRTTNACYTTNCCNGGCAT	This study
B27f	AGAGTTTGATCCTGGCTCAG	Jiang et al. 2006
U1492r	GGYTACCTTGTTACGACTT	Jiang et al. 2006

Table 2-4. Quinone composition of the bacterial prey alone (*Cantina marsupialis*-free) culture and the original culture with *C. marsupialis*

	MQ-6	MQ-7	MQ-8	UQ-7	UQ-8
Culture with *Cantina* (%)	16.33	7.80	9.93	2.35	63.59
Culture without *Cantina* (%)	15.83	8.62	14.94	–	60.61

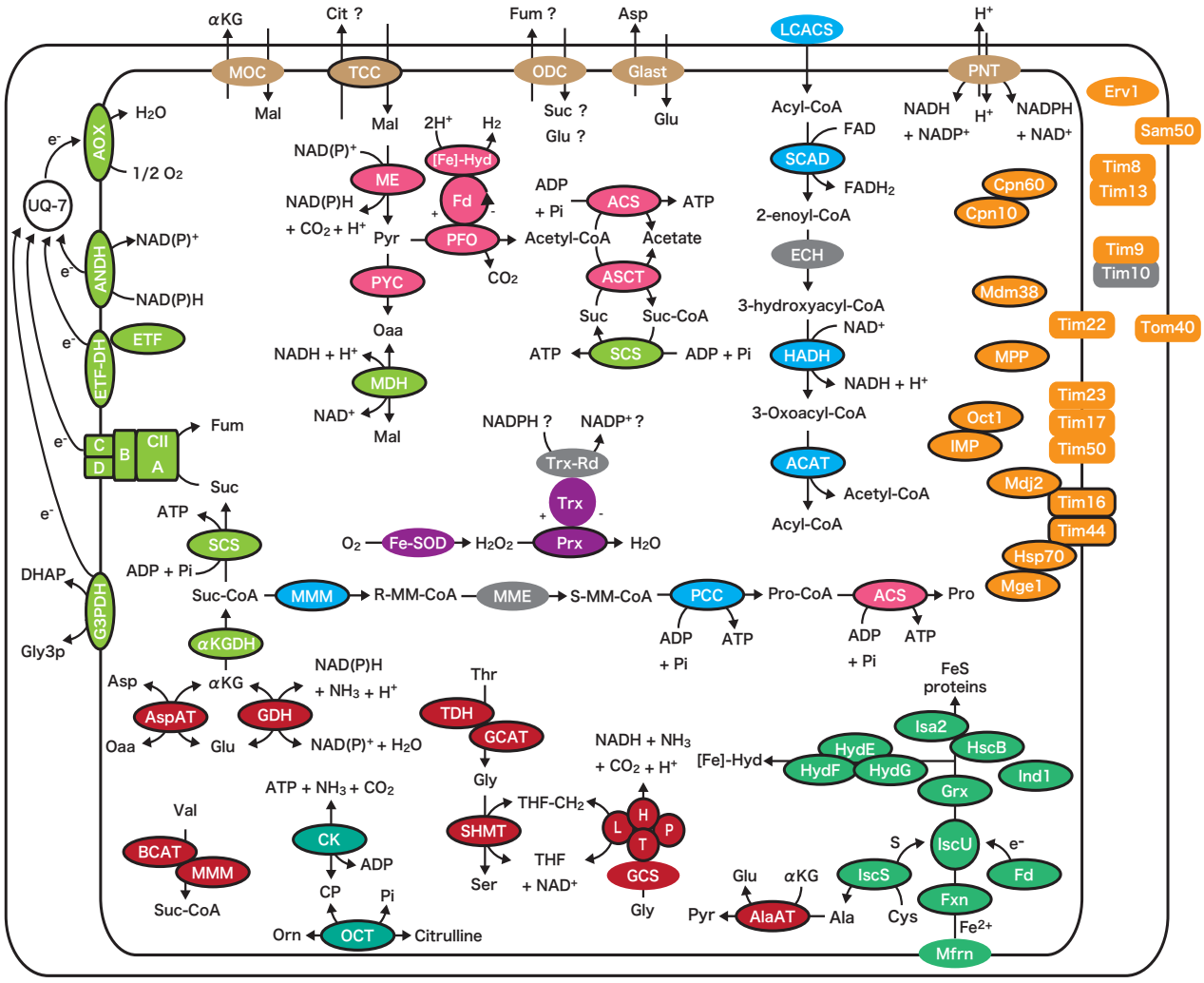

Figure 2-1. Predicted function of mitochondrion-related organelles in *Cantina marsupialis*. Transcripts predicted to encode mitochondrial targeting signals are outlined in black. Electron transport system and TCA cycle (lightgreen): Oxo, 2-oxoglutarate; OxoDH, 2-oxoglutarate dehydrogenase (DH) complex; Suc-CoA, succinyl-CoA;SCS, succinyl-CoA synthetase; Suc, succinate; CII, complex II (succinate DH); Fum, fumarate; Mal, malate; MDH, malate DH; Oaa, oxaloacetic acid; ETF, electron-transferring flavoprotein; ETF-DH, electron-transferring flavoprotein DH; AOX, alternative oxidase; DHAP;

dihydroxyacetone phosphate; Gly3p, glycerol-3-phosphate; G3PDH, glycerol-3-phosphate DH; ANDH, alternative NAD(P)H DH; UQ-7, ubiquinone-7. Pyruvate metabolism (pink): ME, malic enzyme; Pyr, pyruvate; PYC, pyruvate carboxylase; PFO, pyruvate:ferredoxin oxidoreductase; Fd, ferredoxin; [Fe]-Hyd, iron-only hydrogenase; ACS, ADP-forming acetyl-CoA synthetase; ASCT, acetate:succinate CoA-transferase. Iron-sulfur assembly machinery (green): Mfrn, mitoferrin; Fxn, frataxin; Cys, cystein; IscS, cystein desulfurase; Ala, alanine; IscU, iron-sulfur cluster assembly enzyme; Grx, glutare-doxin; HydG, E, and F, hydrogenase maturases G, E, and F; Isa2, iron-sulfur assembly protein 2; HscB, heat shock cognate protein B; Ind1, iron-sulfur protein required of NADH dehydrogenase 1. Amino acid metabolism (Red): Val, valine; BCAT, branched-chain aminotransferase; MMM, methylmalonyl-CoA mutase; AlaAT, alanine aminotransferase; Glu, glutamate; GDH, glutamate DH; Asp, aspartic acid; AspAT, aspartate aminotransferase; Thr, threonine; TDH, threonine 3-DH; Gly, glycine; GCAT, glycine C-acetyltransferase; Ser, serine; SHMT, serine hydroxymethyltransferase; GCS, glycine cleavage system. Fatty acid β-oxidation (blue): LCACS, long-chain fatty acyl-CoA synthetase; SCAD, short-chain acyl-CoA DH; HADH, hydroxyacyl-CoA DH; ACAT, acetyl-CoA acetyltransferase; Pro, propionate; Pro-CoA, propionyl-CoA; PCC, propionyl-CoA carboxylase; S-MM-CoA, S-methylmalonyl-CoA; R-MM-CoA, R-methylmalonyl-CoA. Reactive oxygen species defense mechanism (purple): Fe-SOD, Fe-superoxide dismutase; Prx, peroxiredoxin; Trx, thioredoxin. Protein import machinery (orange): Sam50, sorting assembly machinery 50; Tom40, translocase of the outer membrane of mitochondria 40; Tim8, 9, 13, 16, 17, 22, 23, 44 and 50, translocases of the inner membrane 8, 9, 13 16, 17, 22, 23, 44, and 50, respectively; Mge1, mitochondrial GrpE-related protein 1; Cpn60 and 10, chaperonin 60

and 10; Hsp70, heat shock protein 70; MPP, mitochondrial processing peptidase; IMP, inner membrane protease; Mdm38, mitochondrial distribution and morphology protein 38; Oct1, octapeptidyl aminopeptidase 1; Mdj2, mitochondrial DnaJ homolog 2; Erv1, sulfhydryl oxidase. A candidate urea cycle (dark green): CK, carbamoyl kinase; CP, carbaboyl phosphate; Orn, ornithine; OCT, ornithine carbamoyltransferase. Carriers (light brown): PNT, pyridinenucleotide transhydrogenase; Glast, glutamate-aspartate transporter; ODC, 2-oxodicarboxylate carrier; TCC, mitochondrial tricarboxylate carrier; MOC, mitochondrial 2-oxoglutarate/malate carrier. Undetectable transcripts in RNA-seq data, but predicted components (gray): MME, methylmalonyl epimerase; Trx-Rd, thiroredoxin reductase; ECH, enoyl-CoA hydratase; Tim10 usually associated with Tim9.

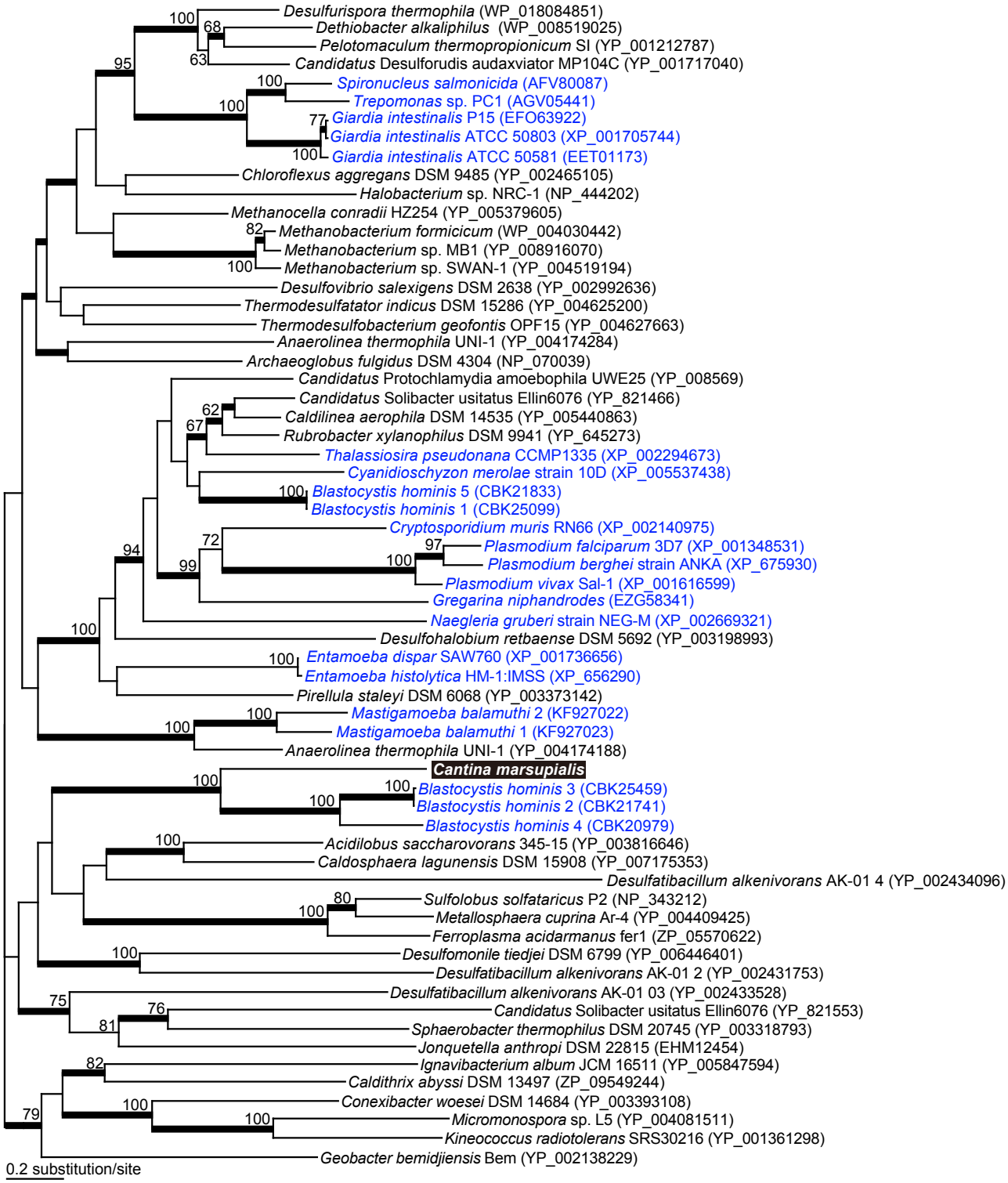

Figure 2-2. Unrooted maximum-likelihood phylogeny of ADP-forming acetyl-CoA synthetase from a broad range of organisms. Bootstrap probabilities are shown for nodes with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Eukaryotes are shown in blue typeface.

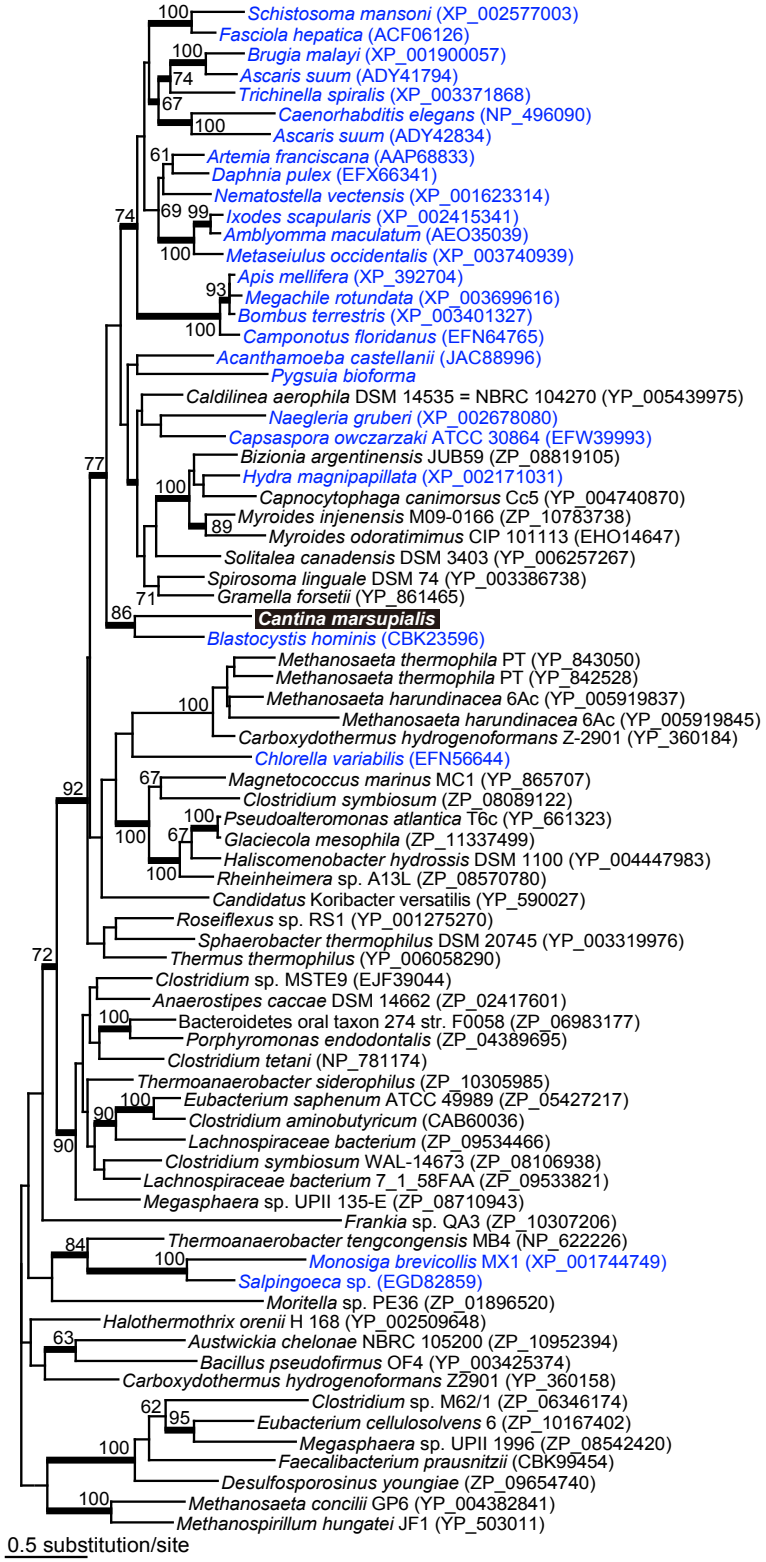

Figure 2-3. Unrooted maximum-likelihood phylogeny of acetate:succinate CoA-transferase (ASCT1B) from a broad range of organisms. Bootstrap probabilities are shown for nodes with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Eukaryotes are shown in blue typeface.

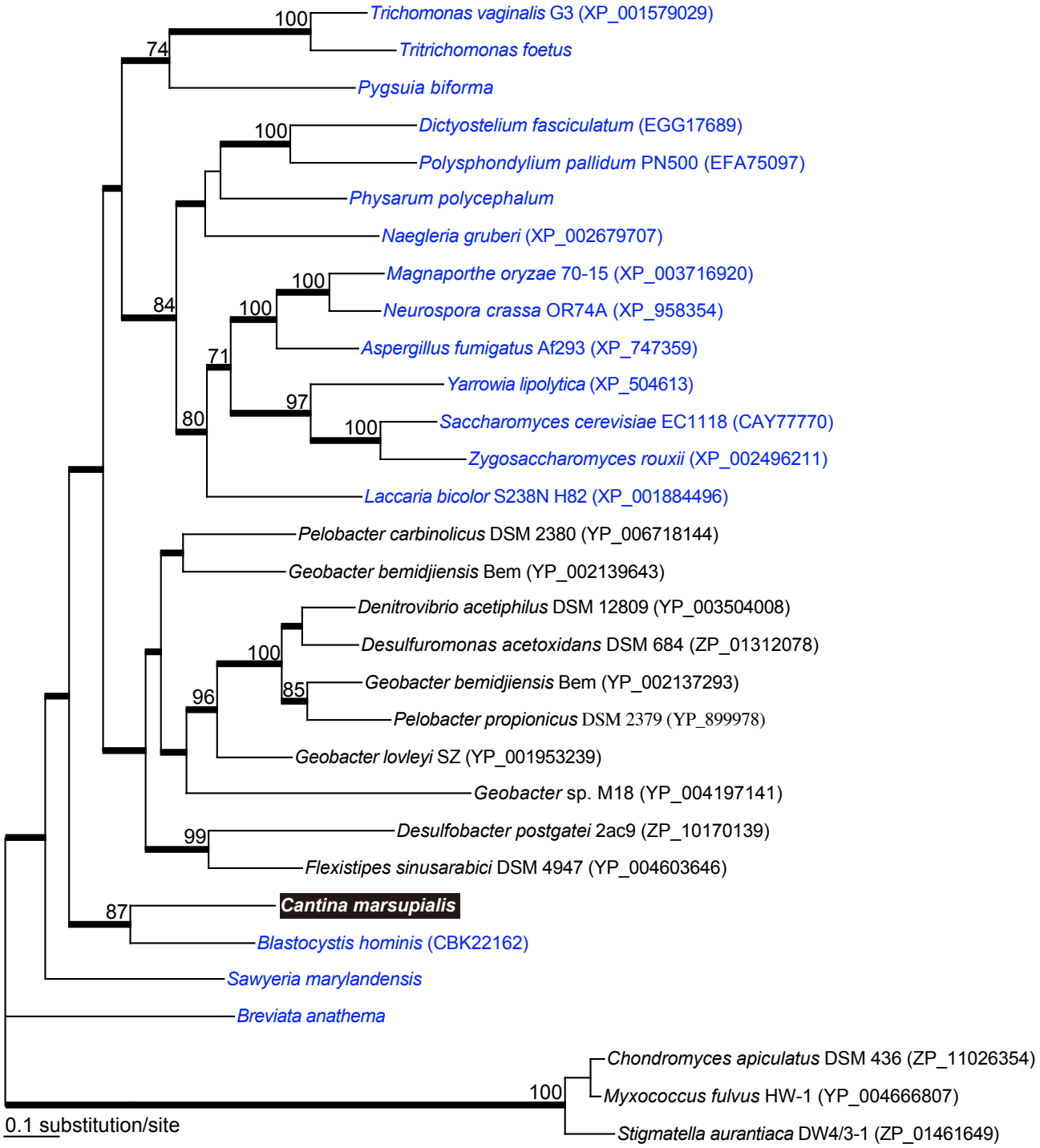

Figure 2-4. Unrooted maximum-likelihood phylogeny of acetate:succinate CoA-transferase (ASCT1C) from a broad range of organisms. Bootstrap probabilities are shown for nodes with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Eukaryotes are shown in blue typeface.

Figure 2-5. Unrooted maximum-likelihood phylogeny of iron-only hydrogenase from a broad range of organisms. Bootstrap probabilities are shown for nodes with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Eukaryotes are shown in blue typeface.

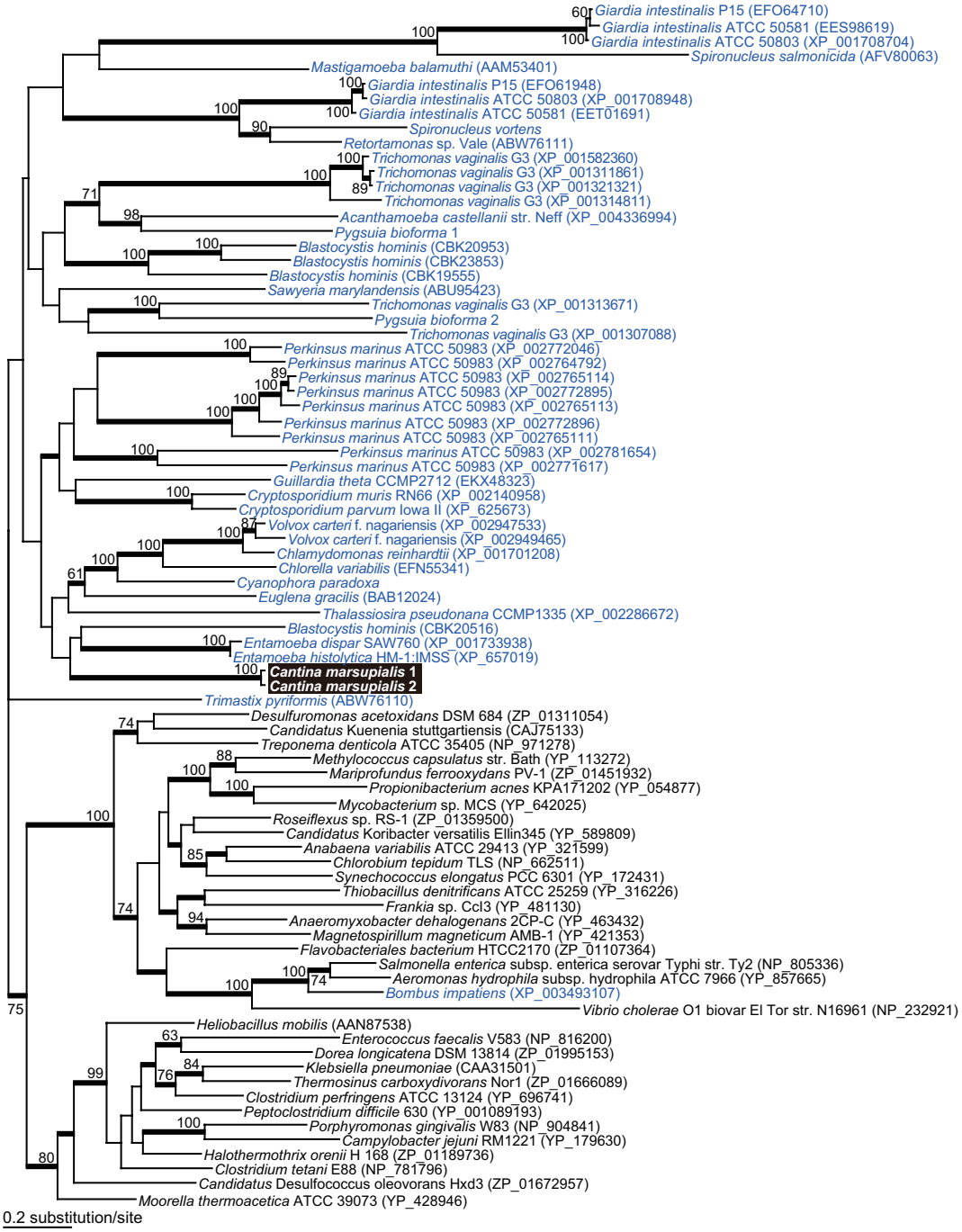

Figure 2-6. Unrooted maximum-likelihood phylogeny of pyruvate:ferredoxin oxidoreductase from a broad range of organisms. Bootstrap probabilities are shown for nodes with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Eukaryotes are shown in blue typeface.

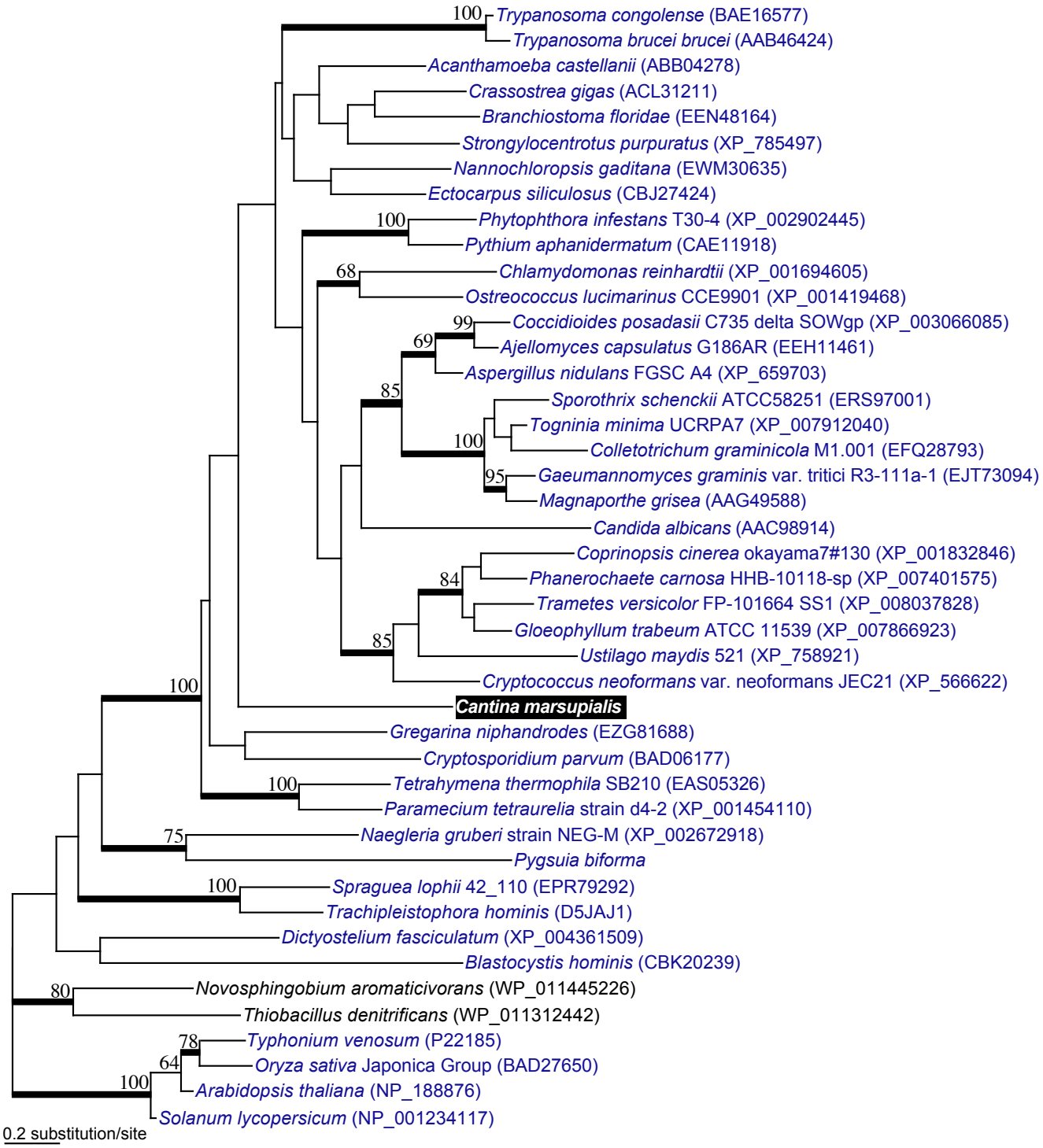

Figure 2-7. Unrooted maximum-likelihood phylogeny of alternative oxidase from a broad range of organisms. Bootstrap probabilities are shown for nodes with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Eukaryotes are shown in blue typeface.

