ORIGINAL ARTICLE

Genome-wide Transcriptional Analysis of *Tetrahymena thermophila* Response to Exogenous Cholesterol

Sebastián R. Najle[a,b,1] (iD), Josefina Hernández[a,1], Eduard Ocaña-Pallarès[b], Nicolás García Siburu[a], Alejandro D. Nusblat[c], Clara B. Nudel[c], Claudio H. Slamovits[d] & Antonio D. Uttaro[a]

a Facultad de Ciencias Bioquímicas y Farmacéuticas, Instituto de Biología Molecular y Celular de Rosario, CONICET, Universidad Nacional de Rosario, Ocampo y Esmeralda s/n, S2000FHQ, Rosario, Argentina
b Institut de Biología Evolutiva (CSIC-Universitat Pompeu Fabra), Pg. Marítim de la Barceloneta 37-49, 08003, Barcelona, Catalonia, Spain
c Facultad de Farmacia y Bioquímica, Instituto de Nanobiotecnología (NANOBIOTEC), CONICET, Universidad de Buenos Aires, Junín 956, C1113AAD, Buenos Aires, Argentina
d Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, B3H 4R2, Nova Scotia, Canada

ABSTRACT

The ciliate *Tetrahymena thermophila* does not require sterols for growth and synthesizes pentacyclic triterpenoid alcohols, mainly tetrahymanol, as sterol surrogates. However, when sterols are present in the environment, *T. thermophila* efficiently incorporates and modifies them. These modifications consist of desaturation reactions at positions C5(6), C7(8), and C22(23), and de-ethylation at C24 of 29-carbon sterols (i.e. phytosterols). Three out of four of the enzymes involved in the sterol modification pathway have been previously identified. However, identification of the sterol C22 desaturase remained elusive, as did other basic aspects of this metabolism. To get more insights into this peculiar metabolism, we here perform a whole transcriptome analysis of *T. thermophila* in response to exogenous cholesterol. We found 356 *T. thermophila* genes to be differentially expressed after supplementation with cholesterol for 2 h. Among those that were upregulated, we found two genes belonging to the long spacing family of desaturases that we tentatively identified by RNAi analysis as sterol C22 desaturases. Additionally, we determined that the inhibition of tetrahymanol synthesis after supplementation with cholesterol occurs by a transcriptional downregulation of genes involved in squalene synthesis and cyclization. Finally, we identified several uncharacterized genes that are likely involved in sterols transport and signaling.

STEROLS are usually regarded as a hallmark of eukaryotic cells. These terpenoid lipids play structural roles in modulating membrane fluidity and permeability and are also involved in key eukaryotic cellular processes, such as phagocytosis and membrane trafficking (Chang et al. 2006). Moreover, sterols play important roles in signaling, either by participating in the formation of lipid rafts or as metabolic precursors of steroid hormones and vitamins. Thus, sterols are considered essential components of most eukaryotic membranes. However, it is known that some eukaryotic species, that are only distantly related to one another, including some ciliates, the anaerobic/microaerophilic excavates *Sawyeria marylandensis*, *Stygiella* (formerly *Andalucia*) *incarcerata,* and

Paratrimastix (formerly *Trimastix*) *pyriformis*, the fungus *Piromyces* sp. and the polychaete worm *Alvinella pompejana*, do not harbor sterols in their membranes (Takishita et al. 2012). Instead, those species produce tetrahymanol, a pentacyclic triterpenoid that can be synthesized in the absence of molecular oxygen (Bloch 1965).

Tetrahymanol was first described in the 1960s in the ciliate "*Tetrahymena pyriformis*" (Mallory et al. 1963). Since then, ciliates from this genus, *Tetrahymena*, have been studied for their capacity to take up different sterols from the growth medium, replacing tetrahymanol in the ciliates' membranes (Conner and Ungar 1964; Conner et al. 1971). Depending on the sterol species, the lipid moiety can be

directly packed or modified before its incorporation into the membrane (Conner et al. 1969; Mallory and Conner 1971; Mallory et al. 1968).

