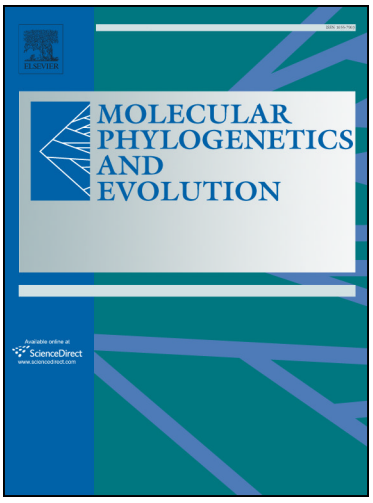


Phylogenomic analysis of integral diiron membrane histidine motif-containing enzymes in ciliates provides insights into their function and evolutionary relationships.

Nicolas G. Cid, María L. Sanchez Granel, María G. Montes, María E. Elguero, Clara B. Nudel, Alejandro D. Nusblat

PII: S1055-7903(17)30098-2
DOI: http://dx.doi.org/10.1016/j.ympev.2017.05.023
Reference: YMPEV 5836

To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 31 January 2017
Revised Date: 24 May 2017
Accepted Date: 24 May 2017

Please cite this article as: Cid, N.G., Sanchez Granel, M.L., Montes, M.G., Elguero, M.E., Nudel, C.B., Nusblat, A.D., Phylogenomic analysis of integral diiron membrane histidine motif-containing enzymes in ciliates provides insights into their function and evolutionary relationships., *Molecular Phylogenetics and Evolution* (2017), doi: http://dx.doi.org/10.1016/j.ympev.2017.05.023

TITLE PAGE

Phylogenomic analysis of integral diiron membrane histidine motif-containing enzymes in ciliates provides insights into their function and evolutionary relationships.

Nicolas G. Cid, María L. Sanchez Granel, María G. Montes, María E. Elguero, Clara B. Nudel, Alejandro D. Nusblat*.

Universidad de Buenos Aires. Consejo Nacional de Investigaciones Científicas y Técnicas. Instituto de Nanobiotecnología (NANOBIOTEC). Facultad de Farmacia y Bioquímica, Junín, 956 (C1113AAD), Buenos Aires, Argentina.

Corresponding author at all stages of refereeing and publication, also post-publication:
* Alejandro D. Nusblat,
E-mail: anusblat@ffyb.uba.ar
Address: Universidad de Buenos Aires. Consejo Nacional de Investigaciones Científicas y Técnicas. Instituto de Nanobiotecnología (NANOBIOTEC). Facultad de Farmacia y Bioquímica, Junín, 956 (C1113AAD), Buenos Aires, Argentina.
Fax: +54-11- 5287-4675

ABSTRACT

The Integral Membrane Histidine Motif-containing Enzymes (IMHME) are a class of binuclear non-heme iron proteins widely distributed among prokaryotes and eukaryotes. They are characterized by a conserved tripartite motif consisting of eight to ten histidine residues. Their known function is the activation of the dioxygen moiety to serve as efficient catalysts for reactions of hydroxylation, desaturation or reduction. To date most studies on IMHME were carried out in metazoan, phototrophic or parasitic organisms, whereas genome-wide analysis in heterotrophic free living protozoa, such as the Ciliophora phylum, has not been undertaken. In the seven fully sequenced genomes available we retrieved 118 putative sequences of the IMHME type, albeit with large differences in number among the ciliates: 11 sequences in *Euplotes octocarinatus*, 7 in *Ichthyophthirius multifiliis*, 13 in *Oxytricha trifallax*, 18 in *Stylonychia lemnae*, 25 in *Tetrahymena thermophila*, 31 in *Paramecium tetraurelia* and 13 in *Pseudocohnilembus persalinus*.

The pool of putative sequences was classified in 16 orthologous groups from which 11 were related to fatty acid desaturase (FAD) and 5 to the fatty acid hydroxylase (FAH) superfamilies. Noteworthy, a large diversity on the number and type of FAD / FAH proteins were found among the ciliates, a feature that, in principle, may be attributed to peculiarities of the evolutionary process, such as gene expansion and reduction, but also to horizontal gene transfer, as we demonstrate in this work. We identified twelve putative enzymatic activities, from which four were newly assigned activities: sphingolipid Δ4-desaturase, ω3/Δ15 fatty acid desaturase, a large group of alkane 1-monooxygenases, and acylamide-delta-3(E)-desaturase, although unequivocal allocation would require additional experiments.We also combined the phylogenetics analysis with lipids analysis, thereby allowing the detection of two enzymatic activities not previously reported: a C-5 sterol desaturase in *P. tetraurelia* and a delta-9 fatty acid desaturase in *Cohnilembus reniformis*.

The analysis revealed a significant lower number of FAD's sequences in the spirotrichea ciliates than in the oligohymenophorea, enphasizing the importance of fatty acids trophic transfer among aquatic organisms as a source of variation in metabolic activity, individual and population growth rates, and reproduction.

Keywords
Ciliates; fatty acid desaturase superfamily; fatty acid hydroxylase superfamily; horizontal gene transfer; phylogenetics; histidine motifs.

Abbreviations
IMHME: Integral Membrane Histidine Motif-containing Enzymes
FAD: fatty acid desaturase
FAH: fatty acid hidroxylase
SCD1: stearoyl–coenzyme A desaturase-1
FA2H: fatty acid 2-hydroxylase
HGT: horizontal gene transfer
C4MO: C-4 methyl oxidases

INTRODUCTION

Proteins containing binuclear non-heme iron centers are widely distributed in nature. They serve as catalysts for essential O_2-dependent reactions required for many fundamental biological processes such as DNA repair, response to hypoxia, detoxification and synthesis of valuable metabolites (neurotransmitters, antibiotics and lipids). The key step in their catalytic function is the activation of dioxygen by the iron center for a subsequent diversified chemistry, mainly consisting of hydroxylation or halogenation of unreactive aliphatic C-H bonds, electrophilic aromatic substitution, desaturation, epoxidation or ring closure (Solomon et al. 2000).

Examples of binuclear non-heme iron proteins include the soluble di-iron-carboxylate enzymes, that perform hydroxilation reactions (methane monooxygenase, MMO; toluene monooxygenase), desaturation (soluble stearoyl-ACP desaturases), and reduction (ribonucleotide reductase) among others (Mitić et al. 2009; Nordlund and Eklund, 1995). Another well-known group of binuclear non-heme iron proteins are the Integral Membrane Histidine Motif-containing Enzymes (IMHME). They are characterized by a conserved tripartite motif consisting of eight histidines: $H(X)_{3-4}H$, $H(X)_2HH$ and $H/Q(X)_2HH$ (Shanklin et al. 1994) and the presence of 4 to 6 membrane spanning helices, as revealed by hydropathy plots (Sperling et al. 2003) and crystal structures (Bai et al. 2015) (Figure 1A, B). The IMHME are vastly distributed among prokaryotes and eukaryotes and display diverse enzymatic activities; they are classified as desaturases, hydroxylases, epoxidases, acetylenases, conjugases, ketolases, decarbonylases and methyl oxidases. Typical examples from prokaryotes are the alkane ω-hydroxylase (AlkB) and xylene monooxygenases (XylM) in *Pseudomonas* (Shanklin and Whittle, 2003), the plastid-type desaturases in Cyanobacteria (Chi et al. 2008) and from eukaryotes the stearoyl-CoA desaturases, front-end and methyl-end fatty acid desaturases, sphingoid C-4 hydroxylase and sterol C-4 methyl oxidase, among others (Bard et al. 1996; Haak et al. 1997; Sperling and Heinz 2001; Stukey et al. 1990) (Figure 1C).

The IMHME require molecular oxygen and reducing equivalents from NADH, delivered by ferredoxin (in cyanobacterial and plastidial enzymes) or by cytochrome b_5 (in the endoplasmic reticulum enzymes) with termination at the di-iron center of the enzyme for their catalysis. The crystal structures of two enzymes belonging to this group have been recently published; a stearoyl–coenzyme A desaturase-1 (SCD1) from human and mouse and a fatty acid 2-hydroxylase (FA2H) from yeast (Bai et al. 2015; Wang et al. 2015; Zhu et al. 2015). The enzymes were co-crystallized with their substrates, thus defining precisely their di-metal catalytic centre and providing structural basis for the regioselectivity and stereospecificity of the reactions involved.

A comparative structural analysis and phylogenomic typing of IMHME proteins from Bacteria and Eukarya classified them in two groups that differ in the spacing between the second and third histidine motifs, thus generating a “large” or a “short” sequence segment between the second and the third trans-membrane region (Figure 1C) (Ternes et al. 2002). Interestingly, the crystal structure from the two representatives of these groups (SCD1 and FA2H) confirmed this topology: both enzymes resemble a mushroom-like architecture

consisting of a stem of four trans-membrane domains that span the endoplasmic reticulum membrane and a catalytic cap domain that faces the cytoplasm, as shown in Figure 1B (Wang et al. 2015; Zhu et al. 2015).

The group carrying the "large" interspaced sequence is composed of enzymes that belong to the fatty acid desaturase (FAD) superfamily (Pfam: PF00487) and include the acyl-CoA, methyl-end and front-end fatty acid desaturases, the delta-8 and delta-4 desaturases of sphingolipids long chain base, and the delta-3 desaturases of the amide-linked acyl residues of sphingolipids, according to the classification of Sperling et al. (2003). These enzymes are involved in the lipid metabolism of fatty acids and sphingolipids, and all of them have desaturase activity, except the sphingolipid delta-4 desaturase that has both desaturase and hydroxylase activity (Ternes et al. 2002). The group with the "short" interspaced sequence comprises enzymes belonging to the fatty acid hydroxylase (FAH) superfamily (Pfam: PF04116). These enzymes are engaged in sphingolipid metabolism (sphingolipid C4-hydroxylase, fatty acid 2-hydroxylase), sterol metabolism (C-5 sterol desaturase, C-4 methyl sterol oxidase, C-25 cholesterol hydroxylase, C-24 sterol deethylase) and hydrocarbon and carotenoid biosynthesis (aldehyde decarbonylase and carotene hydroxylase, respectively). Interestingly, none of the enzymes in this group is involved in fatty acid metabolism. A list of substrates and products illustrating all described enzyme activities is provided as Supplementary Material in Figure S1.

As mentioned, many of these enzymes, particularly desaturases, have been studied in prokaryotes and eukaryotes including metazoans, higher plants, algae, mosses, fungi and few other lower eukaryotes, principally because of their importance in distinctive metabolic pathways (Khozin-Goldberg et al. 2011; Pereira et al. 2003; Uttaro, 2006). To date, genome-wide analysis of IMHME in heterotrophic free living protozoa, such as the group of the ciliates, has not been thoroughly examined, thus remaining a great opportunity, both for fundamental studies and as a source of important biotechnological activities.

Ciliates are acknowledged as single cell organisms of significant structural and evolutionary complexity. These heterotrophic and, in most cases, free-living microbes, together with the dinoflagellates and the parasitic apicomplexa, constitute the three major evolutionary lineages that make up the alveolates, one of the members that compose the SAR supergroup (Adl et al. 2012). Several studies have reported about the composition and metabolism of fatty acids, sphingolipids and sterols in *Tetrahymena* and *Paramecium*, two ciliates model organisms (Arnaiz et al. 2007; Eisen et al. 2006; Ruehle et al. 2016). The presence of Δ6, Δ9 and Δ12 polyunsaturated fatty acids (not acquired from the diet) and some detected modifications in sphingolipids moieties, such as the introduction of a double bond at Δ4 position of sphingoid bases and the hydroxylation at the α-position of amide-linked acyl residues, suggested the activity of IMHME (Kaneshiro, 1997; Kaya et al. 1984; Nozawa et al. 1974). To date, five groups of IMHME enzymes were unequivocally identified; four in *Tetrahymena thermophila*: delta-9 and delta-6 fatty acid desaturases, C-5 sterol desaturase and C-24 sterol deethylase (Nakashima et al. 1996; Nusblat et al. 2009; Rusing et al. 2006; Tomazic et al. 2011) and delta-5 fatty acid desaturase in *Paremecium tetraurelia* (Tavares et al. 2011).

Currently the exploration can be extended, provided that the complete genomes from seven ciliates of two distant phylogenetic classes are available: *Tetrahymena thermophila*, *Ichthyophthirius multifiliis*,

Paramecium tetraurelia and *Pseudocohnilembus persalinus* (class Oligohymenophorea) and *Oxytricha trifallax, Stylonychia lemnae* and *Euplotes octocarinatus* (class Spirotrichea) (Aeschlimann et al. 2014; Arnaiz et al. 2007; Coyne et al. 2011; Stover et al. 2006; Swart et al. 2013; Xiong et al. 2015; Xiong and Miao 2016). With this information, we performed a genome-wide analysis to study the nature and possible evolutionary lineage of the IMHME in the Ciliophora phylum. Combining bioinformatics research with lipid analyses, we identified potential IMHME relevant for lipids modifications. We also analyzed the apparent genetic selective pressure on unknown genes, in an effort to contribute to the knowledge on their genetic stability and functional variability.

MATERIAL AND METHODS

Strains and culture conditions

Paramecium tetraurelia strain d4-2 was obtained as a gift from Thomas G. Doak, Department of Biology, Indiana University. Cells were grown in 250 ml Erlenmeyer flasks in a bacterized Wheat Grass Powder (Pines International, Lawrence, KS) medium (Aury et al. 2006). *Klebsiella pneumonia* was inoculated into the sterile wheat grass infusion the day before used for feeding paramecia, and incubated overnight for bacterial growth, at 30°C. A final concentratrion of 0.4 µg/ml of Stigmasterol (5α-cholest-5,22-dien-24β-ethyl-3β-ol) was added to the final culture medium.

Oxytricha trifallax strain JRB310 was obtained as a gift from Laura Landweber, Department of Ecology and Evolutionary Biology, Princeton University. Cells were grown in 250 ml Erlenmeyer flasks containing Pringsheim buffer with *Chlamydomonas reinhardtii* and *Klebsiella oxytoca* as the food source (Fang et al. 2012). To obtain *O. trifallax* cells devoid from feeding traces, dense cultures were allowed to consume most of the algae present in the culture, then they were fed with cultures of dead (autoclaved) *K. oxytoca* and transferred to 6°C for 3 days. Under these conditions the *Oxytricha* cells clear the culture of the remaining algae and cyst formation is inhibited (Dawson et al. 1984).