>succinate dehydrogenase (ubiquinone) flavoprotein subunit

MLARIGSNIRQVRLATRQFALYNLRVGQTAENMTDYETAYDIFDHDYDVAIVG
GGGGGLRAAMGLSEAGFKTAVISKIPPMRSHTVAAQGGVNAALGNTHEDSWKW
HFYDTVKGSDWLCDQDAAHYMCREAPAAILELEEWGLPFSRNAEGKIYQRAFGG
QSIDYGKGGQARRTCAAADRTGHAMLHTLYGRSLAFDTDYFIEYFAMDLLMNGE
DCVGVVAMSMEDGTIHRFHANNTILATGGCGRTYFSATCAHIVTGDGMAMALRA
GIPLQDMEFVQFHPTGIYGIGCLMTEGCRGEGGILRNSKGEPFMEVYAPNAKDLAS
RDVVSRAMTTEILNGRGCGPDGEYINLHLDHIPDEIIDERLPGIAETARIFTGADVK
KDPVPVIPTVHYNMGGVPTSFIGEVISPTEEDPHRVVKGLYACGETACPSVHGANR
LGANSLLETLVYGRAVAHHIVENNEPGAEKKPLPENAGMESVQNLHEIRFSNGEQ
SCADLRLKMQKTMQHHAAVYRTGELLKAGVKAVREVYHEFPTLKISDRSAVYNT
DLTEALELQNMLGLSRVIIESAENRKESRGAHAREDFTERDDENWMKHTVTHLTE
DHDVKITYRPVRMYTLDEEEVATVPPMVRAY

>succinate dehydrogenase (ubiquinone) iron-sulfur subunit

MLAKQLRLASNVASRAFSLKIAEETAERIKVFRLMRFDPETDKEPYYQSFPVNID
ECGPMYLDALIKIKNEQDATVAFRRSCREGICGSCSVNIDGHNQLSCLVNYEQNTK
PTTIRPLPHLPIIRDLVPDIGLFYEQYKMIDPWLKRKTPKQEGEKEFYQSEEDRALID
GLYECVLCACCSTACPSFWWNSEKSFFGPAVLQQVYRWVIDSRDEFTSERLEELD
DAFKMYKCHTIMSCTKACPKGLNPGQNIEKLKHLFEEAKANGFQKQ

>succinate dehydrogenase (ubiquinone) cytochrome b560 subunit

MLARAPRLAPRIVRSFSRAQVLPKKPAFRPRSPSLGTMPLDHVAISSVSIRMTS
CVIAGLTCGMGGLYLAGIDTPEIIKNTMMKLEEKKLGCVARFTVAFPFVFHFLGM
ARHKMFTMTSGYGLGSMKAINASSLALIGSAFVLSAALASVKIPVKEEAPVKEE

>succinate dehydrogenase (ubiquinone) membrane anchor subunit

MLARNASKAVRVGVKATRSMHTPPSFHYSLKGLYYAEGVHHHVINVVTAVTFA
GCASIAIPCKYINKPMDYVLAAAFPIFAQLNMVNPIVDYVPLLGKKAAKIGIPALR
MTMTGVSLLTFMGLMKLNICGPGITETMRNVFKVKEQ

>electron transfer flavoprotein alpha subunit

MLSLRSNVRSFTRAFGNLNILVVADHSNAVIAPSTLNAVNAANQLGGPVSVLVA
GNNCGDAAQMASKISGVSNVLVAQDPAYEGMIAENITNAVLAAQEAGNYTHIIAP

ASNTGKNVIPRVAVKLDVAAISDISGIVDEETFIRPTYAGNANTTVRSKDTVKVVT
VRPTAFEKADLGEEQAAIADAPAAEKADSGLSKFVKDEIEISDKPDLATAKVVIAA
GRGIKDAEGMKLCEDLADKLDGAIGATRACVDAGLCPNDVQIGQTGKVIAPELYI
GLGISGAIQHVAGIKDSKVIVSINKDPEAPVYQVSDIGLAEDLFAVVPELIEKL

>electron transfer flavoprotein beta subunit
MKVLVGVKRVVDYAVKVRVQPDKMGVMLKNVKMSMNPFCEIAVEEAVRLKES
GVADEVIAVSIGPKQSQEVLRTALAMGVDRGIHVQTDIRTDQDLQPLAVAKLLEK
VVEKEEPGIVLVGKQAIDDDSNQTGQMLAGLLDWPQATYACGLEINAADQKITIE
REVDGGVQTIEAPMPAVVTADLRLNEPRYATLPNIMKAKKKPLETIKADDLGVDC
EPRLEIVSVEEPAQREAGVMVESVDELLSKLKESGVL

>electron transfer flavoprotein dehydrogenase
MLRLSTVFRQNQLRAAARRFSDIVRDVMEYDVVIVGAGPSGLSCAISIMEESEKL
GKELSVCVVEKGHEVGAHILGGNVLNPKALNELFPDWKDMEAPLDNPVKSDQFR
MFFEKRGIPIPAPGLGNHGNYIVSLGQLCRWLGERAEERGVDIFPGFPAQSVINTNG
VIEGIVTADMGRNTEGEETENFVPGMELRAKQTVFAEGCRGSLSEDLMETFNLRE
NCGPQTYGLGIKEVWEVPEEQCDPGHVQHSFGWPLDYHTYGGSFMYHMKPNLV
AVGLVVGLDYKNANMNIYNEFQRWKTHPDVAPVLKDGTCIQYGARALNEGGFQ
AIPKLTVPGGVMVGCSAGFVNVPKIKGSHTAMKTGMIAGETIAKEITAESEYGIEL
SQYEDSVKESWVWKELKSVRNARPAFQYGMIPGIMHSGLSIVAQGIEPWTLKFNH
EDHECTKTLADSHPIEYPKFDGKLTFNLLDNLILSGVNHEHGQPSHLKVKDDAIVK
ELVETFDYPEGKFCPAGVYEMNEETGMPEINAQNCIHCKACDIKAAGNNIKWTVP
EAGGGGPAYEMM

>alternative NADH dehydrogenase
MESIAMMGNGSARFFSKSINKQRIVTIGSGWAAMHLIQRLNSRKYDSVIVSSRD
HFLFTPLLPSVVGASVPEDLTVRKVNRTTYPMPFRLTPAPNTFINSTVVDVNPLEK
YVVCSDGNKVDYDKLVITVGAQPCTFNIPGVDKHAFFFKEEEDGLLLKEKVMKIK
EAAEKEPEREFKICMVGAGPSGVELTAELSDLFADFKNIKLVLIEMAPCVLSMFHE
DLRVDALENLKARNVEVLLKTAVCELKEDVVITKCEGEESTMPYDACVWTGGV
MQRPLIHKLKERFGLPESRGGLPINGFMRIDGLEDVYAAGDCAASGLPPTAQVASQ
QGEFLAKLLNKHNGKEETEKGPLPEFEFNNRGIMSYVGGNKSVVEVMKKPFSGSF
GNMLWAAIQTLNQGTIPSMIRLDWTLLRSRLFGKRFFDSKNDEK

>alternative oxidase

MLTSTRLFTKGLKPACGPALQKFRLFSSTPEGEDDYLLFHPVYDEAGLDVKVT
HRKPKKPTDYIALGILRAVRITFDKLTGYGPNMTEDKFINRCIFLETVAGVPGISGA
LIRHLRSLRSCTRDNGRIHTLLEEAENERMHLLTMLELKQPSKMFKAATYFSQIGF
FAGYTTAFAFFPKTCHRFIGYLEEEAVLTYGRMIDDLESGKLPAWEGMDAPAIAK
DYWKLPEDATFLDTLKAIRADEAVHRHVNHTLGDMSNDDVNPFKKGSVIHD

>glycerol-3-phosphate dehydrogenase

MLSRAAAATALCGLPALYLSRNGSKISCSESGDESYASRAIQINALKNPNEQFDL
VIIGGGITGSGCALEAQTRGLNVALIEKDDFASESSSRSTKLIHGGLRYLEKVIRNR
DWNQYKLVQKALKERGTLLKNAPHISDWLPVMIPIRKFWDVPYFWAVCFFYDMC
AGFPKRSYFVNAKKAIKLFPQLDQTKEKLFGCMVYHDGQMDDARVCMSVAKTA
SKHGAIVANYVEVTELIKDENGIVIGAKCVDKRTDEEFEVKGKQVLSAVGCFSDD
TREMEDENASKMVQPSRGTHLILPGYLCPEKMGLLDPKTSDGRVLFYLPWEGSTL
VGSTDIKCDAVPHIKPLEEEIDWVFNECKKQISSDIELKREDIKSAWCGIRPLVRDPS
KENTQEICRNHVVHTGKNGMVTVSGGKWTTYRAMAEHAIDIVQEKLGLPKDK

>heat shock protein 70 (Hsp70)_1

MLSALCRQGKRMVRQFGAIRGSVIGIDLGTTNSCVAVMSGSEARVIENAEGQRT
TPSVVGFLPGGDRLVGTPAKRQAVTRPTSTFFATKRLIGRRFDDDEIKKAKEMVPY
EIVEAPNGDAWVQHADEKMSPSQIGSMVLNKMKETAETFLGEDVHHAVVTVPAY
FNDSQRQATKDAGRIAGLEVERIINEPTAAAMAYGMTTDEDKVIAVYDLGGGTFD
ISILEISNGCFEVKATNGDTLLGGEDFDNALYEYLAKEYKKTVGEPLNDAVAVQR
VRDAAEKCKRELDGLKQTDISLPFLSATAAGPVHFETNVSQATFEKLVESLINRSM
DPVKKCLKDAGIQKSQIDEVLMVGGMTRTPMVTKKVESFFEKRACKGVNPDEVV
AIGAAIQGGVLKGDVKDVLLLDVTPLSLGIETYGGVFTRLIPRNTAIPTKKSQVFST
AADDQSQVDVKVLQGEREMAADNNQLGQFTLTGIPPAPRGVPQIEVTFDIDADGI
MQISAKDMATNKEQNIVIKSSSGLSEDDIQQMIDDAEKFAEDDAERRKEIELKNEA
TQLANSAQNTLDEHKDSLSEEDIAVIEEAIKDVEMNVQNSSIEDLSPKVEALQKAL
HKIGEAMYKNVNNNEEQTEQ

>heat shock protein 70 (Hsp70)_2

MLKAISTNFGRFKPFLGCSRLFSTDIRGSIIGIDLGTTNSCVAVMEGKDPRVIENS
EGQRTTPSVVAFHKDGTRLVGLPAKRQAVTNPKTTFFATKRLIGRRFEDPEITKAA

SMVPYDIISADNGDAWVKYDGKQMSPAEIASMVLVKMKETAETFLGHDVKHAVI
TVPAYFDDSQRQATKDAGMITGLNVERIINEPTAAAMAYGMSTDIDQVIAVYDLG
GGTFDISILEISEGCFEVKATNGDTLLGGEDFDDALYEHLAKDFKRTEGFELNDPV
AIQRLRDAAEKCKRELDNLKETDIHIPFISANEVGPIHFSTRVSQATFENLSTELVKR
SMGPVAQCLKDAGIEITSIDEILMVGGMTRTPMVTREVEKFFKQKACKGVNPDEV
VAMGAAIQGGVLKGDVKDVLLLDVTPLSLGIETFGGIFSRLIPRNTTIPTKKSEIFST
AADDQSQVEIKVLQGERDLAIDNKTLGRFDLIGIPPAPRGVPQIE

>Fe-S protein assembly co-chaperone HscB protein
MNRLVSSCKQLKLSHFARFLTNNHTGCFAKCWACDKEMPGCDFFCAGCNKIQP
PSNDETDLDFFDLLEMPHSFKVNKKELDKSYRRLQKKLHPDLFSQTSSTEQDYSQ
MQSARINNAYHILRDPHKRAQYMLEIKGRDAFSSEDVRPDPGLLIEIFGYREILDGC
ETDSQREMFEVENNKRIEELEHRIDDAFSVGDLDLVEKLVIRMQYLRRIKDEMND
LG

>frataxin
MLRILRQFPAQSYVRAFSLADGKYHLVADSIIHHLEECLDPVFDTNEDFEMDNA
AGVLTFGTSAGTYVVNKQAPNKQIWLASPISGPKHFDYDEDSEQWIDHQAGVNM
WELLTEELKTMLSLDINDMCGL

>ferredoxin
MLAINKVLRKFPRQAASLARAFSDEIVHMTFVTSDGKKVAAEGKTGENLVEVAH
KFGVQLEGACECSLACSTCHVILDPKSFDSLGYPCEDEDDLLDLAPCLSETSRLGC
QVLLKPELEGMVAELPPMTRNFYVDQ

>scaffold protein Isa2
MLSLLSRKFSVPLVKSVANGAFLSKAIDVLNVRSFSHGTGVGKVEISEMSDEGC
SGPICLTESAAERVRTLLEGNSDAIGIRLGLDIAGCSGNTYSMNYAFNSEDLLDECE
RVQSQGVDVYIDNKALMTIFGTVMDWQENVLSAEFVFSNPNAASTCGCGSSFQV
E

>hydrogenase maturase HydE
MLSRLTRSFGGIRAFSAVAPFERFLGRSIESGQGVDLSFDELVHLLGSKDDELLN
GLYSHANSIATSVFNNRIVFRGLIEVSNVCQKNCSYCGIRRDADPFRYTMEKDEVV

DAALWAHEKQFGSIMIQSGEVVTRKRMDFILECLDEIMEKTTAIDGKGLGISISLGE
LPKEYYDELIRAGARRYLLRIESSNPELYGSLHPNDGQHTFEQRMDALKDLKDAG
FQTGTGVMIGTPNQTLEDLAGDVIFFRDFGVHMIGMGPYVLESSTPLGKVWIEDR
RKRNPGKSDDQILAEYGDWAFETTTKMIALSRCMLPSANIAATTALQTLNPTGREI
AITRGANVMMPNITPTKYRESYQLYQGKTAVKDDPTKSLKKLESEVESVGKTVG
WGEWGDPPQYYGRGSTFEDAKDASHIPLWTYSS

>hydrogenase maturase HydF

MNQVVSFGKRFFANSVLRTHIGVFGAMNAGKSTLMNILTQSETSIVDNTPGTTA
DTKVSVMQFHELGPVKLFDTPGINETGLLGDKKRQKAWKSLKQADISVIVLNPFD
QETINSANAVVDELKKREKTPHVLLVHNLRKSDIETAGPAVDDILEKVENELLPED
FKYESIALDFHTEEAQTRLINFFNRQGTAKAVNKVPLLPPNAMLDFDSTVLMNIPM
DGETPSGRLLRPQSMAQEQLLRKGVNTMAFRMDLGNGRSDDLDLKKSEEMRFRS
MVDNLKNSGLSLIITDSQAMDLVHPWTLDEDGNETIPITTFSIMMINFLTGGRLPTF
VEGLERFKSLKEGDRVLICEACNHNRIQDDIGTVQLPRVFEKRFGDSVSIEHAFGRE
YESKQMSDYDLILHCGGCMLDAQHMQARLADIEQLGVPITNYGTVLSYIQAPEAL
ERVLVPWNKTLTEGLKE

>hydrogenase maturase HydG

MLSKTIRGTVSALKPAINLALRNFAYAWSPELENDHVIPTKDQIINPALINKHLE
ETKKYAGDKERIRDILQIAEERALLKKVDPTQNMGSEYVQGLSLEEAATLLNMDE
NNESLMQDLFDTALTIKREIYGNRIVLFAPLYLANYCMNSCTYCAYRGENKNIERN
AMTQAELIEEVKALQELGHRRALVLTGEHPKYPFDSFLDALKTISEVKTEPYGSIR
RINVEIPPLSVSDFRRLKDTNVVGTYTLFQESYHEGIYKQMHPYGPKSDFEYRLQT
MDRAQIAGIDDVGIGALLGLHDYKYEVMAMMMHGSHLESTYGAGPHTISIPRMR
PADGAPDSEQPPAPVDDNAFRKLVAVIRCAVPYTGMILSTRESEDMRRQLLHLGV
SQFSAGSRTEVGGYVKGDLEGELAGDYNDNVKAGQFSLLDHRCLDDVVKDLLK
DGFVPSWCTACYRLGRTGEEFMKIAKCGEIKNFCHPNALLTLQEYLDDYASDETK
SLGLDVIAQESQVFDKEKVKEIFKTKMQGIKEGERDLCL

>mitoferrin (Mrs3/Mrs4)

MDDDWEEWDPSKGSFVHHMIAGSCAGLMEHAAMFPVDTLKTHLQATGSIKEIK
MLMRRPKKLQNMTVGTMPRPCPNKMMNSSLKNSLKPEWLKQRCKNFNMANRF
RLWRGVSSMFIGCVPSHAAYFSIYESAKQILGANKEGHHPIAAGSAGIIATLAHDAI

ITPMDVIKQRLQLGFYNNVPDAMRTIIRTEGLKAFYVSYPITLAMNVPFAALMVAA
NESSKTILSGGNSENCSMPVYFTSGAIAGAFAGALTNPLDVIKTKLQTQSVICSCPR
LQNTIELEAASNPALNGFKEAAAHIMKTGGYRGFLNGIIPRTVFHSSSWAIAWATY
EFIKKSLRPSKPIDKNKKF

>cysteine desulfurase IscS

MLSRQVAKLARKSALPAIQALRAYSFSIKGESIEGKPAYLDFQATTPMDPRVLDE
MMPYLTEHYGNPHSRTHEFGWEAEAAVEDGRAKVAKVIGADPKEIIFTSGATESN
NLALKGVANFYKKRKNHIVTTQTEHKCVLDSCRHLEREGFDVTYLPVQPNGLINL
QDLDNAMRPDTSIVSIMGVNNEIGVVQPLKEIGEIVKKHKAFFHSDCAQMFGKMPI
DVNDLKIDLMSISGHKIYGPKGIGALYVRRRPRVRLEPLISGGGQERGLRSGTLATP
LVVGIGAAADIALNEMEDDHRWIEFLANKLKNGLHERIPHIQINGDENERYIGNLN
VSFAFVEGESLLMSLKNIAVSSGSACTSASLEPSYVLRALGVNDELAHTSIRFGIGR
FTTEDEIDFTIDLCARHVERLREMSPLWEMVQEGIDLDSIKWDSH

>iron-sulfur clusters assembley enzyme IscU

MLAIRNVLSKPSRFLAAAPRVFRLYHDNVIDHYENPRNVGSLDKNSDNVGTGLV
GAPACGDVMKLQIEVDDNGTITNAKFKTFGCGSAIASSSYATELLLGKNLNDADNI
KNSDIASHLKLPPVKTHCSLLAEEAIKAAAEDYRQKQEAKKSA

>glutaredoxin

MLRVVQRSIRACSVIPALFRASAARNFHSTPFALENKDNKGTHPDFQPKKKLPYD
TDEESLINMVKTHINKYPIMLYMKGTPTQPECGFSYQVVKILQASNVNFSSIDVLK
HPLVAMATRKVSDWETFPQLFVKGEFLGGCDVIGEMYETGELKEVFEEYDLILPPP
EEE

>P-loop NTPase Ind1

MLTSFIKRVSRPAVFSRLFATQDDVFSVLENIEVINNKHNLVSAGLVDSVSIDGD
RVGLSLSFPTPAFPKKDDLVEAAKSSLINGLADVSNVDVNTSIRTPIKKTKIDTMKN
IEHVVLVSSCKGGVGKSTVSVNLARSLANLGARVAIFDADIYGPSLPFMVNVDKK
VRYNDEKTRYLPAESDDMKCQSYGFMAADGSAGEGQAAIMRGPMVSNVIDQLL
KFTEWGELDYLVIDCPPGTGDIHLTLSQMLEISGGVVVTTPQKLSFVDVVKGIEMF
QKLNVPTLGLVENMSYFECEHGTKYRPFGMGHSEEIQERYDVGETFFLPLDGDVS
LSSDSGHCVSQVFPDSPAGQEYKRLAETVVEETLLQSFTGSTVPRVSSVVDKGIVL

RWITDEGAEEKVLNPAEVRRKCRCALCIDEMTGQRTLKDEDVPDDVEAVLVEPK
GNYAVHLQWTDGHTSLMPYRILEEMALTSEN

>pyruvate carboxylase

MLSQIKARQLAPAARSLGFLRETKAPFKKVMAANRGEIAVRIMRATHELGMKS
LGIYSHEDRYTQHRFKADESFKVGAGMSPVAAYLAIDEIIEVAKKHGVEAVHPGY
GFLSENATFAQKLADNGITFVGPTIENLKVFGDKTLARKAAIESGIPVVPGTEEACA
TLEEVKKFTDENGFPVIIKAAFGGGGRGMRVVRAEEELPDAFAGASSEALAAFGN
GTCFVERFVEKPRHIEIQILADKEGNVVHLYERDCSVQRRHQKVIEMAPSRGLPDE
LREQLYADAIKICKHVNYFNAGTVEFLIGPDNKHYFIEVNPRVQVEHTVTEEVTGV
DIVRAQMRLAAGETLEDLGLTQDKIECNNFAMQCRITTEDPQKNFAPDSGVIQVY
RAASGKGVRLDEGPGFTGANITPHYDSLLVKITCSDRDFDRTLIKMRRTLQEYRIR
GIKTNIPFLSNVLNHHDFVKEIPNTSFIETHPEIMETRDPSLNRGTKLLRFLGDMAV
NGPPKALGCSGSPVSVDPSPLPLSTTPKEELKGWRDVYKKDGPEAFAKAIREHPGL
LLTDTSMRDAHQSLLATRVRTKDLKAVAPYYADNMQNLFSLENWGGATFDVAM
RFLKEDPWERLSQLREEIPNIPFQMLLRGANAVGYTSYPDNLIYKFCEKAVDEGM
DVFRIFDSLNYLENMKLGIDAVGAAGGIIEAAVCYTGDVARDGDNKYNLDYYLN
LVRELKEMGTHILGIKDMAGLLKPGAARKLVGAIREEFPDLPIHVHTHDTACTGV
ASMVECAMAGADVVDAAMDSVSGMTSQPSLGALVAALENHERDTGIDFHNAAR
ISQYWEECRGLYQGFESGQKAGSTDVFMHEMPGGQYTNLQFQANQLGLAGRWP
AIKKAYAEANKLMGDIIKVTPSSKVVGDLAQFMVQNELSYDDVMEQAETLSFPQS
VVEYFQGYLGIPFDGFPEPLRSKVLKGKKLPNGKDCFEGRPGAELPAFDFGAATK
NAIKIFPDSNDLDVLSYAMYPQVFEEWKAFETEYGKVDFLPTRSFVEPMKPGDEV
CCPIDEGKDVYIRLQSIGDVDENGDREVTFILNGERRVQKVHDNNAEVSVVSRAK
ADPNNDLHVGSPMPGVCVSCAVKVGDELEVGDNICTLSAMKMETVVTAKAAGKI
KSCPIAVGDNLEGGDMLIEFE

>iron hydrogenase_1

MLTLSSKLLKTHLSRALVARSGLLLRSFSNHHHNPIDIEDARFIDHSSDAIKFDFHD
CIACGFCIAACEEQADVLTFQEVSGLGEIPRTISGEYLADTNCIECGQCANVCPVDC
ITEVDHLTRVENAMADPDKIVVLQAAPSTRVAIAELFGVRPGEIATEKMVDACKK
AGFKFVFDTNFAADLTIQEEVKEFLERFNDPDSVLPMFTSCCPGWINLIEQKYPELI
PHLSTCRSPMMMLGPVIKTFWANKMNYDPSRIYSVALMPCTAKKGESDREEMFM
EDGSRCIDAVLTTRELAKLLKNKGIHTWNELGGSKFDSTLGESTGGGAIFGVSGGV

MEAALREVWQQLTKKPLNDLEMDVFFHDVRGIDEHVKVFELDLKAYGVPRVVR
GAVVHGANHASNLLRSLKDGEDKYGRLDFIEVMACPGGCISGGGQPKHQGEQAP
SRRFKAIYKIDRESSNRYSGANMELSKLYKEFLDKDEHLRHELLHTSYSPKPTKD

>iron hydrogenase_2

MLSLIRASSLRRVVRPAFGAVRFFSSSDSESDWSDSDESGSGSEEGEGARMTDNS
SPSIRFNFEDCIQCGGCIAACEESANVLHFGETEDGDEIPATISGALLNDTECISCGQ
CATACPVGCITEVDSVQKVMEMLDSGKTVVLQTAPAPRVAIAEEFGRPAGEISTG
KMVSAAKQAGFNYVFDTNFAADLTIMEEAHEFIDRVTNGGVLPMFTSCCPGWVN
LVEKRHTELMPNLSSCRSPMMMEGSVIKSYWAEKNGVNLEDIYTVALMPCTAKK
DEITREQMFNEVGPSVDNVLTTREFAKVLKNRGIDWESLSDEGAFDNSLGESSGA
GVIFASSGGVMEAALRSAYEVISGEELPDINIEAARGIEGVKQFTVPIKDMEVKCAV
ISGASNADEFMQKVIAKEDGYDQFHFVEVMACPGGCINGGGQPRGAGEAGVKAR
LESIYSIDADSPVRKSHQNVEVQELYNAFLEKPNSHKAHELLHTSYKNRKVEA

> pyruvate:ferredoxin oxidoreductase_1

MLSATTQIIKKSARVAPVMARGIIKSVDGNEAAAHAAYACSDSSFIYPITPSSTM
GELVDLWRAKGRVNAFGNVMAVSELEHEGGAAGALHGALSAGAMATTFTASQG
LMLMLPNMYKIAGELMPCVVHVAARALAGQALSIFGDHSDVMAARTTGFCLLG
ASTVQEAADLAVVSHIATLEASLPFIHFFDGFRLSHEINKIDLLENEQIKALLPADKI
KAFQDRALSPAHPIQKGTSQGPDIFFQMVESSNDLYNAVPEHVQAAMDKVSAVTG
RPIKLFDYEGAEDAEDVVVVMGAGAPICEEASKYLNARGGKTGVLKVRLFRPWS
VKDFAAALPKSVKRIAVLDRVKENGAVGEPLYEEVATSLMEEKMAGNLDVIVGG
RFGLGSKEFDPAMARACFDNLHAEEPKNHFTVGINDDRTHTSLEVGESFSFVPEGT
KQCQFWGMGSDGTVGANKDAIKIIGDNTDQYAQAYFVYDAKKSGGVTTSHLRFG
PEPITSSYLVQQADFIGCHQPGYLTKYDVTQALSENGVFVLNSNLSDEELFETMPN
AVKKALADKKAKFYVVDAFKVAEEAGLKGRINNVMQTAFFKLANVLPMEEAIGL
LKGAIEKTYGAKGQKIVDMNKKVVDMSLDAVREVSVNAEWSTLPADGVGFAPIG
DKFVDEVVNPIMGLKGDDLPVSVLPRAGVFPTGTAKFEKRGIATDIPEWIKETCTQ
CNQCAVMCPHAAVRPFISSKDESANAPGVWDTLNFRGKEAKDMDFRIQVSPFDCT
GCGVCVAVCPTDSLKFRPAADGIEKESENWEYAVGLENRGSLFTPDTPKNTQFQQ
PLLEFSGACAGCGETPYVKMITQMFGERMVIANATGCSSIWGGTAPSNPYTTTAS
GMGPAWANSLFEDNAEYGYGMRMAQQTRRAQYVNTVNEALEKGNMTAETREK
LSKWVEVSNNGEETLALYKELSAELDAQAGNDEFMTKLAERKDQLIATSQWIFG

GDGWAYDIGYGGLDHVIASGENVNIVVLDTEIYSNTGGQASKSTPMGAIARFAEA
GKEVQKKDLGEIAMTYGHVYVASIAMGADMKQAAKALREAEAYDGPSLILCYSH
CLGQGIAGGMSNGPAQQKQAVKAGYWPLYRFNPELKAQGKNPFVLDSKKANTD
GMMDFLKNENRFAQLMRDDPENAAVLNEQLVQFRTEKVAKYEMMAAAGKKAK
KSKKSKKKN

> pyruvate:ferredoxin oxidoreductase_2

MLSLTTRQFARKSLLAGPIALNNVRSYVKSVDGCEAAAHGAYACSDSSFIYPITP
SSTMGELDDLWRSNGMKNAFGDVLSVSEMQSEGGAAGALHGALSAGAMATTFT
ASQGLMLMLPNMYKIAGELMPCVVHVAARALAGQALSIFGDHSDVMAARTTGF
CLLGASTVQEAADLAVVSHIATLEASLPFIHFFDGFRLSHEINKIDLLENEQIKALLP
ADKIKAFQDRALSPAHPIQKGTSQGPDIFFQMVESSNDLYNAVPEHVQAAMDKVS
AVTGRPIKLFDYEGAEDAEDVVVVMGAGAPICEEASKYLNARGGKTGVLKVRLF
RPWSVKDFAAALPKSVKRIAVLDRVKENGAVGEPLYEEVATSLMEEKITNLDVIV
GGRFGLGSKEFDPAMARACFDNLHAEEPKNHFTVGINDDRTHTSLEVGESFSFVPE
GTKQCQFWGMGSDGTVGANKDAIKIIGDNTDQYAQAYFVYDAKKSGGVTTSHL
RFGPEPITSSYLVQQADFIGCHQPGYLTKYDVTQALSENGVFVLNSNLSDEELFET
MPNAVKKALADKKAKFYVVDAFKVAEEAGLKGRINNVMQTAFFKLANVLPMEE
AIGLLKGAIEKTYGAKGQKIVDMNKKVVDMSLDAIREVSVNAEWSTLPADGVGF
APIGDKFVDEVVNPIMGLKGDDLPVSVLPRAGVFPTGTAKFEKRGIAADIPEWIKE
TCTQCNQCAVMCPHAAVRPFISSKDESANAPGVWDTLNFRGKEAKDMDFRIQVS
PFDCTGCGVCVAVCPTDSLKFRPAADGIEKESENWEYAVGLENRGSLFTPDTPKN
TQFQQPLLEFSSACAGCGETPYVKMITQMFGERMVIANATGCSSIWGGTAPSNPYT
TTASGMGPAWANSLFEDNAEYGYGMRMAQQTRRAQYVNTVNEALEKGNMTAE
TREKLSKWVEVSNNGEETLALYKELSAELDAQAGNDEFMTKLAERKDQLIATSQ
WIFGGDGWAYDIGYGGLDHVIASGENVNIVVLDTEIYSNTGGQASKSTPMGAIAR
FAEAGKEVQKKDLGEIAMTYGHVYVASIAVGADMKQAAKALREAEAYDGPSLIL
CYSHCLGQGIAGGMSNGPAQQKQAVKAGYWPLYRFNPELKAQGKNPFVLDSKK
ANTDGMMDFLKNENRFAQLMRDDPENAAVLNEQLVQFRTEKVAKYEMMAAAG
KKAKKSKKSKKKN

>acetyl-CoA synthase

MLSKLIQNTTFRSVGAFSALRQFSSSKPTIDGFFEGKSYVVLGASTKVGSLGWSI
ASNFNATFEGDSFFVNPRGGELLGKELYKSLNDLPMVPESACICTPAKQAVKDAK

TCIDMGVKNLIIIAGGFAEADEEGFKAQEELVKMCQESGTRLVGPNGMGVFDNNT
GTNTLFISRELLELPEAGPISVITQSGALGCSLMNLMCQHEKKDWVSRFISFGNAAD
VNENDSLEYLGRDENTKQIWTYLEGIRDAPEYLRQLRSTCGSKPVLTLKASRGDA
GAHASGSHSASLAANDEVCDYLLRQAGSLRIETWPEFFQAGVGLLGQSLPKGNRV
GIVTNAGGCGVMAADAVEHNELELPMYSEKTVSEFYDTMPSFFQCVNPMDLTGS
ATTEQYLDATELALKDDNIDSVLLMIQPSGPVLDKPEEFCRKIIERFGDKKFDKPIIP
VIFGGNGDYDKLYNDKLISAGFPMCMSPESGVRTIRMMRDYAAFLEREEKKIEAG
VPVVPEMPTANPEIQKIIDGARADGRKVLLEPESRDIFRMSGLPVPADVLAHTPEEA
VAFWESVKTPLVMKLVSPQVIHKTDEGAVFVNKNNAEDVYETAKHLFDKFEGRD
VRGVLLYEMVPHGTEMVIGMNTDDVFGPLVLVGAGGAMVEILNDATFNMCPTD
RFDADEMLNNLTHQALLEGYRGSPAVNREEMSDMIVKISNLAAAHSEDIKEMDA
NPIIACPDGDHWAVDARILLH

>acetate:succinate CoA-transferase 1B

MLSNIVKRIGVRAFTQAEIHQACAAINRVNIDEKSLSLVEQMRKIKGDKQPIKTS
MAEAVKAVNSGDRVYVHMGTPVALLDELTNYGKDNLKDIELIHLLMMGNAPHT
QHPDVFRSNALFISGNVRGAVGCGESAYTPIMLHQVPHLFEHNVQPCDVSLVQVT
PPNEEGYCSIGYCLDATRSAIDNSKIVIAQINDKLPFCYGEGMIHVSHFDYMCEHNE
PIMDVPPAKIDDNAAKMGKWIADNLVEDGATLQIGIGGIPDAVTANLKNHKHMG
VHTEMFSDGIVDLVDCGAIDGSMKNFDKGLVTTTFVMGTDKTYNFIDRNPMVKF
REVAFTNNPANIARQPKFTAINSAIEVDITGQIVSDSIGTRAFSGFGGQVDFIYGASM
SEGGKAVIALPSSSKGISRIVPMIKPGAGVVTTRGHAHHIVTEHGGVDLFGQNLQE
RARLMISIADPAHQEELHQAAFERLGVLV

>acetate:succinate CoA-transferase 1C

MSAALLSRVGRSSLQSKIMSAKDTVKFFKDGQKMVWSGFTPAGYPKAVPTALAD
HVEQTGEKMGFELFVGASVGAETEDRMASLGMIKRRWPYQTGKNIAKGINNGEIS
MGDTHLSMFAQNIEYGFYSTETPKGKNLDIAIVECTEILPNGGLVLGTGIGMSPQA
VSSADKIIIEVNTSLPVLRGLHDINMEVLPPHRQPYLISRVDDRMGTDYLPIDTDKVI
AVVESTMPDNGRGLGPADEVSQTIGDHILEFFKHEVNAGRLPENLLPLQSGVGNIA
NAVTGGLCHGEFEDLTVWTEVLQDTMLDFFDSGKLKYASSTSLSFSPEGFEKFYN
NLEFYMPKTVLRPQHISNHPELIRRLGVIAMNTPVEIDMYGHANSTLVGGTKMING
LGGSGDFLRNAYLSIMHAPSARGTKTDPTGISTIVPKASHIDHTEHDLDVVVTEQG