Modifications introduced in the incorporated sterols by *Tetrahymena* sp. include desaturations at the C5(6), C7(8), and C22(23) positions in the sterol structure (Conner et al. 1969; Mallory et al. 1968). In addition, 29-carbon (29C) phytosterols, like β-sitosterol or stigmasterol, are modified by removal of the ethyl group at C24, producing 7,22 *bis*-dehydrocholesterol as the end product (Mallory and Conner 1971). These enzymatic activities were shown to be associated with the endoplasmic reticulum and are induced by the presence of sterols in the culture media (Nusblat et al. 2005). The identity of the enzymes responsible for such modifications remained obscure for many years. Recently, by using reverse genetics approaches, three genes involved in the sterol modifications performed by *Tetrahymena thermophila* were identified. Two of them encode enzymes belonging to the fatty acid hydroxylases/desaturases superfamily (FAHD) and were characterized, respectively, as the sterol C5 desaturase (DES5) and sterol C24 de-ethylase (DES24) (Nusblat et al. 2009; Tomazic et al. 2011). Also, a Rieske-type monooxygenase was identified as the enzyme responsible for desaturation at C7(8) in the sterol molecule (DES7) (Najle et al. 2013). Nevertheless, identification of the sterol C22 desaturase, that completes the repertoire of sterol modifying enzymes predicted for this species, remains elusive. It is known that sterol C22 desaturases from many other organisms belong to the cytochrome P450 oxygenase superfamily, named ERG5 in fungi and CYP710A1 in plants (Desmond and Gribaldo 2009). Surprisingly, structural and phylogenetic analysis on the *T. thermophila*-predicted proteome suggested that none of the 44 putative P450 oxygenases belong to the canonical sterol C22 desaturase family (Fu et al. 2009). Hence, identifying a putative sterol C22 desaturase from *T. thermophila* remains challenging.

Concurrently with the incorporation and modification of sterols, tetrahymanol biosynthesis is repressed (Conner et al. 1968). Inhibition of tetrahymanol synthesis by cholesterol was first suggested to occur at two different steps in the pathway, in the steps between acetate and mevalonate and/or during squalene synthesis (Beedle et al. 1974). Later, it was stated that the presence of cholesterol inhibits the actual production of the enzyme squalene synthase (Warburg et al. 1982). Besides these seminal studies, the mechanism by which tetrahymanol production is inhibited by the presence of sterols is still unclear.

Other basic aspects of *T. thermophila*'s response to the presence of extracellular sterols also remain unknown. Recently, it was shown that sterols enter into the cells by phagocytosis (Elguero et al. 2018), but we do not know how the ciliate senses the presence of sterols, how this signal is transduced, and what is the nature of the downstream effectors. To increase our understanding of this peculiar metabolism, we here investigated the transcriptional response of *T. thermophila* to exogenous cholesterol added to the culture medium, using comparative RNA-seq.

MATERIALS AND METHODS

Cell cultures and induction experiments

Tetrahymena thermophila cells from strain CU428 (mpr1-1/mpr1-1, VII) were grown in 250 ml flasks containing 100-ml SPP medium of the following composition (weight/vol): 1% beef peptone, 0.1% yeast extract (both from Britania SA, Buenos Aires, Argentina), 0.2% glucose, and 0.003% iron citrate (Sigma-Aldrich, Saint Louis, MO). The cultures were incubated in a rotatory shaker at 150 rpm and 30 °C. Once the cultures reached the exponential phase of growth (5×10^6 cells/ml), they were divided into two aliquots. One aliquot was induced by adding 5 μg/ml cholesterol (Sigma-Aldrich), from a 5 mg/ml stock solution prepared in absolute ethanol. This corresponds to the "treated" experiment. The other aliquot was used as a control by adding an equal amount of ethanol.

RNA purification

For every biological replicate, 1.5×10^6 cells were collected by centrifugation at 360 *g* for 2 min at 4 °C and the cell pellets were immediately resuspended in 750 μl of RNAprotect cell reagent (Qiagen, Toronto, ON, Canada), chilled on ice, and stored overnight at −20 °C. Total RNA was extracted as described (Xiong et al. 2012) using the RNeasy Protect Cell Mini Kit with the Qiashredder (Qiagen). Total RNA concentrations were determined using Qubit fluorometer (Invitrogen/Thermo Fisher Scientific, Waltham, MA) and RNA integrity was verified using denaturing Tris acetate-ethylenediaminetetraacetic acid (TAE)/formamide gel electrophoresis in 1.2% agarose gels prepared with 1% TAE buffer (0.04 M Tris-acetate, 1 mM EDTA) (Sigma-Aldrich) (Masek et al. 2005).