Cohnilembus reniformis was obtained from the Culture Collection of Algae and Protozoa (CCAP), Scotland, United Kingdom. Cells were grown without shaking at 18°C in 250 ml Erlenmeyer flasks containing 20 ml of sterile seawater. The medium was inoculated with 400 µl of *Escherichia coli* DH5α grown at OD 3.0 and supplied with 250 µg/ml of Amphotericin B. After 4 days a cellular density of $1x10^4$ cells/ml was achieved and cells were harvested at 3,000 × *g* for 10 min at 4°C.

Identification of fatty acids by gas chromatography-MS (GC-MS)

Cells from cultures grown as indicated above were collected by centrifugation at 3,000 × *g* for 10 min at 4°C, washed twice with 20 ml of distilled water, and the lipids were extracted according to Bligh and Dyer (Bligh and Dyer, 1959). The organic phase was evaporated under N_2 stream and saponified with 2 volumes of methanol:chlorhydric acid (5:1) overnight at 80°C. After two-fold extraction with 4 ml hexane, the organic solvent was evaporated under N_2 stream and the residue was resuspended in 1 ml of hexane. Each sample was sowed in a pre-condicionated silica cartridge (Sep-Pak Vac 3cc) and eluted with 5 ml of ethyl acetate. After evaporating under N_2 stream, the samples were resuspended in 50 µl hexane. The composition in fatty acid methyl ester derivatives was analyzed in a Hewlett Packard HP 6890 gas chromatograph equipped with a Zebron ZB-5 column (30-m long, 0.53-mm inner diameter; Phenomenex, California, USA). The increase in temperature in the column was programmed at 10°C/min from 100 to 310°C and subsequently held for 10 min at 310°C. MS was carried out using a HP mass selective detector (Mass Spec Model 5975VL) operated at an ionization voltage of 70eV with a scan range of 50 to 600 atomic mass units (amu). Fatty acid identification was performed by comparision with standards present in NIST database (National Institute of Standards and Technologies).

Detection of C-5 sterol desaturase activity

C-5 sterol desaturase activity was analyzed by the convertion of lathosterol (5α-cholest-7-en-3β-ol) to 7-dehydrocholesterol (5α-cholest-5,7-dien-3β-ol) according to Nusblat et al. (2009). Briefly, 150 ml of a *P. tetraurelia* culture supplemented with 0,4 µg/ml of stigmasterol was grown to reach approximately 4000 cells/ml. Then lathosterol, cholesterol (5α-cholest-5-en-3β-ol) (used as control for C-7 sterol desaturase activity) or ethanol (blank) were added to the culture medium to a final concentration of 20 µg/ml from a 5 mg/ml stock solutions in ethanol and left for 7 days. The cell pellet of 150 ml of culture was resuspended in 2 ml of water and submitted to lipid saponification using 2 ml of 2 M NaOH in methanol-water (1:1, vol/vol) at 60°C for 1 h. After cooling, the sterols were extracted twice with 5 ml of hexane and dried under N_2 stream. β-sitosterol (5α-cholest-5-en-24β-ethyl-3β-ol) was used as the internal standard and added prior to extraction. The sterols were resuspended in ethanol and separated by high-performance liquid chromatography (HPLC) on a ZORBAX Eclipse XDB-C18 column, using methanol-water (95:5, vol/vol) as the mobile phase. Sterol identification was performed by comparison with standards.

Bioinformatics and phylogenetic analyses

The data set comprising the fatty acid desaturases and fatty acid hydroxylases of fully sequenced genomes of *E. octocarinatus, I. multifiliis, O. trifallax, S. lemnae, T. thermophila, P. tetraurelia* and *P. persalinus* was generated combining the sequences retrieved from the following resources: 1) the National Center of Biotecnology Information - Conserved Domains Database (NCBI-CDD), [FADs(cl 00615) and FAHs(cl 01132)] "Related Proteins", for ciliates annotated in the NCBI database (*I. multifiliis, O. trifallax, S. lemnae, T. thermophila, P. tetraurelia*); 2) using publicly available protein secuences from Pfam 30.0 (EMBL-EBI), searching by PFAM (PF00487,PF04116) in ciliates; 3) using publicly available genome data bases of each of the ciliates and selecting proteins with the PF00487 (FAD) or PF04116 (FAH) protein families including: *E. octocarinatus* (http://ciliates.ihb.ac.cn/database/species/eo), *I. multifiliis* (http://ich.ciliate.org/index.php/home), *O. trifallax* (http://oxy.ciliate.org/index.php/home), *S. lemnae* (http://stylo.ciliate.org/index.php/home), *T. thermophila* (http://ciliate.org/index.php/home), *P. tetraurelia* (http://paramecium.cgm.cnrs-gif.fr/) and *P. persalinus* (http://ciliates.ihb.ac.cn/database/species/pp).

Redundant sequences were excluded from the data set sequences using BLASTp tool at NCBI´s homepage with default setting parameters (Substitution matrix: BLOSUM62; word size: 6; gap opening cost: 11; gap extension cost: 1). Pairwise alignments of each individual sequence from the dataset against the entire dataset were retrieved. Those sequences with identity of 100% with another sequence from the dataset were considered as repeated sequences and therefore were excluded from the dataset. CD-HIT with a sequence identity cut-off of 0.99 was also used to exclude redundant sequences. All the sequences were scanned for the tripartite conserved motif (H-x(3,4)- H-x(3,150)- H-x(2,3)- H-H- x(50,250)-[HQ]-x(2,3)-H- H) using ScanProsite tool, and those not matching were discarded.

The OrthoMCL database v5 was used to retrieve the orthologous group of each protein sequence using blastp tool.

For phylogenetic analyses, PROMALS3D with default parameters was used to obtain the multiple alignments (MSA) of the datasets. Subsequently, Block Mapping and Gathering with Entropy (BMGE) with entropy cut off value of 0.9 was used to remove the sequences segments that were poorly aligned. This MSA was used to construct the phylogeny by Maximum Likelihood using PhyML 3.0 optimized by SPR and NNI and a support of 100 bootstraps. The models for the MSA were selected using Prottest 3.0 server (Guindon et al. 2010; Le and Gascuel, 2008). For Horizontal Gene Transfer analyses, MUSCLE tool provided in the MEGA 6.0 package was used to obtain the MSA.

Analysis of dN/dS

For calculation of dN/dS ratios the sequences were first trimmed so that they all started and ended at the corresponding START and STOP codons. Nucleotide alignments for each gene were then obtained using TranslatorX. An unrooted phylogenetic tree for each alignment was constructed using the PHYML 3.0 program. Using both, these alignments and trees, we calculated the dN/dS for the branches of interest by running the codeml program from the PAML package (Bielawski and Yang, 2004) with the parameters model=2; NSsites=0.

RESULTS

A search of proteins associated to the FAD (Pfam: PF00487) and FAH (Pfam: PF04116) superfamilies, performed in November 2016, with the proteome of representative organisms of 26 main lineages of eukaryotes, 31 representatives of eubacteria and 13 of archaea (Table S1), showed that none of the archaea carry orthologs with significant homology to any of the enzymes of the two superfamilies, in agreement with previous bioinformatics (Nie et al. 2014) and analytical lipids data (Tornabene et al. 1979). Conversely, in most eubacteria and eukaryotic organisms, one or both types of orthologs were present with a clear supremacy in the number of FAD proteins with respect to the less abundant FAHs (Figure 2). Moreover, proteins of both groups were found in many Cyanobacteria and Firmicutes, which are considered the earliest division of the bacterial phylum (Ciccarelli et al. 2006, Hug et al. 2016), suggesting that these superfamilies may have been generated early after the division of the eubacteria and archaea/eukaryotic lineages.

In eukaryotes both superfamilies seem to have expanded considerably, as the average number of FAD and FAH proteins are 8.88 and 6.00 respectively, while in eubacteria it is only 2.68 and 1.13 respectively. The low number of IMHME found in several alfa-proteobacteria supports the assumption that the expansion displayed in eukaryotes must have taken place at a later phase than the symbiosis of an archaeon (or a member of a protoeukaryotic sister lineage to archaea) and one bacterium, the ancestor of mitochondria.

Using the Conserved Domain Database (CDD) at NCBI and other resources indicated in Materials and Methods, we identifed 139 FAD/FAH sequences in the complete genomes of the seven ciliates examined. Subsequent analysis discarded 21 sequences for different criteria, such as redundancy and incomplete histidine clusters. From the remaining 118 putative sequences, 70 sequences were structurally linked to the FAD superfamily whereas 48 showed the typical features of the FAH superfamily (Table S2). Noteworthy, the total number of IMHME among the seven ciliates was significantly different, as 7 were found in *I. multifiliis*, 11 in *E. octocarinatus*, 13 in *O. trifallax* and *P. persalinus*, 18 in *S. lemnae*, 25 in *T. thermophila* and 31 in *P. tetraurelia* (Figure 3A).

The pool of putative sequences was classified in 16 orthologous groups from which 11 were related to FAD and 5 to FAH superfamilies (Table 1). Although *P. tetraurelia* stands out for the highest number of sequences (31), *T. thermophila* reveals the largest expanded groups of IMHME, with 8 diverse groups for FADs and 5 for FAHs. The high number of paralogs in *P. tetraurelia* can be explained by the fact that this ciliate has undergone at least two rounds of whole genome duplication since its divergence from the last common ancestor of *T. thermophila* (Aury et al. 2006). In contrast, the spirotrich ciliates *O. trifallax* and *S. lemnae* are the ciliates with the lowest number of orthologous groups, displaying only 2 FAD (alkane-1 monooxygenase and sphingolipid delta-4 desaturase) and 2 FAH groups (C-4 methyl oxidase and fatty acid 2-hydroxylase).

From the 118 putative IMHME´s coding genes present in the ciliates genomes, there is very little information on their proteins and their role in microbial metabolism. As mentioned above, even in the most extensively studied genera, *Tetrahymena* and *Paramecium*, only six genes were identified to date: C-5 sterol desaturase (TTHERM_01194720), C-24 sterol deethylase (TTHERM_00438800) and delta-6 fatty acid

desaturase **(**TTHERM_00138530), all by knockout analysis, delta-9 fatty acid desaturase (TTHERM_00052620) by RNA analysis in *T. thermophila* and two orthologs of delta-5 fatty acid desaturases in *P. tetraurelia,* by heterologous expression in *Saccharomyces cerevisiae* (GSPATT00004542001 and GSPATT00037760001). We have recently identified two more genes in *T. thermophila* by somatic knockout, a delta-12 fatty acid desaturase (TTHERM_00535680) and a fatty acid 2-hydroxylase (TTHERM_00463850) (unpublished results from these authors).

The orthologous analysis also showed 6 groups that were present in only one class (Table 1): three of them contain sequences exclusively of *T. thermophila,* two of *P. persalinus* and one of *P. tetraurelia*. This uneven distribution could be attributed to various reasons; for example, in the case of *T. thermophila* the proteins belonging to the alkane 1-monooxygenase 1 group (OG5_249793) seem likely to have expanded from the alkane 1-monooxygenase group (OG5_156889). A similar pattern is observed with genes belonging to the fatty acid desaturase/cytochromeb_5-like heme/steroid binding domain ortholog group (OG5_139969 and OG5_128446) in *P. tetraurelia* and *P. persalinus* respectively. Another reason of disproportion may be due to gene gain or loss; for example, the unique sequence in OG5_132934 belongs to *Tetrahymena* and corresponds to a C-24 sterol deethylase (with confirmed enzymatic activity), presumably acquired by horizontal gene transfer (HGT) from an evolutionary distant organism, and displaying a desaturation mechanism not resembling any known deethylases from nematodes nor insects (Tomazic et al. 2011).

To get more information on the genes with only putatively assigned biological roles, we examined their rate of evolution as a possible indication on the stringency of their maintenance. For that purpose, we calculated the index of selective pressure on codon usage (dN/dS ratio), according to Bielawski and Yang (2004). Overall, values >1 are indicative for low stringency (or high diversifying codon selection), whereas values <1 indicate stringent selection (or high codon conservation). That is, the lower the dN/dS ratio, the more stringent selection is acting on the gene. We set the range of the scale calculating the dN/dS ratio for an histone gene, which has high selective pressure, and a mitochondrial *Ymf* gene, which is reported to have relaxed selective constraints (0.0122 and 1.5276 respectively). As shown in Table 2, all the set of genes tested displayed dN/dS ratios below 1, indicating a high degree of codon conservation (or a high purifying selection tendency), and suggesting an essentially/conserved biological role though not yet identified. In this context, an extreme case was the C-24 deethylase gene, possessing the highest dN/dS value among all analyzed genes in *Tetrahymena* (0.3345), although it was still within the range indicative for purifying selection.

Additionally, the measurement of the gene expression profiles of the above mentioned genes at different stages of *Tetrahymena´s* life cycle (vegetative growth, starvation and sexual reproduction) available at the TetraFGD database, revealed a high degree of change in their expression during the cell cycle, thus reinforcing the hypothesis of a related biological function. In contrast, no differential expression has been revealed for the *Paramecium* genes (ParameciumDB) identified through bioinformatics analysis, compared to data collected during various physiological conditions (reciliation and/ or regulated exocytosis), albeit more physiological conditions should be analyzed for conclusive results.

A consensus phylogenetic tree generated by the Maximum Likelihood method with the 118 ciliates sequences showed the typical two groups, represented by the FAD and FAH superfamilies (Figure 3B). The phylogenomic classification could separate efficiently the different types of enzymes, according to the orthologous groups referred above, and the enzymatic activity predicted or confirmed. The FAH superfamily consisted of five groups: three related to sterol metabolism [C-4 methyl oxidases (C4MO), C-5 sterol desaturase and C-24 sterol deethylase], one to sphingolipids (fatty acid 2-hydroxylases) and one with unpredictable enzymatic activities made up by sequences of *T. thermophila* and *P. tetraurelia*.

Analysis of the FAH (PfamPF04116) superfamily.

Previous reports identified unambiguously a C-5 sterol desaturasa (TTHERM_01194720) and a C-24 sterol deethylase (TTHERM_00438800) in *T. thermophila* (Nusblat et al., 2009; Tomazic et al., 2011). In spite that no clear ortholog for a C-5 sterol desaturase was was found in the genome mining of *P. tetraurelia* in our analysis, we investigated this activity in cultures, together with a C-7 sterol desaturase activity, previously reported by Najle et al. (2013). For this purpose we supplemented cultures with lathosterol (5α-cholest-7-en-3β-ol) or cholesterol (5α-cholest-5-en-3β-ol), to evidence C-5 and C-7 sterol activity (respectively) by their biotransformation to 7-dehydrocholesterol (5α-cholest-5,7-dien-3β-ol) in both cases.