LADLRGLHPRARAKEIITKCAHPDYIDQLLDYYNMAEKKCYAMGAGHEPQDLSQ
VFKMHENLANPEIMSMKATWEH

>succinyl-CoA synthetase alpha subunit

MLARSSKIIRSFSSTAKVWVDKNTKLMVQGFTGGQGTFHSQNSIEYGTQVVGGT
NPKKAGTMHLDRPVFATVEECKKETGANATVLFIPPPLAAASIMEAIEAEMELIVCI
TEGIPQQDMVKVRYALDRQEKSRLIGPNCPGIIKPGECKIGIMPAYIHTPGKIGIVSR
SGTLTYEAVNQTTKFGLGQSTVVGIGGDPFNGTNFIDVLEKFRDDPQTEGIIMIGEI
GGEAEEDAMAWWAEHGDANKPIVGFIAGRTAPPGRRMGHAGAIISGGKGDANSK
ISAMESVGAVVVDSVSDLGEAMFKRMSDL

>succinyl-CoA synthetase beta subunit

MLSRVVRPLVRATGVRNLNLHEYQAKYLLEDYNVRCQKGKAAATPEEAAAVA
EWILTENPAADVVCKAQVHAGGRGKGHFNTGYEGGVQLIKSPEEALATSKEMLG
NYLITKQTTAEGQFVSKVLLNEGITINNELYLAILMDRSMDGPVVVASTEGGMEIE
EVAEHTPEKIFKEPIDIMKGIQAEQTERIARALGFESEKQIKDCQQQIEGLYNLFIGT
DATQVEINPLAVGGVPGFDTDLVYCVDAKLNFDDFAAFRHQELFEQRDVTQEDA
RDVLADDLGLNYIGLDGNIGCMVNGAGLAMATMDVIKLHGGEPANFLDVGGSA
TKEMCCEAFKLLNEDTQVEAMLVNIFGGIMRCDIIAEGIVAAVQEVGLRFPLVVRL
EGTNVEAGKKIIEEAALPGVISATDLNDAAKKAVKAAEESRQ

>NADP-dependent malic enzyme

MATVAHTNVILTNKQVLRNPLYNKGTAFTEEERKHLHLTGLLPPAVQSWDLQAK
RALAMVRSKTTPLEKYIFLSDLQDRDEDLFFKVLIENVKELMPIVYTPTVGEACQK
FSHILRHPRGLFISIKDKGHIAEILANHPQRDIAAIVFTDGERILGLGDQGAGGMGIPI
GKLSLYTACAGVNPAKCLPVTIDVGTNRETLLEDEMYIGLKQNRVRGAEYDELID
EFIMEAQKRWGEKVLLQFEDFGNLNAFRLLETYRTKCCTFNDDIQGTASVVVGGL
YASIPVTGKAIDQHKFLFMGAGEAGVGIADLIAAAIMETTGKTIDEARKQIFLFDSR
GLVCKSRTGLAHHKLAYAHDVPQQTSFLEAIKEMKPTGIIGVSATPDVFTEDVCTE
MGKLNERPIIFALSNPTHKAECTPTTAYTCTQGRALYASGSPFDPVEYEGKTLIPGQ
GNNAYVFPGLALGVIAAGAKRIPDELFLIASKALAEKVTEEHLSHSTLYPPLNQIRS
VSAYIAMKVAEKVYDLGLATIPRPDDILAAIESEMYDASEKSFI

>propionyl-CoA carboxylase alpha chain

MFAKTAKPVQNFVRSVISPAGEKLFDKVLIANRGEITCRVIRSCQALGIKTVAVYSE
ADAEANHVKMADEAVCVGPAASAESYLVMDKIIDAVKQTGADAVHPGYGFLSE
NQAFATALAENDVAFIGPSNYAIEAMGDKIQSKIVAQAAGVNCIPGDNRVIKDAEE
AVVVANQVGYPVMIKASAGGGGKGMRIAYDDAECREGFVLSTEESKASFADDRV
FIEKFIEDPRHIEIQLLADKFGNVVPFPERECSVQRRNQKVLEESPSCLLDDETRVA
MGEQACALAKEVGYLTAGTVEFLCDKNKNFYFLEMNTRLQVEHPVTEYISGVDL
VEQMIRVACGYPLPDELVNGPRPLPIKGWAIESRVYAEDPFRGFLPSTGRLMSYAE
PTGTIEEGFRIDTGVTEGSEISMFYDPMICKLITHGKDREEALVKMQKALDTYIIRG
VNHNVPFLRDVITAPRFQSGKITTGYIAEEYPEGFSGVTLVGEKRARAAAVAAIME
EMRLNKATTIDGQLRTFVPKVAETMFVTMDADEVFEVTLSHSESGIKAVVGENTF
DFQGIDWAVNSPLFSAEVSGVDFDAQLMNRLPEGYALYMDGAFMNCAVRNAKE
HEYAQYMVPKPEIDFSKMLVSPMPGTLISIDVEEGDSVEPGQKVCVVEAMKMQN
VLVAEKKGIVKKVCAGPGETLACDQLIVEYEN

>propionyl-CoA carboxylase beta chain

MLASSLRTLRKAPIAAFSTAIPTPFERKVIFNERLNAAREESKLGGGAKRIAKQHE
RGKLTARERIDILLDEGSFREVDALKKHNCHEFNMEKQRPAGDGVVTGYGSINGR
LSYVYSQDFTVFGGSLSKAHAEKICKIMDSAMKQGAPVIGLNDSGGARIQEGVESL
AGYSDVFQRNVLASGVVPQISLICGPCAGGAVYSPAMTDFIFMARDTSYMFVTGP
EVVKTVTGEDVTQEELGGATTHARKSGVIHNAFDWDVEMIKQTREFFDYLPLNN
KQEVPQVECTDPTDRYEVALNQIVPSDPNVPYNMKEVVKKIVDDQDFFEIQPEFA
KNIIVGFSRMNGRTVGIVGNQPCELAGCLDNNASIKGARFVRFCDAFNIPIISLIDVP
GFLPGTDQEHNGIIRNGAKLLYAYCEATVPKLAVITRKNYGGAYCVMSSKHLRGD
VNYAWPSAEIAVMGAKGAVEIIFRGQDLVKAEAEYTDKFANPEVAAARGYVDDV
IEPATTRQRLCEDLEILRSKQLENPWRKHGNIPL

>methylmalonyl-CoA mutase

MLSSVSKLSIRAARAFSAVPINQQWADMVQKELKGKKTAEQLVTKMPEGINMK
PLYTAEDVANIKKVDQLPGEFPFTRGVYASMYTARPWTVRQYAGFSTAEESNAFY
RKNLAMGQQGLSVAFDLATHRGYDSDHPRVVGDVGMAGVAVDSVEDMKILFDK
IPLDKMSVSMTMNGAVLPILAFYVVAAEESGVDQKLLTGTIQNDILKEFMVRNTYI
YPPKPSLRVIEDIMGYLGTNMPKFNSISISGYHMQEAGADAMLELAFTLADGVEY
VKCAQKAGLNVDQVAPRLSFFFGIGMNFFMEIAKLRAARQLWAKTMKDLGAENP

KSMLMRTHCQTSGYSLTEQDPYNNVMRTTIEAMAAVMGGTQSLHTNALDEALG
LPTEFSARIARNTQLVIQKETNITNVADPWGGSYLIEALTNELVEGAEEIMKEVEEL
GGMANAIESGMPKRRIEASAAAKQARVDSGMDAIIGVNTCQLEQEEPFDVLQIEN
SAVRQAQIDRLTSIKATRDEEAVQAALAKLNASSKMNESTSAGDHEHNLLGLAVE
AARLRATLGEISDALKAEWGEHNPQQEIVTGAYSTQFQESSMESEFTETVDMVKK
FEEEHGRRPRIYIAKMGQDGHDRGAKVTASGYADLGFDVDVGPLFQTPEEAAQA
AIDADVHVVGASSLAAGHKTLIPAMIEELKKRGGEHITVIAGGVIPPVDYDFLYDA
GVSAVFGPGTRIPHSAQVMIDIINKKN

>acetyl-CoA C-acetyltransferase
MNIAPIARSFAKLAARDAVIVSAVRTPIGIMGGDLASLTAPQLGSIAMREAVARA
GIDASEVQECIFGNVLTAGSGQAPARQAVIGAGLNLDTPTTTINKVCASGLKSVM
MAAAMIKAGEQECIIAGGMESMSNSPYVLPDARNGARYGHKTMIDTVINDGLWD
PYNNMHMGMCAEHVAAEHGFSREDQDAFALESYRRANEAWAAGRFDKEVVPV
EVPVRRGDPKIIRMDQEPGNLRVDKVAGLRPAFTKDGTVTAANASGINDGAAALI
VMSYERCQALGLTPMARIVSTADAAIEPINFPVAPEPAVRRALSKAGMEISDIDFHE
VNEAFSVVTLANSKLLGLDMDTVNVNGGAVAMGHPIGASGARVLTTLLHVLEQQ
DASLGHASICNGGGGAAAMIVERL

>hydroxyacyl-Coenzyme A dehydrogenase
MLSRVQVGLRQFAQVQKAAVIGLGLMGHGVVQTIAESKIPVIACEINDAAIERG
MSMIDSSLQQVVNRNVKKDKMTAEEGKAYIEEVKSFITTTTDKKDLKDCDIVIEA
VIENMDLKKSIIKEVAGILKPEGIIATNTSSLKVTEIAEASGRPESVLGVHFFNPVQM
MKLVELVQTDVTNSELFQDAQDFVSKLGKTPVPCKDTPGFIVNRLLIPYIGEAIGLL
ERGDASAKDIDVAMKLGAGVPMGPIQLADYVGLDTCLFILQGWVQDYPNDPSFR
VPKLLEELVEAGKLGRKTGEGFFKWEGNKCLM

>short-chain specific acyl-CoA dehydrogenase
MLSTIARGAARVVSKHVVRGMSILAQLPEEHEMIRQMCRDYAENELMPIAGMT
DKEHKYPAQQIKEMGELGLMGISVGDDKGGVGMDYLAYAIAMEEISRGCASAGV
IMSAHNSLYCYPVETFGTPEQHEKFLAPYASGEKLGSFALTEPGNGSDAAKATTTA
ERKGDHYVLNGTKAWITNAHQADAHIVIATSDKSLKHKGISAFLVETDTPGFTLGP
CEDKMGIRGSSTSNLILDNVEVPVENRLGEEGEGFKLAMMTLDAGRIGIAGQALGI
AQAALDCAVQYSQERKTFNMPIWRHQMIQQKLADMATELDAARLLTWRAAAM

KDAGLPFSKEAAMAKLKASETSTFVAHEALQILGGMGYVSDMPAERHTRDARITE
IYEGTSEICRIVIASNLIKENPL

>long chain fatty acid CoA ligase_1
MGSSQSIPEQCSYIVPNSQDGINGAIHRHPTEPEGLVERMMPKGEVETVYDTMKRG
YGMHKNAPCLGYRPFVNGIAQNYKWISYQQVMDKLTNIGQGIMSLKLFANVDK
MRIVGIYSKNTPEWTMTQMATYRHKGCIVPLYDTMTPENLAFIVKQTELSTIFSSG
ENLGKLAELKKNFAEETASLKNVVVMSEYKETDKKMVIDSGLSCHTMEELLKIGE
DNPCVPEYAERPDVALICYTSGTTSFPKGALLTHRNIMSISSGAGIQHVGFVLNSD
MVHLSYLPLCHMYEQWLHVMCFMYGGRVGYGQGDPRKIPDDIKTLRPTIFASVP
RVYNRIYDKMTQKIADMKGLKKMMISRALKTKLANLDAGRGQNHRLWDRLIFS
KVRKELGFDRCVFMLTGSAPMAPEIMRLLRVMFCCPILEGYGLTETSAGATVNEF
GVRSVGHVGGPAINNELKLVSVEDMGYSVEDTNHNGIECLGRGEVCIRGFNIFPG
YFRDEEKTRETIDEDGWLRTGDIGVWLPNGCLKIVDRKKNIFKLAQGEYVACERV
ENAYTRCPFVGQVFVTGDSTRDFAITFIIPDEEYTMEYFKKHSLLTGSDFKTIIQSKE
FRELVFAGLEVLEKEDKLRGFEKAKNVRFDSELWSADNALLTPTFKLKRPQLNTK
YKLMIKEMYEEGPIRCVRTRAPTSTGAPSNRV

>long chain fatty acid CoA ligase_2
MQDVEFAYEHTEGVYRNSCLKKEFKQGRFLTADLLSELDGCKTAYECFNHGVNV
NPNNPCLGTRFKQSDGKFGPYEFKTYAEVSSDVNQFGSGLLSESCVEKGSFIGIYSR
NNYEWVVVDQACNAHSLVSVPLYSTLSPEHLTFIIHQTEMSVVCGAKELIHNVIEH
CINKEDNKVTLLIQFEDDIDPKMNTLASEHNVKLTSLAQMKRFGKSNPIPHTPPSA
DDLATCCFTSGTTGVPKGAMLTHLNIVSTLGGALLQHLEISSTDTHISYLPSAHMM
ERVFLLAFMHVGAKIGFYTGDVKNLMADISALAPTIFLSVPRLYNRIHQKIMNGVN
EAGGLKKFLFNRALKTKLKNMRNNEQLTHWLWDRIIFKKVQKVLGGRIRFMLTG
SAPLSEEVKGFMHAVFGVRVYEGYGLTETTAGSFCSPLSPIDFSSVGSPLPQLEYKV
DSVPEMEYLSTDSPPRGEVCIRGNSVFKGYFKRPDLTAEVLDEDGWFHTGDIGKIN
SNGTLSIIDRKKNIFKLSQGEYICPEKIENIYGNSPLLAQVFIYGDSLQSTLLGIIVPDE
EASKSWANNNDKSGLSFKELCLDQDFKDAVIKEMKTCARVSKLHGFEQAKDILLE
SEPFSVENGLLTPTFKLKRTELQKAYGKDIEQMYSKMS

>long chain fatty acid CoA ligase_3
MGSQQSTNMWSVPVNESEPTYRRSSSSPDELVKNLPGCDMDNIYEIFQRSAERFPK
NNCLGARAGSEFVFSTYEQVEQKTIDYSNGIAKLGLLADVEGMKIMAHFAKNCPE
WTISQIASYRQSGTIVPLYDTLSTENLVFILNQCEATTIVVGVDQAARLVEIYESHK
DQIKVKNVVFMKEPTEEIAKKAGEVGLTTYTFEQVCQIGASEKIEPQFAGRDGVAL
ICYTSGTTGNPKGVMMTHGNVISVIGAGYRHCQGMTQDDVYLSYLPLAHMYEQ
WLLLFVLTEGGSVGYYSGDTRILVEDIKILRPTYFVSVPRVFNRIYDKMAQKISKLK
GLKKMMVSKALNSKVNKMKKGGQNTHWLWDKLVFAKIRKELGFDRTRFMLTG
SAP

>long chain fatty acid CoA ligase_4
MGCAQSNEFDHAAQLMNDDHEYSVEVSKGIWRDAAHKDELMTIIPGTDDAATL
WELFNFYVRSRPDSQMYGERVPLGGDKFGEFEFDSPLEFSKKVSKLIGGMRKDLS
FLKPKSKVGVYARNCRDYVCMLQACWNQNWTVVPIYDTFGAESVKYVLEHSEIE
AIFCLPEKLNKLESYIQESDSQVQAIVIMDSLAKGVAAAEKESKSDKKEIVVSMDD
ATTGPVPTFDMSALMEEGEEYDLTKGVATADDWAFIMYSSGTTGTPKGVIQTHK
MMIAATAGFRVRLAKLVTDNQPEIVSFLSYLPLAHAYENVVQLFVLSCGGCVGFF
SGNIRKIVDDLQVLKPLAMVGVPRVFQRMETVIRQKFAAKGPVSRALIGHAIGKSI
NAARKGKKHSSVWDKLVFKKVQAAFGGRLKIMVSGSAPIAPTLIEFMRVCCGVIF
VEGYGLTET

>long chain fatty acid CoA ligase_5
MGGCNSTENFDYFGALMSDKTEYSYEVEKGVYRNISSKGKDLTEFVEGTDDAAT
LWELFNAVVNKRSNGLMYGERIPNAEGVKGDYDFLTCKQSAERVARIIGGMQKR
FSFLKKGSKVGIYGKNCTDWVLAAIACWNQGWTIVPIYDTFGQEAVKHVINHSEM
EIILTTAENLPKLSKYLDGDCPSIQGVITFGNRTIPSAKVLVMEAEDHAMIKDLEVP
EKFDVPICSLNNLVDSNYPYDLAQGKAEREDLAFIMYTSGTTGLPKGVMQTHRM
YVSSVGGFCIQLADMIKELTADGGAMIVPSYLPLAHAFENILQMFVLGAGGRIGFY
GGDVRKLVDDIKTLKPTLMAGVPRVYQRMEQVIRQNFDSKKGIAKMLINRALNV
QSKAVLKKKSRSGLYDKLVFSKVKMAFGGKLKYMITGSAPIAPALLTFMKTNCGI
RFLEGYGLTETAAAHTVMDPSDNNCGTIGPQIPCAETKLVSVGELDYTIDDKPCPR
GEICLRGPHVFKGYYKDPEKTAEVLDEDGWFHSGDIGRMNPNGTISIIDRKKNIFK
LAQGEYVAAEKIENVLLRSDLVGQIFIYGDSLQSFLVTIVTPDPTTFIPACKKLGLEV

IPYGEEGWKARFQELCKAPEAIKMVLTDLTALGTSAKLMRFELPKKVYLEGEVNE
LNQGFSVENDLLTATFKTKRPQMKKHYKKIIDEMYGQ

>aminomethyltransferase (glycine cleavage system_T_protein)
MLSRTVGATRVSVRAFSQLEKTACFDWHVNNGGKMVPFAGYELPIDYKGTSIL
KSHMHTRTEGCASVFDVSHMGQIRWHGKDAIDFLETMVVGDLKALDVDSACLSL
VTNENGGILDDTVITKMDGCINMVVNGATKFDDMKHFDEYLAAYQAKGGDCSY
EYYHEKQLLALQGPGAVDTLASYIDAADEKRLRTMPFMSAGDFTVAGIPARVTRC
GYTGEDGFEISIEWDQAENLMDVLCSGETVLPAGLAARDSLRLEAGLCLYGHDLN
EEITPNMGALMWTIPKSRRADKRFLGSETIIAQTAKGAVPKKRIGLMVEGGVARE
GVEIWNEKQTKKIGEVTSGTMSPCLKKPIAMGYVAKGFTKTGKKLAVKMRGKFK
NATVTAMPFVPAGYYRGEN

>glycine cleavage system protein H
MLSRAATAALRTVRNVRSARTFATFFTKDHEWMKVEGGIAQIGITDFAQEQLG
DIVFVELPGVGDECVAQEAFCSIESVKAASEVYSPISGIVEEINEVLDGEPELVNSAA
ETDGWLIKVALAEEGIPEDMMNEEAYAAYLEECHH

>glycine dehydrogenase (decarboxylating)
MLRRITKVARPFVRNLQAFSKVDAFTPSDVFLYRHIGPNEEETKEMLKTLKFDSL
DQFTDATVPKDIHLNHDLLEGALSESEAQTLLMSKAKKNKILKSFIGQGYHETITP
AVIRRNVFENPGWYTAYTPYQAEISQGRLEALLNFQTVITELTGFPMSNCSLLDEA
TSAAEAMKMCFDQSKGKKNKFFVDVNVFPQTLDVIRTRAECLGIEVVVGDVNTD
MDLDNVCGVILQYPNAVGEVSDFSVIANKIHEAKSMLVTAVEPLSLSILKSPASFG
ADIAVGASQRLGIPMGFGGPSAGFLATSKKFSRKMPGRIIGITRDSRGKPALRMAM
QAREQHIRRDKATSNICTAQALLASMAGMYAVYHGPKGLQDIAQRIHKMACITAE
ALIQGGLEIVNGKSFFDTLTVKVADADAVLANGVEKNMNFARIDDTTVSITFDEAT
KKHQVQDLISAFGVNVDIDSVQVETDCIESFMRDDEILTQEVFNTYHSETAMMRY
LNHLERKDLSLNYSMISLGSCTMKLNAASELEPLSWPEFSNMHPHVPADQAEGYL
DMIADLNAKLSEVTGFAQVSQQPNSGANGEYAGLVTIRRYFQEIGEDHRNVCLIPT
SAHGTNPASASMAGMKIIPVKNDANGNIMVEDFRAKAEKYKDNLACAMITYPSTF
GVFEDTIRELCEIIHENGGQVYMDGANMNAQVGLTSPGHIGADVCHLNLHKTFAI
PHGGGGPGVGTIGVAEHLVKFLPGHAVVGDRMPIVNGTTGQVSASPFGSAMVLPI
TWMYLTMLGKDVKRATQAAIMNANYMATRLEDHYDILFRAQGANGHEFIMDFR

PFKKFGITEKDVAKRLMDYGFHSPTMSWPVPGTLMCEPTESENKFEVDRLCDALI
QIRGEIEDVIEGKIDPEDNMLVNSPHTMDMVMGNEWNHPYSREEAAFPTDWVRA
HKFFPTVGRCDDVWGDRNLDVVVHQCSSYLED

>serine hydroxymethyltransferase
MLARVASGVLRQKMGVRAISEASRNWATSMNKSLNETDPELEAILEKEKQRQR
ESLVLIASENFTSKSVCEALGSVMSNKYSEGYPGARYYGGNEFIDMAENLCIDRAL
TAFRLNSEEWGVNVQTLSGSPANFQAYTAVLQPHDRILSLDLPHGGHLSHGFQTP
KAKVSKVSLFFESMPYRLNEETGFIDYDEVERSAELFRPKLIVAGASAYARKIDYA
RMRKIADKVGAYLLADMAHISGLVAAGVMPSPFDYCDIVTTTTHKTLRGPRGAMI
FFKKGVKEVKKDGTEVKWDLEEKINSTVFPGLQGGPHNHTISALATALKQCTTPE
YAEYQTQVLMNAQVMSEGLQKKGYKLVSDGTDNHLILVDLKPMGVDGAAVELI
LEKASIAVNKNTVPGDKSALRPGGLRIGSPALTTRGFTEEDFEKVVDFIDRGVKIA
MELSEKVGKNKPKKFKKTVGALKDGDIPELDELKNEVNAFASEFPGICF

>glycine C-acetyltransferase
MMQRFARSAVVGFRNFNKGSAILKEALENSIQDFKNEGVYKSERIITSMQAAEINV
AGTDEPVLNFCANNYLGLSDNADLITAAKNTLDTHGFGLSSVRFICGTQDIHKELE
GKIAEFHGMEDTILYASCFDANGGVFEGVLTPEDCVISDALNHASIIDGIRLCKAKR
ARYDHMDMQHLESLLKENMDKRIRMIVTDGVFSMDGDVAPMKEICDLADKYDA
IVFMDESHSTGFFGATGRGTDEHCGVQGRVDIINSTMGKALGGATGGYTTASKEV
VDILRQKSRPYLFSNSLCPPVVGASIAVFDRLMESSELVEKVHENTLRFRDRMEEA
GFKILGCRDHPIAPVYLGDARLAAEFADAMLKKGIYVIGFSFPVVPRGEARIRVQIS
AAHSLDQVDRAVDAFIEVGKEKGVIA

>threonine dehydrogenase
MQQENKMLAKTTKAMSQMARSFSKKTPKVLFTGCGGQIGSEFVPFLREKYGK
ENVIASDVKACPEISAEGPFAYIDVCDRDALSRVVAENEVDVIIHFAALLSAVGER
NPSLALEVNVRGFENVLEVAKNHGCKLFAPSSIAAFGPTTPADNTPDLTVMRPTT
MYGVSKVYLELLGEYYHKRFGVDFRSIRYPGIISNKVLPGGGTTDYAVEIYYEALK
QGKYTSFLDEGTILPMMYMPDCLKATYDLIEADNDRLTQRTYNVGGFSFAPEDIA
ESIKKIYPDFECKYEPDFRQEIAESWPRSLDDSKAKEDWDWAPDFDLDSMSKDMF
KHLKIKLDL

>alanine transferase

MLPFVRTFGKLTPETLNAAVLKAQYAVRGPIVVRAKEIADQLASKGPEAGFDFD
SVISCNIGNPQALGQKPNTFMSQVLALVQYPDLLKNEACKSAFPADAIDRANKYV
GQIPDQIGAYSHSQGVNVVRQEVAEFIEARDGFPADPESIYLTDGASSGVQMVIQS
LIRDENDGLLVPVPQYPLYSASIALCDGSFVGYELDETQGWCMDMDRVKASLKE
ARDNGKTVRALVVINPGNPTGQCMPEENIREVIQLCKDEGLVLMADEVYQENIYT
EKPFHSFKKVMMGMGAGIADEVELVSFHSTSKGFLGECGKRGGYMELCNFDAGV
VEQLYKLASVRLCSNVIGQLTVGLMVNPPKEGSESYDGYTAERSGILDSLARRAK
KLVAALNDLPGVSCNDAEGAMYAFPQITLPPKAVEDAASKGLAADAMYCMTML
EKTGIVVVPGSGFGQEEGSYHFRTTFLPSENDIDAVINRLSNFHIEFMEKYA

>branched-chain amino acid aminotransferase

MLSRVSQVARKGLMRAFATGFELEGLKVVSNPEPKTKLPNEELTFGTSFADHM
FSVDWTEGHGWHDAQIIPHGDLPLNPAAAALHYGMQCFEGMKAYIDSEDQIRMF
RPELNMARLNDSMKRLYLPQFDEEGLLGCIKELLKKDKSWIPKGEGYSLYIRPTGI
STHGFLGVGPSKEAKIFVITSPVGPYYAAGFKPVKLWAETSYVRAWPGGTGNAKV
GGNYGPTIMPAVKAAEKGAQQVLWLFGEEKYITEVGAMNFFVLWENEDGEKELI
TAPLDGTILPGVTRRSVLELAKSWGEFKVSERSYTMPELAKALDENRVIEAFGTGT
AAIVTPVDGIMYDDKEYNVPLDPEKEDAVCGKITQRVWDTIVGIQYGDIEGPEGW
SVVV

>3-hydroxyisobutyryl-CoA hydrolase

MSKASVIGKVQKGISVLTLNNAKQMNAVSNAMCSNLREVIQEQNETNSRALILEG
SGKAFCAGGDLKSIRDEGMNPNSDSPARQFYGNMYNVSYMLHSNEKPVLPILTGF
TMGAGVGLSAHTKFSVATENTILGIPQMVIGLIPDDGASHSLPKMPYNIGYHLGLT
GSRLSGKDVVSAGYASHFISSGKIEEFKAELANEMELRGREKPENVMRDVLNKFN
EAVTESEKAKFIDTFGPCEEIYDLASTPFENDEILEESMCQGNIRTLEFIDDIMKQISN
PSTTDSAQKLLKKTKKTIDTLSPTSLSLTLHLLPFCEQYGLTLKQVLGMEHDCVQEI
LNRPDFFEGIRCTLVDREDTPKWSPAAHADVSLKEFHEMLIRWRDLYWHMVAY

>3-hydroxyisobutyrate dehydrogenase

MAMKIGFVGTGVMGQHMAKHCLQKVPETTLSVFNRTKAKADGLLALGAKW
CDSPKEVAKECDIVFSIVGYPKDVREVAMGENGILNGLREGGIFVDMTTSEPSLAK

EIFAFGETKKISCIDAPVSGGDVGAQNGTLSIMVGGEKDAIDTVMPYFEAMGKNIN
HMGPAGSGQHTKMTNQILIASGMIGVCEGLVYAYKSGLDMEQAIRAVEAGAAGS
WSISNLGPRILKRDFDPGFFVEHFVKDMGIALDEAKRMNLCLPGLSLAHQLYLSV
MAQGNARDGTQALMKAIEAMNSVECPKYDL

>aspartate aminotransferase

MLARVTRQVARPFSMFASIPAAPADPIIGLSEAYNKDTFAQKVNLGVGAYRDNE
GNPVVLSAVRAAEKRILDKKMNHEYPGMQGVMDYLEVALKFAYGADSVPLEEN
RVSSLQTISGSGACRVAAEFMHEFGNGAPIYFPNPTWGNHGAIFGAGQMKMESYT
YYNPETCGLDFEGMITSLKEIPAGSFILLHACAHNPTGVDPTPEQWDEISEVIKQQG
VRPFFDCAYQGFASGDAEKDAYSIRKFVEDGHNIMLSQSFSKNFGLYGQRAGALS
VVCEDEKEQAAVTSQLKRLGRAMYSMPPVHGARIVSEVLHDDRLTRKFYEECKE
MADRIIDMRSLLRKNLEEINPRNWQHITDQIGMFAFTGLDKEQCAELVNDKHVYL
TANGRISMAGINSNNVEYVAKSINDVCRD

>glutamate dehydrogenase

MLRISSALKASKPSLNAVRFLTTVPDDEPRFLEMVQQNFDLASKASGMEEDVLA
FLKTADSALRVNFPLKRDDGSIEVVEGFRVQHNHIFQPCKGGFRFSEHVDLNETEA
LASLMTYKCAVVDVPFGGAKGGVKINPKKYSSNELASIVRRYTMELNRYGFIGPA
IDVPAPDIATGAREMALMKDTYQMLFGYNDLNAAGCCTGKPVGHGGVDGRTEA
TGLGVFFSTRTYMNNPEIMEPMGMTCGVEGKTVAIQGFGNVGSWAAHYFHEAGA
KVVSVSDSDGSLFNYDGIDVPALVEYKTQNNCLKGFPGAKFEQDNELVLFADCDV
LVPAAMECTIHRDNAGMIKAKILSEGANGPCTPAAEEIINNNGGIIIPDMLANAGGV
TVSYFEWLKNLSHIDFGRLTKRWEEKSKTRIVEALREGTAQEMEADFLSGPSERDI
VHSGLEETMIVASERVMKAAKELGVTQRVGAFTIALKKLDDGLYNAGITI

>ornitine aminotransferase

MLARSTRSVSRLGVRCITESERLIALEDRYGAHNYHPIPAVLSKGEGVHVWDVE
GKHYFDFLSAYSAVNQGHCHPKIIGALKEQAERLTLTSRAFHNDCLGEFEKYVTE
MFGYDKILTMNSGVEACETAVKLARRWGYDVKGIPTNQAKVVFAENNFWGRSL
AAVSASTDPTSFQGFGPFMPNFEVIPYNDISALEKMLESDPNIAAYMCEPIQGEAGV
VVPDEGYLKQVRELCTKHNVLWIADEVQTGIARTGKMLAVDYDAVKPDILVLGK
ALSGGVFPVSGALANDEVMLTIKPGEHGSTYGGNPLACKVAHAALEVIHEEKLAE

NAFNMGKILRESLRSIDTDVIELVRGRGLLNAMVVKPKNGKDAMDVCMAMKEN
GLLAKPTHGDIIRFAPPLVINEEQVAECCDIITKSVMEVFA

>translocase of the outer mitochondrial membrane 40 (Tom40)
SHSFCKYITISICVKTMAQQPGKEKLFDSADDSIKNPGNFEQISQEPRSLFELDAFDG
LRLDMVRSLTPRFMATANCLLGSVMIPGGSMCRLGVNAAPTDNNLVMLSADKTG
RQDYRWHFNHKQFSSKVNVQLGEQKMFSAEVERKGKTHTMGWKMMNDKMMS
MHYLQALTPKLVLGADYIWNRAENKSVMSLQGKYGTPMYYTIFSFLPTQQMLQS
SVVRKLNERVTMGAQMTYVHPQGTANANLGVQFNFKHAKLSCSVDSDGKLNSV
LMDQLMPGLTGMFSATLDHAEQGCKFGFGIQFGQ