Library preparation and Illumina RNA sequencing

Purified total RNA was sent to McGill University and the Genome Quebec Innovation Centre in Montreal, Canada. The RNA quality was determined with a Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA) to ensure a minimum of 400 ng/ml of RNA and an RNA integrity number (RIN) greater than 8.0 for reliable RNA sequencing. Libraries were prepared using the TruSeq RNA sample preparation kit (Illumina Inc., San Diego, CA). A total of six libraries resulting from three biological replicates of control and treated conditions were multiplexed in two flow cell lanes and sequenced on the Illumina HiSeq 2000 instrument with a paired-end sequencing protocol (2×100 bp). Base calls were made using the Illumina CASAVA pipeline, with base quality encoded in phred 33. The raw sequence reads have been deposited in NCBI's Gene Expression Omnibus (GEO) and are available at GEO series accession number GSE130336.

Data processing and differential gene expression analysis

Quality of raw reads was evaluated using FastQC (v0.11.5) (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and MultiQC (v0.9) (Ewels et al. 2016). Sequence reads were trimmed using Trimmomatic (v0.36) (Bolger et al. 2014) [ILLUMINACLIP:2:30:10, LEADING:25, TRAILING:25, SLIDINGWINDOW:10:30 MINLEN:50; only the nondefault parameters are specified], to remove sequencing adaptors and library-specific barcode sequences and filtered with a quality score threshold < 30. To estimate transcript abundances, the de-multiplexed reads were pseudoaligned to the *T. thermophila* reference genome (June 2014 release, available at http://ciliate.org/) using kallisto in pair-end mode, with 50 bootstrap replicates (Bray et al. 2016). Differential gene expression analysis was performed with the DESeq2 package (v1.16.1) (Love et al. 2014) implemented in R (v3.4.3). Transcript abundance estimates obtained with kallisto were imported into DESeq2 with tximport R package (Soneson et al. 2016). To assess the quality of the RNA-seq sequencing run, the regularized-logarithm (rlog) transformation was applied to the un-normalized count matrices and used to calculate Euclidean distances between samples and to perform principal component analysis (PCA) (Fig. S1). An adjusted P-value < 0.05 and log_2 fold change ≥ 1 were set as thresholds to consider genes as significantly differentially expressed between conditions.

Phylostratigraphic analysis

For phylostratigraphic analysis, we chose a taxon sampling of 31 species (see Table S1) that represents the whole of eukaryotic diversity, with an increased representation of taxa more closely related to *T. thermophila*. Proteins from *T. thermophila* and other eukaryotic species were clustered into orthogroups (i.e. groups of sequences that descend from a single gene in the last common ancestor of their corresponding species) with the OrthoFinder pipeline (v.2.1.3) (Emms and Kelly 2015), using DIAMOND (v0.9.13.114) (Buchfink et al. 2014) [-more-sensitive] for the all-vs.-all protein alignments. We then used Count (v10.04) (Csűrös 2010) to reconstruct, under a Dollo parsimony framework, the history of all the orthogroups (OG) including *T. thermophila* sequences in order to map their origin to the corresponding ancestors. For that, we used as input the OrthoFinder output file, including the information of the OG content per species (properly formatted) as well as a phylogenetic tree of the species used in the study (based on (Adl et al. 2012; Derelle et al. 2015; Gao et al. 2016; He et al. 2016)). For those differentially regulated genes that originated from the ancestor of ciliates to *T. thermophila*, we investigated whether the OGs in which they have been clustered may have been acquired in the ciliate from prokaryotes through lateral gene transfer (LGT). To do this, we used BLASTP (v2.3.0) (Altschul et al. 1990) [-evalue 1e-3] to align all the sequences of those families against a combined dataset including the

previously used eukaryotic dataset, as well as prokaryotic sequences from UniProt reference proteomes (Release 2016_02) (The UniProt Consortium, 2016). We considered that an OG potentially originated through LGT if the best scoring BLASTP hit among all the members of that OG corresponded to a prokaryotic sequence.