As shown in Figure 4A and in Figure S3 (Supplementary Material), there was an increase in 7-dehydrocholesterol recovered in cultures either with lathosterol or cholesterol, with respect to the cultures with no sterols supplementation (0,096 and 0,032 ug respectively), thus indicating a C-5 and a C-7 sterol desaturase activity in this ciliate. Remarkably, the C-7 sterol desaturases from *Tetrahymena* and *Paramecium* are not IMHME but belong to the Rieske type iron-sulfur enzymes (Najle et al. 2013).

It is doubtful the assignment of the proposed C4MO activity in the ciliates, as no C4 de-methylated sterols were detected in cultures of *T. thermophila* supplemented with various C4 mono or di-methyl sterols added. Moreover, single knockouts on the four putative C4MO genes in *T. thermophila* did not yield indication on their possible biological role, as no appreciable change in growth parameters, cellular behavior, morphology, movement, or in the sterol or fatty acid composition could be detected (Tomazic et al. 2014).

On the other hand, the fatty acid 2-hydroxylase orthologous group (OG5_129723) carries sequences from all the spirotrichean ciliates available, plus only one oligohymenophorean (*T. thermophila*). The proposed enzymatic activity in this group of orthologs was confirmed by the generation of a *T. thermophila* knock-out strain (TTHERM_00463850) in which the 2-hydroxylation activity of the amide-linked acyl residues in sphingolipids was lost (unpublished results from this author). It is interesting to note that, although α-hydroxy fatty acids in the sphingolipids composition of *P. tetraurelia* have been reported (Kaneshiro et al. 1997), no genes of this organism has members in this precise branch, thus suggesting that the gene/s might be in the sister branch of unknown sequences (OG5_187540), which does include six *P. tetraurelia* orthologs.

Analysis of the FAD (PfamPF00487) superfamily.

The IMHME belonging to the FAD protein superfamily retrieved from the ciliates genomes correspond to enzymes of six different activities according to the OrthoMCL database: front-end fatty acid desaturases,

methyl-end fatty acid desaturases, stearoyl-CoA fatty acid desaturases, sphingolipid delta-4 desaturase (all of them described in ciliates) and two not studied to date: acylamide-delta-3(E)-desaturases and alkane-1 monooxygenases.

The acylamide delta-3(E)-desaturase group is composed of four paralogous sequences of *T. thermophila* and one of *P. persalinus*. This enzyme, which introduces a double bond at the C-3 position of the hydroxyl-fatty acyl moiety of sphingolipids (see Figure S1), has been identified only in fungal species. Previous reports on the composition of lipids in *Tetrahymena*, and a new check recently performed in our group, did not reveal the presence of any delta-3 unsaturated compounds, neither as free fatty acids nor in the fatty acyl moiety of sphingolipids, or triglycerides, generating doubts on the possible role of these genes in the organism. In spite, the genes display high purifying selection tendency, according to the analysis of selective pressure, and high expression rates and regulation at different stages of the life cycle (Table 2) suggesting an important, though unknown, biological role. A more restricted phylogenetic analysis performed with sequences retrieved by BLAST searches plus stearoyl-CoA desaturases and fatty acid 2 hydoxylase sequences, used as out group, showed three main branches (Figure 5A): one group of uncharacterized ciliates sequences, one group of fungal acylamide delta-3 desaturases and a group of bacterial uncharacterized sequences. The absence of eukaryotic sequences and the patchy phylogenetic distribution suggest that fungal and ciliates genes were acquired via HGT from bacteria as the most parsimonious scenario. Moreover, the largest similarity of the eukaryotic sequences with proteins from Bacteroidetes points this bacterial phylum as a possible donor.

The presence of a branch made up of 24 sequences from the seven ciliates related to alkane-1 monooxygenases (OG5_156889 and OG5_249793) is also intriguing. These enzymes are abundantly present in bacteria, where they are responsible for the initial oxidation of alkanes (Maeng et al. 1996), but this activity has not been reported in eukaryotes. A BLAST analysis with all the ciliates AlkB sequences retrieved only eight other eukaryotic sequences: a unicellular opisthokonta, two red algae, one brown algae, one discobid, one hacrobian and one apusozoa, none of them from the SAR supergroup shared by the ciliates. Moreover, a phylogenetic analysis grouped the AlkB of ciliates and the above mentioned eukaryotic sequences together with AlkB bacterial proteins, and far from the other FAD groups (Figure 5B), suggesting a probable acquisition through HGT from bacteria of various phyla (most likely Bacteroidetes, Proteobacteria a/o Spirochaetes). The same hypothesis arises from the analysis displayed in Figure 3, which suggests that the acquisition of AlkB should have occurred in an early evolution stage of the ciliates, before the diversification of the main classes, followed by a gene expansion in different lineages. The high purifying selection tendency and the highly regulated expression in different stages of the life cycle shown for those genes suggests a yet unknown but important biological role (Table 2).

In relation to FADs, the phylogenomic analysis showed significant differences among the two classes of ciliates (Figure 6); while the oligohymenophorea ciliates carry many sequences corresponding to the three types of desaturases, namely stearoyl CoA, front-end and methyl-end fatty acid desaturases, the spirotrich ciliates display only one sequence corresponding to a putative front-end desaturase, found in the genome of

E. octocarinatus. Phylogenetic analysis with 21 front-end desaturases from ciliates plus a set of delta-4, delta-5, delta-6, and delta-8 desaturases from diverse organisms, grouped the Euplotes sequence (EO8932.g28) in a branch with other ciliates sequences, without assigning a clear regioselectivity to this group (Figure 5C). As a possible explanation for this unexpected result, we propose that spirotrichs could have lost the stearoyl CoA and methyl-end fatty acid desaturases after the differentiation of the Oligohymenophorea and Spirotrichea classes, whereas the front-end fatty acid desaturases could have been lost at some point during the evolution of the subclass Hypotrichian, in which *O. trifallax* and *S. lemnae,* but not *E. octocarinatus* (subclass Euplotia) are included.

To test the feasibility of this hypothesis we analyzed the total fatty acids content of *O. trifallax,* and compared that to the lipid composition of a mixture of bacteria (*Klebsiella oxytoca*) and algae (*Chlamydomonas reinhardtii*) that were used to feed cultures. As shown in Figure 4B, the fatty acids composition of the cells of the ciliate (devoid from their food) consisted of 16:0, 18:0, and 18:1 fatty acids, whereas the mixture of algae and bacteria (devoid of the ciliate) were mainly 16:0, 18:1, and 18:3 fatty acids. Thus, there was a change in the profile of saturated and unsaturated fatty acids between both samples, with a higher content of saturated fatty acids in *O. trifallax* (56.1%) and of unsaturated fatty acids (61.0%) in the bacteria/algae mixture, mainly due to the (18:3) fatty acid, which decreased from 44.3% in the bacteria/algae mixture to only 8.5% in the ciliate. A similar ratio between saturated/ unsaturated fatty acid was reported in other ciliates with different lifestyles, such as the axenically grown *T. thermophila, P. tetraurelia* and *Pleuronema sp.* (bacterivorous) and in *Fabrea salina* (feeding on algae) (Harvey et al, 1997; Kaneshiro, 1987; Nozawa et al. 1974). Thus, our analysis showed that the major fatty acids present in *O. trifallax* are derived from the feeding mixture, from which they were incorporated with no further desaturation. Together with the absence of presumed fatty acid desaturase genes, these findings suggest a possible gene loss in the spirotrichean ciliates from the common ancestor, although more studies will be required to confirm this hypothesis.

Among the Oligohymenophorea, *P. tetraurelia* carries a large set of genes with a number of proposed but not confirmed enzyme activities, such as two paralogous genes of putative stearoyl-CoA delta-9 desaturase, four delta-6 fatty acid desaturases, one delta-12 fatty acid desaturase and other six front-end fatty acid desaturases, that include the only two delta-5 fatty acid desaturases known presently with confirmed enzymatic activities (Figure 3 and Figure 5C). In principle, this large inventory of desaturases can explain the lipid composition of this ciliate, which includes palmitic (C16:0), stearic (C18:0), oleic ($C18:1^{\Delta 9}$), linoleic ($18:2^{\Delta 9,12}$), Gamma-linolenic ($18:3^{\Delta 6,9,12}$) and arachidonic ($20:4^{\Delta 5,8,11,14}$) as the major fatty acids present in the cells (Kaneshiro, 1987). As shown in Figure 5C, the four putative delta-6 fatty acid desaturases sequences of *P. tetraurelia* group with a well characterized acyl-lipid delta-6 fatty acid desaturase from the ciliate *T.thermophila* (Rusing et al. 2006), suggesting a similar activity.

The two paralogous genes assigned to stearoyl-CoA delta-9 desaturases (Figure 3) provided that *P. tetraurelia* has a direct requirement for oleic acid ($C18:1^{\Delta 9}$), not replaceable by stearic acid (Soldo et al.

1967), suggesting that the delta-9 desaturase genes may not be active in this role, in spite of their significant transcription levels during autogamy, reciliation and exocytosis (Table 2).

The main fatty acids identified so far in *T. thermophila* were the same as in *P. tetraurelia* except for arachidonic acid, ruling out the possibility of an active delta-5 fatty acid desaturase (Nozawa et al. 1974). These results agree with our genome mining of this ciliate, which revealed the presence of two paralogs stearoyl-CoA 9 desaturases, two delta-6 fatty acid desaturase, one delta12 fatty acid desaturase and only one front-end fatty acid desaturase of unknown regioselectivity (Figure 5C).

The ciliates *I. multifiliis* and *P. persalinus* lack both stearoyl-CoA (delta-9) desaturase and delta-6 fatty acid desaturase putative genes, while they have retained a possible delta-12 desaturase and other front end fatty acid desaturases of unknown activity (Figure 5C).

Interestingly, *P. persalinus* genome contains a putative sequence that codifies for a ω3/Δ15 fatty acid desaturase (PPERSA_00051670) which is not found in the other ciliates or in any eukaryotic organism from the SAR supergroup. The gene is actively expressed during vegetative growth, as shown by RNA seq-analysis (*P. persalinus* database). A phylogenetic tree with sequences retrieved by BLAST from ω3/Δ15 and ω6/Δ12 fatty acid desaturases from *P. persalinus* as well as other ω6/Δ12 fatty acid desaturases from other organisms plus ω3/Δ15 fatty acid desaturases identified in cyanobacteria, land plants, chlorophytes, fungi and the unusual fat-1 enzyme from *Caenorhabditis elegans* (Figure 5D), failed to enclose the *P. persalinus* sequence into any well defined branch, ruling out a probable origin by HGT while suggesting, as the most probable scenario, a derivation from its own ω6/Δ12 fatty acid desaturases. A remarkable and highly significant separation between the ω3/Δ15 and the ω6/Δ12 desaturases exist in the ciliate as well in cyanobacteria, land plants and chlorophytes, indicating that these two activities were separated a very long time ago in these organisms compared to the relatively recent invention of fungi and *C. elegans*.

With the aim to see the widespread of this newly described ω3/Δ15 activity in the group, we analyzed the lipid composition of *Cohnilembus reniformis*, another scutociliate of the same order Philasterida (Figure 3A). The main fatty acids consisted of 16:0(30.8%), 18:0(26.3%), $18:1^{\Delta 11}$(21.0%), $18:1^{\Delta 9}$(9.6%) and $16:1^{\Delta 9}$(8.8%) (Figure 4C). No polyunsaturated fatty acids were detected in the samples, ruling out the presence of any front- or methyl- end desaturases including a ω3/Δ15 fatty acid desaturase. Nevertheless, the presence of oleic acid is striking because it suggests an active delta-9 fatty acid desaturase in *C. reniformis* (Figure S2) not supplied by the *E coli* cells that were used to feed the ciliate. In effect, *E. coli* cells only contain palmitoleic acid ($16:1^{\Delta 9}$) and vaccenic acid ($18:1^{\Delta 11}$) as the only unsaturated fatty acids (Heipieper, 2005; Magnuson et al. 1993). Altogether, these results confirm that ciliates, even if evolutionarily close, may display a very different make-up of desaturases and coding genes, possibly influenced by their specific environment (marine, freshwater) a/o lifestyle (free-living, parasite).

DISCUSSION

The increasing amount of genome sequence data available, along with the advancements in culturing techniques and the development of new genetic tools, provides an excellent opportunity for exploring genes in microbes not previously examined. In particular, genome-wide analyses have contributed to shed light on the evolutionary lineages of IMHME, helped to elucidate diverse metabolic pathways and discovered new enzymatic activities (Chi et al. 2008, 2011; Nayeri and Yarizade, 2014).

Due to the beneficial effects on human health of polyunsaturated fatty acids, like EPA and DHA, fatty acid desaturases have attracted great attention from the biotech and genetic engineering sectors with a focus on enhanced PUFA production (Warude et al. 2006).

Interestingly, the ciliates and the archaeplastides are one of the phyla with the higher number of this kind of integral membrane enzymes across the tree of life. Our analysis on the genetic diversity of IMHME suggested that FAD and FAH superfamilies may have generated early in eubacteria, after the division of the eubacteria and archaea/eukaryotic lineages, and expanded largely after their eukaryotic origin.

In the case of the archaeplastides, the number of proteins members of the FAD and FAH superfamilies is one of the highest retrieved among eubacterias; their expansion could be due to a primary endosymbiosis of plastids during the uptake and retention of a cyanobacterium (Keeling, 2013). In ciliates, the FAD/FAH superfamilies expansion is noticeable particularly in *P. tetraurelia* and *T. thermophila*, a process already reported for other gene families including K+ channel proteins, kinases, membrane trafficking and motor proteins found in these ciliates (Bright et al. 2010; Eisen et al. 2006). On the other hand, the low number of these proteins in the oligohymenophorea *I. multifiliis* and *P. persalinus* could be explained by their parasitic lifestyle. In effect, these ciliates lost many of the ciliate-specific genes during their extensive genome reduction, remaining only 8,096 genes in *I. multifiliis* and 13,186 in *P. persalinus*, numbers significantly smaller than in other free-living ciliates, such as *T. thermophila* and *P. tetraurelia,* which harbor 24,725 and 39,642 genes respectively (Xiong et al. 2015). Other examples of genome reduction in parasitic organisms might be seen among the FAD/FAH proteins in the phylum Apicomplexa (*Toxoplasma gondii*, *Neosporacaninum, Eimeria maxima and Plasmodium vivax*) (Keeling and Slamovits, 2005).