>sorting assembly machinery 50 kDa subunit
MNHNSVSMNSISSVKISGNKRTKSTLIEYYLQDVYKSEPENLDVELELAYMDMMA
SGLFKEVLFVRKDETSPNSPDLHVRVIENKTTGINIGAQIKAGGETVFGASGTVTNA
LGNGEIFTLESSVGHRESSSFSASLAIPHAFYLPILQQTHLYNNIANFKNSSSFHLKT
AGASVSFQSERHPCEGVHEVTLKAEQRDVQPELSERSEDGGVVFDASRQVVSQRG
MSTKTSIGYEYLFNNEELPETISVYDKDEDENIDALDIWSKFRANLELAGLMGDVS
FLKYGFKYEKWAALAEGLPKVSVGAGVWFRHIIPKGNVRLPLCDRLFADNPMWL
RGFQPRGMGPASSSFTPSGSTEPKRDILGGNLSMGMFGLMDIPITGILNAHLFSNFG
TVVNSPKDLASPAQWRGSIGGGIALNLGPIRLEVNYIPYKRSSDGDHIVKRMFDISI
GASF

>sulfydryl oxidase_Erv1
MLFGKITKFVCLAVPFVVGVCGSVNLYDNTIKQSAKNAINCEDPICFSRRELFEKM
MHIDTKKKVEKSDFEEGSESVLDVLPRDLPPRKPWCPVDRDELGQMSWTILHSFA
AYYPDEPTEEDKQAAVAFMQSFTHLYPCVHCRSDFKEAIAANPPRVETREEFSLW
MCEMHNYVNKKLGLKVVKCNIEDLDTRWRYGAANCWELPEGMEENEKVEKKD
LNTRT

>mitochondrial import inner membrane translocase subunit Tim8
KINSRQLIVFQYFSVKMSSLDPKVAQQLQAIEQQAQASMVSEMLKDMTDSCYEKC
VRNPAAKLDKVCLKNCMQRYQEVIQVVAHTMEKRQSSQQ

>mitochondrial import inner membrane translocase subunit Tim13
MNGGGMNDQQRMMMAIQQKLQVQKAIQDISNKAFKYCIDRPSDSLSNKENKCV
ANFVERYLDTSVFVTQSLMAQQQQQHH

>mitochondrial import inner membrane translocase subunit Tim44_1
LRHYLSDRNLHQNYYSCRPNFDEIESLLGPKNPFKIFLMLPALRNAAHTACRARRF
QQRHFSILQEQIKAEKSHKTATEKDLDLISLNKHVDYDKHIDPALYVQSEFAQEKV
RGVFEMIGKTTYEEANVLQKMVEESESKKEPISIGFVFKPFTKTFGWLKDFLVKGF
DDLKGKNKTFAFNTRARYEVKTSRIAEKRATIKQNIANMEERGLVLTEQSLSPWD
TILNILNKTPVINSIVAEAEKALKSEKVVKITNAVKDTVEDVREGYETNQNPWVYR
AVNTWDSIFTKTETSAALSEIRKMDPKMDLDHMVIKLEYEIFPKVVRAFVECDFEP
IAQYCTEEGFEQIGASMQARIDAGLTLDPEILDIRPFEVVSASRREKGPPVVLVLAV
LQQVHCIKDREGNIIEGAEDCIHANYYMLAMNRIFDSEVGGFEWKIREMMIIGKVD
YI

>mitochondrial import inner membrane translocase subunit Tim44_2
MLSILRRQVTKQLRIGGFQRQFGILSDLKNAIPKPDEVNNKKTSDEPEIDVEAEI
DEAKEKVRKVLSDMRKTATESAKMFADKASQASETSKEAMGDAKEKFDSNIRDK
IKAPKQVSSLFSKIGSTPAFSWMSEQIKGGLQDLKGDFSAEVRYENYKKDLEKKRE
MLKDRVANTVDTGLIVSDHQKGSTWESFLTILNKTPIINHMVAEAEKALKSEQVQ
KVKNQLKDKVEDARQVYETSQNPWVYRFSYAWDSVFTESETGAALSEIKKMDPS
VNMDAFLVEMEKEIVPTVVEAFLCGDLEPIAKYCTEEAFEQIGAPMQARLEEGISM
DPTILDHRQYEMVSASRREKAPPVVLCQCTVQQIHCIKDKDGNVIEGDEDRIQANY
YMFAMNRLYNPDIREFEWKIREMMVVGRMDYL

>mitochondrial import inner membrane translocase subunit Tim50
KLEIRCKMFGGSVRAFARSFSTKAKLASEALKVSNSQAKTSFGKSVLKASMGLLG
AGAVIGAPTLLFSEFELPEKFPMRDEIENLRMRAQPLKNRFMEMFDFFVEPASDDL
LPDFAGNFPMRTLVIALEDNLLHEEWTHEHGIIWLKRPGVEQFLQKMCQYYELVL
FSNKGFGAVEGCVGKLDPYGFFQHRLFRDSTLYKKGRHIKDLSRLNRELNKVIML
EVNPADCMQNENVLTLSKYEGDRMDNELFEIEAFLESLVMENVDDVREALKQYA
PPRGKTFGEHFAEMQKVKRVQQVQSKSHGLVGALQKVAGAVASTTEVQAPASTP
KSKRNRRRSVRD

>mitochondrial import inner membrane translocase subunit Tim16
MSGALFRVFAQAIITGTSVVARALMVSFQQAIAQSGSGEVAAKAAQRAVHHKM
GNMDEDQAMKILNLTKDSTLEEREDNFQRFFDANDPNNDGSFYIQSKFVRAKEFL
EWLDEEQKKAAEEAVKDAHKAHEEGLDEESLKDEETDSVEEVKEDEIKEETKQEP
KVDEEGNAFEYADEFADESEEEEKETPKDAEDVIQL

>mitochondrial import inner membrane translocase subunit Tim9
MNLDSLPDAQKEKFIDWANQTQMKDTLRMYNDVVERCFDKCVHQFRSPALDDN
EKKCLKNCAELFLKHTQRVGQRFAEFQMMQQAGGAPNEEK

>mitochondrial import inner membrane translocase subunit Tim17
MDSGRLPCPHRILDDVGGAFAMGCIGGGVWHLFKGARNAPSGGRIVGAFQAARF
RAPALGGNFAVWGGLYSSFDCTLSYLRKKEDPWNSIVAGGLTGGVLAGRSGFKIV
AQNAVVGGVLLGLIEGLGILMNRVFSPQYQMGGLKQKGGVDLHAPQTLAPPTVH
RNVNAL

>mitochondrial import inner membrane translocase subunit Tim22_1
MEQILQHAPYWLQQRIMDPTSCSGKATMASFTGFTMGIGLGVFFGTFEGAHGEIV
GNTMLQQLNNGFRRSLTMTLKRSASFAPNFAVVGALLTGLECKLEKHRAKSDIW
NPVIASAVAGSAFGILPNVRRGAKVMAKKGAIFGTAFVAFSLFMHYGIDYFREVG
MRNQKRFG

>mitochondrial import inner membrane translocase subunit Tim22_2
MADEQQSPLPFGQEKGALNHPVKGLSSDEKLKSLPWKQHPQIYMAKVFESCAFR
GVTSAIVGGGLGIGVGGLIAAYKTLTPPVPVPGQPAPPTVPFKQRLKLCPPAMKLR
AKSWGKNFMIVGGVYAGMECFIERLRAKHDVKNSILAGCATGAALAYQAGPLG
MTVGCIGFAAFSGIIDKVMER

>mitochondrial import inner membrane translocase subunit Tim23
MSGEQSWDEDLGDKFDAIGSENRGIGSDFRIAPEALGLGINIDPAKSGVGSLFGVV
DQDFIQTDARKWHERMFYRTGTLYLSGIVSGGIYGFVRGVRRAPNKRFKIKINNIL
NASGRKGADVGNCVGAAVLMYCFSRHLVDKVITDEHFVMPEFARASTAGFLTGL
GYKSLAGPRGAIVAGILGASLMGSLTFVNTYVCSKFKFARSVSRMFN

>mitochondrial GrpE-related protein 1 Mge1

MISLLSRTRPLTRFLGASNARFGVSFFSTEEVKQTEEVVEPEIEENSEDPNDTIEAQ
LAEFENEIKELNNTKMRLLAEMENVRRIAKKDVLNAKEYAISSFAKRLLDVADNL
DRAIEAVPTDALENETPLMNTLYNGIKMTQTEMYKVFASEGIQKFGEIGDEFDPQR
HDALLSMDCPEGETPNTIAQVAKTGFSMKKRILRPAQVCVFGA

>mitochondrial DnaJ homolog 2 mdj2

MFTLRQACSALVKPQVRHLMVKSTAPVMSRLYPSSAPIIRRHYHPSNPQEIVPLLIG
VGVATVCIGGSKLIAYLDEIEKNPPKPKPKKTAFQNFYNGGFKDEMDRREAALILG
CREHADPKKVRQRYQALLKKNHPDAGGSTHIAAKINEAKDVLLGGGASHTGM

>mitochondrial processing peptidase alpha subunit

MLKSLSSCTRMFSRAAAAVPLNQAIENSPVLKPFSQLPQAKARMTVLPNGLR
VISRDEHSPLFSVDFITSSGAKFETSSTKCCSTVADALAHQIDISSHGNLDTRLDIDS
MRYHIHVVKDHVENAIDSLAHVVKNQVDGSWDNSIVEKAKRTVIARNKKNEEED
HGTLIGMLSRECAFDGPMGNVMYCPTDDIMNVTSESIKEFLKGTMTADRSAIAAT
GIDHDMFVYLVQKSFGFLESGKGILNPPSVYKHKHTILEAPAKPRISVEEIPNFPFVS
VMFQGPGWNINEIMPFQVLKSLIGGGSAFSAGGPGKGMHCRAYKHILQRHQFVDS
CHFLGDIYKETGLFGFQASVYPGYENHMINVIGQEMYTLKDQPPTEEELQRSKNQF
GSQYLQALELAEVESEDMARQALVWNTYYDQDRSEEHTX

>mitochondrial processing peptidase beta subunit

MLRKFSKLTRAKSLFRHFSSKFPEYLLSQPATEVTTINNGMTVASEGSHGQTAT
VGVYIKAGSKFENDSNSGVAHFLEHMNFKGTEMMSINNFEAYFENAGAQINAYTS
REYTVYNVRCLKKDIKESLTIIADMLQNSRITSPQVNSERYTILREKEEVERIPEEV
MFDLVHEASYPNSSLGFRILGPEENINTITRKQIIDFVEELYIGPRMVVAGAGAIDHE
ELVALTESTFGKIKEEGPHNEVIANSKQIFESTMKRELDEHIPVAHTMCMWEAPEW
ISPDTIPFMLFGQILGEYDSTGMSSNLGITKLSTDLSSSSSCLRYSTFFHQYEGSGLFG
FYLLSDPINPEEPFEKCYREFNRLSNGITETELAAAKEHLKTQIVGNIGDSNGCCDDI
GRQMLFYGRRISLAELFMRIDEIELDTLHDVAKKYISNKIPAQACLGSISKVPDQEM
MQKWSDAL

>mitochondrial inner membrane protease

MLSKCVNKNVVRNLKFLFGIGSATYIVRDNVLEFGICQGLSMFPTLNSSSPLRDV
VFCEKLSYKFSSIQRNHVVMCYLPSNKNTRIVKRVTGLPGDIVIDKYGRERPIPKGY
IWVEGDNKRNSFDSRMFGPMPSALVLGRVRCRIFPFNKIASF

>chaperonin CPN60

MLARFCRSCVRGFAKDLHFGVDARNRMLAGVNKLADAVSTTLGPKGRNVVIQ
QSFGSPKITKDGVTVARAIEFEDPQENLGAQLVKEVAGKTNDIAGDGTTTATVLTR
AIFSEGCKAVASGVNPMDLRRGMTKAVDSVVAHLEGQAIEISSHEHVSQVATISA
NNDKEIGDLIANAMDRVGKEGVITVQEGKTMHDLLEVVEGMKLDRGFVSPYFIN
DTKTQKCEFEHPYIVMFDKKISNVNALVPALEIAMNENRPIVLIAEDIESEALALLV
VNKLRNGMKVCAVKAPGFGDNRKATLQDMAVLTGGEVVSEEAGMKLDEIQAFQ
FGTCKKISISKEETIILNGAGDSEKLEDRCNLLRSQIETVASTYEREKLQDRLAKLSG
GVAVIKVGGASEVEVGEKKDRIDDALNATRAAVEQGIVPGGGTALLNASLILDSL
AEEMENFDQKMGVDIVKKAIRVPAHRISQNAGDEGSVVVGKLLEKTDERIGYNA
QRGEYVDMIDAGIIDPVKVVITALIDASSVSSLMATTECMITDIPTEGSSAPMPPGM
GGMGMPGMM

>10kDa chaperonin

MSGLLKKFLPLANRVLIKRVLPITQTASGVLLPSSAQKKGNEGEVISVGPGMID
MNGIRIPMSLKEGDKVLLPEYGGMKIDFDGIEENTEFQLFREEDILGKFE

>cell division protease ftsH_1

GKTLLARAIAGEAGVPFFYACGSEFDEVFVGLGAKRVRQLFQDAKKHAPCIIFLDE
IDAVGGGQRQVKDQAAMRMTLNQLLTEMDGFESNNGIIVIGATNFSESLDKALTR
PGRFDRHVQVPLPHQKGREEILQLYGKKITLDKKADLKSLAKRSAGASGADLFNI
MNTGALRAAQRGANAVTQKDLDEAFDHVLMGPQRKSLVLTPEDKYHTAIHESG
HAVASLFTKGADTIHKATIMPRGPALGMVQYLGQEEHRNFTREQFMAQLDTAMA
GRAAEDIKWGIDKLTGGCSNDIQQATQIARMMVCKLGMCEEEFGLMVPNEHSSPE
TKNKIDKAVSKLLTESYDRTKALLTAKWKQVEHLAKTLVDHETLDVNEVNEVVF
QGKIHGKRVRI

> cell division protease ftsH_2

IDEIDAVARKRNKGSFSGGNDERENTLNQLLVEMDGFKSQEGVVVLAGTNRKDIL

DTAILRPGRFDRQINLELPDLESRGEIFKVHLAKLKLKHDADLVAERMAALTPGMS
GADIANVCNEAAIFAARRNRKKVSLRDFENATDRVVGGLEKRSGFLSPEEKHRVA
VHESGHAVTGWFLKYASPLLKVTIIPRASGALGFAQYLPKELHLHSREELEHWICA
ALGGRVAEELLCGCHTSGAQDDLSKITQIAHQMVTGFGMSDVLGAVSYPNGDPF
MSMKPYSEHTGQIIDEEVRRIVTEAEAHTRELLESKKDLLLELAKELEAKETLTEV
EIEAILGSRPAGTPDSLRKYVRATVEDSDEEDDGTSSSDYDDSSDSDENENDEQISE
HDDDDEEEIEEYEEIDDKGKNNSEESDLYCSKEETK

>octapeptidyl aminopeptidase 1 oct1

MRRFSTISHLAKSARRTSQRLDSLMTIINTCSPKETVEVADEISNQLCLGSDLANF
VASTYHTENGRKEAGEILEHLHEIIHIINSNESVSNCLQAALWNKDSLPMEHIAVAE
SLLKESTHFKNANIEELQGLFHHHRALQAGYLNNLRQTQEIAIPQELFHMFHEAYH
GYITKMDPNASHFFVELPSHIKQEIIQSCPFRELREVIFKATDEFAERNIKMLSVLVN
IRKNLANSLDYGSWKEYNHDALFLSETKEIQNILMGCAESVMEKYKDEREMLNEE
MKKQMNGQESEIKPWDEAFLANEVMRQAGGIPIDAVTFPLNYVLDSLFNVLASTF
KIQVTPVEMAADECWANQGDLRKYEATCLESGKPLGTVFVDPFPRSGKFHNSALF
TIECGKTIPGFLSESSQPERQLPVMSIVANVGEVDAPNKWNQPLFKPEVTTLFHEFG
HAVHVLRSQTELQHFSGSRGAPEFGEIPALLMEKLGSHEKIMQSCGESNHKELISA
MLKSSNDSSELFGATATLEQVLQSLFDLELSGPTPPKDIKECWSNLRNELLPDSFYS
PECDYSWIGRFPHLLTYGGSVYSYLLADISASDIWTSVFNEDPSNSIAGDKVNKML
SFGSSVPPSKCLSDLLGREVNGQAFLSC

>distribution and morphology protein 38

MLASRAVVRLFTTQTPKIVSVSGFDRIRTIRHKIAHEAKHYWTGSKLFAADVKIS
SKLLKRSLQGFPLSRRERRQLIRTTSDVFRLIPFSVFVLVPFAELLLPFAIKFFPNMLP
TAFEDVDKKQQELKQQLQRRLTVARVIQDTLKQMSTTLQNKGDDVEYFQDLIDK
CRAGLPVDAEDIVKVSQLFKDDITLDNMNRSQLISLCSLMGVPRFGNDFLLRFHLR
KKLRDIQKDDLEIMFEGVQSLTVAELQIACAERGMRAHGLSPMGYRKQLEEWLD
LACSKKVPASLLLLSRAIGFSRFDEFSEEEKISVADSIASLDDDIVREAIVDHAGSGS
ADLLINEMKLKSLKLQNEMINRDYQEAAASGLSVVEDDDNAVLSLKEAEELIMGT
GNIAIVEANQVKRLLESDKELEIDDLIADVHAGDDDIIVKLRSAFRTLTQKLNADIQ
IIESGGIQRIKSDILGRVETDILKKMVFEHLKDSHSDESVDMILEELFENLEDQDYM
TVELFAKRIEQLALHK

>2-oxoglutarate dehydrogenase E1 component
RKPLIVASPKSLLRNPDCISDLSDFSEGTMFQEVLGDISIEKPDQVQCHIFCTGKIYY
ELNQYRKKHGLNNFAITRIEQLSPFPYDKIMEQLEKYPDAVLRWVQEEPMNMGG
WEYIRPRLCLAARTSELKRCGVEYRGRKSSAATAVGYGEIHDKEQKHVVESGFDI
PEGLNEVSCTK

>2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase)
MLRLVQSKLRKIRLGPIFRSFSIFEVKVPENVGNMAFESWCVNVGDHVTSGEVI
SVIGGDESLIEIRSPSAGNVIKIFAQPGDGLDKGEPLLWLDLDENEENVITQENHSIN
ESVNRQQKNEKCSHTISVNKLH

>dihydrolipoamide dehydrogenase
MLAISQRFAQSRGLARLFSEAYDVCVIGGGPGGYVCSIKAAQLGLKTVCIESRGT
LGGTCLNEGCIPSKALLHSSHMYEQADKEFKHVGIKVDNLDIDVPQLMKNKDKV
VKGLISGIESLFKANGVSYVKGLGSFGEDGKIIAQLNDGSEEIIEAKNVVIATGSKPT
SIPNLEIDNEKGRIIDSTGALSLGKVPKSLVVVGGGVIGLELGSVWRRLGAEVTVVE
FMDRILPTMDNEVSAATMRALKKQGMKFMLGTKVTNSNVGEDDVSLAVENVKN
GKSSVLDTEVVLVSTGRRPYHDDLKLSDAGIALNERGFVDIDDHWRTNVDGVFAI
GDAVPGPMLAHKAEEEGVAVANLIAGKPAHVNYDVIPGVVYTNPEVAAVGKTEE
ELKVDKIAYTKGSFPFLANSRAKANGETDGMVKILADKETGKVLGVHMVGSGVS
EMIHEGALAIEKGLTAADIAHMCHAHPTMGEAVKEAALSTFDKAIHMPPPKARAS
RGKKK

>malate dehydrogenase
MQALRSVTQRFPYSVAILGAAGGIGQPLALLQKLNPKISEVRLFDVASCKGIAADL
SHIATPANVEGFDSAAEAVKDADFVVIPAGVPRKPGMTRDDLFNTNASIVSSLVAD
VAKNSPNAFILLITNPINSVVPIAANVLKEHGLPTNKLMGVTSLDKLRAKTFYSDL
VGKNPSQVDIEVIGGHAGCTILPVFSVHKDYNKLETEEIEALTHRVQFGGDEVVQA
KEGTGSATLSMAVAGNRLVSNLINACEKAEGRIEVAYIENPASETGFFASSFRVDK
TGIKEIMPVPTLSEFESSLLAEATEVLKTQVKKGLDFKMSA

>thioredoxin
MNAVPLKKKPIQIDGGAHLRNVLTHNSGKLVVCNWTTYWCECCRQIEKELTQFC
QDNEDLVCVNIEATDNETLVGQAGIHVYPTFTFFFNGEKLTDYSFG

>peroxiredoxin

MLKVGDSLIGKGLKLQMARTFDDAVADGFKDTDLDEILKEKKVAIFALPGAFTGV
CSKAHVPSWRDNAEALKEAGIDTIVCISVNDAFCMNAWKKNVDPENKILFLGDPI
VAFGKHTGLTAEMGSAMGTRSHRYSLLIDNGVVTMLNSEKGIQDLEVSNGEYML
AALRK

>Fe-superoxide dismutase

MFLIIILLSAFSFYNTFAKMELPTMKYELTGLEPVLTEDILTYHYGKHFAGYINNVN
KLTDNKVTFDTLMDTIMEADGGLFNNAAQVFNHKFYFEGMGPVDKVKKEPTGA
LMEMIVRDFGSFEEFKALFSKAAATHFGSGWAWLVEKAGKLEIVTTHDAGHPVR
EGLGKPILTCDVWEHAYYTAYHNARADYIKNWWRIVDWEVAEAQL

>superoxide dismutase

MFFCGITLYSMSKPHEHFTIPPLPYEFGGLEPFIGESTMFVHHGKMLSSCVSKLNQL
MDIHGDKYIKHGKPIGLLDLVLSAPEDSELYQNAAQVWSHDFFFNSLKPYDRDNC
RPGYVLLDKIVRDFGSLKDMKKLFNEKACSHFGSAYVFLLERDDGTLCIECCHDA
NNPIRLGLGNPLICCDIWEHAWIESKTSVESYLEKFWSVVNWGFAEKNLLKEFRD

>ornithine carbamoyltransferase-carbamate kinase

MLSAVTRMNCMQRMLPAFASAFSTGPMHVTKISDMSPQQVQQVLDEAIKMKA
NPEKYNKVLEDKTLLMLFEKPSLRTRVSFETGMTELGGHAIFYSIGDSPLGKKETF
SDTAQVLNRYVCATMARVKSRNDITELAKYSDIPVINGLDDWAHPCQTLTDLLTI
LEKKGTFEGLKMAFVGDIHNNVTYDLMRGACLMGFEISVSGPEGEGYEVDPELIK
ECEELSATHGGKFNITHNPAEAVKDCDVVYTDSWMSYGIPEDAMAARKEKFMPY
QVTEELLTNAKDDVIFMNCLPALRGMEQTAGVIDGPHSVVFDQAENRLHSQKALL
MYLVHGFNYSVDEAQRILVALGGNAMKKPSDLGTAEEMMANIDLATEQLSELVA
HGLDLTITHGNGPQVGNIFLQNQKCSPEIPAMNLTVCGAKSQGMLGYFLQQSMQN
NLTKRGIEKCVSSLVTQTVVDHEDPAFQSPSKPVGQFYTEAEAQELMKQGKNLVE
DAGRGWRVVVPSPKPVKVVEGEAVKTLVDGGHIVIATGGGGIPVIMKNGQLEGR
EAVIDKDMTAVKLAEQVKADALLLLTDVEKVALDWGKDTQRDVDRMTIKEARE
FLAEGHFGKGSMGPKIEAAVQWVEQTGNPCIISSLDQAINALDGKTGTHVVA

>mitochondrial carrier (MC) family_1
MFDIKDTISGAFGSLCLVAGGLPFDTAKSRIQTQHTNVYKGPIDCIIKMARHEGIGS
LWKGFTPAFSSACLENMVLFSGIGIVRRLFMKLNGTKEISLFQEALSGSTAGIVSAT
VICPPEMIKVKMQCQKNVGTKTIGHNAVIYRSSFDCLKQVAKAEGLKGIFKGLSPL
LARDIPFHFFFFGTYETVNFLLTKVSKTRKSKAELSPAHLFFSGGLAGGFAWAIVYP
FDTCKSEMQKSIKTESFGKVLKGFLKEGGIKRLYCGYTAALLRAYPANASLLVGY
ELCQRFFKWFETA

>mitochondrial carrier (MC) family_2
MSVKKQIKWSVEELAAGGLAGGLSRFIIAPLDILKIRFQLQPLPISRHVVDARYQSIF
QSYRNLIKEEGFKTLWRGNLAAELLWISYTAIQFAAFDRISRIVESDSIITKSLNNIT
CGAIAGTSATIATYPFDVMRTRFASQGIPKVYSSMPNMIAQIAKHEGFAGFYKGM
GPTLVQIAPFIGIKFAMYGPLKRYSNSHLPTRFSAFDTILAGSIAGFAAKVVVYPLD
TVKKRLQMQGIVRDSSYGVLPNYSGTMDCFKTVWRQEGLRGFFKGIQPSAVKAA
VAAAVTFCAYEDIRKFIHEKNAGRRAWLK

>mitochondrial carrier (MC) family _3
MSSIPVVAQSAKIQHESHHESPLAKFVISGALSFWFELGFGYPAEFMKIAKQTTTVS
YPNILKRAVSQKGIVGLLDGFFPWGSIYAVLKGAMFGVGHSISKNCLDGRIHDSFA
EIAAGGMGGYVQGLSLSPVMLLKTRVMTNPAFRTSGSVMDTAAKSCLLGKQIVK
QEGVRVLMKGADVFAAKRFADWTSRFVFVEVVEQGMKKYLHRTNLSRGEKCGA
ALAGGTLSALCTLPMDVIVAQLQCQNSGGVQEVSAYQIVKKQFKTGGAKHLLLW
SYRGFVARVGHVALTTFLMKTISTDVYKYYKEHFGK

>mitochondrial carrier (MC) family_4
MFSKFTSNSQMKGVHNPLAFSVKNQSAESPIAKFAISGLVSWSFEFLLAGHFVEFL
KVEKQATGLSYKQITRNITQHKGFVGIFDGFFPWGSIYSVMKGCMFGWGQAVGR
NFFDGKVSDSTANVLSGGVGGFVQGITLSPTLLLKTRVMTDPRFRVSGGTLQTTIA
SCKIGTEICRTEGAATLMKGSMVFATKRLSDWTTRFLFVELIQNQLKTMINDRPLN
QFEKCSAAFAGGAVSAFVTLPIDVLVATFQKASKAGVKMQVTDIWKEQLRKGGV
GQLIKFSTAGFGARVGHVALTTMFLKTVSSHVYDRYKAHKEQKALQWATAQ

>mitochondrial carrier (MC) family_5
MFPKSMIAPSAQPVHDASYYAKCMLGGVLSCGLTHTAICPLDVVKCNMQTNPKG
YPNLTQGFKNMKAEGGLPKLTRGWAPTLIGYSMQGLCKFGFYEIFKDLYANVAG
PELAHQYRTSLYLAASASAELIADVALCPMESVKVRVQTTPAELKFPTKLAPATK
AILADEGVSGLFKGLSPLWMRQVPYTMVKFACFERVVEAFYTYAFKKPKAEYAK
STQLSITFASGYIAGVFCAIVSHPADTVVSKLNQEKGGSIGGIVKNLGMKGIWRGL
GARIIMIGTLTGLQWWIYDSFKTAMGMGTTGGSAPVEKKEE

>mitochondrial carrier (MC) family_6
MSSEKVNLAFARRFASASLAGAFTKTAVNPLERLKIVIQINPQPIRFAVKNILIKESP
LAFWKGNLSNVIRAVPAKGLLLVSNDVCKQQLGMNPLYSGAVAGVISTFCTYPLD
LIRTRVAGVVCSRYQGIFPAIRLTLVEEGITGLWRGCTPTLLGSLPYTAIQFAAYDY
FCKQAGKSEAFKNNRVAIAVIAGGLAGFVSQTITHPNDTMRRLMQMEVEKASYF
QILSKTLKASGVRALYAGVAANLLRIIPNTAFQFGFYEFFKKALDV

>mitochondrial carrier (MC) family _7
MEFEEDVEELAEFVHLEWKDLNKWKFLTLNPLFFTVLRGALYPMSLIKTRLQVQT
GNEVYKGVFDAAKSIAKTEGLRGFYRGFLTSTYYIASGNIYIASYELARQLVYERE
LPFSNGKTVTIPMANFFAGATSSIITQSVVVPIDIVTQRLMMQGQKGMGEVLKARD
IFRSIYRNEGMSGFMRGYSTSLLTYVPNSAIWWSAYGVVSSKLYNMFKDKLPSDN
AAKYSTQALGGSIAGLTAAVLTNPMDVIRTRHQVLTADVARVSIKDTVSSLLKTE
GAKGLLKGVCAKMLSQAPSGAMVCFIYETTKRLSQISNDEQAPF

>mitochondrial 2oxodicarboxylate carrier
MSQTKAVVANDPSKRPFWKTLVAGGVAGSAEICIMYPTDVIKTRHQLKSGAGEG
MITTFKNILKNEGTPALYRGIISPIFAEAPKRAIKFSTNEKYKKLTCKKDGSLNGFRA
GIAGAMAGMSEVVVNCPFEVVKVRMQAKEAAGLYKSTGDCAMQLLKKEGPIAL
YRGAEPQLWRNAAWNGVYFGLIGSIRSAYPMKKDVSNGGKLFYNFCTGVVGGTI
ATVFNTPFDVIKSRMQNQGAGVAKYVWTIPSAMTIVKEEGVRALYKGFGPRIVRL
GPGGGIMLVAFDFISKLLERF

>ABC transporter A family protein
DDIDEEIAASSKTIDKYDFDFLDESLINQSDLSDEFSSNDEAYILNNDTETHIRSIGED
NNEDGAVDMTLPDQHMYNWIDFGHENSILIDEVLDDEDEVLNKTQNDENHKRYD

FKTCVDDPYIMSASQKLPTATMNSLEPPVGDTVIEVAATIVSLALFPALVAEVIQER
RGNLLSLMKLQGLKQSAFWVSTYMNCFMSTLLVQTLWYIFCYIMNITVYVETNW
ITIMTMNITWSHACVGMVFITSALLVRSSMSSFFCYLYLTISSGACPVLVAVVGSFN
HNLLYWPSMAQSQIMFILVDQRVQDTDVLNQANIALLLSGTVFIVVGLYLHQIIAS
PEDPEKKHPLFFLNFIWVFICGFFKMVWYILKSVACCCIHICKKRKPDFSKLEVPLL
SPLEREDSNGSNDVYLEHLRALEEGTKNLPVRILNLTKAFKSSGWNNKHLAVNKL
SLTVQPKECLGLLGPNGSGKTTTLRMLSGHETITSGDAWICNAHVVKRRKQAAKS
LGICPQYDGVSYLYIKYIIT