Protein functional annotation

Functional information about *T. thermophila* proteins was obtained with eggNOG-mapper (v1.0.3) (available at http://eggnog-mapper.embl.de) using DIAMOND as a search method against the "euk bact arch viruses databases" (Buchfink et al. 2014). The Gene Ontology (GO) terms (Ashburner et al. 2000), also obtained from eggNOG-mapper annotation, were used to perform enrichment tests for up/downregulated datasets against the whole predicted proteome with Blast2GO (v4.1.9) (Conesa et al. 2005), using Fisher's Exact Test [FDR ≤ 0.05]. We also assessed the overrepresentation of KEGG pathways (Kanehisa et al. 2012), PFAM protein domains (Bateman et al. 2004) and InterPro protein domains and features (Hunter et al. 2009) in the differentially expressed dataset using STRING (v10.5) (Szklarczyk et al. 2015) [FDR ≤ 0.05].

Real-time quantitative PCR analysis

Quantification of *DES7* mRNA levels by RT-qPCR was used to determine the timing of the maximal induction of transcription (see Results). Reverse transcription was performed in a 20-μl volume reaction with 2 μg of total RNA, which was previously treated with RQ1 RNase-Free DNase (Promega, Madison, WI), Oligo(dT)18 Primer and M-MLV Reverse Transcriptase (both from Invitrogen/ Thermo Fisher Scientific), according to the manufacturer's instructions. Real-time PCR amplification was carried out in a Mastercycler Realplex (Eppendorf, Hamburg, Germany), in 20-μl reaction tubes containing 10 ng of template cDNA, 0.5 μM of gene-specific primers (Table S2) and 13 μl of commercial mix for RT-qPCR (Biodynamics SRL, Buenos Aires, Argentina), which contains Taq DNA polymerase, dNTPs, Mg^{++} buffer, and green fluorophore. Cycling procedures were 95 °C for 2 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 20 s, and 72 °C for 20 s (measuring step). A melting curve was added as a final step, in order to ensure the specificity of the amplification. Gene expression levels were normalized to the abundance of the 17S rRNA gene as a reference, and relative quantification was performed following the modified ΔΔCt method (Pfaffl 2001). All conditions were examined in quadruplicate, each of them with two technical replicates.

Construction of vectors for dsRNA production and RNA interference by feeding

DNA fragments homologous to the target genes (TTHERM_00129290 and TTHERM_00085010) were PCR-amplified and cloned into the pCR-Blunt-TOPO II vector

(Invitrogen/Thermo Fisher Scientific) and their sequences were controlled by Sanger sequencing. The primers used are listed in Table S2. Inserts were sub-cloned into the EcoRI site of plasmid pLitmus28i (New England BioLabs, Ipswich, MA), which possesses two convergent T7 promoters flanking the cloning site, for production of double-stranded RNA (dsRNA). These vectors were used to transform the RNAse-III-deficient *Escherichia coli* strain HT115 (F-, *mcrA*, *mcrB*, IN [*rrnD-rrnE*]1, lambda-, *rnc4*:: *Tn10* [DE3 lysogen: *lacUV5* promotor-T7 polymerase]), used to feed the ciliates (Timmons and Fire 1998). Bacteria were grown in Luria-Bertani broth to reach an OD600 of about 0.8 units, as described previously (Najle et al. 2013) with slight modifications. Production of dsRNA was induced by adding IPTG to a final concentration of 0.5 mM. After 4 h of induction, bacteria were collected and re-suspended in glucose-depleted modified SPP medium (1% beef peptone, 0.1% yeast extract, and 0.003% iron citrate, wt/vol) and subsequently used to feed *T. thermophila* KO640, a *DES7* mutant strain (Najle et al. 2013). The use of this mutant simplifies the analysis of sterol profiles, as it converts cholesterol (cholesta-5-en-3β-ol) into 22-dehydrocholesterol (cholesta-5,22-dien-3β-ol) as the only unsaturated product. IPTG was also added to the modified SPP medium to ensure the continued production of dsRNA by the bacteria. After two successive days of feeding with the induced bacteria, *Tetrahymena* cultures were supplemented with 10 μg/ml cholesterol and fed bacteria for another two days. As a control, cells were fed with induced bacteria transformed with the empty vector (without insert). On the 5th day, cultures were harvested, lipids were extracted, and their sterol profiles were analyzed by GC/MS.