Noteworthy, the diversity on the number and type of FADs and FAHs in ciliates may not only be attributed to some features of the evolutionary process (i.e, gene expansion and reduction), but also to HGT events during different stages of the evolution of this phylum. For instance, the alkane 1-monooxygenase, found in all the oligohymenophorea and spirotrichea ciliates analyzed, seems to have been acquired early after the ciliates diverged from the alveolate common ancestor, probably in the Precambrian era, while the putative acylamide delta-3(E)-desaturase, found in *T. thermophila* and *P. persalinus*, seem to have been acquired later, during the evolution of the oligohymenophorea linage. More recently the C-24 sterol deethylase, present only in *Tetrahymena*s species (*T. thermophila, T. borealis, T. malaccensis,* and *T. elliotti*) suggests that it was acquired during the evolution of the family Tetrahymenidae (Tomazic et al., 2011).

HGT events have been previously reported in ciliates and fungi (Fitzpatrick, 2012; Reyes-Prieto et al. 2008), and, as a matter of fact, they could well originate xenolog genes, with different enzymatic activity as a

result of different evolutionary histories, a process not yet described in ciliates but already reported in other organisms (Kooning, 2005).

Our analysis of IMHME identified twelve different enzymatic activities in the genome mining of the seven fully sequenced ciliates, from which four were newly assigned activities: sphingolipid Δ4-desaturase, ω3/Δ15 fatty acid desaturase, a large group of alkane 1-monooxygenase, and acylamide-delta-3(E)-desaturase. However, a considerable uncertainty exists about their real activity. For example, the sphingolipid Δ4-desaturases from *T. thermophila* and *P. tetraurelia* seem to be *bona fide* because their genes are actively expressed and their lipid products have been identified in both ciliates (Kaneshiro, 1997; Nozawa et al. 1974). In the case of the ω3/Δ15 fatty acid desaturase from *P. persalinus* and the alkane 1 monooxygenases present in all the ciliates, there are no clues on their enzymatic activity and remain to be tested. Whereas for the putative gene that codifies for the acylamide-delta-3(E)-desaturase, a xenolog hypothesis is proposed, as no product of this activity could be identified in *T. thermophila*, even during active gene transcription. A similar observation was reported for a group of C4MO in *T. thermophila*, also under active gene transcription conditions (Tomazic et al. 2014).

Through the detection of a C-5 sterol desaturase activity in *P. tetraurelia* reported in this work, we could assign this activity into the OG5_187540 orthologous group. It is interesting to note that although only one sequence of *T. thermophila* falls in this group (Q23JL0), the activity is not abolished with the gene disruption (unpublished results from these authors). On the other hand, a full knockout phenotype is achieved with the disruption of another sequence (Q22AK3) belonging to the related orthologous group OG5_129040 (Nusblat et al. 2009).

The analysis revealed a significant lower number of FA desaturases sequences in the spirotrichea ciliates than in the oligohymenophorea. Only one sequence corresponding to a front-end desaturase was found in the Euplotian *E. octocarinatus*, while none was identified in the hypotrichian *O. trifallax* and *S. lemnae*. Phylogenetic analysis of this desaturase rules out an adquisition by HGT, suggesting as the most parsimonius explication, that desaturases have been lost in different periods during the spirochitrean class evolution. Therefore, these ciliates must acquire the unsaturated fatty acids from the diet (i.e, phagocytosis of algae and bacteria) as we showed by the analysis of the lipid content in *O. trifalax* and its foodstuffs. This phenomenon of fatty acids trophic transfer has been shown to be crucial for aquatic organisms because they affect metabolic activity, individual and population growth rates, and reproduction. Additionally, as demonstrated by several authors, in zooplankton species the fatty acid composition of a predator usually resembles that of its prey (Ederington et al. 1995, Kainz et al. 2004; Lees and Korn, 1966).

Another striking observation is the disparitiy in the metabolic make up among ciliates, as very large species-specific differences in fatty acid and amino acid metabolism have been reported. Moreover, the metabolism of polyunsaturated fatty acids seems to differ more profoundly than the metabolism of other fatty acid classes or amino acids between ciliated protozoa (Boechat and Adrian, 2005).

Recently, during the reviewing process of this manuscript, the genome of the Heterotrich *Stentor coeruleus* became available (Slabodnick et al. 2017). Our analysis identified 16 IMHME putative sequences, 2 linked

to the FAD superfamily and 14 to the FAH superfamily (Table S3). The identification of 2 paralogous sequences of alkane 1-monooxygenase and 11 paralogous sequences of C-4 methyl oxidase, reinforce the idea that these genes could have been acquired early after the ciliates diverged from the alveolate common ancestor. Another interesting observation is the absence of any IMHME involved in the lipid metabolism of fatty acids or sphingolipids.

Due to this high diversity, it is difficult to predict the gene content of non-sequenced ciliates. In this line, *C. reniformis* and *P. persalinus* offer a good example; in spite of their phylogenetic relatedness, a landmark marker, the delta-9 fatty acid desaturase, is only present in one of them (Fig 6). As an advantage, this diversity could contribute to generate new activities, different time/stage gene expression or even cellular localizations, allowing the generation of new pathways, metabolites and thus adaptation to new environments of these unicellular heterotrophic eukaryotes, making them very valuable organisms for research and biotech.

Acknowledgements

We thank K. Hellingwerf (University of Amsterdam, The Netherlands) for the critical reading of this manuscript. A.D. Nusblat and C.B. Nudel are members of the Argentinean Research Council CONICET, Argentina. This work was supported by Grants PIP 11220120100213, ANPCYT-PICT 2013-0701 and UBACYT 20020130100680BA.

References

Adl, S. M., Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., ... &Heiss, A. (2012). The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, *59*(5), 429-514.

Aeschlimann, S. H., Jönsson, F., Postberg, J., Stover, N. A., Petera, R. L., Lipps, H. J., ... & Swart, E. C. (2014). The draft assembly of the radically organized Stylonychia lemnae macronuclear genome. Genome biology and evolution, 6(7), 1707-1723.

Arnaiz, O., Cain, S., Cohen, J., &Sperling, L. (2007). ParameciumDB: a community resource that integrates the Paramecium tetraurelia genome sequence with genetic data. *Nucleic Acids Research*, *35*(suppl 1), D439-D444.

Aury, J. M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M., ...& Arnaiz, O. (2006). Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. *Nature*, *444*(7116), 171-178.

Bai, Y., McCoy, J. G., Levin, E. J., Sobrado, P., Rajashankar, K. R., Fox, B. G., & Zhou, M. (2015). X-ray structure of a mammalian stearoyl-CoA desaturase. *Nature*, *524*(7564), 252-256.

Bard, M., Bruner, D. A., Pierson, C. A., Lees, N. D., Biermann, B., Frye, L., ...& Barbuch, R. (1996). Cloning and characterization of ERG25, the Saccharomyces cerevisiae gene encoding C-4 sterol methyl oxidase. *Proceedings of the National Academy of Sciences*, *93*(1), 186-190.

Bielawski, J. P., & Yang, Z. (2004). A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. Journal of Molecular Evolution, 59(1), 121-132.

Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry And Physiology*, *37*(8), 911-917.

Boëchat, I. G., & Adrian, R. (2005). Biochemical composition of algivorous freshwater ciliates: you are not what you eat. *FEMS Microbiology Ecology*, *53*(3), 393-400.

Bright, L. J., Kambesis, N., Nelson, S. B., Jeong, B., & Turkewitz, A. P. (2010). Comprehensive analysis reveals dynamic and evolutionary plasticity of Rab GTPases and membrane traffic in Tetrahymena thermophila. *PLoS Genet*, *6*(10), e1001155.

Chi, X., Yang, Q., Lu, Y., Wang, J., Zhang, Q., Pan, L., ... & Yu, S. (2011). Genome-wide analysis of fatty acid desaturases in soybean (Glycine max). *Plant Molecular Biology Reporter*, *29*(4), 769-783.

Chi, X., Yang, Q., Zhao, F., Qin, S., Yang, Y., Shen, J., & Lin, H. (2008). Comparative analysis of fatty acid desaturases in cyanobacterial genomes. *Comparative and Functional Genomics*, *2008*.

Ciccarelli, F. D., Doerks, T., Von Mering, C., Creevey, C. J., Snel, B., & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*, *311*(5765), 1283-1287.

Coyne, R. S., Hannick, L., Shanmugam, D., Hostetler, J. B., Brami, D., Joardar, V. S., ...& Kumar, U. (2011). Comparative genomics of the pathogenic ciliate Ichthyophthirius multifiliis, its free-living relatives and a host species provide insights into adoption of a parasitic lifestyle and prospects for disease control. *Genomebiology*, *12*(10), R100.

Darriba, D., Taboada, G. L., Doallo, R., &Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics, 27(8), 1164-1165.

Dawson, D., Buckley, B., Cartinhour, S., Myers, R., & Herrick, G. (1984). Elimination of germ-line tandemly repeated sequences from the somatic genome of the ciliate Oxytricha fallax. Chromosoma, 90(4), 289-294.

Ederington, M. C., McManus, G. B., & Harvey, H. R. (1995). Trophic transfer of fatty acids, sterols, and a triterpenoid alcohol between bacteria, a ciliate, and the copepod Acartiatonsa. *Limnology and Oceanography*, *40*(5), 860-867.

Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., ... & Delcher, A. L. (2006). Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. *PLoSBiol*, *4*(9), e286.

Fang, W., Wang, X., Bracht, J. R., Nowacki, M., & Landweber, L. F. (2012). Piwi-interacting RNAs protect DNA against loss during Oxytricha genome rearrangement. *Cell*, *151*(6), 1243-1255.

Fitzpatrick, D. A. (2012). Horizontal gene transfer in fungi. *FEMS Microbiology Letters*, *329*(1), 1-8.

Gao, F., Warren, A., Zhang, Q., Gong, J., Miao, M., Sun, P., ...& Song, W. (2016). The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). Scientific Reports, 6.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307-321.

Haak, D., Gable, K., Beeler, T., & Dunn, T. (1997). Hydroxylation of Saccharomyces cerevisiae ceramides requires Sur2p and Scs7p. *Journal of Biological Chemistry*, *272*(47), 29704-29710.

Harvey, H. R., Ederington, M. C., & Mcmanus, G. B. (1997). Lipid composition of the marine ciliates Pleuronema sp. and Fabreasalina: shifts in response to changes in diet. *Journal of Eukaryotic Microbiology*, *44*(3), 189-193.

He, D., Fiz-Palacios, O., Fu, C. J., Fehling, J., Tsai, C. C., &Baldauf, S. L. (2014). An alternative root for the eukaryote tree of life.Current Biology, 24(4), 465-470.

Heipieper, H. J. (2005). Adaptation of Escherichia coli to ethanol on the level of membrane fatty acid composition. *Applied and Environmental Microbiology*, *71*(6), 3388-3388.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., ... & Suzuki, Y. (2016). A new view of the tree of life. *Nature Microbiology*, *1*, 16048.

Kainz, M., Arts, M. T., & Mazumder, A. (2004). Essential fatty acids in the planktonic food web and their ecological role for higher trophic levels. *Limnology and Oceanography*, *49*(5), 1784-1793.

Kaneshiro, E. S. (1987). Lipids of Paramecium. *Journal of Lipid Research*, *28*(11), 1241-1258.

Kaneshiro, E. S., Jayasimhulu, K., Sul, D., & Erwin, J. A. (1997). Identification and initial characterizations of free, glycosylated, and phosphorylated ceramides of Paramecium. *Journal of Lipid Research*, *38*(12), 2399-2410.

Kaya, K., Ramesha, C. S., & Thompson, G. A. (1984). Temperature induced changes in the hydroxy and non-hydroxy fatty acid-containing sphingolipids abundant in the surface membrane of Tetrahymena pyriformis NT-1. *Journal of Lipid Research*, *25*(1), 68-74.

Keeling, P. J. (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annual Review of Plant Biology*, *64*, 583-607.

Keeling, P. J., & Slamovits, C. H. (2005). Causes and effects of nuclear genome reduction.*Current opinion in genetics & development*, *15*(6), 601-608.

Khozin-Goldberg, I., Iskandarov, U., & Cohen, Z. (2011). LC-PUFA from photosynthetic microalgae: occurrence, biosynthesis, and prospects in biotechnology. *Applied Microbiology and Biotechnology*, *91*(4), 905.

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics 1. *Annu. Rev. Genet.*, *39*, 309-338.

Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. Molecular Biology and Evolution, 25(7), 1307-1320.

Lees, A. M., & Korn, E. D. (1966). Metabolism of unsaturated fatty acids in protozoa. *Biochemistry*, *5*(5), 1475-1481.

Lynn, D. (2008). *The ciliated protozoa: characterization, classification, and guide to the literature*. Springer Science & Business Media. ISBN: 978-1-4020-8238-2.

Maeng, J. H., Sakai, Y., Tani, Y., & Kato, N. (1996). Isolation and characterization of a novel oxygenase that catalyzes the first step of n-alkane oxidation in Acinetobacter sp. strain M-1. *Journal of Bacteriology*, *178*(13), 3695-3700.

Magnuson, K., Jackowski, S., Rock, C. O., & Cronan, J. E. (1993). Regulation of fatty acid biosynthesis in Escherichia coli. *Microbiological Reviews*, *57*(3), 522-542.

Mitić, N., Schenk, G., & Hanson, G. R. (2009). Binuclear non-heme iron enzymes. In *High Resolution EPR* (pp. 269-395). Springer New York.

Najle, S. R., Nusblat, A. D., Nudel, C. B., & Uttaro, A. D. (2013). The Sterol-C7 desaturase from the ciliate Tetrahymena thermophila is a Rieske Oxygenase, which is highly conserved in animals. Molecular Biology and Evolution, 30(7), 1630-1643.

Nakashima, S., Yutong, Z. H. A. O., & Nozawa, Y. (1996). Molecular cloning of Δ9 fatty acid desaturase from the protozoan Tetrahymena thermophila and its mRNA expression during thermal membrane adaptation. *Biochemical Journal*, *317*(1), 29-34.