>ATP-binding cassette subfamily B

MLSRGRHVLSLSRVATGKVPTFARGVVGPSGPPLPSFGSADKLVNPLSKKIIKEV
WGHVWNPKVASAATKWSLRSRLLVVVGLMLGGKLIIVQVPGIFKKIVDDLSNETP
EDKQKRKEDLLKYGIPLALLLSYGVAKAGASGFQELRNAIFATLAQRTIREISRDLF
LHLHSLGLKFHLNRQTGALSRVLDRGSRSVEYVLRSTMFNVFPTIFEIGVVTTMLG
MQCGPQFAGVTLGTVTAYTAYTIAMTQWRTKFRRQMNSLENEASSKAIDSLLNY
ETVKYFNNEEHEANQYTKTLKGYQNAALKTMTSLSALNFGQQLIFAGGLTAQMIL
AATQVVNGNMSIGDLVLVNGLLFQLSMPLNFIGSTFRDLRQALIDMEAMFELGAT
TPDIKDDDDATELEVSHADLQLRDLEFAYEEDNSEDNVVLKGLNVDIKGGSKVAF
VGSSGAGKSTLLRLLFRFFDPQNGSILIDGQDSKKVTMKSLRKNIAVIPQHTVLFND
SIYYNIAYGNPEASEEEVHRAAKMAQLHDSVMRMPLGYKTQVGERGLKLSGGEQ
QRVSIARAILKNAPILLCDEATSSLDSSTEKNIMGAIDHIAKDRTTIMIAHRLTTIKN
CDEIFVLDDGKVLEHGTHEELLAKDGKYADLWFAQLLSDSMKSPCPAVDSAAKF
DKRAEDIEIDDLTGCSKCPSRGSGCGPR

>ABC transporter E family member

MSSDTVRIVIVDPEKCKPKKCNHQCKRSCPVVKMGKQCITVNKKSSAAEVSEVLC
IGCNICVKQCEFDALKIINLPKNLESQTSHRYSANGFKLHRLPMPRVNEVLGLVGT
NGIGKSTAIKILAGRLQPNLGRFDEIPDWDMILKNYRGSSLQNYFKNILEEKITCMV
KPQYVDQIPKAVRGQVDALLSKKNQRGMKEKLCSTLDLNHVLDREIEQLSGGEL
QRFAIALVAVRNCNVYMFDEPSSYLDVKQRLRAAETIRSLLTIDHDVYVLCVEHD
LSVLDYLSDYICCLYGQPGAYGVVTMPFSVREGINVFLAGFIPTENMRFRNESLTF
KVAQSAGNDRSTGQEKKDEGETRAGVYKYPAMTKTLGGAKGRAK

>ABC transporter F family member

MARKKSGRMAAKMKAMKLAEAKKVEEAATKKTEDKKDAFDLINCTFAENEDKI
ERNVRDIQVTNFNVNLAGLELLVNAELKLSYGCRYGLLGLNGSGKSTLLKVIGRR
LIPIAKSIDIFHLDQEVPGSEMSALEAVLEVDEEKKVLEDEADEITESGDDSDEATD
RLMQIYERLDELCAADAESRAAKILHGLGFSPAMQIKKCKDFSGGWRMRIALARA
LFVNPTCLLLDEPTNHLDMEAVVWLERYLAKFQKILLLVSHSQDFLNNVCSHTIHL
QQQKLKYYGGNYDTYVQTRAELEQNQAKRYAWEQEQIRHMKEYIARFGHGSAK
LARQAQSKEKTLEKMIQAGLTEAVAKDKVLNLKFEDPLPLPPPVLQMQEVTFGYE
PTDLLYQNVDFGVDLESRIVLVGPNGAGKTTLLRLISQELMPLDGQVRHHPHLKL
CRYNQHYADVLDLEQTPIEFMMAKFTKQLDLVQTRQFLGRYGLSGKDQMTKIKY
LSNGFKCRIVFAYIAKQNPHIMLLDEPTNNLDIETIDALAKAIKDFKGGIVLVSHDM
RLIEQVADQIFICDDRTITHFKGSIADYKKVLADKVEQDL

>tricarboxylate carrier_1

MEGVRLPAPIRRLINPRSSGTTSAEVEVRPAIDASFYDQSTFIGRVRHFFEITDMRT
LFANKAKLAECEKIVNSYREGDMSVSFAEARTAKKTLDAVLHPDTGKPILLPFRM
SFQVPGNLVMFAGLMNGTSFGATIAWQIINQTFNVGVNFCNRNASSSLTNKQILST
YFAACGGSVAVALGVRQMVKNVKPAFKAVATRAVPFFGVCAAHFLNVGLMRRA
EMEEGVEITDSEGTIHGKSKICGKRAILQTVASRIALAALVMTTPPLIMSGLTKAGL
VPKKFRLPIDIALIGAMIWAGNPFAIGVFPQKYGIDTAKLEPEFQGLKMSDGTPITQ
VTYNKGL

>tricarboxylate carrier_2

MESYKQQFKSFVDIVKTECQIKMNGILPLEVPNVSQATYGGRCLHFLKMMNPMY
MATMNKSAKKQHDEMMASVEKGAEIDESFRAKKDLSHATIHPQSGEPLPYLTRM
TAFPIVSVPNVLGMLLAAPTVGNLIFWQWLNQSYNAYFNYANRNLAGSMTTADV
AKSYGLAAVSACSIAVGLNGLVKKLSTVFSPAIMGSVGILIPYLAVGGAGSINAFV
MRKSELTTGVGLTNSAGEMEGNSQIAAKKCMTEVVLTRMAIPLPILLVPPIVRKAII
AAKMMPKSKIGTVLMDVGVVVACLYGALPLAVGMFPQNGSIKRKDVESQFNPEG
CDGEEMLYYNKGI

>tricarboxylate carrier_3

MNQAESELIEKKSNNFMEKYDQSTFWGRACKFIEQIDPWLALVSKKEALKAKEIIE
SVKDNPSEIDNYSEKEFDRARRIYLNSVHPQTGHLVPKLGRMSSYALINLPISAGMI

FSAPIVGNIILWQWVNQSLNAVFNFCNRNTENGKEGDDTKGKRSAIGSEVWKSYF
LAASTAVGAGLGFNSLLRFTHNMPPAFSSVMRVMVPFVAVATAGCVNAFTMRRN
ELVEGIPLEFEDEQTTDLKSKKAAKICMSQVMLSRIVLPFPMLIGPPLVINALKYMK
LFPSGNLAKKGLEAFIMGMFVYGALPIAVGLFPQRGAIRLEKLEPRFSDLVKETRC
DEDLKNFKLYFNKGL

>LAO/AO transport system ATPase
MLGKLKLKSLAFIGQRSFAEILPRTQQLIDGILSGSRSHLSRGITLTESKKASDIEQA
NFLLNECLKGKQAKGSKSFRVGIAGPPGAGKSTFIEALGTQLAQAGTRVAVLAIDP
SSVRSGGSILGDMTRMVELSVHENAYVRPSPSRCTLGGVAEATNDVASLCEACDY
NIILVETVGLGQSEVAVDDTVDMVMLVVPPANGDELQGVKKGIVEISDLVVVNKS
DGDFVNAAKKSQSDFTHALHMIRHKNPNWEPKVRRCSAIKKGESVSKVWEVCEE
FNKTMSENGEIEKKRSKQRTQWMWDQLKRQVLARVAENKTVSAVAQEVENRLN
EGVVTSRQAAKDILDAYFKTGICIDEE

>glutmate-aspartate transporter
EPVNSSCLNSQSFTLRILKGFQKMSQAVEIQKTKPKKVPLPAKFVVGGVAGIVGTT
AIFPIDMVKTRLQILNPQGMKMYSSPLHVVKSLLKHEGIAGFYRGLIPNLIGVTPEK
AIKLAVNEILREVLEDKNGNITFKKEVLAGMGAGTCQVIATNPMEIVKIRLQVQST
LPAAEQMSAKQVVSHLGLRGLYRGAAVTMMRDVPFSMMFFPGYANLNAWFEK
KLGRQALWASLLSGCLAGAFASGAVTPMDVIKTRLQVKGGAKLYNGFLDCFSKV
VANEGPSALFRGLVPRACVQAPLFGIALLAYEMQKRYMFRNEQ

>sodium/potassium-transporting ATPase subunit alpha
MSDLDKRRSSAEIVAVANELRQRSQMRSKSKFSFKKKKEEVNLKENIEMDEHQIP
MQDLVARLDSNLTSGLTDAQVAAKQEREGHNRLSPPETTPEWLKLFRQLTGFFAL
LLWLGSFLCLIGFVLKGEADNLYLAIVLAAVVTITGIFSYMQERNSSNLMDSFKNM
MPTVTVVTRNGKDVEINAIDLCRGDIVKVKGGDKVPADIRVISCTDDFEVDNACL
TGESEPQKRAPECSDVNPLETKNLIFFGTLIPKGRCTGIVVSIGDRSVMGRIAALAT
NVDADMTPIGKEIHHFILIVSGVAITLGVGFFIIGLMLDTDMITNLVFMIGIIVANVP
EGLLATVTVCLTLTAKRMASKSVLVKNLEGVETLGSTTCICSDKTGTLTQNVMTV
ANLLFDNTIFSTELVANDRWALYDDKNVTLQKFQRCMTLCNNATFDENSKYEEII
GADGKKTVNRDVSVEFMSEKILGDGSKQLTINWKTLGDGSESALIKLVQNKQDV
VETRKLSKKLCEIPFNSANKYQVSIHQLGESNNTVLVMKGAPERIISRCSHVLMNG

EEVELTEELKNEIEEKQMTLQAEGRRVLGFCEKELPDEYTPDYEFNIDTRNFPMGE
ASIEQAGDIVPDNKQLQKLTFIGLAALIDPPRVQVPPAVSKCKTAGIKVVMVTGDH
PVTAKAIAKEVGIIWGKTSDDIIADNKKAGAEDPSHPDYVDPCLAPAIVVAGKEFS
VDTAKETWDYYLDHTQIVFARTSPQQKLLIVENCQERGEIVAVTGDGVNDSPALK
KADIGVAMGIMGSDVSKEAADMILLDDNFASIVAGIEEGRLIFDNLKKSIAYTLSSN
IPEISPFLLFITLQIPLPLSTVLILCVDLGTDMVPAISMAWETAEADIMRRPPRNSSVD
RLVTKKLVVFAYLQIGVIQACAGFFTYFVVLNDYGYPPSVLFGLGAFDNWGKQIM
YCKFEDGMELVNKMGTAYTIPTGQTIADYFDTAIQEGFAMWDGEEVEDCVFAAK
NYLGRGSTGNIDLNDAATFLGGTEDFELPSLEAIEALYKAEYYPYVPYKAVTSPFW
KTEWLKQNVDDSDIPGFGSADELLTFQHQLPGVFRIETAGIYTSDTSGGLKAMTEA
EDALDLFTFTDTTPIGNRVYREATFVNDAANFKYHMTTLEGNVAYTNVASRMAQ
KEALHHAQCAFFCSIVVVQWADLMICKTRWLSIHHQKMTNVAMNFGLLFETILA
CCICYFPAFNVAVKTRDIRLVHWFCGMPFSVLIFLYDEVRKFVMRKTTQVTEDKE
TGRVISKYGWLAKNTYY

>carnitine/acyl carnitine carrier
MDNLKHYFVGSAGGISAILVGQPFDYLKVRVQTSKKMYTGPMDALFQTVKTEGII
GLYRGMSAPLLGVAPIFGVSFWGFTAAKQFLLTVREFPTCQNLSIPDVAAAGGLSS
LVTSSIITPVERIKCLLQTQNLGNRKPIYRGAIDCGRLLTKSEGLSSLYRGFSATLLR
DSLGFAAYFAGYEMAKRYLSKDKEQKPYHIVLAGSFAGLSYAMLAMPADIIKSR
MQSGISKNGFVTVATEVMGDYGLKGFYKGFAPLALRAMPAAAACFLGIEAASKF
L

>mitochondrial carrier protein
MPNTFINIAAGTCGGIAQTLTGHPFNTLKVRLQTSNAFSGPMDCLRTLVKKEGAM
AVYKGVSSPLVGLGAINAVLFMANEFAKQHLTNDQGYLSMPRVVAAGVFAGLA
QSFVVCPFELVMVRLQTQFVHFNGRKVYTGPIDCAKKLIHNNGIKGIYQGMGSTII
REIPECGTFLAGYEYAKRLLMKYNHGRDNSGIQLLAGGFAGVFSQVVQLPFDTVK
SRIQTSLQPESFVKKTKALIKSDGFMGFYKGLGPVLLRAFPANAATFVAFEFTKKH
LEERFVMS

>mitochondrial 2-oxoglutarate malate carrier protein_1
MAQQQKTLPSWAKFACGGLGGMTGWLFVHPFDTLKVRMQLLSEGGKKLEGGAL
KQMLSIVKSEKVPGMYRGLSAALMRQATYTTIRLGLYDVVREAWAGSLPPSELPF

YKKASIGLLAGAIASFACCPVEVSLVRMQADGRLPVADRRGYKNVFNALYRIGAE
EGAATYFRGATPTVARAMVVCMTQVAMYDQMKTVYKNYLGMKDGVPLHFAG
ALSAGLIYSYCSLPLDTAKTRMQNQVVSAEGKLMYRSIPQTLGMIAKQEGPTALW
RGFSSYFARSGGHTVFMFLCLEQYRKLFRHMLNVQQ

>mitochondrial 2-oxoglutarate malate carrier protein_2
MPRIANVLEPYCCGGTAAIIAASTIHPIDLAKTRIQVLGLTAKPGDAKPAAIPILKQI
FKSEGVPGLYAGLSASVMRQAVYGTARLGLHRSFSNKLEEMRGGKIPLWQKIGSS
FVSGAIASTIGNPCDLAMVRMQADSLKPIAERRGYKNVFDAITRVVKEEGLFTLW
RGCSPTVYRAIAMNVGMLATFDQAKDMLTPLMGDRLITTLTSSALAGFGCAFMSL
PFDMMRTRLMNMKPLPDGSMPFKNTLDVAVKTVSREGVPALWRGFVPFYLRCAP
HAMIILCIMESVRNVYRTVTASH

>solute carrier family 25 member 27
MASIGKELSLFELAKQSFIQCSLGGMSCGVATVATNPIDVVKVRMQLANKVPGQK
LGMGRTFLNVAKTESVQGLYKGLFASILRSFSYSAIRLGLYGPAKSIVGADSDPTLL
KKMLSGSISGSIAAGIANPIELVKVRLQADCARGNEKLYRGIVDAFVKITRTEGLM
TFKTGLFPHICRGAVITCSQVATYDHVKCFLRDRMEVKEGVGLHFLSSLVAGLVV
TTASSPLDVIKTNMMHAGKFNNAFECIQHCRKTGGVMSFFKGWLPNYT

>solute carrier family 25 member 39_1
MESRSNRICSSIVASGIVSVFVNPFDVLRVRMQNTMSKSTLSVIKKKCACGRMFYV
PIQCCKKSTGQNSGAKLVSTILKAEGLGGLWSGTLATMFTMIPQSMIYYSLYDELK
PAMESTILPGFLVPAVAGGFSRSIAATVVSPFELIRTNLQANGSELGIWGTAKREMK
TSGVKALWRGLIPTLGRDIPFSVIYWSAFEMIRPHLQKNRKPNNQRSSVSSFFTGVL
AGSIATVSTHPFDIIKTRRQLGLYGLESTLPKTCSLPALVKQLFLDEGIAGLTVGLGP
RLLRVSIACGIMISSYDLIKSMIQKSTGAEENQKM

>solute carrier family 25 member 39_2
MPIFASETSRGNRICSSIMAAFMTIQVVNPLDVIRVRLQNMGSKNIQMAVKDCTTC
SGLRQNSLKVLGSVIKSEGPATLWAGALTAMYAAIPSTMIYFTMYDELKMTFGKT
NLPTFFVPAVAGGLARIVTTTAVSPFELIRTRLQAEGNIGMMNIIKEEAKRGGLWR
GFLPTLYRDVPFSVLYWTSYEAIRPLLGEKQQQSSGKRGLASFLTGCSAGSLSAIIT
HPVDIIKTRRQSTAGGPTSTAVLTRQLMKEVKVNRQIRKLVRKKKKKTFAG

>pyridine nucleotide transhydrogenase
GGADMPVVVTVLNSYSGWALAAEGFMLHNPLLSVVGSLIGMSGFILSHIMCKAM
NRSILSVILGGLGAPVANVIEVADGAKATMATVEHAAERLMSANKVMIVPGYGM
AVSKAQWALADIANDLIAQGKEVSFGIHPVA

>aldehyde dehydrogenase
MLSSVIRSSCRSGVRAFASFMDLDPFRISKENPVKLYNLVGGRWKVSKSYIDIPDP
MNGDPIIKMPNTKVEELDEFFESAASCPKSGMHNPLKNPERYLMYGEVCHKAAA
MLEKPEVADFFAKAIMRVMPKSYPQAAGEVKVTKQFLKNFAGDNVRFAACGQH
TAGDHQGQMANGYRWPFGPVAIVSPFNFPLEIPVLQLMGSLFMGNKPTIKAASTV
SYVLEMFIRMLIEDCGLPASEIDLLHCGGRVVGELIAEGPFRVSQFTGSSGVSEQLA
SSTRGKVRVEDAGFDWKILGPDVQEFDYVAWQCDQDAYACTGQKCSAQSILFMH
ENWAEAGIEERLKEHASKRNLADLTVGPVLSVSTKEMLDHTDKLLAIPGARVCW
GGKELKNHTIPEIYGAMEPTAVFVPLDEMLKPENFELCVTEIFGPYQVVTEYKEGE
VDKVLEALERMSHHLTAGVVSNDPIFTNHILANTVNGTTYSGIRARTTGAPQNHW
FGPAGDPRGAGIGTTDAIRLVWSCHREIVTDIGPVPAEFEPKCT

>nadph:adrenodoxin oxidoreductase
MIHANMIPSSLVRHFSKLKVAIVGSGPSGLYFARQLIQAKGQDNSEINVIDKFPVP
GGLIRSGVSPDHQQARNVLNEFKKMFLTNKDSISFFGNVEVGEDITIEELRKTHHA
VVLATGCEGDRKLNIPGEDLSNVFSARDVVRWYNGYNPDLEINLTDISSVCIIGQG
NVALDVARIFSKSYEELKDTEIASYALEALKNSNIQEVNVLGRRGHIQTAFTMKEI
RELNVLENAKMIVYPDDIERGLNDASLHEALESRACAKKKKFFNDISQEFEKDNE
KHPKNLKMRFLVSPYQILGDDDSNSNTVGYLSLKRNELSGEPFKQVAVSTDETILL
KCGMVISSVGYKGEAMEGVPFDETKSVIPNKNGRVLQKPVDTKNNKTNQEEFVE
GLYCTGWLKRGPKGVIGTNIGDSQDTLVSLLEDFGDSLPHLPDTRLEDILDKKFVN
FEEFMKIDAAEIENRRRGSNAPRERFSEAEMLKFLDKKQNNQ

>hypothetical protein_1
MNILRFAQTRPFAFNIISTSVLSFFGDGIAQQFDSDEGYSLKRGLRMGLFGMLVS
GPPMYCWFKYINQKIVGRSMRDVMKKVLLTQSVVSPIMNSTFFGYIEFWKQIDAE
VFDTKDMIESYRARIKGDLIRAMSISTGFWIPAQMVNFKFTPSFLQLPVTAVYNVI
WCAIISYIGYRKKQKLKNE

>hypothetical protein_2

MLSRFCFKLGRKCQTRAFSGLNKDIWSPERDEYFKKINDLVEKFPATSLARAVM
DREHALQSAAKALNDNNDKELKQILKPFLPFSVEDDVKQKVHERELQLKPAILSE
MSKIMSRIPRSYRHEVRKRAAVFMPLAVVDGVPSIILTKRSNTISRHRGQICFPGGM
LDTADTSIIATACREMEEEIGIVPRQIMGILRCDWTEIESLVGVGVTPVVGTIGDLNR
CDVRLNPHEVERIFTIPLADVNNEALWLHREYAPPVFQGPDCVIWGLTAYVLDKF
REMVLAKVPGQWNTIPTSSIHHDKCGCDTCVPPSECSESEDIQDNQIVEEPNKFEK
VTSSKN

>hypothetical protein_3

MLRNFARVTAIAAGCLGSYACIEEFPSFFTEKFPKTPEFMREDIPLEQKLLMLEKT
KDITVIDPKKHNFFIELDNHLIWDSLYSDDHLHNYKLYVMENEIWAKGSLGKSLC
GHQSILHGGFTAAMLDEFFGALFYSLKVGHGFTGNLNVSYRKPIKTESDVCLRGW
VEKIDGRKYTLKANVFDMDGNLMATATALFIKVNVVPQDLLKPKATNDGKKN

>hypothetical protein_4

MLSRCFARQFGKLSLLSSLSMKPHPEKAHKGGEDAAFVEQSIVGVFDGVGGWQ
EHGVDPADYSRQLKKGSLEYIRNHKQYSLKEVLAHAHDSAKDVVGSSTACLARLI
PEETKLETLNLGDSGFFYFKRDDENNRFEVIYAAEEQRYGFNCPAQLGTGWELTA
NEAASQDITPSHGDILLMATDGMFDNVFVFEIEDLLSNSELLSLKDMKNVTKDDV
QRMTDEFAAELTELTFEKASRDHGFSPFAIEAKQFNIPFEGGKLDDITIILSAIVDED
TLA

>hypothetical protein_5

MLSSSSLRGTRILLRSRPQVARQFSISGLKSTNPLVKLGRSYNKCLDKRPVVTQ
MFTSLVLFGAGDVVAQVYAPSEKGFDWPRLMRCSFFGLTTVGILGRLHYKFMEY
FCTSLFKFNPKLIPFIKIVLEQWVYWAPFLMSQYHFQLSLMEGNSVKDSVARVKRE
LMPTLKANWTFWPFIQFFNFKFVPVRQQLNFVLVASLVWTTYLSWRFPPVKEGEQ
AAIEDVKDTQEEVLIEEIESDSESESETEEVEVFAKEQPEILSSETVTPSGEEFVEPIKP
ADVEVTPSGDDEVEVEIVEEDIISAPAEGIVVEEEHVEEN

Figure 2-8. Deduced amino acid sequences of proteins putatively operating in MROs. Mitochondrial targeting signals predicted by MitoProt are shown with bold.

Chapter 3

Complex evolution of two types of cardiolipin synthase in the eukaryotic lineage stramenopiles

Abstract

The phospholipid cardiolipin is indispensable for eukaryotes to activate mitochondria, and it was previously reported that two phylogenetically distinct types of enzyme synthesizing cardiolipin, one with two phospholipase D domains (CLS_pld) and the other with a CDP-alcohol phosphatidyltransferase domain (CLS_cap), are patchily and complementarily distributed at higher taxonomic (e.g., supergroup) levels of eukaryotes. Stramenopiles, one of the major eukaryotic clades, were considered to exclusively possess CLS_cap. However, through my present surveys with genome or transcriptome data from a broad range of stramenopile taxa, species with both CLS_cap and CLS_pld and species with only CLS_pld or CLS_cap were discovered among this group. Because these homologues of CLS_cap and CLS_pld retrieved from stramenopiles were likely inherited from the last eukaryotic common ancestor, it is reasonable to assume that a common ancestor of all stramenopiles harbored both CLS_cap and CLS_pld. Furthermore, based on the robust organismal phylogeny of stramenopiles unveiled with large-scale phylogenetic analyses, the earliest diverging lineage of stramenopiles (including bicosoecids, placidids, etc.) was found to comprise species with both CLS_cap and CLS_pld along with species with only either CLS_cap or CLS_pld. These findings suggest that a common ancestor of the most basal stramenopile lineage retained these two vertically inherited enzymes and that differential losses of either CLS_cap or CLS_pld occurred in this lineage. On the other hand, in the other stramenopile lineage composed of Ochrophyta, Pseudofungi, and Labyrinthulomycetes (to the exclusion of the most basal lineage), only CLS_cap was found, and therefore a common ancestor of these three groups likely lost CLS_pld. Based on these

findings, the evolution of CLS_cap/CLS_pld in stramenopiles appears to be more complex than previously thought.

3-1. Introduction

Cardiopilin (CL) is an anionic dimeric phospholipid with two phosphate residues and four types of fatty acyl chain (Fig. 3-1). It is widely held that CL in mitochondria plays a pivotal role in stabilizing and/or regulating various mitochondrial proteins such as the respiratory chain complexes/supercomplexes (e.g., Fry and Green 1981; Eble et al. 1990; Robinson et al. 1990). Furthermore, mitochondrial CL is involved in organizing structure characteristic of mitochondria such as cristae, as well as in aging and apotosis (e.g., Schug and Gottlieb 2009; Sakamoto et al. 2012). It is also known that bacterial CL is essential to activate proteins associated with energy metabolism and to properly localize membrane proteins (e.g., Yankovskaya et al. 2003; Gold et al. 2010; Arias-Cartin et al. 2011). Based on these findings, it was originally proposed that CL is exclusively located in mitochondrial (inner) and bacterial plasma membranes (White and Frerman 1967; Zinser et al. 1991), but recently this type of phospholipid has also been found in the plasma membranes of a very limited number of archaeal species and in the peroxisomal membranes of eukaryotes (Lattanzio et al. 2002; Koga and Morii 2007; Wriessnegger et al. 2007; Lobasso et al. 2008).

CL is known to be biosynthesized by either two phylogenetically distinct enzymes: CL synthase (CLS) with two phospholipase D domains, CLS_pld, which synthesizes CL from two molecules of phosphatidylglycerols (PGs) (e.g., Nishijima et al. 1988) or CLS with one CDP-alcohol phosphatidyltransferase domain, CLS_cap, which produces this lipid using a PG and a cytidine diphosphate diacylglycerol (CDP-DAG) as substrates (Fig. 3-2) (e.g., Chang et al. 1998; Tuller et al. 1998). It has been believed that

CLS_pld and CLS_cap function in bacteria and eukaryotes (mitochondria), respectively (e.g., Schlame 2008). In contrast to the bacterial-type CL, mitochondrial "immature" CL synthesized by CLS is further remodeled (reacylated), resulting in mature CL generally possessing the same fatty acids at sn-1, 2 sites in one molecule. This eukaryotic CL maturation pathway consists of two steps: in the first step, immature CL is deacylated into monolysocardiolipin (MLCL) with either CL-specific phospholipase (CLD) or calcium-independent phospholipase A2 (iPLA2) beta/gamma (Beranek et al. 2009; Malhotra et al. 2009; Zachman et al. 2010). After the deacylation in the first step, MLCL is then reacylated to generate mature CL with either CoA-independent tafazzin (TAZ) or acylCoA:lysocardiolipin acyltransferase 1 (ALCAT1) (Cao et al. 2004; Gu et al. 2004).

Recently, exceptions to the above-mentioned "simple" hypothesis regarding CLS phylogenetic distribution, in which CLS_pld and CLS_cap are exclusively found in bacteria and eukaryotes, respectively, have been found: actinobacteria and proteobacteria were found to contain CLS_cap-like proteins (Sandoval-Calderón et al. 2009; Tian et al. 2012). In addition, Tian et al. (2012) reported that the eukaryotic supergroups Amoebozoa, Excavata, and Alveolata, a subgroup of the supergroup "SAR" (Adl et al. 2012), have only CLS_pld (without phylogenetic affiliation to any particular bacterial homologues), while the supergroups Opisthokonta (including animals and fungi) and Archaeplastida (including land plants) along with another SAR subgroup stramenopiles possess only CLS_cap (closely related to alpha-proteobacterial homologues). It should be noted that Tian et al. (2012) analyzed a broad range of eukaryotes including non-model organisms other than animals, yeasts, and plants, resulting in the detection of CLS_pld in eukaryotes. It is also remarkable that both CLS_cap and CLS_pld are patchily and complementarily distributed

at higher taxonomic levels of eukaryotes. Based on these findings, Tian et al. (2012) proposed the following evolutionary scenarios: a primitive eukaryotic cell prior to acquisition of mitochondria inherited CLS_pld from its ancient bacterial ancestor, and then mitochondria originating from an endosymbiotic event of an alpha-proteobacterium took CLS_cap to the last eukaryotic common ancestor (LECA), meaning that the LECA harbored both CLS_cap and CLS_pld. Subsequently, either CLS_cap or CLS_pld was differentially lost in various eukaryotic lineages (Fig. 3-3).

In chapter 2, I analyzed the metabolic characteristics of mitochondrion-related organelles (degenerated mitochondria) in the free-living anaerobic stramenopile *Cantina marsupialis* based on RNA-seq data. Through those analyses, I found that this stramenopile species has the gene encoding CLS_pld rather than CLS_cap, which stramenopiles exclusively possess according to Tian et al. (2012). Stramenopiles are a huge monophyletic assemblage of eukaryotes comprising Ochrophyta (composed of a wide variety of photosynthetic lineages including diatoms and brown algae) and two heterotrophic groups, Bigyra (including bicosoecids, labyrinthulomycetes, placidids, and *Blastocystis*) and Pseudofungi (composed of oomycetes, hyphochytriomycetes, and *Developayella*) (Cavalier-Smith and Chao 2006). Nevertheless, Tian et al. (2012) surveyed the presence or absence of CLS_cap/CLS_pld in a very limited number of phylogenetically narrow representatives of stramenopiles (five species of oomycetes, two species of diatom, one species of brown alga, and *Blastocystis*). Therefore, it remained uncertain whether many other stramenopile lineages contain CLS_cap or CLS_pld and whether only *Cantina* exceptionally possesses CLS_pld among stramenopiles.

In this chapter, using genome or transcriptome data, I examined the presence or

absence of the two types of CL biosynthesis enzyme, CLS_cap and CLS_pld, together with the CL maturation (remodeling) enzymes, CLD, iPLA2, TAZ, and ALCAT1, in stramenopile species more comprehensively in terms of phylogeny than in the case of Tian et al. (2012). Consequently, several species of stramenopiles other than *Cantina* were also found to have CLS_pld. I further traced the evolution of CLS_cap/CLS_pld within the radiation of stramenopiles based on the organismal phylogeny of this large eukaryotic group as resolved by large-scale phylogenetic analyses.

3-2. Materials and methods

3-2-1. Cultures

The strains *Cafeteria roenbergensis* NIES1012, *Wobblia lunata* NIES1015, and *Developayella elegans* NIES1388 were purchased from the Microbial Culture Collection of the National Institute for Environmental Studies (NIES, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan). These cells were grown according to the instructions from the NIES.

3-2-2. Analyses of RNA-seq data from cultured strains

Cells of *C. roenbergensis*, *W. lunata*, and *D. elegans* were harvested by centrifugation at 4,170 g for 60 min at 4°C. Total RNA was isolated from the harvested cells using TRIzol reagent (Thermo Fisher Scientific, USA). Construction of cDNA libraries and paired-end sequencing with Illumina HiSeq 2000 (100 bp per read) were performed by Hokkaido System Science Co., Ltd. Raw sequencing reads were deposited in the Sequence Read Archive under the accession numbers DRA004177–DRA004179.

154–158 million raw sequencing read data were filtered using TRIMMOMATIC software version 0.30 (Lohse et al. 2012) to remove adapter sequences and low-quality bases. Filtered sequences were then assembled into 23,271–49,186 transcript contigs using the TRINITY package (release 2013-02-25) (Grabherr et al. 2011).

3-2-3. Sequence data collection

In addition to RNA-seq datasets from *C. roenbergensis*, *W. lunata*, and *D. elegans* generated in this study and from *Cantina marsupialis* generated in chapter 2, publicly available genome or transcriptome datasets from 80 taxa of eukaryotes (including 55 taxa of stramenopiles) were downloaded to find orthologous sequences for phylogenomic analyses and/or to identify the genes encoding enzymes for cardiolipin (CL) biosynthesis and maturation. Taxon names for which sequence datasets were collected are listed in Table 3-1.