Identification of sterols by GC/MS

Cells from cultures with added sterol were collected by centrifugation at 3,000 *g* for 5 min at 4 °C, washed twice with 20 ml of distilled water, and the lipids were extracted according to (Bligh and Dyer 1959). The organic phase was evaporated to dryness under an N_2 stream, and the lipids were saponified with 10% KOH in methanol for 40 min at 60 °C. After a 2-fold extraction with 2-ml hexane, the organic solvent was evaporated under the N_2 stream, and the residue was re-suspended in 50 μl of distilled pyridine. Sterols were then derivatized by adding 100 μl of acetic anhydride and incubating the mix for 40 min at 80 °C. The composition of the steryl acetate ester derivatives was analyzed by running samples through an SPB-1 column (L × I.D. 30 m × 0.25 mm, d_f 0.25 μm; Supelco/Sigma-Aldrich) in a Shimadzu GC2010 Plus gas chromatograph. The column was temperature programed at 5 °C/min from 160 to 320 °C and subsequently held for 10 min at 320 °C. Mass spectra were obtained with a GCMS-QP2010 Plus mass detector, operated at an ionization voltage of 70 eV with a scan range of 20–600 atomic mass units. The retention times and mass spectra of all peaks were compared with those of authentic standards (Sigma-Aldrich) and those available in the National Institute of Standards and Technology (NIST) mass spectral library.

RESULTS

Timing of sterol metabolism induction for RNA-seq

We had previously characterized the sterol C7 desaturase (DES7) of *T. thermophila* (Najle et al. 2013). DES7 is responsible for the one-step conversion of cholesterol into pro-vitamin D3 in the ciliate, and its expression is known to be induced by the presence of sterols in the culture media. Hence, we used the *DES7* expression profile in the presence of cholesterol as a reference to determine the time-point for our RNA-seq experiment. Quantification of *DES7* mRNA levels by real-time quantitative PCR (RT-qPCR) indicated that maximal induction of transcription was reached between two and three hours after the addition of 5 μg/ml cholesterol to a culture of *T. thermophila* at mid-exponential growth phase (Fig. 1A). *DES7* was induced 4.7 times (log2 fold change = 2.18) two hours after cholesterol supplementation. We thus sampled our cultures following an induction period of 2 h. This timing was similar to that reported for *Drosophila melanogaster*, a sterol auxotroph, in which cholesterol triggers a rapid and coordinated transcriptional response (Horner et al. 2009).

Whole-genome transcriptome profiling

We performed triplicate RNA-seq of poly-A-purified RNA samples for each *T. thermophila* growth condition: growth in conventional SPP medium (control) and SPP medium, plus cholesterol (treatment) (Material and Methods). The RNA sequencing generated a total of ~498 million (raw) reads in the 6 libraries, of which ~487 million passed trimming and quality filters. About 462 million of the filtered reads (95%) were pseudoaligned onto the *T. thermophila* reference CDS FASTA file using kallisto (Bray et al. 2016), covering 24,944 of the originally annotated 26,997 open reading frames. The quality assessment of our RNA-seq is shown in Fig. S1 (Material and Methods).

Differentially expressed genes (DEGs) were identified using DEseq2 (Love et al. 2014) (see Material and Methods for details). Genes with an absolute $\log_2$ fold change ≥ 1 and adjusted *P*-value ≤ 0.05 were classified as significantly differentially expressed. Using these criteria, we identified 356 genes that showed at least a two-fold variation in expression between the control and treated samples. Of these, 179 genes were upregulated and 177 genes were downregulated in the treated samples (Fig. 1B, C and Table S3).