Nayeri, F. D., &Yarizade, K. (2014). Bioinformatics study of delta-12 fatty acid desaturase 2 (FAD2) gene in oilseeds. *Molecular Biology Reports*, *41*(8), 5077-5087.

Nie, Y., Chi, C. Q., Fang, H., Liang, J. L., Lu, S. L., Lai, G. L., ... & Wu, X. L. (2014). Diverse alkane hydroxylase genes in microorganisms and environments. *Scientific Reports*, *4*, 4968.

Nordlund, P., & Eklund, H. (1995).Di-iron—carboxylate proteins. *Current Opinion In Structural Biology*, *5*(6), 758-766.

Nozawa, Y., Iida, H., Fukushima, H., Ohki, K., & Ohnishi, S. (1974). Studies on Tetrahymena membranes: Temperature induced alterations in fatty acid composition of various membrane fractions in Tetrahymena pyriformis and its effect on membrane fluidity as inferred by spin-label study. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *367*(2), 134-147.

Nusblat, A. D., Najle, S. R., Tomazic, M. L., Uttaro, A. D., &Nudel, C. B. (2009). C-5 (6) sterol desaturase from Tetrahymena thermophila: gene identification and knockout, sequence analysis, and comparison to other C-5 (6) sterol desaturases. *Eukaryotic Cell*, *8*(8), 1287-1297.

Pereira, S. L., Leonard, A. E., &Mukerji, P. (2003). Recent advances in the study of fatty acid desaturases from animals and lower eukaryotes. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, *68*(2), 97-106.

Reyes-Prieto, A., Moustafa, A., & Bhattacharya, D. (2008). Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. *Current Biology*, *18*(13), 956-962.

Ruehle, M. D., Orias, E., & Pearson, C. G. (2016). Tetrahymena as a unicellular model eukaryote: genetic and genomic tools. *Genetics*, *203*(2), 649-665.

Rusing, M., Kiy, T., & Dominitzki, A. (2006). *U.S. Patent No. 7,135,623*. Washington, DC: U.S. Patent and Trademark Office.

Shanklin, J., & Whittle, E. (2003). Evidence linking the Pseudomonas oleovorans alkane ω-hydroxylase, an integral membrane diiron enzyme, and the fatty acid desaturase family. *FEBS Letters*, *545*(2-3), 188-192.

Shanklin, J., Whittle, E., & Fox, B. G. (1994). Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochemistry*, *33*(43), 12787-12794.

Slabodnick, M. M., Ruby, J. G., Reiff, S. B., Swart, E. C., Gosai, S., Prabakaran, S., ... & Gunawardena, J. (2017). The Macronuclear Genome of Stentor coeruleus Reveals Tiny Introns in a Giant Cell. Current Biology, 27(4), 569-575.

Soldo, A. T., & Wagtendonk, W. (1967). An analysis of the nutritional requirements for fatty acids of Paramecium aurelia. *The Journal of Protozoology*, *14*(4), 596-600.

Solomon, E. I., Brunold, T. C., Davis, M. I., Kemsley, J. N., Lee, S. K., Lehnert, N., ...& Zhou, J. (2000). Geometric and electronic structure/function correlations in non-heme iron enzymes. *Chemical Reviews*, *100*(1), 235-350.

Sperling, P., & Heinz, E. (2001). Desaturases fused to their electron donor. *European Journal of Lipid Science and Technology*, *103*(3), 158-180.

Sperling, P., Ternes, P., Zank, T. K., & Heinz, E. (2003). The evolution of desaturases. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, *68*(2), 73-95.

Stover, N. A., Krieger, C. J., Binkley, G., Dong, Q., Fisk, D. G., Nash, R., ... & Cherry, J. M. (2006). Tetrahymena Genome Database (TGD): a new genomic resource for Tetrahymena thermophila research. *Nucleic Acids Research*, *34*(suppl 1), D500-D503.

Stukey, J. E., McDonough, V. M., & Martin, C. E. (1990). The OLE1 gene of Saccharomyces cerevisiae encodes the delta-9 fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase gene. *Journal of Biological Chemistry*, *265*(33), 20144-20149.

Swart, E. C., Bracht, J. R., Magrini, V., Minx, P., Chen, X., Zhou, Y., ... & Jung, S. (2013). The Oxytricha trifallax macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoSBiol*, *11*(1), e1001473.

Tavares, S., Grotkjær, T., Obsen, T., Haslam, R. P., Napier, J. A., & Gunnarsson, N. (2011). Metabolic engineering of Saccharomyces cerevisiae for production of eicosapentaenoic acid, using a novel Δ5-desaturase from Paramecium tetraurelia. *Applied and Environmental Microbiology*, *77*(5), 1854-1861.

Ternes, P., Franke, S., Zähringer, U., Sperling, P., & Heinz, E. (2002). Identification and characterization of a sphingolipid Δ4-desaturase family. *Journal of Biological Chemistry*, *277*(28), 25512-25518.

Tomazic, M. L., Najle, S. R., Nusblat, A. D., Uttaro, A. D., & Nudel, C. B. (2011). A novel sterol desaturase-like protein promoting dealkylation of phytosterols in Tetrahymena thermophila. *Eukaryotic Cell*, *10*(3), 423-434.

Tomazic, M. L., Poklepovich, T. J., Nudel, C. B., & Nusblat, A. D. (2014). Incomplete sterols and hopanoids pathways in ciliates: Gene loss and acquisition during evolution as a source of biosynthetic genes. *Molecular Phylogenetics and Evolution*, *74*, 122-134.

Tornabene, T. G., & Langworthy, T. A. (1979). Diphytanyl and dibiphytanyl glycerol ether lipids of methanogenic archaebacteria. *Science*, *203*(4375), 51-53.

Uttaro, A. D. (2006). Biosynthesis of polyunsaturated fatty acids in lower eukaryotes. *IUBMB life*, *58*(10), 563-571.

Wang, H., Klein, M. G., Zou, H., Lane, W., Snell, G., Levin, I., ... & Sang, B. C. (2015). Crystal structure of human stearoyl-coenzyme A desaturase in complex with substrate. *Nature Structural & Molecular Biology*, *22*(7), 581-585.

Warude, D., Joshi, K., &Harsulkar, A. (2006). Polyunsaturated fatty acids: biotechnology. *Critical Reviews in Biotechnology*, *26*(2), 83-93.

Williams, T. A., Foster, P. G., Cox, C. J., &Embley, T. M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. Nature, 504(7479), 231-236.

Xiong, J., Wang, G., Cheng, J., Tian, M., Pan, X., Warren, A., ...& Miao, W. (2015). Genome of the facultative scuticociliatosis pathogen Pseudocohnilembus persalinus provides insight into its virulence through horizontal gene transfer. *Scientific Reports*, *5*.

Xiong, J. and Miao, W. (2016). Genome Database and Website of Euplotes octocarinatus. http://ciliates.ihb.ac.cn/database/species/eo (accessed 15.05.17).

Zhu, G., Koszelak-Rosenblum, M., Connelly, S. M., Dumont, M. E., & Malkowski, M. G. (2015). The crystal structure of an integral membrane fatty acid α-hydroxylase. *Journal of Biological Chemistry*, *290*(50), 29820-29833.

Figures and tables legends:

Figure 1: IMHME belonging to fatty acid hydroxylase and desaturase superfamilies. **A.** Scheme of the dimetal center showing the locations of the coordinating His (red) and Asn (yellow) residues in conserved motifs of the fatty acid α-hydroxylase (green) and the mammalian stearoyl-CoA desaturase (blue). **B.** Topology diagram of both types of enzymes showing the transmembrane domains, the residues involved in metal coordination and the short (FAH) and large (FAD) sequence segments. **C.** Schematic representation of enzymes belonging to fatty acid hydroxylase superfamily (green) and the fatty acid desaturasesuperfamily (blue), based on phylogenetic analyses.

Figure 2: Numbers of enzymes belonging to FAH and FAD superfamilies (green and blue ovals respectively) in representative organisms of the tree of life (two primary domains hypotheses) based on phylogenomic analyses of He et al. 2014; Hug et al. 2016; Keeling 2013 and Williams et al. 2013.

Figure 3: A. Scheme of the Ciliophora phylum based on phylogenetic analyses from Gao et al. 2016. Green and blue ovals show the numbers of enzymes belonging to FAH and FAD superfamilies respectively. **B.** Phylogenetic analysis and ortholog groups classification of ciliates IMHME. The phylogenetic tree of the 118 amino acid sequences selected was created using maximum-likelihood method with 100 bootstrap replicates using PhyML software (Guindon et al., 2010). The VT model of amino acid substitution was used to analyze amino acid sequence evolution, calculated with the Protest model evolution software (Darriba et al. 2011). The parameters proportion of invariant sites and across-site variation (gamma) were estimated and used in the analysis. The bar indicates percentage of substitutions. Sequences in bold have experimental evidence.

Figure 4: A. Recovery of 7-dehydrocholesterol from *P. tetraurelia* cultures grown with stigmasterol plus lathosterol or cholesterol added. **B. I**. Fatty acids content of *O. trifallax* (devoid from their food) and the content in the mixture of algae and bacteria. **II.** Composition of saturated and unsaturated fatty acids between both samples. **C.** Main fatty acids content of *C. reniformis* feeded with *E. coli* cells. The average of three independent (A, B and C) experiments are shown. The following symbols were used to represent p values: ** $p < 0.01$ and *** $p < 0.001$.

Figure 5: Phylogenetic analysis of ciliated sequences belonging to: the orthologous group OG5_156889 named acylamide-delta-3(E)-desaturase (A), the orthologous groups OG5_156889 and OG5_249793 named alkane-1 monooxygenase (B), front-end fatty acid desaturases (C) and methyl-end fatty acid desaturases (D). The phylogenetic trees were created using maximum-likelihood method with 100 bootstrap replicates with PhyML software (Guindon et al., 2010). The LG (A, B and C) and Blosum62 (D) models of amino acid substitution were used to analyze amino acid sequence evolution, calculated with the Protest model evolution software (Darriba et al. 2011). The proportion of invariant sites and across-site variation (Gamma)

were estimated and used in the analysis. The C-5 sterol desaturase from *T. thermophila* (Q22AK3) was used as outgroup in A, the fatty acid 2-hydroxylase from *T. thermophila* (Q23PP8) was used as outgroup in B and D; while the omega 6 fatty acid desaturase from *B. juncea* was used as outgroup in D. Ciliates sequences are shaded in light grey and those identified with experimental evidence in dark gray. The bars indicate percentage of substitutions. Taxonomic abbreviaton of the eukaryotic organisms shown in the AlkBtree: Ap Apusomonada; Ch, Ar Chlorophytes, Archaeplastida; Cy, Ar Cyanidiophytes, Archaeplastida; Fi, Op Filasterea, Opisthokonta; Ha, Ha: Haptophytes, Hacrobians; He, Ex: Heterolobosea, Excavata.

Figure 6: Fatty acid desaturases in ciliates. The schematic diagram shows: gene identification with experimental evidence (black circle), putative gene identification by bioinformatic analysis (grey circle), gene absence by bioinformatic analysis (white circle) and measured enzyme activity (asterisk) in the Oligohymenophorea and Spirotrichea classes. The evolutionary relationships of ciliates are based on phylogenetic analyses from Gao et al. 2016.

Figure S1: Examples of enzymes with their substrates and products belonging to FAD (stearoyl-CoA 9 desaturase, delta-12 fatty acid desaturase, delta-6 fatty acid desaturase, sphingolipid-4 desaturase, sphingolipid-3(E)-desaturase) and FAH (sterol-C5 desaturase, sterol-C4 methyl-oxidase, fatty acid-2 hydroxylase, C-24 sterol deethylase, alkane 1-monooxygenase) superfamilies.

Figure S2: A. GC-MS analysis of methyl ester derivatives of total fatty acids isolated from *C. reniformis* cells feeded with *E. coli*. The main fatty acids identified are palmitic acid (16:0), stearic acid (18:0), vaccenic acid ($18:1^{\Delta 11}$), oleic acid ($18:1^{\Delta 9}$) and palmitoleic acid ($16:1^{\Delta 9}$). **B**. Mass spectra of methyl oleate (*cis*-9-octadecenoate) in which the molecular ion (*m/z* = 296) and ions representing loss of the elements of methanol (m/z = 264 or [M-32]+), the loss of the McLafferty ion (m/z = 222) as well as the McLafferty ion per se (m/z = 74) were identified. The compounds were identified with the National Institute of Standards and Technology library.

Figure S3: HPLC analysis of sterols extracted from *P. tetraurelia* culures grown with stigmasterol plus lathosterol, cholesterol or ethanol. For quantification, β-sitosterol (5α-cholest-5-en-24β-ethyl-3β-ol) was added in all cases as an internal standard. Absorbance was recorded at 210 nm for all sterols and 285 nm for sterols displaying conjugated double bonds (5,7-diene derivatives).