3-2-4. Phylogenomic analyses

Genome/transcriptome data as mentioned above were used as inputs for an in-house pipeline, described below, for the creation of single protein datasets and, subsequently, the phylogenomic data matrix. The organismal data were individually screened for orthologues using either blastp or tblastn, depending on the data type, with the reference orthologue sequences used as queries in BLASTMONKEY from the Barrel-o-Monkeys toolkit (http://rogerlab.biochem.dal.ca). If the sequence dataset was nucleotides, then the tblastn hits were translated to amino acid residues. Blastp was then used to screen these putative orthologues against the OrthoMCL database, and the output

for each gene from each organism was compared against a dictionary of orthologous OrthoMCL IDs. Those putative orthologues that did not match orthologous IDs were designated as paralogues and removed. The remaining orthologues from each organism were combined and aligned using MAFFT-LINSI (Katoh and Standley 2013). Ambiguously aligned positions were identified and removed using Block Mapping and Gathering with Entropy (BMGE) (Criscuolo and Gribaldo, 2010). For each individual protein alignment, maximum-likelihood (ML) trees were inferred in RAxML v7.2.6 (Stamatakis 2006) using an LG model with four categories of among-site rate variation, with 10 ML tree searches and 100 ML bootstrap replicates. To test for undetected paralogy or contaminants, I constructed a consensus tree (ConTree) representing phylogenetic groupings of well-established eukaryotic clades (Brown et al. 2013). The resulting individual protein trees that placed taxa in conflicting positions relative to the ConTree with more than 70% ML bootstrap support, with a zero-branch length, or with extremely long branches were checked manually. All problematic sequences identified using these methods were removed from the dataset. The resulting protein alignments were then re-trimmed for ambiguously aligned positions using BMGE and concatenated into the separate supermatrix with 72,013 sites (245 proteins) of 84 taxa using an in-house script employing alvert.py from the Barrel-o-Monkey's toolkit. The ML tree for the phylogenomic matrix was estimated from 60 independent searches using RAxML under the LG model with four categories of among-site rate variation and an empirical amino acid frequency model, selected by likelihood-ratio tests and the Akaike information criterion. Topological support was assessed using 1,000 RAxML bootstrap replicates. Because RAxML is not equipped with models that account for heterogeneous site-specific

features of sequence evolution, I employed C-series models in IQ-TREE v1.3.3 (Nguyen et al. 2015) on the phylogenomic dataset. The best-fitting model available under ML analyses that I was capable of running with computational constraints was LG + Γ4 + C10 + F with class weights optimized from the dataset using the exchangeabilities from the LG Q-Matrix (LG + Γ4 + FMIX {empirical, C10pi1-C10pi10}) (Wang et al. 2014). Branch support was estimated from 1,000 ultrafast bootstrap replicates in IQ-TREE.

3-2-5. Identification of genes encoding proteins associated with CL biosynthesis and maturation

To screen homologues of two types of CL synthase (CLS_cap and CLS_pld), CL-specific phospholipase (CLD), calcium-independent phospholipase A2 (iPLA2), CoA-independent tafazzin (TAZ), and acylCoA:lysocardiolipin acyltransferase 1 (ALCAT1) from stramenopiles, blastp or tblastn was conducted against genome or transcriptome datasets from 59 stramenopile taxa using CLS_cap from *Phytophthora infestans* T30-4 (GenBank accession number EEY65382), CLS_pld from *Theileria parva* (EAN31681), CLD from *Ectocarpus siliculosus* (CBN79068), iPLA2 from *E. siliculosus* (CBN77084), TAZ from *P. infestans* T30-4 (XP_002904711), and ALCAT1 from *E. siliculosus* (CBJ33474) as queries with default settings. The top three retrieved hits for each protein from each taxon were further subjected to blastp homology searches against the Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg) database to remove possible paralogues of individual proteins. Among the sequences surveyed against the KEGG database, those assigned to K numbers, K08744, K06131/K06132, K13535, K16343, K13511, and K13513, were regarded as orthologues

of CLS_cap, CLS_pld, CLD, iPLA2, TAZ, and ALCAT1, respectively. Among these screened orthologues, those with top hits that were bacterial homologues were potentially derived from bacteria co-cultured with stramenopile species investigated in this study. Thus, I considered only orthologues for which top hits were eukaryotic homologues as originating from stramenopile species and utilized these in subsequent phylogenetic analyses.

3-2-6. Phylogenetic analyses of proteins for CL biosynthesis and maturation

I conducted phylogenetic analyses of six proteins, CLS_cap, CLS_pld, CLD, iPLA2, TAZ, and ALCAT1. The deduced amino acid sequences of these six proteins from stramenopiles identified in this study were separately aligned with the corresponding sequences from phylogenetically diverse organisms using MAFFT. The alignments were inspected by eye and manually edited. I then excluded ambiguously aligned sites from the datasets prior to phylogenetic analyses. The analyzed datasets had the following dimensions: CLS_cap, 76 taxa, 141 sites; CLS_pld, 47 taxa, 364 sites; CLD, 42 taxa, 305 sites; iPLA2, 74 taxa, 236 sites; TAZ, 56 taxa, 266 sites; and ALCAT1, 103 taxa, 209 sites. The alignment data are available on request to the corresponding author. For each single-gene dataset, the ML phylogenetic tree and corresponding bootstrap support values (1000 replicates) were calculated using RAxML. The ML tree was selected from 20 heuristic tree search initiated from randomized parsimony starting trees. In ML bootstrap analyses, a single tree search per replicate was performed. For the datasets of CLS_pld and CLS_cap, Bayesian analyses were also performed using MrBayes version 3.2.3. (Ronquist and Huelsenbeck 2003). Six parallel Metropolis coupled Markov chain Monte Carlo

(MCMCMC) runs, each consisting of three heated and one cold chains with default chain temperatures, were run for 1,000,000 generations. Log-likelihood scores and trees with branch lengths were sampled every 1,000 generations. The first 250,000 generations were excluded as burn-in, and the remaining trees were summarized to obtain Bayesian posterior probabilities. Convergence of parallel MCMCMC runs was judged by the average standard deviation of split frequencies (ASDSF). For both ML and Bayesian analyses, the LG model with four categories of among-site rate variation, which was selected as the most appropriate model with Aminosan (Tanabe 2011), was applied.

3-3. Results

3-3-1. Phylogeny of stramenopiles

The phylogenetic tree based on 245 protein sequences from 84 eukaryote taxa (including 59 stramenopile taxa) is shown in Fig. 3-4. In this tree, Ochrophyta comprising 10 photosynthetic classes showed monophyly with 100% bootstrap probability (BP) of both RAxML and IQ-TREE. In the Ochrophyta clade, the three lineages composed of Pelagophyceae and Dictyochophyceae, of Bolidophyceae and Bacillariophyceae, and of Raphidophyceae, Xanthophyceae, and Phaeophyceae each formed a monophyly with 100% BP of both RAxML and IQ-TREE. The lineage of Pseudofungi composed of 17 oomycete taxa and *Developayella elegans* (Bigyromonadea) was monophyletic with 97% BP of RAxML and 99% BP of IQ-TREE. The sister relationship between the lineages of Ochrophyta and Pseudofungi was supported with 100% BP of both RAxML and IQ-TREE. Four taxa of Labyrinthulomycetes formed a monophyletic clade with 100% BP of both RAxML and IQ-TREE, and this lineage was nested with the Ochrophyta/Pseudofungi

clade with 80% BP of RAxML and 79% BP of IQ-TREE. A subset of Bigyra, *Cantina marsupialis*, *Wobblia lunata*, *Blastocystis hominis*, and two species of *Cafeteria* formed a clade with 100% BP of both RAxML and IQ-TREE, and this clade branched most basally in the stramenopile radiation. In this basal lineage, the anaerobic species *B. hominis* was grouped with *Cafeteria* sp. Caron Lab Isolate and *W. lunata* with 98% BP of RAxML and 100% BP of IQ-TREE, while the anaerobic species *C. marsupialis* was clustered with *Cafeteria roenbergensis* with 98% BP of RAxML and 100% BP of IQ-TREE. Two species of *Cafeteria* were not related in this phylogeny. Either of them may be misidentified based on inadequate morphological assessment.

3-3-2. Presence/absence of the genes encoding proteins associated with CL biosynthesis and maturation

The presence or absence of CLS_cap, CLS_pld, CLD, iPLA2, TAZ, and ALCAT1 in 59 stramenopile taxa examined in this study is summarized in Table 3-2, and the distribution of stramenopile CLS_cap/CLS_pld homologues is mapped on the organismal phylogenetic tree (Fig. 3-4). Only the CLS_cap homologue was identified in 36 taxa that are phylogenetically broad in the stramenopile lineage (Table 3-2). However, only the CLS_pld homologue was detected in *C. marsupialis*, and both homologues of CLS_cap and CLS_pld were retrieved from *Cafeteria* sp. Caron Lab Isolate and *W. lunata*. Neither CLS_cap nor CLS_pld was recovered from the remaining 20 taxa. The homologues of the CL deacylation enzymes CLD and/or iPLA2 (beta or gamma) were detected in 44 taxa, while these two deacylation enzyme homologues were not recovered from the remaining 15 taxa. The homologues of the CL reacylation enzymes TAZ and/or

ALCAT1 were found in 47 taxa, but these reacylation enzyme homologues were not retrieved from the remaining 12 taxa. A significant fraction of the homologues in question from stramenopiles was missing. As almost all datasets surveyed in this study were a transcriptome or "draft" (incomplete) genome, I potentially missed the target sequences. Also, the absence of the enzymes associated with deacylation and/or reacylation of CL may mean that the CL maturation (remodeling) process is not necessarily indispensable for all stramenopiles.

3-3-3. Phylogenies of proteins for CL biosynthesis and maturation

In CLS_cap (Fig. 3-5) and CLS_pld (Fig. 3-6) trees reconstructed in this study, the monophyly of eukaryotic homologues was supported with 90% RAxML BP/0.99 Bayesian posterior probability (PP) and 100% RAxML BP/1.00 Bayesian PP, respectively. Stramenopile homologues of both CLS_cap and CLS_pld did not form monophyletic lineages in the respective eukaryotic clades, but such topology was not statistically supported. In the diatom *Fragilariopsis cylindrus*, there were duplicates of the CLS_cap genes branched to each other, suggesting that gene duplication occurred in this species. *W. lunata* was also found to contain two copies of the CLS_pld genes, but these copies were not related to each other in the eukaryotic radiation. The CLS_pld copies of *W. lunata* were possibly generated by gene duplication at the latest point before the divergence of *W. lunata* and *Cafeteria* sp. Caron Lab Isolate. In the phylogenies of all four CL maturation enzymes, CLD, iPLA2, TAZ, and ALCAT1 (Figs. 3-7–10), deep phylogenetic relationships could not be resolved as in the case of Tian et al. (2012) and the phylogenetic status of the respective stramenopile homologues was not clear, making it difficult to infer

the evolutionary processes of CL maturation enzymes from stramenopiles.

3-4. Discussion

3-4-1. Deep phylogenetic relationships in stramenopiles

In contrast to previous phylogenetic analyses based on the small-subunit ribosomal RNA gene (e.g., Massana et al. 2014; Yubuki et al. 2015) and seven genes (Riisberg et al. 2009), this phylogenomic analyses based on 245 protein sequences successfully resolved deep relationships of stramenopiles as follows: 1) the monophyly of Pseudofungi composed of oomycetes and *Developayella* was well supported (although another Pseudofungi lineage, hypochytridiomycetes, was not included in my analyses). 2) The lineage of Labyrinthulomycetes, a member of Bigyra, was sister to the large clade comprising Ochrophyta and Pseudofungi. 3) Other members of Bigyra including bicosoecids (*Cafeteria* spp.), placidids (*Wobblia lunata*), *Cantina marsupialis*, and *Blastocystis hominis* formed a robust clade, and this clade diverged the earliest in the stramenopile radiation. 4) In this basal clade, two anaerobic protists, *Blastocystis* and *Cantina*, were not closely related to each other. Considering the second and third findings, the heterotrophic stramenopile group Bigyra is unlikely to be a natural assemblage, although Cavalier-Smith and Chao (2006) argued that Opalozoa (including *Blastocystis*), Bicoecia (including bicosoecids and placidids), and Sagenista (including labyrinthulids) are classified into this group based on single gene phylogenetic analyses. The fourth finding is also remarkable. Chapter 2 revealed that several enzymes involved in anaerobic pyruvate metabolism (such as ADP-forming acetyl-CoA synthetase and acetate:succinate CoA-transferase) occur solely in *Blastocystis* and *Cantina* among stramenopiles

investigated and that the respective homologues from these two distantly related anaerobic stramenopiles are very closely related to each other, suggesting evolutionary scenarios associated with the lateral gene transfer of these enzymes required for anaerobic metabolism, as previously discussed in chapter 2.

3-4-2. Evolution of CLS_cap and CLS_pld in the stramenopile radiation

Tian et al. (2012) reported that all nine representatives of stramenopiles they investigated possessed CLS_cap but lacked CLS_pld, suggesting that CLS_pld was lost during the course of evolution from the last eukaryotic common ancestor (LECA) having both CLS_cap and CLS_pld to a common ancestor of all stramenopiles. Similarly, only CLS_cap was detected in 36 stramenopile taxa in this study. However, I also found several cases contradicting the theory of Tian et al. (2012): only CLS_pld was recovered from *Cantina marsupialis*, and both CLS_cap and CLS_pld were found in *Cafeteria* sp. Caron Lab Isolate and *Wobblia lunata*. Notably, all these CLS_pld-bearing species were members of the most basal lineage within the stramenopile radiation. Based on these findings, new evolutionary scenarios regarding CLS_cap/CLS_pld in the stramenopile group could be proposed, as shown in Fig. 3-11. The common ancestor of all stramenopiles probably harbored both CLS_cap and CLS_pld. It is reasonable to assume that both types of CLS were inherited from the LECA, as suggested by Tian et al. (2012), because monophylies of these two types of CLS from a broad range of eukaryotes were respectively robust (Figs. 3-5 and 3-6). It is also highly likely that a common ancestor of the most basal stramenopile lineage retained both CLS_cap and CLS_pld and that *Cafeteria* sp. Caron Lab Isolate and *W. lunata* inherited these two CLSs. In contrast, *C. marsupialis* and *Cafeteria*

roenbergensis may have lost CLS_cap and CLS_pld, respectively (only CLS_cap was detected in the latter). However, it cannot be fully ruled out that I failed to retrieve CLS_cap from *C. marsupialis* and CLS_pld from *C. roenbergensis*, because my analyses of these two species were based on transcriptome data rather than complete genome data. *Blastocystis*, another member of the basal stramenopiles, may have lost both types of CLS in the course of evolution (see below). On the other hand, a common ancestor of Ochrophyta, Pseudofungi, and Labyrinthulomycetes (to the exclusion of the most basal stramenopile lineage) may have lost CLS_pld, as CLS_cap was exclusively detected in as many as 35 taxa belonging to these three lineages.

Tian et al. (2012) argued that each eukaryotic supergroup (such as Opisthokonta, Amoebozoa, Archaeplastida, or Excavata) has either CLS_cap or CLS_pld. Nevertheless, in this pioneering study, only CLS_cap was identified in stramenopiles, a subgroup of the supergroup SAR (composed of stramenopiles, Alveolata, and Rhizaria), while only CLS_pld was retrieved from alveolates. Considering this situation, Tian et al. (2012) doubted the close relationship between stramenopiles and alveolates. However, I newly found CLS_pld in several species of stramenopiles. In addition, there are multiple lines of reliable evidence that SAR is a monophyletic clade (e.g., Hackett et al. 2007; Brown et al. 2013), and therefore I here propose the hypothesis that a common ancestor of stramenopiles and alveolates (or of SAR) possessed both CLS_cap and CLS_pld, with the subsequent loss of CLS_cap in the alveolate lineage. Thus, the doubt regarding the classification and relationship of stramenopiles and alveolates posed by Tian et al. (2012) could be dispelled.

3-4-3. CL in anaerobic protists

Based on biochemical evidence, *Trichomonas vaginalis* and *Giardia intestinalis*, representative anaerobic protists with functionally and morphologically degenerate mitochondria, so-called mitochondrion-related organelles (MROs), lack CL (Rosa et al. 2008; Guschina et al. 2009). In addition, Tian et al. (2012) reported that no genes encoding enzymes associated with CL biosynthesis and maturation were detected in the anaerobic MRO-bearing protists *Entamoeba* and microsporidians along with *Trichomonas* and *Giardia* using their genome or transcriptome data. On the other hand, the presence of CL in the anaerobic parasitic stramenopile *Blastocystis* and the anaerobic commensal ciliate inhabiting the cockroach hindgut *Nyctotherus ovalis*, both of which have MROs, was demonstrated (Keenan et al. 1992; Voncken et al. 2002). Furthermore, Stairs et al. (2014) identified the gene for CL biosynthesis (CLS_cap) in the free-living and MRO-bearing anaerobic protist *Pygsuia biforma* belonging to the eukaryotic lineage Breviatea. Similarly, in this study, the free-living anaerobic stramenopile *Cantina* with MROs was found to possess the gene for CL biosynthesis (CLS_pld). It should be noted that CL/CLS-lacking anaerobic protists (such as *Trichomonas*, *Giardia*, *Entamoeba*, and microsporidians) are completely devoid of membrane-associated electron transport chain (ETC) complexes in their MROs, while MROs in CL/CLS-bearing ones (such as *Cantina*, *Blastocystis*, *Nyctotherus*, and *Pygsuia*) retain one or two ETC complexes (Stechmann et al. 2008; de Graaf et al. 2011; Stairs et al. 2014; Noguchi et al. 2015). Given that CL is responsible for the stabilization and activation of ETC complexes (e.g., Fry and Green 1981; Eble et al. 1990; Robinson et al. 1990), the presence or absence of CL or its synthase in anaerobic protists may be reasonable. Curiously, I detected neither CLS_pld nor CLS_cap from

Blastocystis at the genome level, as in the case of the survey at the transcriptome level by Tian et al. (2012). How this stramenopile parasite produces or obtains CL is unknown.

References

Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EA, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick RS, Schoch CL, Smirnov A, Spiegel FW (2012) The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**:429–493

Arias-Cartin R, Grimaldi S, Pommier J, Lanciano P, Schaefer C, Arnoux P, Giordano G, Guigliarelli B, Magalon A (2011) Cardiolipin-based respiratory complex activation in bacteria. *Proc. Natl. Acad. Sci. USA* **108**: 7781–7786.

Beranek A, Rechberger G, Knauer H, Wolinski H, Kohlwein SD, Leber R (2009) Identification of a cardiolipin-specific phospholipase encoded by the gene CLD1 (YGR110W) in yeast. *J. Biol. Chem.* **284:**11572–11578

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

Cao JS, Liu YF, Lockwood J, Burn P, Shi YG (2004) A novel cardiolipin-remodeling pathway revealed by a gene encoding an endoplasmic reticulum-associated acyl-CoA:lysocardiolipin acyltransferase (ALCAT1) in mouse. *J. Biol. Chem.* **279**:31727–31734

Cavalier-Smith T, Chao EE (2006) Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista). *J. Mol. Evol.* **62**:388–420

Chang SC, Heacock PN, Mileykovskaya E, Voelker DR, Dowhan W (1998) Isolation and characterization of the gene (CLS1) encoding cardiolipin synthase in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**:14933–14941

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

de Graaf RM, Ricard G, van Alen TA, Duarte I, Dutilh BE, Burgtorf C, Kuiper JW, van der Staay GW, Tielens AG, Huynen MA, Hackstein JH (2011) The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. *Mol. Biol. Evol.* **28**:2379–2391

Eble KS, Coleman WB, Hantgan RR, Cunningham CC (1990) Tightly associated cardiolipin in the bovine heart mitochondrial ATP synthase as analyzed by 31P nuclear magnetic resonance spectroscopy. *J. Biol. Chem*. **265**:19434–19440

Fry M, Green DE (1981) Cardiolipin requirement for electron transfer in complex I and III of the mitochondrial respiratory chain. *J. Biol. Chem.* **256**:1874–1880

Gold VA, Robson A, Bao H, Romantsov T, Duong F, Collinson I (2010) The action of cardiolipin on the bacterial translocon. *Proc. Natl. Acad. Sci. USA* **107**:10044–10049

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol*. **29**:644–652

Gu ZM, Valianpour F, Chen SL, Vaz FM, Hakkaart GA, Wanders RJA, Greenberg

ML (2004) Aberrant cardiolipin metabolism in the yeast taz1 mutant: a model for Barth syndrome. *Mol. Microbiol.* **51**:149–158

Guschina IA, Harris KM, Maskrey B, Goldberg B, Lloyd D, Harwood JL (2009) The microaerophilic flagellate, *Trichomonas vaginali*, contains unusual acyl lipids but no detectable cardiolipin. *J. Eukaryot. Microbiol.* **56**:52–57

Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rümmele SE, Bhattacharya D (2007) Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol. Biol. Evol.* **24**:1702–1713

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

Keenan TW, Huang CM, Zierdt CH（1992) Comparative analysis of lipid composition in axenic strains of *Blastocystis hominis*. *Comp. Biochem. Physiol. B* **102**:611–615

Koga Y, Morii H (2007) Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol. Mol. Biol. Rev*. **71**:97–120

Lattanzio VM, Corcelli A, Mascolo G, Oren A (2002) Presence of two novel cardiolipins in the halophilic archaeal community in the crystallizer brines from the salterns of Margherita di Savoia (Italy) and Eilat (Israel). *Extremophiles* **6**:437–444

Lobasso S, Lopalco P, Mascolo G, Corcelli A (2008) Lipids of the ultra-thin square halophilic archaeon *Haloquadratum walsbyi*. *Archaea* **2**:177–183

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* **40**:W622–627

Malhotra A, Edelman-Novemsky I, Xu Y, Plesken H, Ma JP, Schlame M, Ren MD

(2009) Role of calcium-independent phospholipase A(2) in the pathogenesis of Barth syndrome. *Proc. Natl. Acad. Sci. USA* **106**:2337–2341

Massana R, del Campo J, Sieracki ME, Audic S, Logares R (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J*. **8:**854–866

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**:268–274

Nishijima S, Asami Y, Uetake N, Yamagoe S, Ohta A, Shibuya I (1988) Disruption of the *Escherichia coli* cls gene responsible for cardiolipin synthesis. *J. Bacteriol.* **170**:775–780

Riisberg I, Orr RJ, Kluge R, Shalchian-Tabrizi K, Bowers HA, Patil V, Edvardsen B, Jakobsen KS (2009) Seven gene phylogeny of heterokonts. *Protist* **160**:191–204

Robinson NC, Zborowski J, Talbert LH (1990) Cardiolipin-depleted bovine heart cytochrome c oxidase: binding stoichiometry and affinity for cardiolipin derivatives. *Biochemistry* **29**:8962–8969

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574

Rosa ID, Einicker-Lamas M, Bernardo RR, Benchimol M (2008) Cardiolipin, a lipid found in mitochondria, hydrogenosomes and bacteria was not detected in *Giardia lamblia*. *Exp. Parasitol.* **120**:3215–3220

Sakamoto T, Inoue T, Otomo Y, Yokomori N, Ohno M, Arai H, Nakagawa Y (2012) Deficiency of cardiolipin synthase causes abnormal mitochondrial function and

morphology in germ cells of *Caenorhabditis elegans*. *J. Biol. Chem.* **287;** 4590–4601.

Sandoval-Calderón M, Geiger O, Guan Z, Barona-Gómez F, Sohlenkamp C (2009) A eukaryote-like cardiolipin synthase is present in *Streptomyces coelicolor* and in most actinobacteria. *J. Biol. Chem*. **284**; 17383–17390.

Schlame M (2008) Cardiolipin synthesis for the assembly of bacterial and mitochondrial membranes. *J. Lipid Res.* **49;** 1607–1620.

Schug ZT, Gottlieb E (2009) Cardiolipin acts as a mitochondrial signalling platform to launch apoptosis. *Biochim. Biophys. Acta* **1788**; 2022–2031.

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

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

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

Tanabe AS (2011) Kakusan4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Mol. Ecol. Resour.* **11**; 914–921.

Tian HF, Feng JM, Wen JF (2012) The evolution of cardiolipin biosynthesis and maturation pathways and its implications for the evolution of eukaryotes. *BMC Evol.*

Biol. **12**; 32.

Tuller G, Hrastnik C, Achleitner G, Schiefthaler U, Klein F, Daum G (1998) YDL142c encodes cardiolipin synthase (Cls1p) and is non-essential for aerobic growth of *Saccharomyces cerevisiae*. *FEBS Lett.* **421**; 15–18.

Voncken F, Boxma B, Tjaden J, Akhmanova A, Huynen M, Verbeek F, Tielens AG, Haferkamp I, Neuhaus HE, Vogels G, Veenhuis M, Hackstein JH (2002) Multiple origins of hydrogenosomes: functional and phylogenetic evidence from the ADP/ATP carrier of the anaerobic chytrid *Neocallimastix* sp. *Mol. Microbiol.* **44**; 1441–1454.

Wang HC, Susko E, Roger AJ (2014) An amino acid substitution-selection model adjusts residue fitness to improve phylogenetic estimation. *Mol. Biol. Evol.* **31**; 779–792.

White DC, Frerman FE (1967) Extraction, characterization, and cellular localization of the lipids of *Staphylococcus aureus*. *J. Bacteriol.* **94;** 1854–1867.

Wriessnegger T, Gubitz G, Leitner E, Ingolic E, Cregg J, de la Cruz BJ, Daum G (2007) Lipid composition of peroxisomes from the yeast *Pichia pastoris* grown on different carbon sources. *Biochim. Biophys. Acta* **1771**; 455–461.

Yankovskaya V, Horsefield R, Törnroth S, Luna-Chavez C, Miyoshi H, Léger C, Byrne B, Cecchini G, Iwata S (2003) Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* **299**; 700–704.

Yubuki N, Pánek T, Yabuki A, Čepička I, Takishita K, Inagaki Y, Leander BS (2015) Morphological identities of two different marine stramenopile environmental sequence clades: *Bicosoeca kenaiensis* (Hilliard, 1971) and *Cantina marsupialis* (Larsen and Patterson, 1990) gen. nov., comb. nov. *J. Eukaryot. Microbiol.* **62**;

532–542.

Zachman DK, Chicco AJ, McCune SA, Murphy RC, Moore RL, Sparagna GC (2010) The role of calcium-independent phospholipase A(2) in cardiolipin remodeling in the spontaneously hypertensive heart failure rat heart. *J. Lipid Res.* **51**; 525–534.

Zinser E, Sperka-Gottlieb CD, Fasch EV, Kohlwein SD, Paltauf F, Daum G (1991) Phospholipid synthesis and lipid composition of subcellular membranes in the unicellular eukaryote *Saccharomyces cerevisiae*. *J. Bacteriol.* **173**; 2026–2034.

Table 3-1. Taxa investigated in this study.

Groups			Taxon names	Resources (GenBank accession No. / EMBL accession No. / MMETSP ID / URL)
Stramenopiles	Ochrophyta	Pelagophyce	*Pelagophyceae* sp. CCMP2097	GenBank（SRX141952）
Stramenopiles	Ochrophyta	Pelagophyce	*Aureococcus nophagefferens*	http://genome.jgi.doe.gov/Auran1/Auran1.download.html
Stramenopiles	Ochrophyta	Pelagophyce	*Pelagomonas calceolata*	MMETSP0886
Stramenopiles	Ochrophyta	Pelagophyce	*Pelagococcus subviridis*	MMETSP0882
Stramenopiles	Ochrophyta	Pelagophyce	*Aureoumbra lagunensis*	MMETSP0890
Stramenopiles	Ochrophyta	Pelagophyce	*Chrysocystis fragilis*	MMETSP1165
Stramenopiles	Ochrophyta	Dictyochophyceae	*Florenciella* sp. RCC1587	MMETSP1324
Stramenopiles	Ochrophyta	Dictyochophyceae	*Dictyocha speculum*	MMETSP1174
Stramenopiles	Ochrophyta	Dictyochophyceae	*Pseudopedinella elastica*	MMETSP1068
Stramenopiles	Ochrophyta	Dictyochophyceae	*Rhizochromulina marina*	MMETSP1173
Stramenopiles	Ochrophyta	Pinguiophyceae	*Pinguiococcus pyrenoidosus*	MMETSP1160

(Continued next page)

Table 3-1. Continued

Groups			Taxon names	Resources (GenBank accession No. / EMBL accession No. / MMETSP ID / URL)
Stramenopiles	Ochrophyta	Pinguiophyceae	*Phaeomonas parva*	MMETSP1163
Stramenopiles	Ochrophyta	Bolidophyceae	*Bolidomonas* sp. RCC1657	MMETSP1321
Stramenopiles	Ochrophyta	Bolidophyceae	*Bolidomonas pacifica*	MMETSP1319
Stramenopiles	Ochrophyta	Bolidophyceae	*Bolidomonas* sp. RCC2347	MMETSP1320
Stramenopiles	Ochrophyta	Bacillariophyceae	*Thalassiosira oceanica*	GenBank (GCA_000296195)
Stramenopiles	Ochrophyta	Bacillariophyceae	*Thalassiosira pseudonana*	http://genome.jgi.doe.gov/Thaps3/Thaps3.home.html
Stramenopiles	Ochrophyta	Bacillariophyceae	*Phaeodactylum tricornutum*	http://genome.jgi.doe.gov/Phatr2/Phatr2.home.html
Stramenopiles	Ochrophyta	Bacillariophyceae	*Fragilariopsis cylindrus*	http://genome.jgi.doe.gov/Fracy1/Fracy1.download.html
Stramenopiles	Ochrophyta	Bacillariophyceae	*Pseudonitzschia multiseries*	http://genome.jgi.doe.gov/Psemu1/Psemu1.download.html
Stramenopiles	Ochrophyta	Chrysophyceae	*Mallomonas* sp. CCMP3275	MMETSP1167
Stramenopiles	Ochrophyta	Chrysophyceae	*Ochromonas* sp. BG1	MMETSP1105

(Continued next page)

Table 3-1. Continued

Groups			Taxon names	Resources (GenBank accession No. / EMBL accession No. / MMETSP ID / URL)
Stramenopiles	Ochrophyta	Chrysophyceae	*Dinobryon* sp. UTEXLB2267	MMETSP0019
Stramenopiles	Ochrophyta	Chrysophyceae	*Spumella elongata*	MMETSP1098
Stramenopiles	Ochrophyta	Chrysophyceae	*Chyrsophyceae* sp. CCMP2298	GenBank (SRX141983)
Stramenopiles	Ochrophyta	Chrysophyceae	*Paraphysomonas bandaiensis*	MMETSP1103
Stramenopiles	Ochrophyta	Chrysophyceae	*Paraphysomonas vestita*	MMETSP1107
Stramenopiles	Ochrophyta	Eustigmatophyceae	*Nannochloropsis gaditana*	www.nannochloropsis.org
Stramenopiles	Ochrophyta	Xanthophyceae	*Vaucheria litorea*	MMETSP0945
Stramenopiles	Ochrophyta	Phaeophyceae	*Ectocarpus siliculosus*	http://bioinformatics.psb.ugent.be/genomes/view/Ectocarpus-siliculosus
Stramenopiles	Ochrophyta	Raphidophyceae	*Heterosigma akashiwo*	MMETSP0292
Stramenopiles	Ochrophyta	Raphidophyceae	*Chattonella subsalsa*	MMETSP0947
Stramenopiles	Pseudofungi	Oomycetes	*Albugo laibachii*	http://protists.ensembl.org/Albugo_laibachii/Info/Index

(Continued next page)

Table 3-1. Continued

Groups			Taxon names	Resources (GenBank accession No. / EMBL accession No. / MMETSP ID / URL)
Stramenopiles	Pseudofungi	Oomycetes	*Pythium aphanidermatum*	http://protists.ensembl.org/Pythium_aphanidermatum/Info/Index
Stramenopiles	Pseudofungi	Oomycetes	*Pythium arrhenomanes*	http://pythium.plantbiology.msu.edu/download.shtml
Stramenopiles	Pseudofungi	Oomycetes	*Phytophthora capsici*	http://genome.jgi.doe.gov/Phyca11/Phyca11.download.html
Stramenopiles	Pseudofungi	Oomycetes	*Phytophthora parasitica*	http://www.broadinstitute.org/annotation/genome/Phytophthora_parasitica/MultiHome.html
Stramenopiles	Pseudofungi	Oomycetes	*Phytophthora infestans*	http://www.broadinstitute.org/annotation/genome/phytophthora_infestans/MultiHome.html
Stramenopiles	Pseudofungi	Oomycetes	*Phytophthora sojae*	http://genome.jgi.doe.gov/Physo3/Physo3.download.html
Stramenopiles	Pseudofungi	Oomycetes	*Phytophthora cinnamomi var. cinnamomi*	http://genome.jgi.doe.gov/Phyci1/Phyci1.download.html
Stramenopiles	Pseudofungi	Oomycetes	*Phytophthora ramorum*	http://genome.jgi.doe.gov/Phyra1_1/Phyra1_1.download.html
Stramenopiles	Pseudofungi	Oomycetes	*Hyaloperonospora parasitica*	http://protists.ensembl.org/Hyaloperonospora_arabidopsidis/Info/Index
Stramenopiles	Pseudofungi	Oomycetes	*Pythium vexans*	http://pythium.plantbiology.msu.edu/download.shtml
Stramenopiles	Pseudofungi	Oomycetes	*Pythium ultimum var. sporangiiferum*	http://pythium.plantbiology.msu.edu/download.shtml

(Continued next page)

Table 3-1. Continued

Groups			Taxon names	Resources (GenBank accession No. / EMBL accession No. / MMETSP ID / URL)
Stramenopiles	Pseudofungi	Oomycetes	*Pythium ultimum*	http://www.broadinstitute.org/annotation/genome/Saprolegnia_parasitica/GenomeDescriptions.html#PytUlt1
Stramenopiles	Pseudofungi	Oomycetes	*Pythium irregulare*	http://pythium.plantbiology.msu.edu/download.shtml
Stramenopiles	Pseudofungi	Oomycetes	*Pythium iwayamai*	http://pythium.plantbiology.msu.edu/download.shtml
Stramenopiles	Pseudofungi	Oomycetes	*Saprolegnia declina*	http://www.broadinstitute.org/annotation/genome/Saprolegnia_parasitica/GenomeDescriptions.html#Sap_diclina_VS20_V1
Stramenopiles	Pseudofungi	Oomycetes	*Saprolegnia parasitica*	http://www.broadinstitute.org/annotation/genome/Saprolegnia_parasitica/GenomeDescriptions.html#Sap_parasitica_CBS_223_65_V2
Stramenopiles	Pseudofungi	Bigyromonadea	*Developayella elegans*	GenBank (DRX044599)
Stramenopiles	Bigyra	Labyrinthulomycetes	*Aurantiochytrium limacinum*	http://genome.jgi.doe.gov/Aplke1/Aplke1.download.html
Stramenopiles	Bigyra	Labyrinthulomycetes	*Aurantiochytrium limacinum*	http://genome.jgi.doe.gov/Aurli1/Aurli1.download.html
Stramenopiles	Bigyra	Labyrinthulomycetes	*Thraustochytrium* sp. LLF1b	MMETSP0198
Stramenopiles	Bigyra	Labyrinthulomycetes	*Schizochytrium aggregatum*	http://genome.jgi.doe.gov/Schag1/Schag1.download.html

(Continued next page)

Table 3-1. Continued

Groups			Taxon names	Resources (GenBank accession No. / EMBL accession No. / MMETSP ID / URL)
Stramenopiles	Bigyra		*Blastocystis hominis*	http://www.genoscope.cns.fr/externe/GenomeBrowser/Blastocystis/
Stramenopiles	Bigyra	Placidida	*Wobblia lunata*	GenBank (DRX044598)
Stramenopiles	Bigyra	Bicosoecida	*Cafeteria* sp. Caron Lab Isolate	MMETSP1104
Stramenopiles	Bigyra	Bicosoecida	*Cafeteria roenbergensis*	GenBank (DRX044597)
Stramenopiles	Bigyra		*Cantina marsupialis*	GenBank (DRX027417)
Rhizaria			*Bigelowiella natans*	http://genome.jgi.doe.gov/Bigna1/Bigna1.download.html
Rhizaria			*Paulinella chromatophora*	GenBank (SRX015452)
Rhizaria			*Guttulinopsis vulgaris*	GenBank (PRJNA82891)
Alveolata			*Tetrahymena thermophila*	http://www.GenBank.nlm.nih.gov/genome/?term=Tetrahymena+thermophila
Alveolata			*Toxoplasma gondii*	GenBank (PRJNA82891)
Alveolata			*Perkinsus marinus*	GenBank

(Continued next page)

Table 3-1. Continued

Groups	Taxon names	Resources (GenBank accession No. / EMBL accession No. / MMETSP ID / URL)
Alveolata	*Symbiodinium* sp. KB8	http://medinalab.org/zoox/
Alveolata	*Oxyrrhis marina*	MMETSP0044
Cryptobiontes	*Roombia truncata*	GenBank (PRJNA73793)
Cryptobiontes	*Goniomonas avonlea*	MMETSP0114
Cryptobiontes	*Guillardia theta*	http://genome.jgi.doe.gov/Guith1/Guith1.home.html
Glaucophyta	*Cyanophora paradoxa*	http://cyanophora.rutgers.edu/cyanophora/home.php
Glaucophyta	*Gloeochaete witrockiana*	MMETSP0308
Rhodophyta	*Galdieria sulphuraria*	GenBank
Rhodophyta	*Rhodella maculata*	MMETSP0313
Rhodophyta	*Compsopogon coeruleus*	MMETSP0312
Rhodophyta	*Chondrus crispus*	EMBL (CAKH01000001–CAKH01003241)

(Continued next page)

Table 3-1. Continued

Groups	Taxon names	Resources (GenBank accession No. / EMBL accession No. / MMETSP ID / URL)
Viridiplantae	*Arabidopsis thaliana*	GenBank
Viridiplantae	*Chlamydomonas reinhardtii*	GenBank
Viridiplantae	*Micromonas pusilla*	GenBank
Viridiplantae	*Physcomitrella patens*	GenBank
Haptophyta	*Pavlovales* sp. CMP2436	GenBank (SRX127438)
Haptophyta	*Emiliania huxleyi*	http://genome.jgi.doe.gov/Emihu1/Emihu1.home.html
Haptophyta	*Isochrysis* sp. CCMP1324	MMETSP1129
Haptophyta	*Pleurochrysis carterae*	MMETSP1136

Table 3-2. Presence/absence of genes encoding proteins associated with cardiolipin biosynthesis and maturation.

			Protein names						
			CLS						
Groups		Taxon names	_cap	_pld	CLD	iPLA2_β	iPLA2_γ	TAZ	ALCAT1
Ochrophyta	Pelagophyce	*Pelagophyceae* sp. CCMP2097	-	-	-	-	-	-	-
Ochrophyta	Pelagophyce	*Aureococcus nophagefferens*	-	-	-	-	+	-	+
Ochrophyta	Pelagophyce	*Pelagomonas calceolata*	-	-	-	-	+	-	-
Ochrophyta	Pelagophyce	*Pelagococcus subviridis*	-	-	+	-	+	-	-
Ochrophyta	Pelagophyce	*Aureoumbra lagunensis*	+	-	-	+	+	+	-
Ochrophyta	Pelagophyce	*Chrysocystis fragilis*	-	-	-	-	+	-	+
Ochrophyta	Dictyochophyceae	*Florenciella* sp. RCC1587	-	-	-	-	+	-	+
Ochrophyta	Dictyochophyceae	*Dictyocha speculum*	+	-	+	-	+	-	+
Ochrophyta	Dictyochophyceae	*Pseudopedinella elastica*	+	-	-	-	+	-	-
Ochrophyta	Dictyochophyceae	*Rhizochromulina marina*	-	-	-	+	+	+	-

(Continued next page)

Table3-2. Continued

Groups		Taxon names	Protein names						
			CLS						
			_cap	_pld	CLD	iPLA2_β	iPLA2_γ	TAZ	ALCAT1
Ochrophyta	Pinguiophyceae	*Pinguiococcus pyrenoidosus*	-	-	+	-	+	-	-
Ochrophyta	Pinguiophyceae	*Phaeomonas parva*	+	-	-	-	+	-	-
Ochrophyta	Bolidophyceae	*Bolidomonas* sp. RCC1657	-	-	+	-	-	-	+
Ochrophyta	Bolidophyceae	*Bolidomonas pacifica*	-	-	+	-	-	-	+
Ochrophyta	Bolidophyceae	*Bolidomonas* sp. RCC2347	-	-	+	-	+	+	+
Ochrophyta	Bacillariophyceae	*Thalassiosira oceanica*	+	-	+	-	+	+	+
Ochrophyta	Bacillariophyceae	*Thalassiosira pseudonana*	+	-	+	-	+	+	+
Ochrophyta	Bacillariophyceae	*Phaeodactylum tricornutum*	-	-	+	-	+	-	-
Ochrophyta	Bacillariophyceae	*Fragilariopsis cylindrus*	+	-	-	-	+	-	+
Ochrophyta	Bacillariophyceae	*Pseudonitzschia multiseries*	+	-	-	-	+	-	+

(Continued next page)

Table3-2. Continued

Groups		Taxon names	Protein names						
			CLS						
			_cap	_pld	CLD	iPLA2_β	iPLA2_γ	TAZ	ALCAT1
Ochrophyta	Chrysophyceae	*Mallomonas* sp. CCMP3275	+	-	+	-	+	-	-
Ochrophyta	Chrysophyceae	*Ochromonas* sp. BG1	-	-	+	-	+	-	-
Ochrophyta	Chrysophyceae	*Dinobryon* sp. UTEXLB2267	+	-	-	-	+	+	+
Ochrophyta	Chrysophyceae	*Spumella elongata*	+	-	-	+	-	+	-
Ochrophyta	Chrysophyceae	*Chyrsophyceae* sp. CCMP2298	-	-	-	-	-	-	-
Ochrophyta	Chrysophyceae	*Paraphysomonas bandaiensis*	-	-	-	-	-	+	+
Ochrophyta	Chrysophyceae	*Paraphysomonas vestita*	-	-	-	-	-	-	-
Ochrophyta	Eustigmatophyceae	*Nannochloropsis gaditana*	+	-	-	-	+	-	+
Ochrophyta	Xanthophyceae	*Vaucheria litorea*	-	-	-	-	+	+	+
Ochrophyta	Phaeophyceae	*Ectocarpus siliculosus*	+	-	+	-	+	+	+

(Continued next page)

Table3-2. Continued

Groups		Taxon names	Protein names						
			CLS						
			_cap	_pld	CLD	iPLA2_β	iPLA2_γ	TAZ	ALCAT1
Ochrophyta	Raphidophyceae	*Heterosigma akashiwo*	+	-	-	-	+	+	+
Ochrophyta	Raphidophyceae	*Chattonella subsalsa*	+	-	-	-	+	+	+
Pseudofungi	Oomycetes	*Albugo laibachii*	+	-	-	-	-	+	+
Pseudofungi	Oomycetes	*Pythium aphanidermatum*	+	-	+	-	-	+	+
Pseudofungi	Oomycetes	*Pythium arrhenomanes*	+	-	+	-	-	+	+
Pseudofungi	Oomycetes	*Phytophthora capsici*	+	-	+	-	-	+	+
Pseudofungi	Oomycetes	*Phytophthora parasitica*	+	-	-	-	-	+	+
Pseudofungi	Oomycetes	*Phytophthora infestans*	+	-	+	-	-	+	+
Pseudofungi	Oomycetes	*Phytophthora sojae*	+	-	+	-	-	+	+
Pseudofungi	Oomycetes	*Phytophthora cinnamomi var. cinnamomi*	+	-	+	-	-	+	+

(Continued next page)

Table3-2. Continued

Groups		Taxon names	Protein names						
			CLS						
			_cap	_pld	CLD	iPLA2_β	iPLA2_γ	TAZ	ALCAT1
Pseudofungi	Oomycetes	*Phytophthora ramorum*	+	-	+	-	-	+	+
Pseudofungi	Oomycetes	*Hyaloperonospora parasitica*	-	-	+	-	-	+	+
Pseudofungi	Oomycetes	*Pythium vexans*	+	-	+	-	-	+	+
Pseudofungi	Oomycetes	*Pythium ultimum var. sporangiiferum*	-	-	-	-	-	-	-
Pseudofungi	Oomycetes	*Pythium ultimum*	+	-	+	-	-	-	+
Pseudofungi	Oomycetes	*Pythium irregulare*	+	-	+	-	-	+	+
Pseudofungi	Oomycetes	*Pythium iwayamai*	+	-	-	-	-	+	+
Pseudofungi	Oomycetes	*Saprolegnia declina*	+	-	-	-	-	+	+
Pseudofungi	Oomycetes	*Saprolegnia parasitica*	+	-	-	-	-	+	+
Pseudofungi	Bigyromonadea	*Developayella elegans*	+	-	-	-	-	+	+

(Continued next page)

Table3-2. Continued

Groups		Taxon names	Protein names						
			CLS						
			_cap	_pld	CLD	iPLA2_β	iPLA2_γ	TAZ	ALCAT1
Bigyra	Labyrinthulomyc	*Aplanochytrium kerguelense*	+	-	+	-	-	-	+
Bigyra	Labyrinthulomyc	*Aurantiochytrium limacinum*	+	-	+	-	-	-	+
Bigyra	Labyrinthulomyc	*Thraustochytrium* sp. LLF1b	+	-	-	-	+	-	+
Bigyra	Labyrinthulomyc	*Schizochytrium aggregatum cinnamomi*	+	-	+	-	-	-	+
Bigyra		*Blastocystis hominis*	-	-	-	-	-	+	+
Bigyra	Placidida	*Wobblia lunata*	+	+	-	-	-	-	+
Bigyra	Bicosoecida	*Cafeteria* sp. Caron Lab Isolate	+	+	-	-	-	+	+
Bigyra	Bicosoecida	*Cafeteria roenbergensis*	+	-	-	-	+	+	+
Bigyra		*Cantina marsupialis*	-	+	-	-	-	+	+

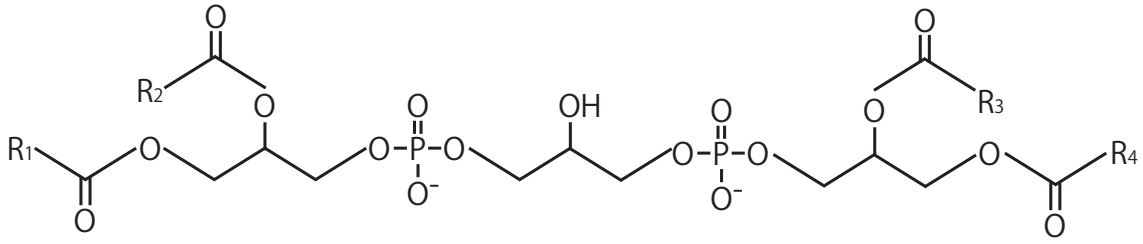

Fig. 3-1. Structure of cardiolipin. Two phosphatidyl residues are linked by a central glycerol brige.

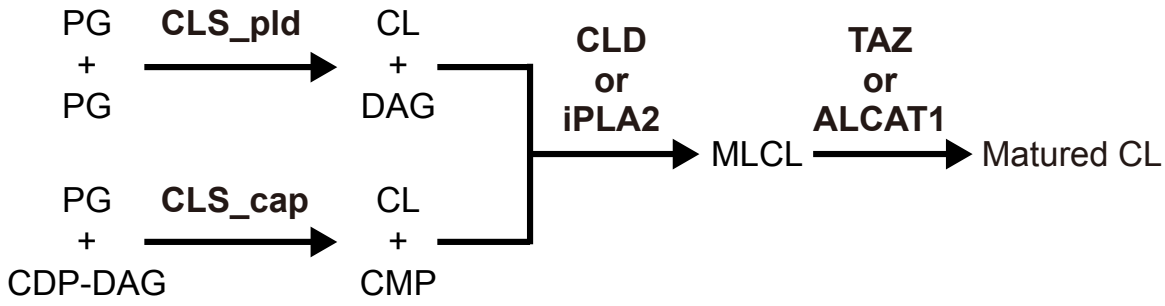

Fig. 3-2. Two types of biosynthesis pathways of cardiolipin. PG, phosphatidylglycerol; CDP-DAG, cytidinediphosphate-diacylglycerol; CMP, cytidine monophosphate; MLCL, monolysocardiolipin.

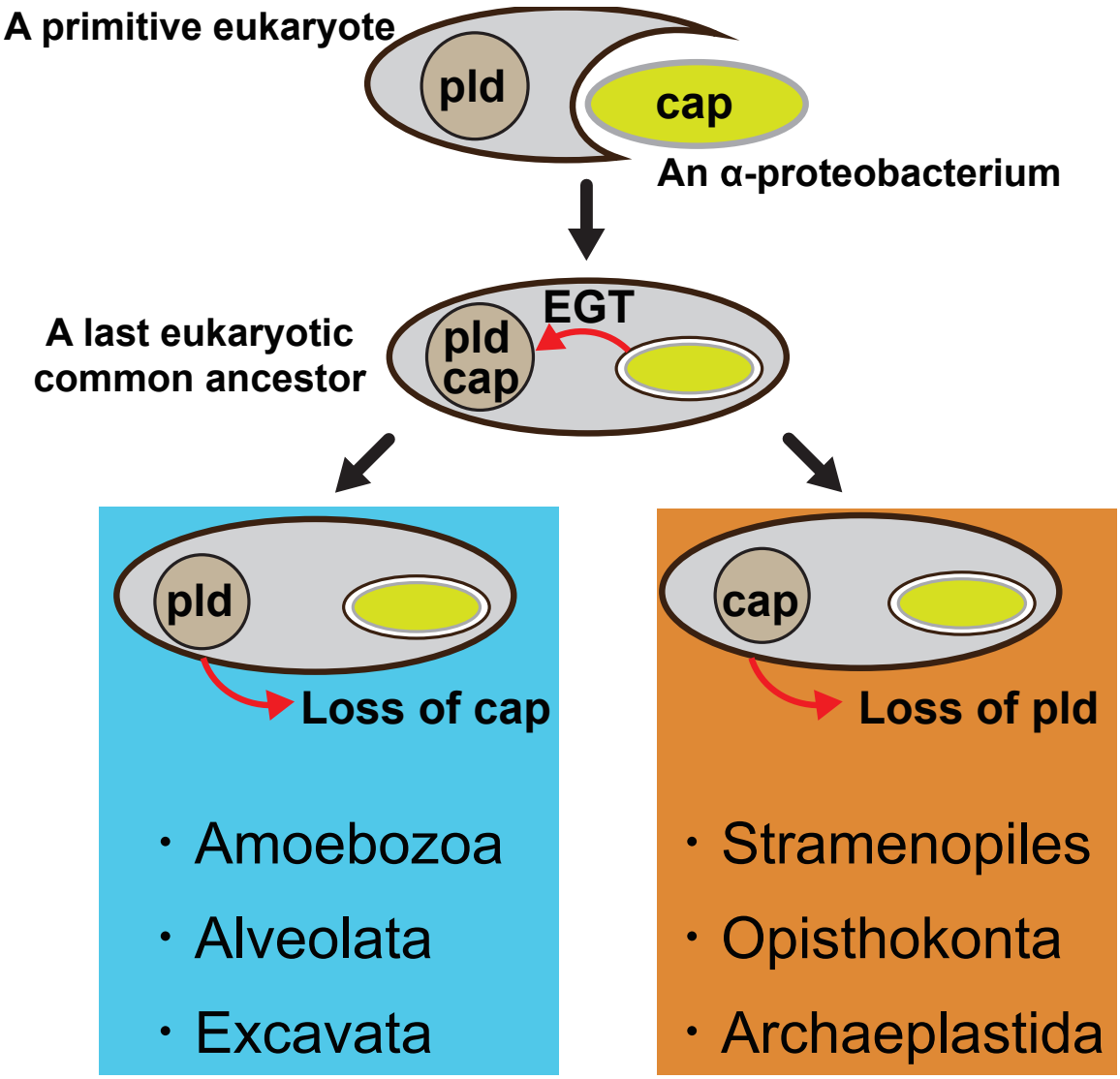

Fig. 3-3. An evolutionary scenario of two types of cardiolipin synthases in eukaryotes proposed by Tian et al. (2012)

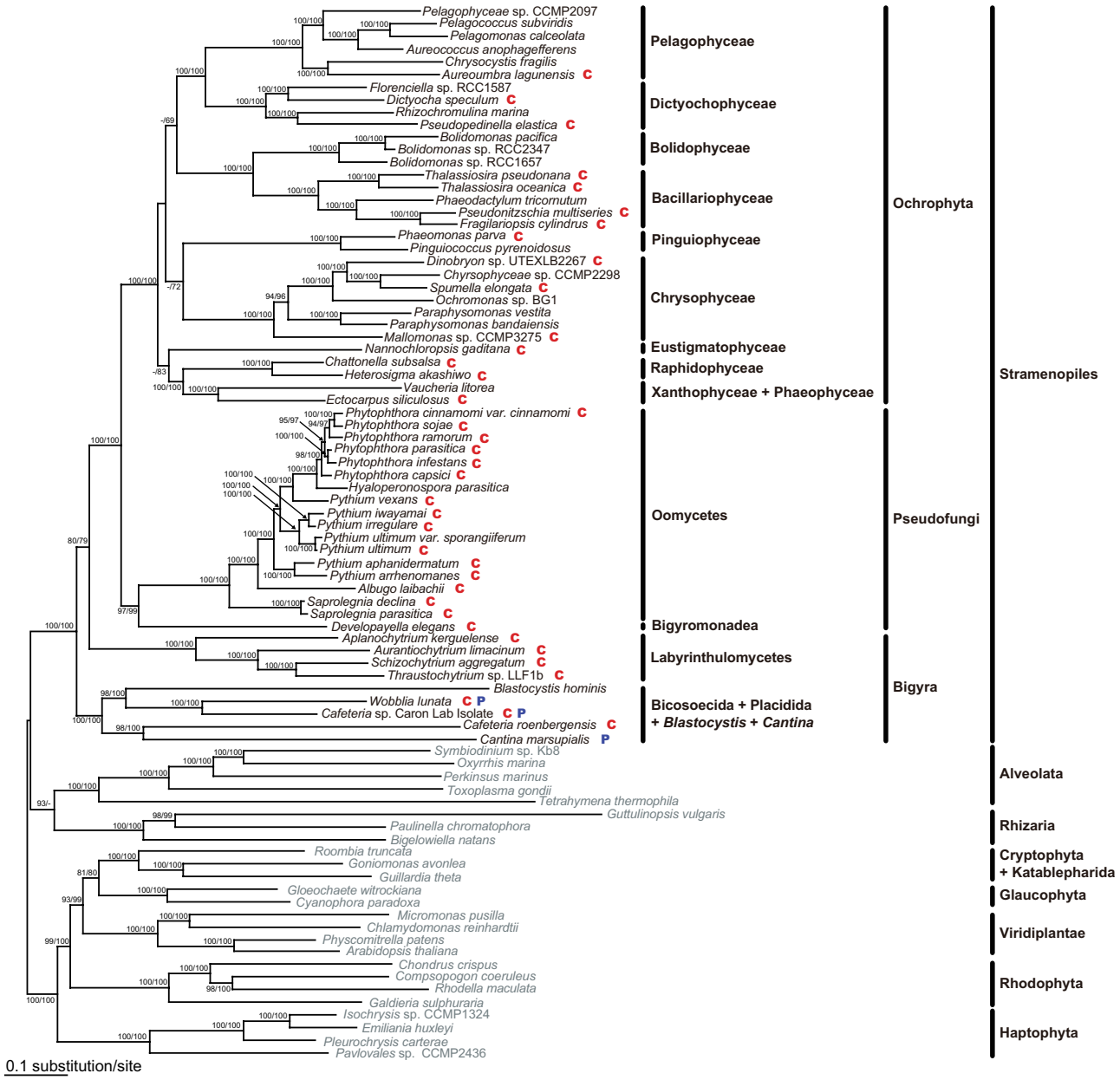

Fig. 3-4. Maximum-likelihood (ML) phylogeny based on the 245-protein dataset from 84 eukaryotic taxa including a broad range of 59 stramenopile taxa inferred by RAxML under the LG + Γ4 model. ML bootstrap probabilities of RAxML (left) and IQ-TREE (right) are shown for nodes with support over 60%. Stramenopiles marked with a red "C" and a blue "P" represent taxa having CLS_cap and CLS_pld, respectively.

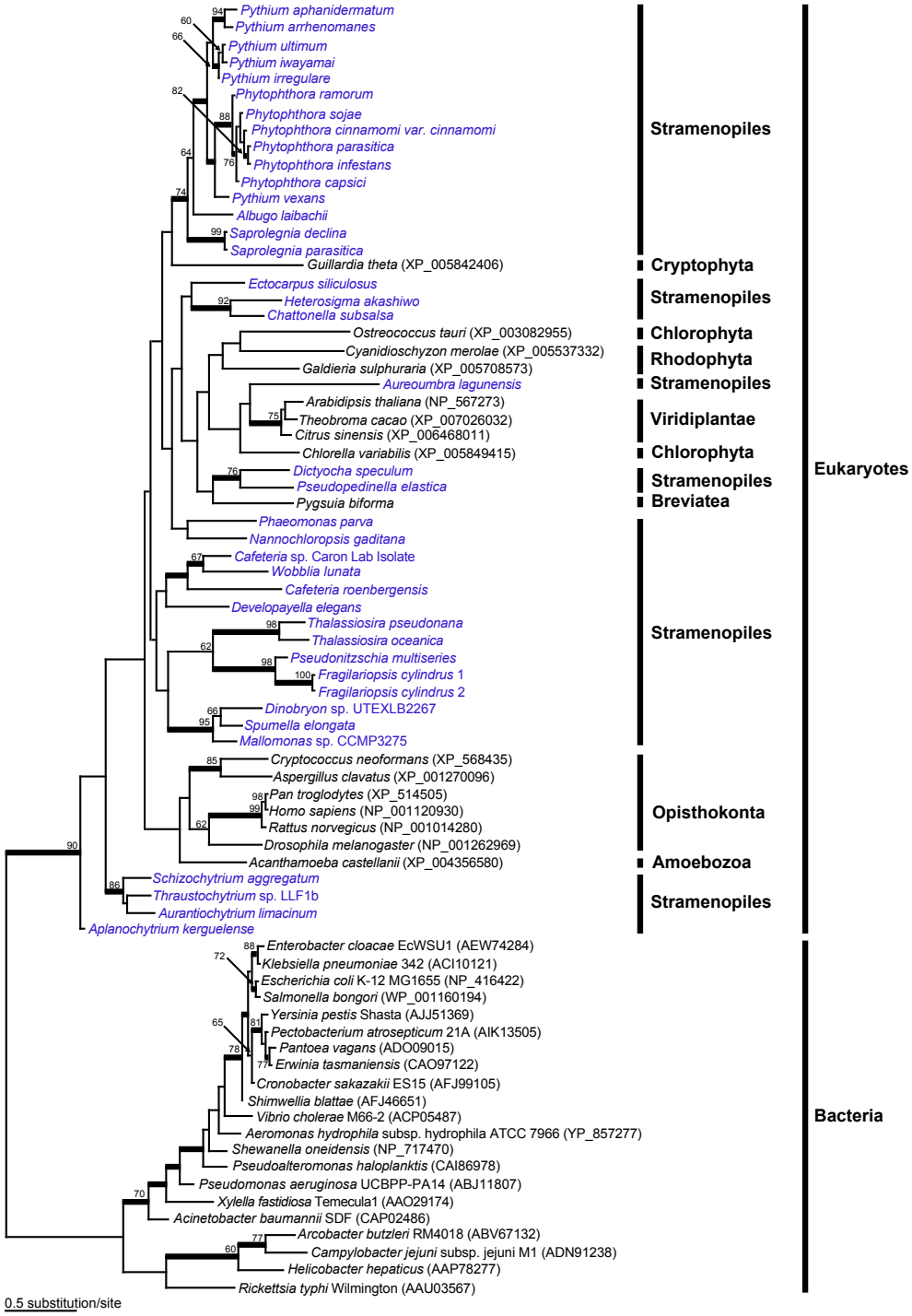

Fig. 3-5. Maximum-likelihood (ML) phylogeny of the cardiolipin synthase CLS_cap from a broad range of organisms inferred by RAxML under the LG + Γ4 model. ML bootstrap probabilities are shown for nodes with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Stramenopiles are shown in blue typeface.

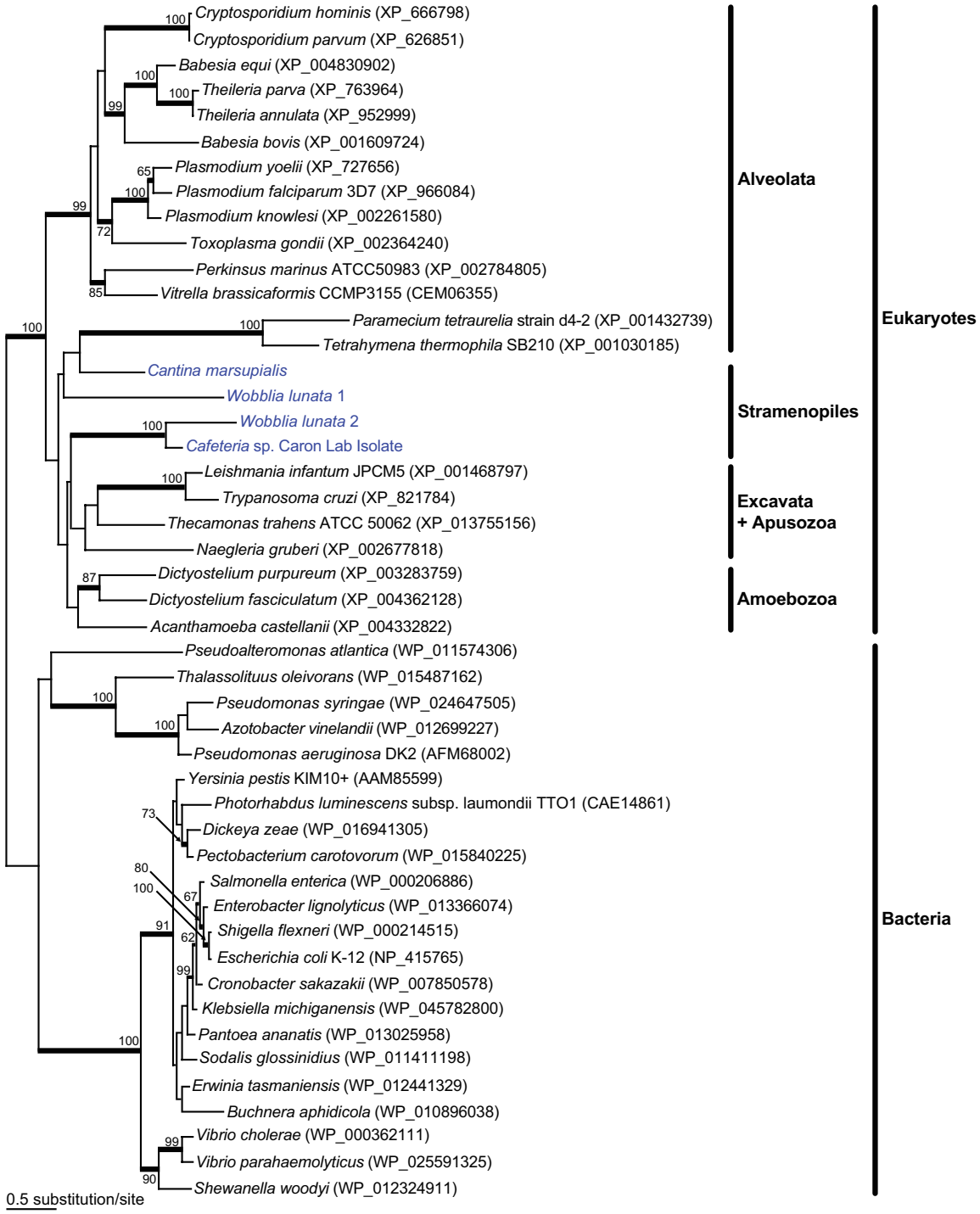

Fig. 3-6. Maximum-likelihood (ML) phylogeny of the cardiolipin synthase CLS_pld from a broad range of organisms inferred by RAxML under the LG + Γ4 model. ML bootstrap probabilities are shown for nodes with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Stramenopiles are shown in blue typeface.