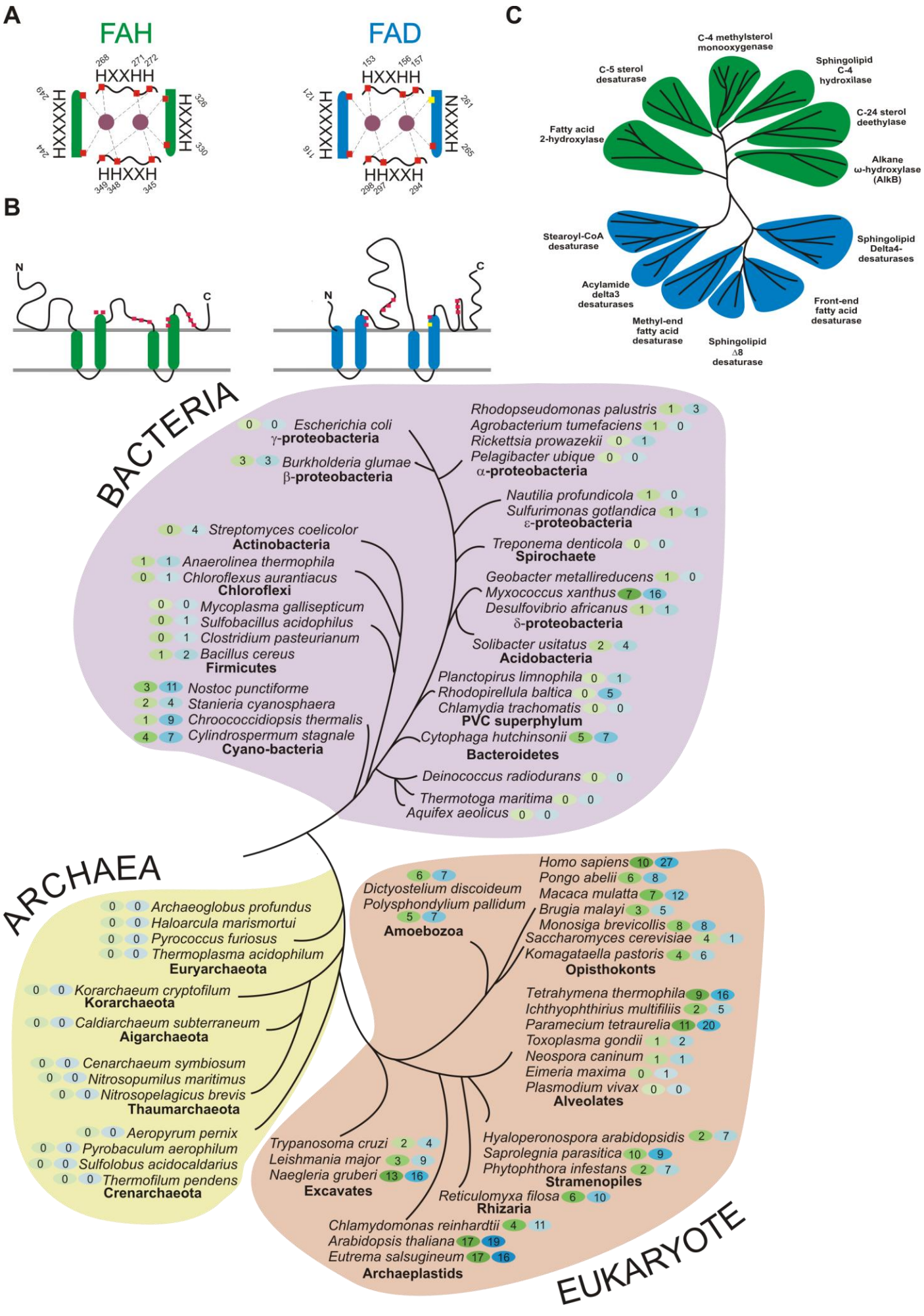
Table 1: Orthologous groups classification of the 118 putative sequences of ciliates. The pool of putative sequences was classified in 16 orthologous groups from which 11 were related to FAD and 5 to FAH superfamilies.

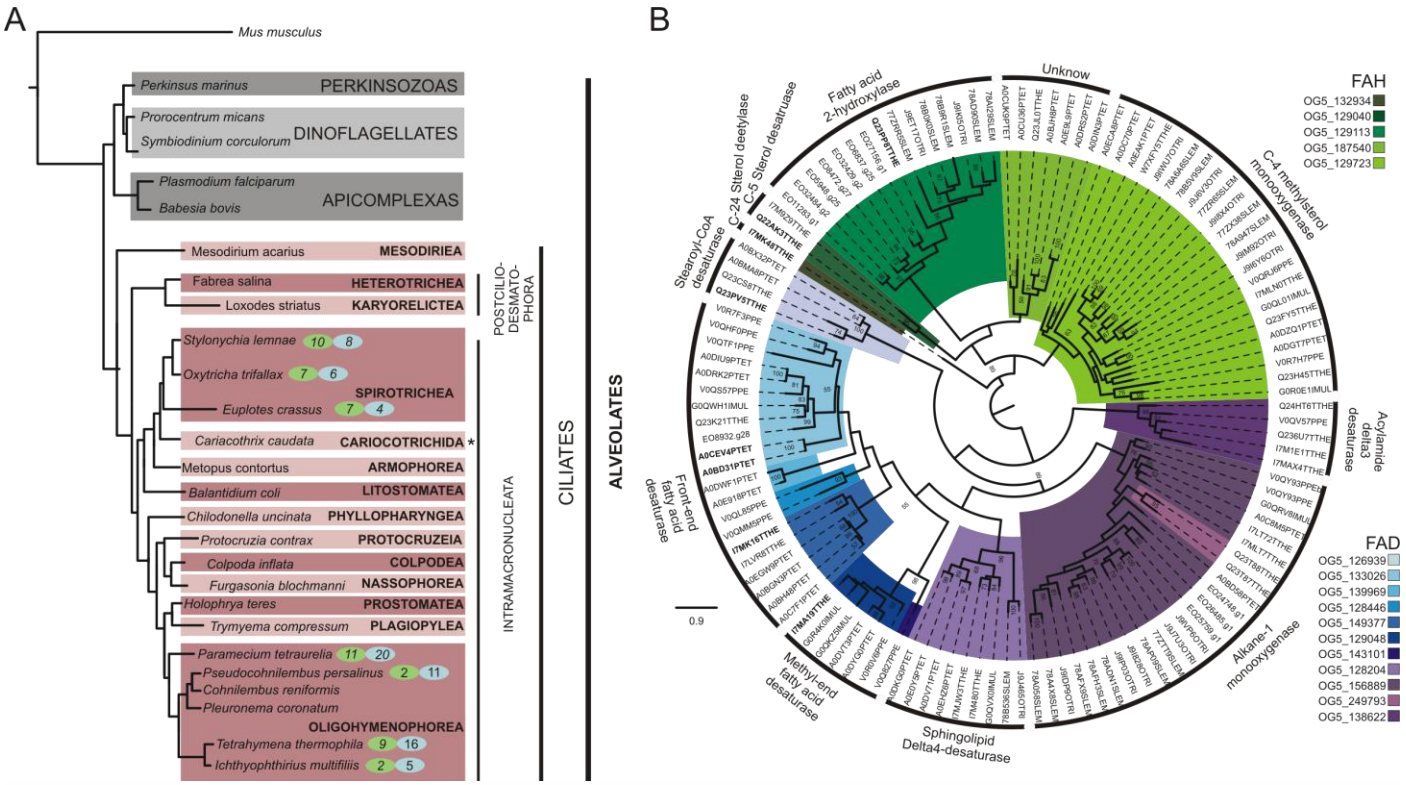
Table 2: Genetic variation of selected IMHME putative genes based on the analysis of dN/dS ratio and gene expression profiles. dN/dS ratios were calculated as described in Materials and Methods. Gene expression peak shows the stage of máximum expression. Note: The asterisks indicate no gene expression at different stages available. In parentheses, expression values normalized (Signal Intensity, AU) retrieved from the Tetrahymena Functional Genomics Database, the ParameciumDB, the Euplotes octocarinatus Genome Database and the Pseudocohnilembus persalinus Genome Database.

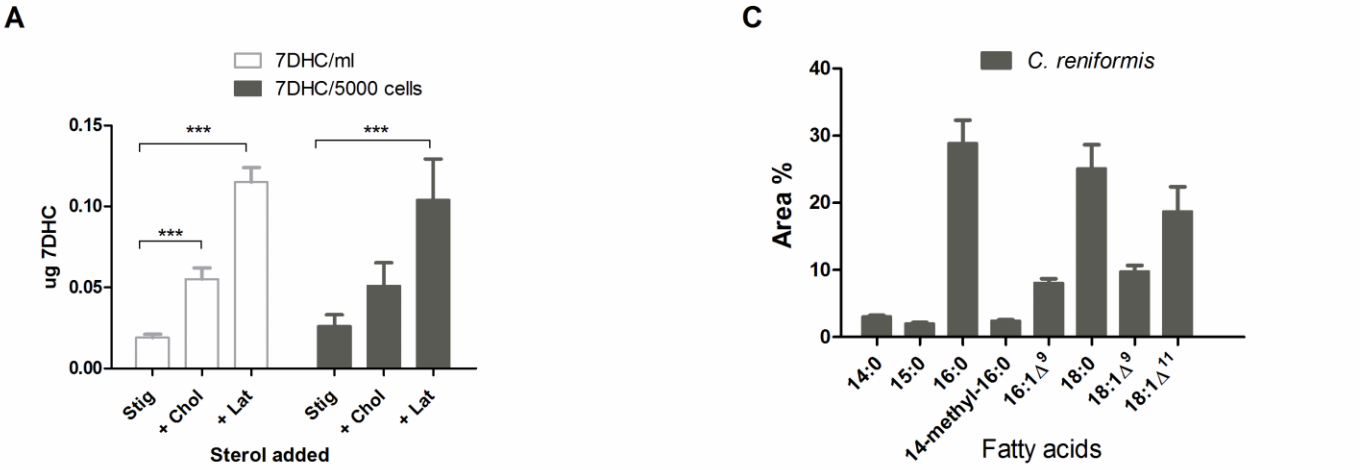
Table S1: Number of enzymes belonging to FAD and FAH superfamilies of representative organisms of the tree of life including 26 main lineages of eukaryotes, 31 representatives of eubacteria and 13 of archaea.

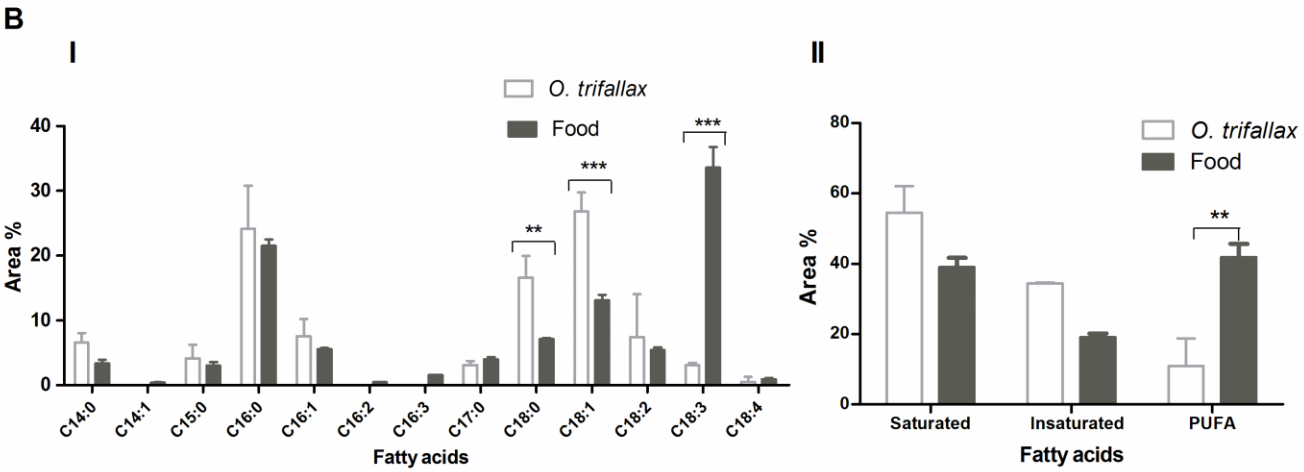
Table S2: Detail of the 118 ciliated IMHME putative sequences selected from the genomes of *E. octocarinatus, I. multifiliis, O. trifallax, S. lemnae, T. thermophila, P. tetraurelia* and *P. persalinus*. 70 sequences were structurally linked to the FAD superfamily whereas 48 showed the typical features of the FAH superfamily.

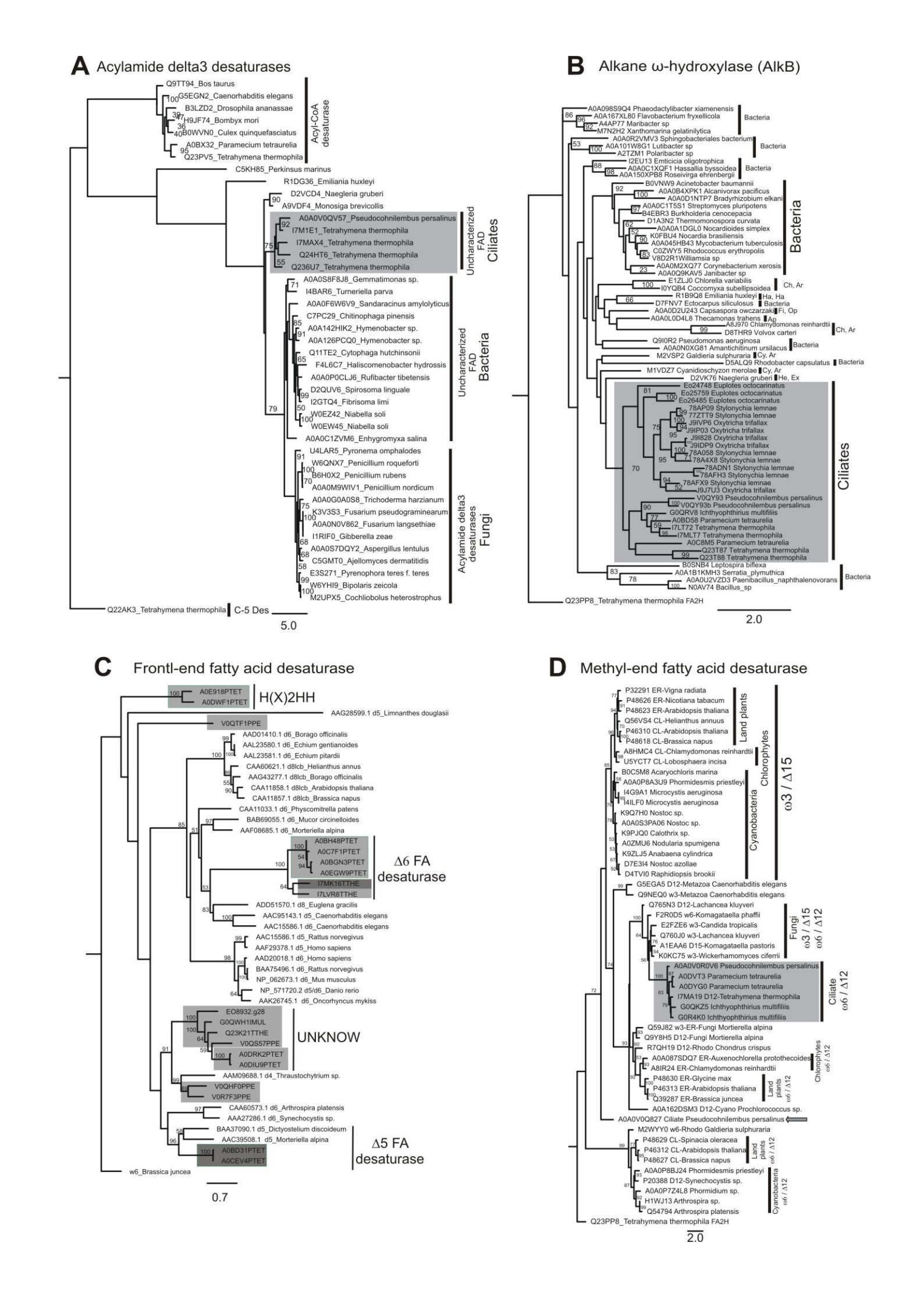
Table S3: Detail of the 16 IMHME putative sequences selected from the genomes of Heterotrich *S coeruleus*. 2 sequences were structurally linked to the FAD superfamily whereas 14 showed the typical features of the FAH superfamily. The FAD sequences fall in only one orthologous group while the FAH sequences fall in 3 different orthologous groups.


A
FAH
HXXHH
HXXXH
HXXH
HHXXH
FAD
HXXHH
HXXXH
HXXN
HHXXH
B
N
C
C-4 methylsterol monooxygenase
C-5 sterol desaturase
Sphingolipid C-4 hydroxilase
Fatty acid 2-hydroxylase
C-24 sterol deethylase
Alkane ω-hydroxylase (AlkB)
Stearoyl-CoA desaturase
Sphingolipid Delta4-desaturases
Acylamide delta3 desaturases
Front-end fatty acid desaturase
Methyl-end fatty acid desaturase
Sphingolipid Δ8 desaturase
BACTERIA
Escherichia coli
γ-proteobacteria
Burkholderia glumae
β-proteobacteria
Rhodopseudomonas palustris
Agrobacterium tumefaciens
Rickettsia prowazekii
Pelagibacter ubique
α-proteobacteria
Nautilia profundicola
Sulfurimonas gotlandica
ε-proteobacteria
Treponema denticola
Spirochaete
Geobacter metallireducens
Myxococcus xanthus
Desulfovibrio africanus
δ-proteobacteria
Solibacter usitatus
Acidobacteria
Planctopirus limnophila
Rhodopirellula baltica
Chlamydia trachomatis
PVC superphylum
Cytophaga hutchinsonii
Bacteroidetes
Deinococcus radiodurans
Thermotoga maritima
Aquifex aeolicus
Streptomyces coelicolor
Actinobacteria
Anaerolinea thermophila
Chloroflexus aurantiacus
Chloroflexi
Mycoplasma gallisepticum
Sulfobacillus acidophilus
Clostridium pasteurianum
Bacillus cereus
Firmicutes
Nostoc punctiforme
Stanieria cyanosphaera
Chroococcidiopsis thermalis
Cylindrospermum stagnale
Cyano-bacteria
ARCHAEA
Archaeoglobus profundus
Haloarcula marismortui
Pyrococcus furiosus
Thermoplasma acidophilum
Euryarchaeota
Korarchaeum cryptofilum
Korarchaeota
Caldiarchaeum subterraneum
Aigarchaeota
Cenarchaeum symbiosum
Nitrosopumilus maritimus
Nitrosopelagicus brevis
Thaumarchaeota
Aeropyrum pernix
Pyrobaculum aerophilum
Sulfolobus acidocaldarius
Thermofilum pendens
Crenarchaeota
EUKARYOTE
Dictyostelium discoideum
Polysphondylium pallidum
Amoebozoa
Homo sapiens
Pongo abelii
Macaca mulatta
Brugia malayi
Monosiga brevicollis
Saccharomyces cerevisiae
Komagataella pastoris
Opisthokonts
Tetrahymena thermophila
Ichthyophthirius multifiliis
Paramecium tetraurelia
Toxoplasma gondii
Neospora caninum
Eimeria maxima
Plasmodium vivax
Alveolates
Trypanosoma cruzi
Leishmania major
Naegleria gruberi
Excavates
Hyaloperonospora arabidopsidis
Saprolegnia parasitica
Phytophthora infestans
Stramenopiles
Reticulomyxa filosa
Rhizaria
Chlamydomonas reinhardtii
Arabidopsis thaliana
Eutrema salsugineum
Archaeplastids


A
Mus musculus
Perkinsus marinus
PERKINSOZOAS
Prorocentrum micans
Symbiodinium corculorum
DINOFLAGELLATES
Plasmodium falciparum
Babesia bovis
APICOMPLEXAS
Mesodirium acarius
MESODIRIEA
Fabrea salina
HETEROTRICHEA
Loxodes striatus
KARYORELICTEA
POSTCILIO-DESMATO-PHORA
Stylonychia lemnae 10 8
Oxytricha trifallax 7 6
SPIROTRICHEA
Euplotes crassus 7 4
Cariacothrix caudata
CARIOCOTRICHIDA *
Metopus contortus
ARMOPHOREA
Balantidium coli
LITOSTOMATEA
Chilodonella uncinata
PHYLLOPHARYNGEA
Protocruzia contrax
PROTOCRUZEIA
Colpoda inflata
COLPODEA
Furgasonia blochmanni
NASSOPHOREA
Holophrya teres
PROSTOMATEA
Trymyema compressum
PLAGIOPYLEA
Paramecium tetraurelia 11 20
Pseudocohnilembus persalinus 2 11
Cohnilembus reniformis
Pleuronema coronatum
OLIGOHYMENOPHOREA
Tetrahymena thermophila 9 16
Ichthyophthirius multifiliis 2 5
INTRAMACRONUCLEATA
CILIATES
ALVEOLATES
B
Unknow
FAH
OG5_132934
OG5_129040
OG5_129113
OG5_187540
OG5_129723
C-4 methylsterol monooxygenase
Acylamide delta3 desaturase
FAD
OG5_126939
OG5_133026
OG5_139969
OG5_128446
OG5_149377
OG5_129048
OG5_143101
OG5_128204
OG5_156889
OG5_249793
OG5_138622
Alkane-1 monooxygenase
Sphingolipid Delta4-desaturase
Methyl-end fatty acid desaturase
Front-end fatty acid desaturase
Stearoyl-CoA desaturase
C-24 Stterol deetylase
C-5 Sterol desaturase
Fatty acid 2-hydroxylase
0.9


A
7DHC/ml
7DHC/5000 cells
ug 7DHC
0.15
0.10
0.05
0.00
***
***
***
Stig
+ Chol
+ Lat
Stig
+ Chol
+ Lat
Sterol added
C
C. reniformis
Area %
40
30
20
10
0
14:0
15:0
16:0
14-methyl-16:0
16:1Δ9
18:0
18:1Δ9
18:1Δ11
Fatty acids


B
I
O. trifallax
Food
Area %
40
30
20
10
0
**
***
***
C14:0
C14:1
C15:0
C16:0
C16:1
C16:2
C16:3
C17:0
C18:0
C18:1
C18:2
C18:3
C18:4
Fatty acids
II
O. trifallax
Food
Area %
80
60
40
20
0
**
Saturated
Insaturated
PUFA
Fatty acids


A
Acylamide delta3 desaturases
Q9TT94_Bos taurus
G5EGN2_Caenorhabditis elegans
B3LZD2_Drosophila ananassae
H9JF74_Bombyx mori
B0WVN0_Culex quinquefasciatus
A0BX32_Paramecium tetraurelia
Q23PV5_Tetrahymena thermophila
Acyl-CoA desaturase
C5KH85_Perkinsus marinus
R1DG36_Emiliania huxleyi
D2VCD4_Naegleria gruberi
A9VDF4_Monosiga brevicollis
A0A0V0QV57_Pseudocohnilembus persalinus
I7M1E1_Tetrahymena thermophila
I7MAX4_Tetrahymena thermophila
Q24HT6_Tetrahymena thermophila
Q236U7_Tetrahymena thermophila
Uncharacterized FAD Ciliates
A0A0S8F8J8_Gemmatimonas sp.
I4BAR6_Turneriella parva
A0A0F6W6V9_Sandaracinus amylolyticus
C7PC29_Chitinophaga pinensis
A0A142HIK2_Hymenobacter sp.
A0A126PCQ0_Hymenobacter sp.
Q11TE2_Cytophaga hutchinsonii
F4L6C7_Haliscomenobacter hydrossis
A0A0P0CLJ6_Rufibacter tibetensis
D2QUV6_Spirosoma linguale
I2GTQ4_Fibrisoma limi
W0EZ42_Niabella soli
W0EW45_Niabella soli
Uncharacterized FAD Bacteria
A0A0C1ZVM6_Enhygromyxa salina
U4LAR5_Pyronema omphalodes
W6QNX7_Penicillium roqueforti
B6H0X2_Penicillium rubens
A0A0M9WIV1_Penicillium nordicum
A0A0G0A0S8_Trichoderma harzianum
K3V3S3_Fusarium pseudograminearum
A0A0N0V862_Fusarium langsethiae
I1RIF0_Gibberella zeae
A0A0S7DQY2_Aspergillus lentulus
C5GMT0_Ajellomyces dermatitidis
E3S271_Pyrenophora teres f. teres
W6YHI9_Bipolaris zeicola
M2UPX5_Cochliobolus heterostrophus
Acylamide delta3 desaturases Fungi
Q22AK3_Tetrahymena thermophila
C-5 Des
5.0
B
Alkane ω-hydroxylase (AlkB)
A0A098S9Q4 Phaeodactylibacter xiamenensis
A0A167XL80 Flavobacterium fryxellicola
A4AP77 Maribacter sp
M7N2H2 Xanthomarina gelatinilytica
A0A0R2VMV3 Sphingobacteriales bacterium
A0A101W8G1 Lutibacter sp
A2TZM1 Polaribacter sp
I2EU13 Emticicia oligotrophica
A0A0C1XQF1 Hassallia byssoidea
A0A150XPB8 Roseivirga ehrenbergii
Bacteria
B0VNW9 Acinetobacter baumannii
A0A0B4XPK1 Alcanivorax pacificus
A0A0D1NTP7 Bradyrhizobium elkani
A0A0C1T5S1 Streptomyces pluripotens
B4EBR3 Burkholderia cenocepacia
D1A3N2 Thermomonospora curvata
A0A0A1DGL0 Nocardioides simplex
K0FBU4 Nocardia brasiliensis
A0A045HB43 Mycobacterium tuberculosis
C0ZWY5 Rhodococcus erythropolis
V8D2R1Williamsia sp
A0A0M2XQ77 Corynebacterium xerosis
A0A0Q9KAV5 Janibacter sp
E1ZLJ0 Chlorella variabilis
I0YQB4 Coccomyxa subellipsoidea
Ch, Ar
R1B9Q8 Emiliania huxleyi
Ha, Ha
D7FNV7 Ectocarpus siliculosus
Bacteria
A0A0D2U243 Capsaspora owczarzaki
Fi, Op
A0A0L0D4L8 Thecamonas trahens
Ap
A8J970 Chlamydomonas reinhardtii
Ch, Ar
D8THR9 Volvox carteri
Q9I0R2 Pseudomonas aeruginosa
A0A0N0XG81 Amantichitinum ursilacus
Bacteria
M2VSP2 Galdieria sulphuraria
Cy, Ar
D5ALQ9 Rhodobacter capsulatus
Bacteria
M1VDZ7 Cyanidioschyzon merolae
Cy, Ar
D2VK76 Naegleria gruberi
He, Ex
Eo24748 Euplotes octocarinatus
Eo25759 Euplotes octocarinatus
Eo26485 Euplotes octocarinatus
78AP09 Stylonychia lemnae
77ZTT9 Stylonychia lemnae
J9IVP5 Oxytricha trifallax
J9IP03 Oxytricha trifallax
J9I828 Oxytricha trifallax
J9IDP9 Oxytricha trifallax
78A058 Stylonychia lemnae
78A4X8 Stylonychia lemnae
78ADN1 Stylonychia lemnae
78AFH3 Stylonychia lemnae
78AFX9 Stylonychia lemnae
J9J7U3 Oxytricha trifallax
V0QY93 Pseudocohnilembus persalinus
V0QY93b Pseudocohnilembus persalinus
G0QRV8 Ichthyophthirius multifiliis
A0BD58 Paramecium tetraurelia
I7LT72 Tetrahymena thermophila
I7MLT7 Tetrahymena thermophila
A0C8M5 Paramecium tetraurelia
Q23T87 Tetrahymena thermophila
Q23T88 Tetrahymena thermophila
Ciliates
B0SNB4 Leptospira biflexa
A0A1B1KMH3 Serratia_plymuthica
A0A0U2VZD3 Paenibacillus_naphthalenovorans
N0AV74 Bacillus_sp
Bacteria
Q23PP8_Tetrahymena thermophila FA2H
2.0
C
Frontl-end fatty acid desaturase
A0E918PTET
A0DWF1PTET
H(X)2HH
AAG28599.1 d5_Limnanthes douglasii
V0QTF1PPE
AAD01410.1 d6_Borago officinalis
AAL23580.1 d6_Echium gentianoides
AAL23581.1 d6_Echium pitardii
CAA60621.1 d8lcb_Helianthus annus
AAG43277.1 d8lcb_Borago officinalis
CAA11858.1 d8lcb_Arabidopsis thaliana
CAA11857.1 d8lcb_Brassica napus
CAA11033.1 d6_Physcomitrella patens
BAB69055.1 d6_Mucor circinelloides
AAF08685.1 d6_Morteriella alpina
A0BH48PTET
A0C7F1PTET
A0BGN3PTET
A0EGW9PTET
Δ6 FA desaturase
I7MK16TTHE
I7LVR8TTHE
ADD51570.1 d8_Euglena gracilis
AAC95143.1 d5_Caenorhabditis elegans
AAC15586.1 d6_Caenorhabditis elegans
AAC15586.1 d5_Rattus norvegivus
AAF29378.1 d5_Homo sapiens
AAD20018.1 d6_Homo sapiens
BAA75496.1 d6_Rattus norvegivus
NP_062673.1 d6_Mus musculus
NP_571720.2 d5/d6_Danio rerio
AAK26745.1 d6_Oncorhyncus mykiss
EO8932.g28
G0QWH1IMUL
Q23K21TTHE
V0QS57PPE
A0DRK2PTET
A0DIU9PTET
UNKNOW
AAM09688.1 d4_Thraustochytrium sp.
V0QHF0PPE
V0R7F3PPE
CAA60573.1 d6_Arthrospira platensis
AAA27286.1 d6_Synechocystis sp.
BAA37090.1 d5_Dictyostelium discoideum
AAC39508.1 d5_Morteriella alpina
A0BD31PTET
A0CEV4PTET
Δ5 FA desaturase
w6_Brassica juncea
0.7
D
Methyl-end fatty acid desaturase
P32291 ER-Vigna radiata
P48626 ER-Nicotiana tabacum
P48623 ER-Arabidopsis thaliana
Q56VS4 CL-Helianthus annuus
P46310 CL-Arabidopsis thaliana
P48618 CL-Brassica napus
Land plants
A8HMC4 CL-Chlamydomonas reinhardtii
U5YCT7 CL-Lobosphaera incisa
Chlorophytes
B0C5M8 Acaryochloris marina
A0A0P8A3U9 Phormidesmis priestleyi
I4G9A1 Microcystis aeruginosa
I4ILF0 Microcystis aeruginosa
K9Q7H0 Nostoc sp.
A0A0S3PA06 Nostoc sp.
K9PJQ0 Calothrix sp.
A0ZMU6 Nodularia spumigena
K9ZLJ5 Anabaena cylindrica
D7E3I4 Nostoc azollae
D4TVI0 Raphidiopsis brookii
Cyanobacteria
ω3 / Δ15
G5EGA5 D12-Metazoa Caenorhabditis elegans
Q9NEQ0 w3-Metazoa Caenorhabditis elegans
Q765N3 D12-Lachancea kluyveri
F2R0D5 w6-Komagataella phaffii
E2FZE6 w3-Candida tropicalis
Q760J0 w3-Lachancea kluyveri
A1EAA6 D15-Komagataella pastoris
K0KC75 w3-Wickerhamomyces ciferrii
Fungi
ω3 / Δ15
ω6 / Δ12
A0A0V0R0V6 Pseudocohnilembus persalinus
A0DVT3 Paramecium tetraurelia
A0DYG0 Paramecium tetraurelia
I7MA19 D12-Tetrahymena thermophila
G0QKZ5 Ichthyophthirius multifiliis
G0R4K0 Ichthyophthirius multifiliis
Ciliate
ω6 / Δ12
Q59J82 w3-ER-Fungi Mortierella alpina
Q9Y8H5 D12-Fungi Mortierella alpina
R7QH19 D12-Rhodo Chondrus crispus
A0A087SDQ7 ER-Auxenochlorella protothecoides
A8IR24 ER-Chlamydomonas reinhardtii
Chlorophytes
ω6 / Δ12
P48630 ER-Glycine max
P46313 ER-Arabidopsis thaliana
Q39287 ER-Brassica juncea
Land plants
ω6 / Δ12
A0A162DSM3 D12-Cyano Prochlorococcus sp.
A0A0V0Q827 Ciliate Pseudocohnilembus persalinus
M2WYY0 w6-Rhodo Galdieria sulphuraria
P48629 CL-Spinacia oleracea
P46312 CL-Arabidopsis thaliana
P48627 CL-Brassica napus
Land plants
ω6 / Δ12
A0A0P8BJ24 Phormidesmis priestleyi
P20388 D12-Synechocystis sp.
A0A0P7Z4L8 Phormidium sp.
H1WJ13 Arthrospira sp.
Q54794 Arthrospira platensis
Cyanobacteria
ω6 / Δ12
Q23PP8_Tetrahymena thermophila FA2H
2.0