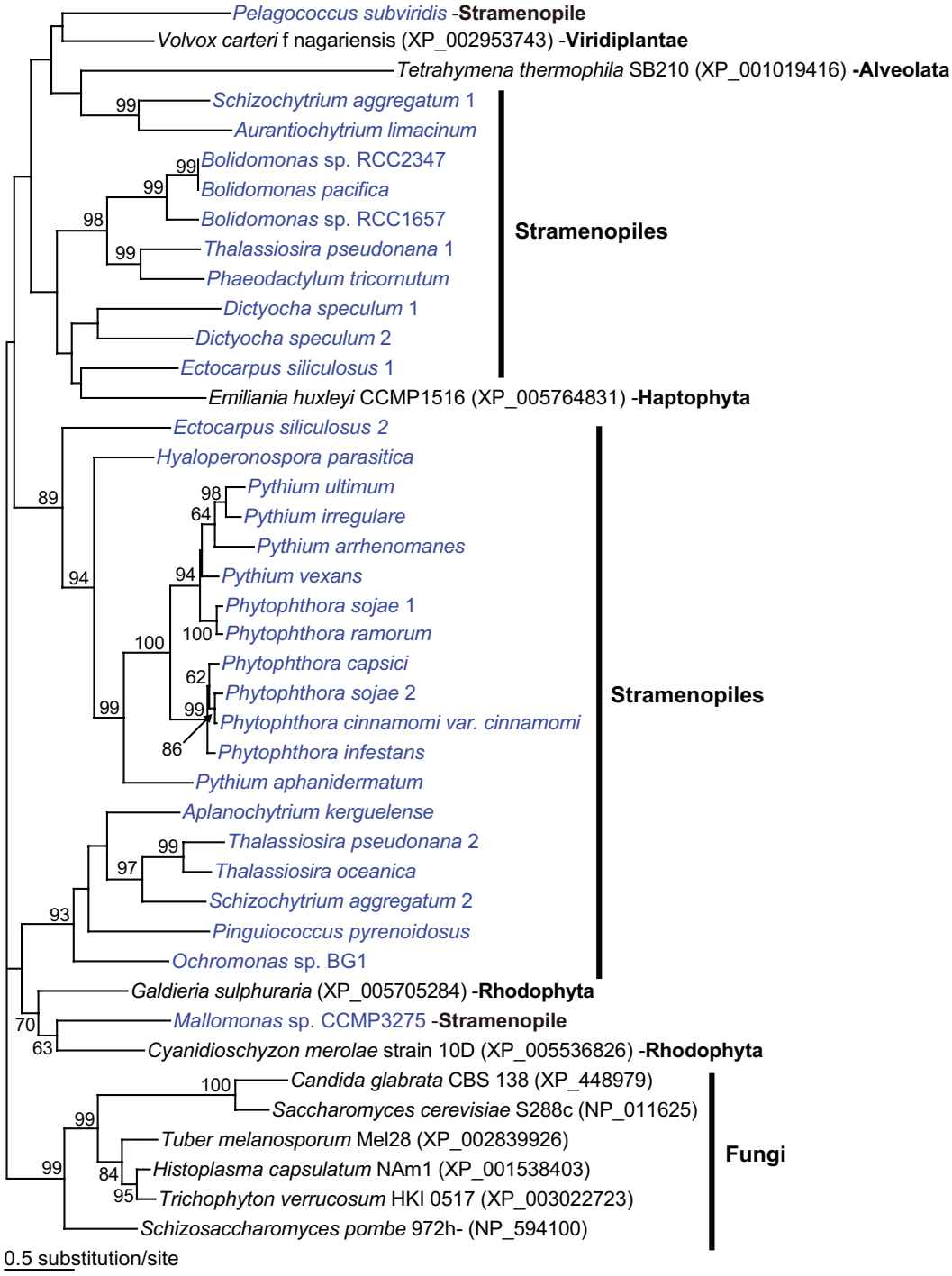

Fig. 3-7. Unrooted maximum-likelihood (ML) phylogeny of cardiolipin-specific phospholipase from a broad range of eukaryotes inferred by RAxML under the LG + Γ4 model. ML bootstrap probabilities are shown for nodes with support over 60%. Stramenopiles are shown in blue typeface.

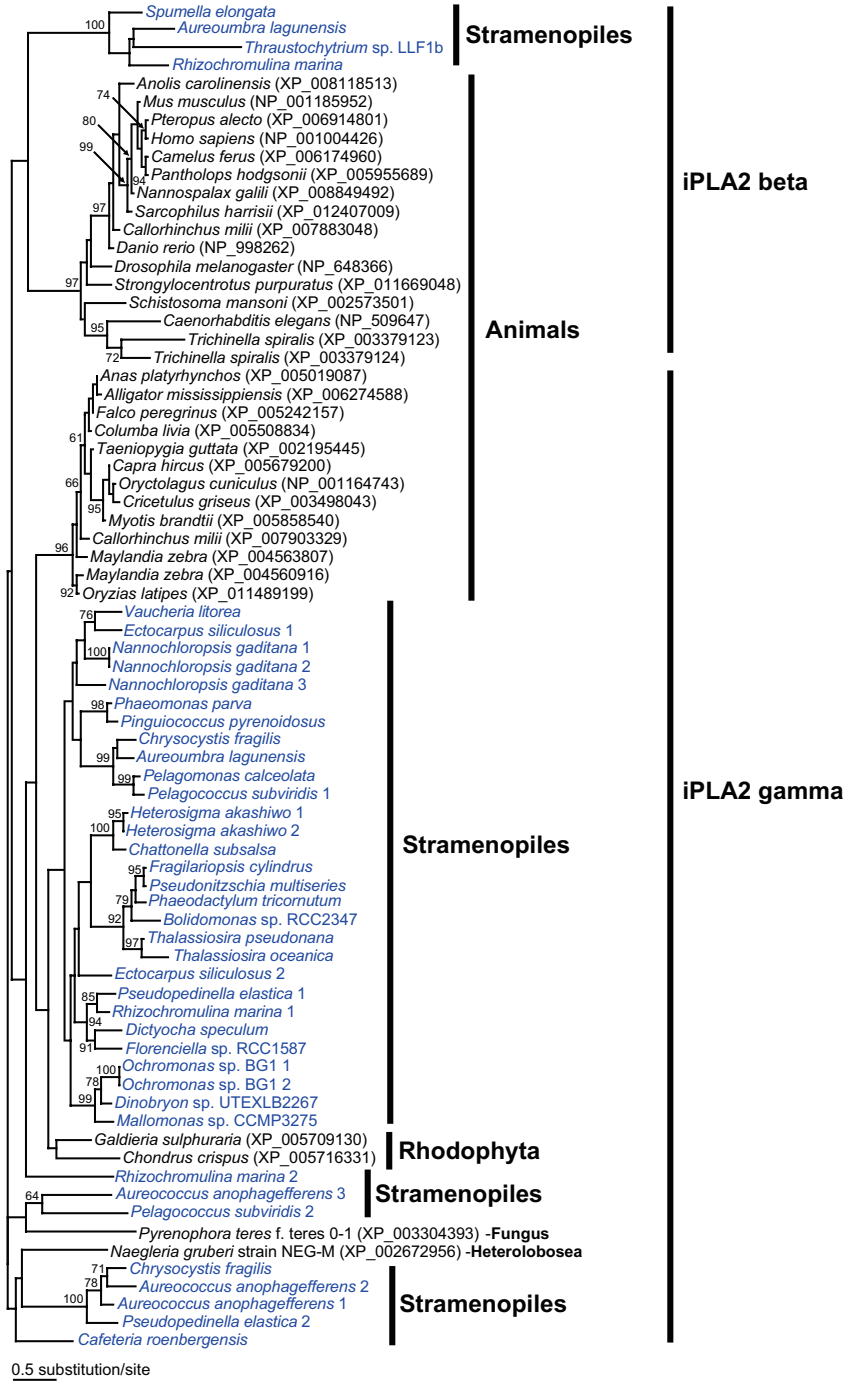

Fig. 3-8. Unrooted maximum-likelihood (ML) phylogeny of calcium-independent phospholipase A2 beta/gamma from a broad range of eukaryotes inferred by RAxML under the LG + Γ4 model. ML bootstrap probabilities are shown for nodes with support over 60%. Stramenopiles are shown in blue typeface.

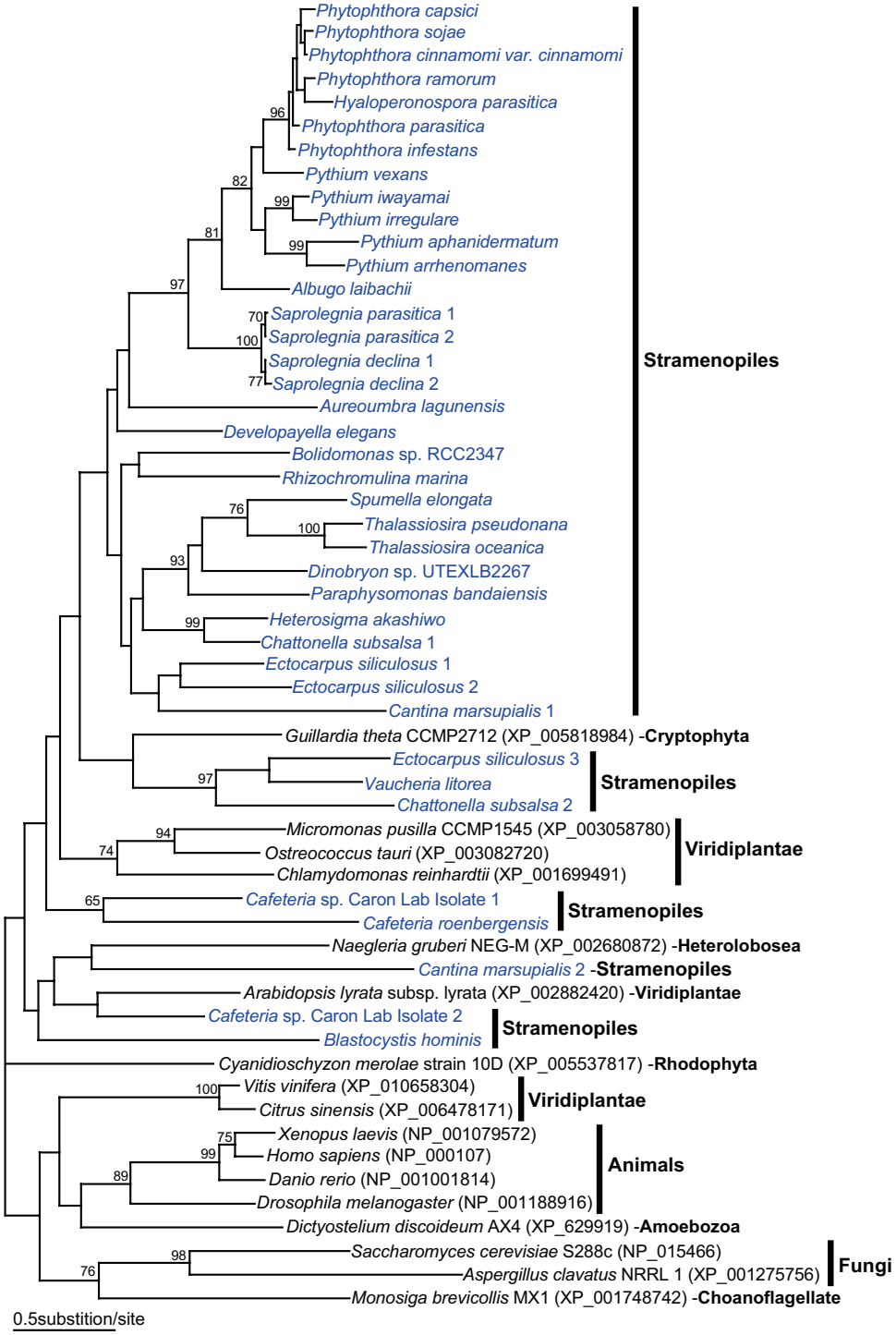

Fig. 3-9. Unrooted maximum-likelihood (ML) phylogeny of CoA-independent tafazzin from a broad range of eukaryotes inferred by RAxML under the LG + Γ4 model. ML bootstrap probabilities are shown for nodes with support over 60%. Stramenopiles are shown in blue typeface.

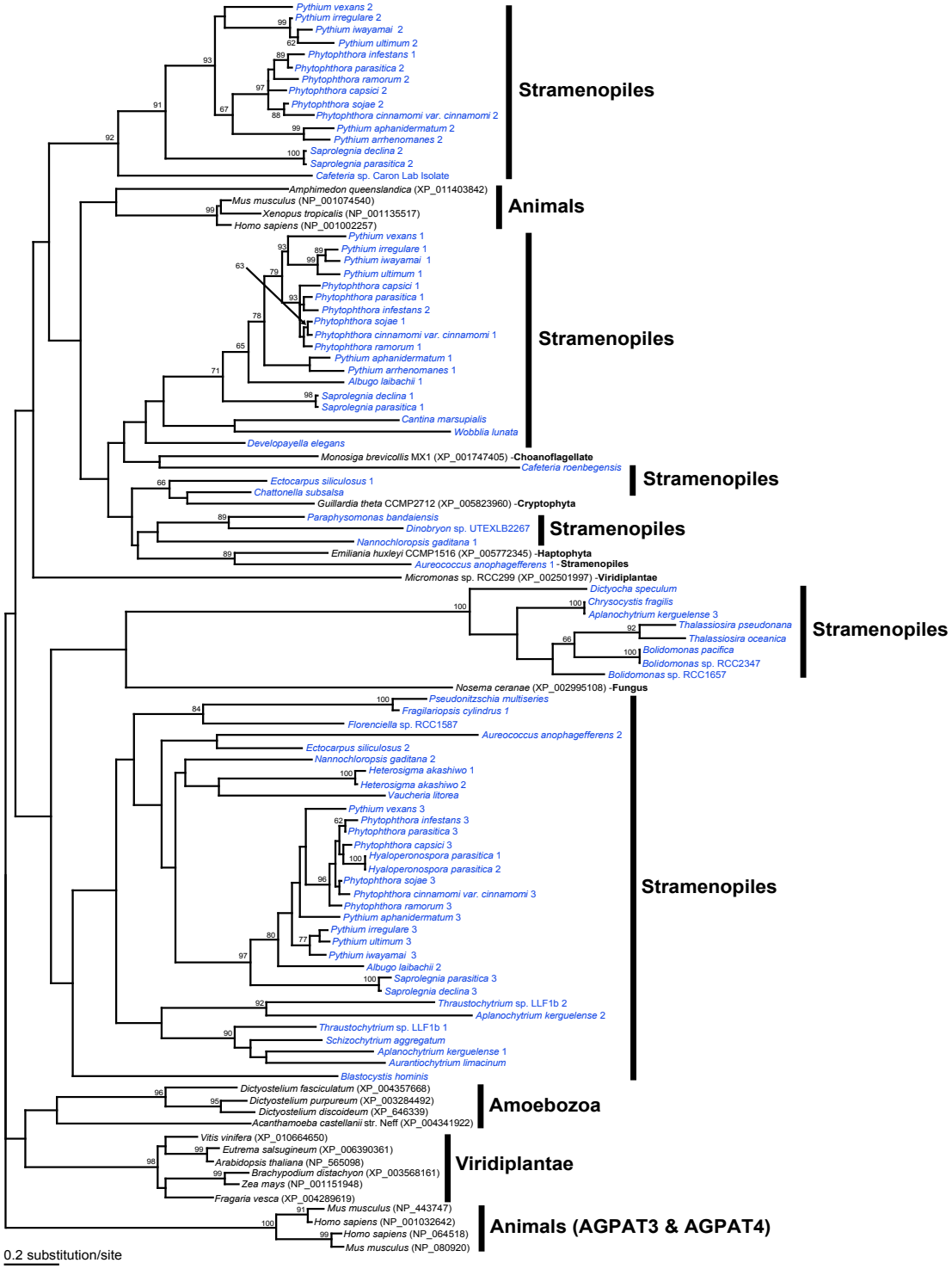

Fig. 3-10. Unrooted maximum-likelihood (ML) phylogeny of acylCoA:lysocardiolipin acyltransferase 1 from a broad range of eukaryotes inferred by RAxML under the LG + Γ4 model. ML bootstrap probabilities are shown for nodes with support over 60%. Stramenopiles are shown in blue typeface.

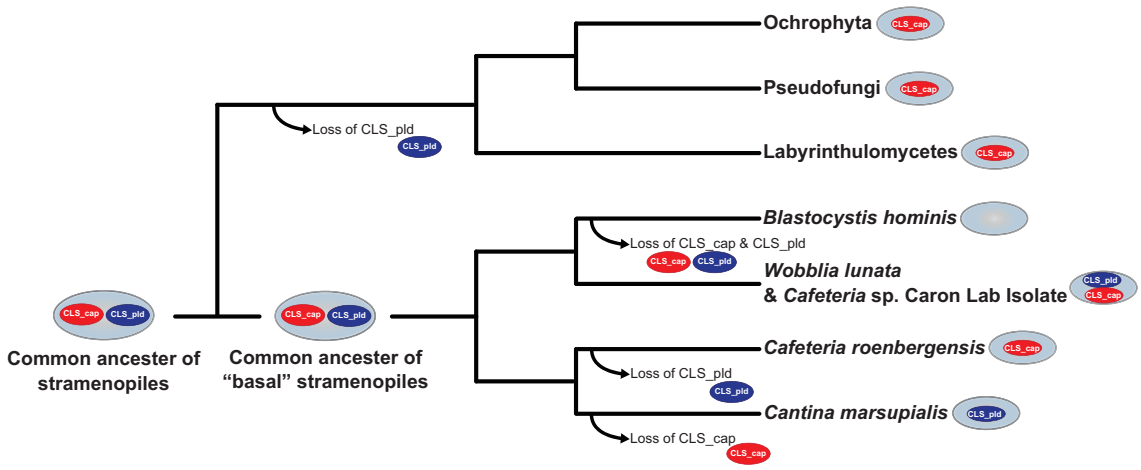

Fig. 3-11. Hypothetical evolutionary scenarios of CLS_cap and CLS_pld in the lineage of stramenopiles.

Chapter 4

General discussion

In this doctoral dissertation, I focused on the two following subjects to elucidate the molecular evolution associated with mitochondria in the major eukaryotic lineage stramenopiles, and their general conclusions are presented here.

1) Metabolic capacity of mitochondrion related organelles in the free-living anaerobic stramenopile *Cantina marsupialis*

Müller et al. (2012) proposed that the mitochondrial family of organelles (i.e., a canonical mitochondrion and its derived organelles) can be divided into five classes in the context of function. Among these classes, a hydrogen-producing (class 3) mitochondrion is defined as an organelle with a proton-pumping electron transport chain (ETC) and iron-only hydrogenase ([Fe]-Hyd). Mitochondrion-related organelles (MROs) in *Cantina marsupialis*, together with those of the anaerobic protists *Mastigamoeba* and *Pygsuia* (Gill et al. 2007; Nývltová et al. 2015; Stairs et al. 2014), possess [Fe]-Hyd and therefore could produce molecular hydrogen. However, there is no evidence indicating the proton-pumping ETC exists in MROs in *Cantina*, *Mastigamoeba*, and *Pygsuia*, although these organelles seem to have some components of the membrane-associated electron transport system including the ETC complex II. Therefore, these MROs cannot be classified into any classes according to the criteria proposed by Müller et al. (2012), although *Mastigamoeba* is classified into class 4 in this review. If MROs in *Cantina*, *Mastigamoeba*, and *Pygsuia* are placed in class 3, the definition of this class may have to be slightly revised by omitting the term “proton-pumping”. In the present study, several unique biochemical characteristics were suggested in MROs in *Cantina*. Among these characteristics, it should be especially noted that *Cantina* unlikely possesses the low redox potential electron carrier, rhodoquinone (RQ), and fumarate hydratase and that the

incomplete oxidative TCA cycle may operate in its organelles. These findings allow us to reconsider the evolution associated with mitochondria that could have occurred during the course of adaptation to hypoxic/anoxic environments (Fig. 4-1). According to the evolutionary scenario for degenerate mitochondria suggested by de Graaf et al. (2011), the successive evolutionary steps from an aerobic (class 1) mitochondrion to a class 3 mitochondrion are as follows: 1) the acquisition of the low redox potential electron carrier, rhodoquinone (RQ), and the employment of the RQ-associated reverse TCA cycle (fumarate respiration); 2) the acquisition of [Fe]-Hyd; and 3) the loss of complexes III, IV, and F1Fo ATPase. However, the order of the two former events could be reversed or the first step could be skipped in the process of evolution, because it is likely that *Cantina* MROs have [Fe]-Hyd, but lack RQ. Based on my present findings, the evolutionary path leading to degenerate organelles in association with adaptation to oxygen-depleted environments may not necessarily be uniform in different organisms.

2) Complex evolution of two types of cardiolipin synthase in the eukaryotic lineage stramenopiles

Tian et al. (2012) pointed out that the respective major eukaryotic groups have one of the two types of cardiolipin synthase (CLS), CLS_cap or CLS_pld. They hypothesized that the last eukaryotic common ancestor (LECA) possessed both CLS_cap and CLS_pld and several differential losses of these two enzymes occurred at an early stage of eukaryotic evolution. For example, they argued that CLS_cap was exclusively found in the eukaryotic group stramenopiles, suggesting that CLS_pld was lost prior to the emergence of the common ancestor of all stramenopiles. However, in my study, one

species with only CLS_pld (*Cantina marsupialis*) and two species with both CLS_cap and CLS_pld (*Cafeteria* sp. Caron Lab Isolate and *Wobblia lunata*) as well as many species with only CLS_cap were found among stramenopiles, and all three CLS_pld-bearing species were positioned in the most basal stramenopile lineage. Given these findings, it is highly likely that the common ancestor of all stramenopiles harbored both CLS_cap and CLS_pld and that these two enzymes were differentially lost in this group. My present findings could also simultaneously support the hypothesis that both CLS_cap and CLS_pld existed in the LECA as originally proposed by Tian et al. (2012). The evolution of CLSs is not as simple even in a single eukaryotic clade, as suggested by the results of this study (Fig. 4-2). Therefore, the evolutionary history of the enzymes in question in the eukaryotic domain as a whole is also possibly more complicated than previously thought, and major eukaryotic groups other than stramenopiles may have to be individually examined with comprehensive taxon sampling. In that case, a robust organismal phylogeny would be the key to trace the evolution of CLS_cap/CLS_pld precisely in each eukaryotic group.

Stramenopiles are extremely diverse in the context of ecology, physiology, and phylogeny, being one of the most flourishing eukaryotic groups on earth. Such diversity may be at least partially caused by the flexibility of their mitochondria. MROs in *Cantina* could be an example to understand such flexibility. To infer the origin and early evolution of eukaryotes, the eukaryotic domain as a whole has to be considered. In addition, to consider this domain as a whole, the information on each major eukaryotic clade has to be individually accumulated. In that sense, my study regarding the CLS evolution in the stramenopile lineage will certainly contribute to figure out the origin and early evolution of

eukaryotes.

References

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

de Graaf RM, Ricard G, van Alen TA, Duarte I, Dutilh BE, Burgtorf C, Kuiper JW, van der Staay GW, Tielens AG, Huynen MA, Hackstein JH (2011) The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. *Mol. Biol. Evol.* **28**:2379–2391

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

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

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

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

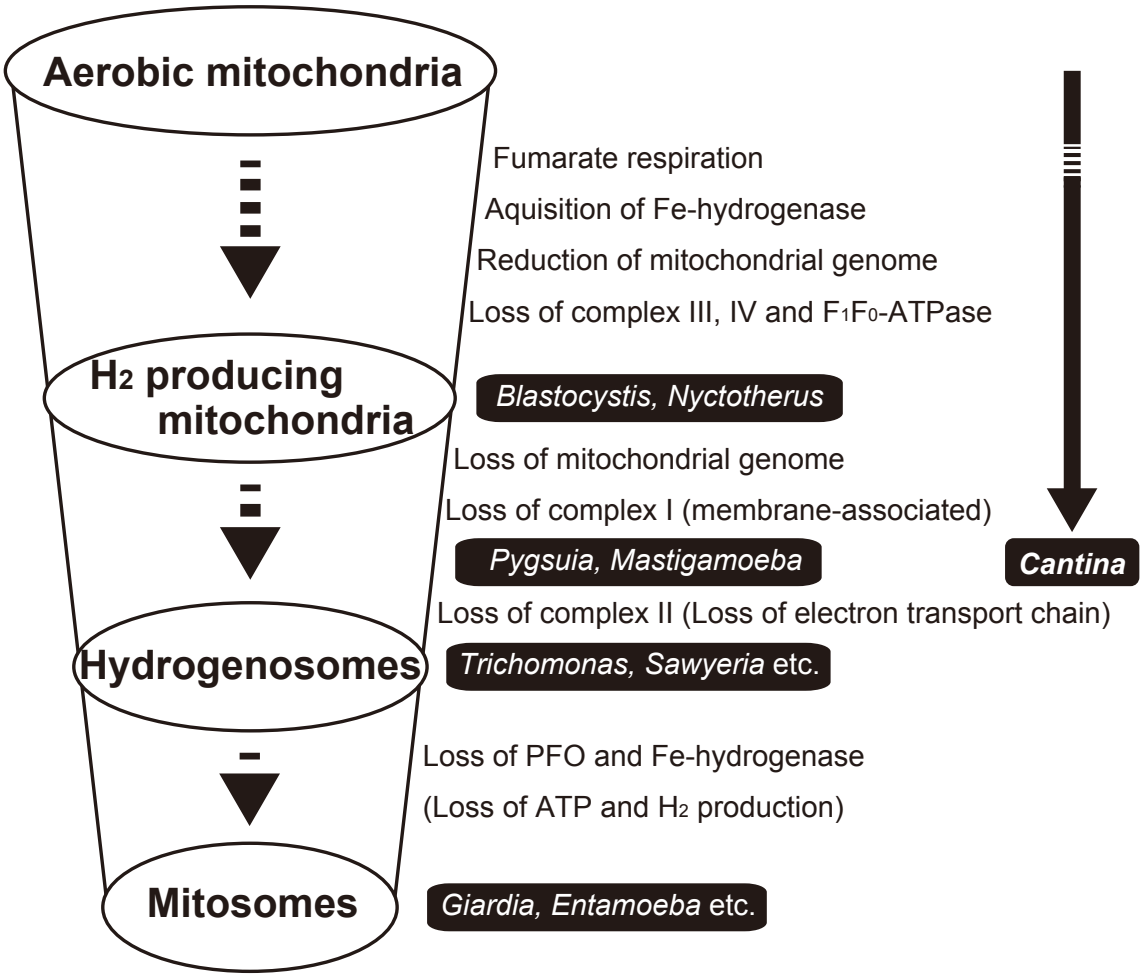

Fig. 4-1. The evolutionary scenario suggested in this study.

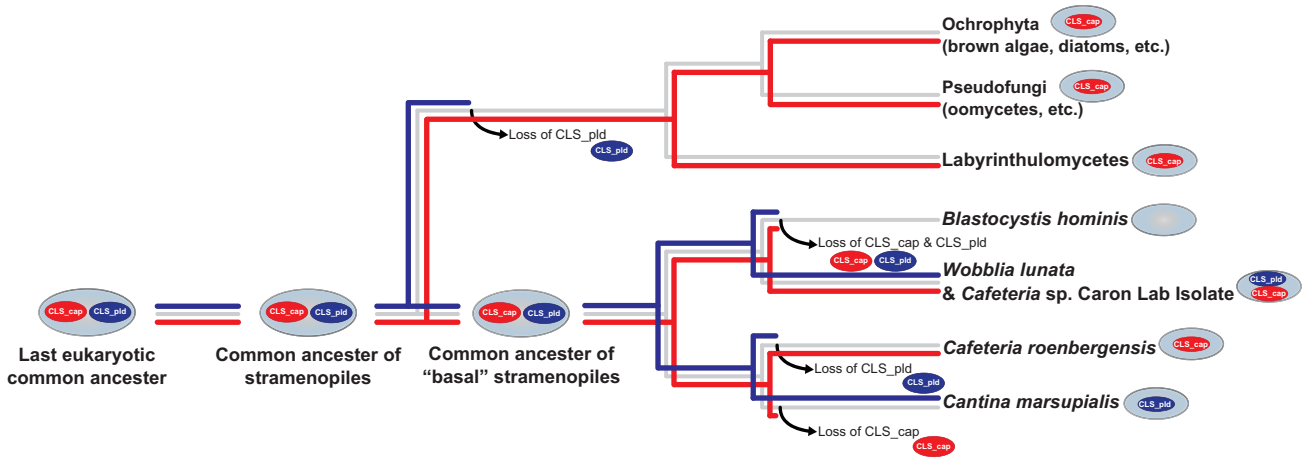

Fig. 4-2. Hypothetical evolutionary scenarios of two CLSs in the stramenopile lineage.

Acknowledgements

I would like to express my graditude to my supervisor Dr. Kiyotaka Takishita (professor of Tokyo University of Marine Science and Technology). I also wish to thank Drs. Yoshihiro Fujiwara (professor of Tokyo University of Marine Science and Technology), Takao Yoshida (associated professor of Tokyo University of Marine Science and Technology), Shinji Tsuchida (associated professor of Tokyo University of Marine Science and Technology), Katsunori Fujikura (former professor of Tokyo University of Marine Science and Technology), Tadashi Maruyama (former professor of Tokyo University of Marine Science and Technology), Tetsuo Hashimoto (University of Tsukuba), and Yuji Inagaki (University of Tsukuba) for their useful advice. I especially thank Dr. Akinori Yabuki (Japan Agency for Marine-Earth Science and Technology) for his advice, technical assistance, and sample provision. I'm grateful to Mr. Shigeru Shimamura (Japan Agency for Marine-Earth Science and Technology) and Dr. Takuro Nakayama (University of Tsukuba) for help of handling RNA-seq data. I also wish to thank Dr. Miyuki Nishijima (Japan Agency for Marine-Earth Science and Technology) for help of HPLC analyses. I also wish to thank Drs. Jan Andersson (Uppsala University) and Courtney Stairs (Dalhousie University) for providing alignments for phylogenetic analyses. I also wish to thank Mr. Euki Yazaki (University of Tsukuba) for help of phylogenetic analyses. I also wish to thank Drs. Goro Tanifuji (University of Tsukuba) and Matthew Brown (Mississippi State University) for help of phylogenetic analyses and data preparation. I'm also grateful to all members of the department of marine biodiversity research in Japan Agency for Marine-Earth Science and Technology for their kindly support to my research.