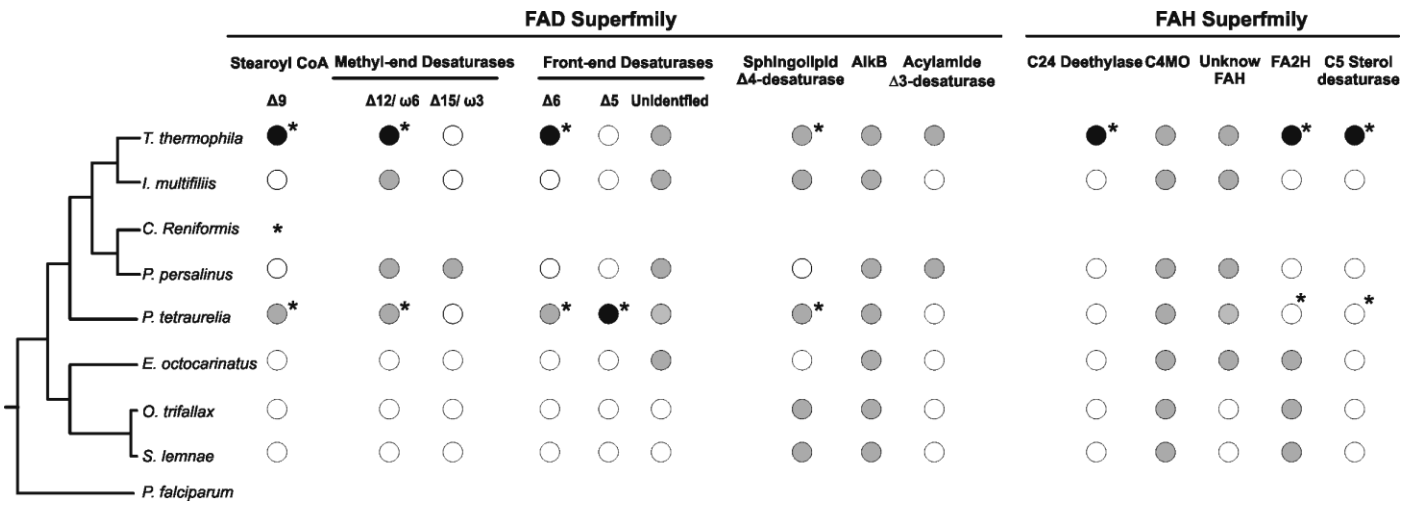


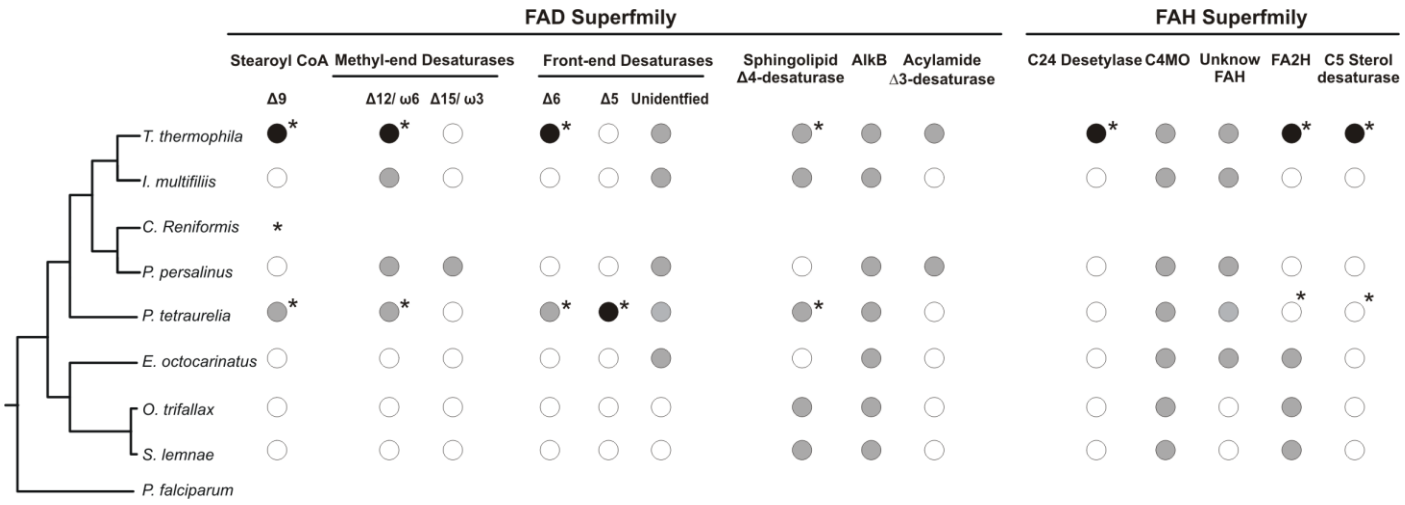
TABLE 1

Superfamily	OrthoMCL Detail	OrthoMCL Group	Organism (N°seq)	
			Spirotrichea	Oligohymenophorea
FAD	Stearoyl-CoA 9-desaturase	OG5_126939		*T. thermophila* (2)
				P. tetraurelia (2)
	FA Desaturase / Cyt b5-like Heme/Steroid binding domain	OG5_133026	*E. octocarinatus* (1)	*T. thermophila* (1)
				I. multifiliis (1)
				P. persalinus (4)
				P. tetraurelia (4)
	FA Desaturase / Cyt b5-like Heme/Steroid binding domain	OG5_139969		*P. tetraurelia* (2)
	FA Desaturase / Cyt b5-like Heme/Steroid binding domain	OG5_128446		*P. persalinus* (2)
	delta-6 Fatty acid desaturase	OG5_149377		*T. thermophila* (2)
				P. tetraurelia (4)
	delta-12 fatty acid desaturase	OG5_129048		*T. thermophila* (1)
				I. multifiliis (2)
				P. persalinus (1)
				P. tetraurelia (2)
	omega-3 fatty acid desaturase	OG5_143101		*P. persalinus* (1)
	Sphingolipid Delta4-desaturase (DES)	OG5_128204	*O. trifallax* (1)	*T. thermophila* (2)
			S. lemnae (1)	*I. multifiliis* (1)
				P. tetraurelia (4)
	Alkane-1 monooxygenase	OG5_156889	*E. octocarinatus* (3)	*T. thermophila* (2)
			O. trifallax (5)	*I. multifiliis* (1)
			S. lemnae (7)	*P. persalinus* (2)
				P. tetraurelia (2)

	Alkane 1-monooxygenase 1	OG5_249793		*T. thermophila (2)*
	Acylamide-delta3(E)-desaturase	OG5_138622		*T. thermophila* (4)
				P. persalinus (1)
FAH	C24 Deethylase	OG5_132934		*T. thermophila* (1)
	C5 Sterol Desaturase	OG5_129040		*T. thermophila* (2)
	Fatty acid 2 hydroxylase	OG5_129723	*E. octocarinatus* (7)	*T. thermophila* (1)
			O. trifallax (2)	
			S. lemnae (5)	
	Fatty acid hydroxylase superfamily	OG5_187540		*T. thermophila* (1)
				P. tetraurelia (6)
	C4 Methylsterol Oxidase	OG5_129113	*O. trifallax* (5)	*T. thermophila* (4)
			S. lemnae (5)	*I. multifiliis* (2)
				P. persalinus (2)
				P. tetraurelia (5)

TABLE 2

UniprotID	Gene ID	OMCL group	Protein detail	dN/dS	dN	dS	p-values	Gene expression peak
I7MK48	TTHERM_00438800	OG5_132934	C24-deetylase	0.3345	0.0281	0.0841	0.348	Starved (18000)
I7M1E1	TTHERM_00129290	OG5_138622	Δ3-acyl amide desaturase	0.0509	0.1833	3.6013	0.000	Conjugation (1200)
I7MAX4	TTHERM_00331050	OG5_138622	Δ3-acyl amide desaturase	0.0299	0.2680	8.9611	0.000	Conjugation (220)
Q236U7	TTHERM_00085010	OG5_138622	Δ3-acyl amide desaturase	0.0187	0.1726	9.2425	0.000	Conjugation (350)
Q24HT6	TTHERM_01133950	OG5_138622	Δ3-acyl amide desaturase	0.0284	0.3109	10.9426	0.000	Starved/Conjugation (180/190)
A0A0V0QV57	PPERSA_00063800	OG5_138622	Δ3-acyl amide desaturase	0.0165	0.3211	19.4438	0.000	Growth (439)*
A0DWF1	GSPATT00021010001	OG5_139969	Front-end desaturase	0.0262	0.1539	5.8638	0.800	No differential expression (1085)
A0E918	GSPATT00024516001	OG5_139969	Front-end desaturase	0.0802	0.1065	1.3284	0.300	No differential expression (4275)
I7LT72	TTHERM_00648790	OG5_156889	Alkane monooxygenase	0.0007	0.0606	82.5049	0.000	Starved (22000)
I7MLT7	TTHERM_00444260	OG5_156889	Alkane monooxygenase	0.0013	0.1063	82.4189	0.011	Starved (6000)
Q23T87	TTHERM_00666960	OG5_249793	Alkane monooxygenase	0.0205	0.1955	9.5353	0.127	Starved (1000)
Q23T88	TTHERM_00666950	OG5_249793	Alkane monooxygenase	0.0124	0.1881	15.1390	0.246	Starved (1600)
EO8932	Contig8932.g28010	OG5_133026	FAD / Cytb5-like Heme	0.0019	0.1629	85.2852	0.457	Growth (345)*
Q23CS8	TTHERM_00052620	OG5_126939	Δ9 Fatty acid desaturase	0.0087	0.1626	18.7711	0.555	Growth (19000)
Q23PV5	TTHERM_00463280	OG5_126939	Δ9 Fatty acid desaturase	0.0176	0.1493	8.4639	0.969	Growth (55000)
P69152	TTHERM_00498190	OG5_127916	Histone H4	0.0122	0.0001	0.0120	0.000	Conjugation (55000)
Q09F96	ymf77	No Ortholog	Ymf77	1.5276	0.1888	0.1236	0.000	Not available


FAD Superfmily
FAH Superfmily
Stearoyl CoA
Methyl-end Desaturases
Front-end Desaturases
Sphingolipid Δ4-desaturase
AlkB
Acylamide Δ3-desaturase
C24 Desetylase
C4MO
Unknow FAH
FA2H
C5 Sterol desaturase
Δ9
Δ12/ ω6
Δ15/ ω3
Δ6
Δ5
Unidentfied
T. thermophila
I. multifiliis
C. Reniformis
P. persalinus
P. tetraurelia
E. octocarinatus
O. trifallax
S. lemnae
P. falciparum

HIGHLIGHTS.

- High diversity of non–heme IMHME has been identified in seven ciliates genomes.
- Gene expansion, reduction and HGT events have modify the content of ciliates IMHME.
- A C-5 sterol desaturase activity in *Paramecium tetraurelia* was detected.
- A delta-9 fatty acid desaturase in *Cohnilembus reniformis* was detected.