Functional annotation of DEGs

In order to characterize the function of the DEGs, we annotated the corresponding protein sequences using egg-NOG-mapper (Huerta-Cepas et al. 2017) (Table S4). We
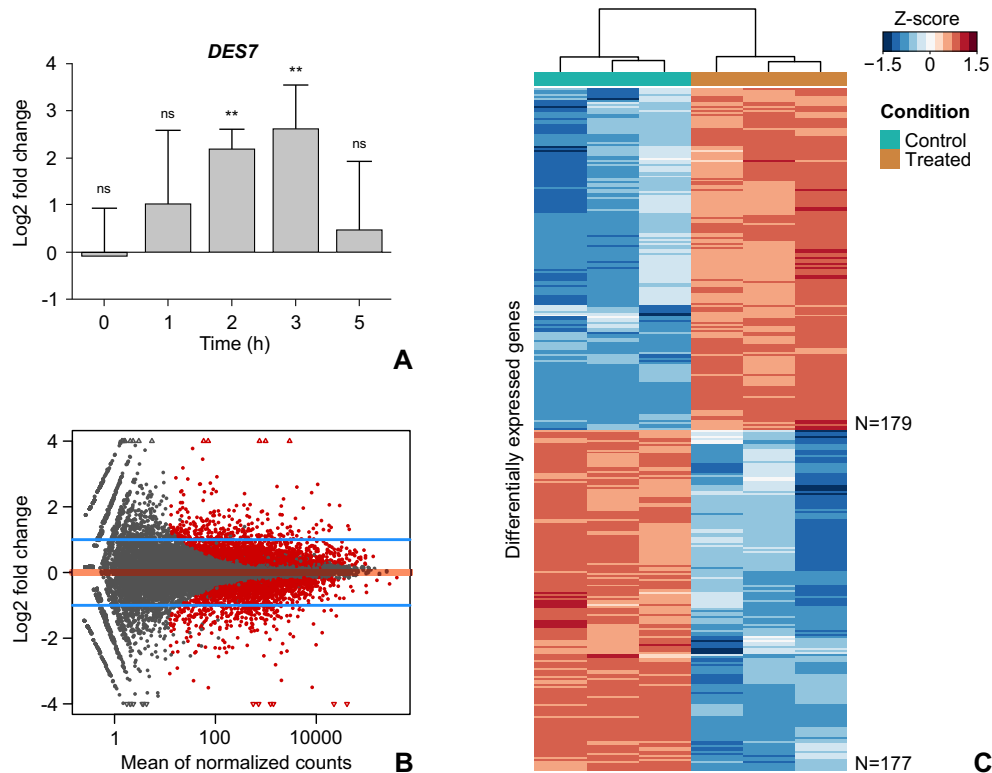
Figure 1 Experimental design and summary of RNA-seq results. **A**. Expression of *DES7* in response to cholesterol as measured by RT-qPCR. Asterisks indicate statistical significance according to a *T*-test (unpaired, one-tailed) between treatment and control without cholesterol (*P* values, *0.5, **0.01, ***0.001; *n* = 4). **B**. MA-plot showing changes in gene expression after 2-h treatment with cholesterol. The $\log_2$ fold change for treated vs. control samples is plotted on the *y*-axis and the mean of counts, normalized by size factor, is shown on the *x*-axis. Each gene is represented with a dot. Genes with an adjusted *P*-value < 0.05 are highlighted in red. Blue lines denote the fold change thresholds used as a cutoff to consider genes as differentially expressed between conditions. **C**. Heat map display of the differentially expressed genes between conditions. CONTROL: without sterols added, TREATED: with the addition of 5 µg/ml cholesterol for 2 h.

retrieved functional annotations for 244 (68.54%) out of the 356 DEGs of our dataset. This result implies that a significant proportion of DEGs had either no hit in the egg-NOG database or a hit to a nonsupervised orthologous group of unknown function (labeled "no annotation" in Table S4). Specifically, 70 out of 179 (39.10%) upregulated genes and 42 out of 177 (23.73%) of downregulated genes were not functionally annotated (Fig. 2A). These results were not surprising, given the large number of genes of unknown function in ciliates (Aury et al. 2006; Eisen et al. 2006).

Gene ontology (GO) annotations also obtained from egg-NOG-mapper were used to perform GO enrichment analysis. We retrieved significant results only for the group of downregulated genes (Fig. S2 and Table S5). These enriched GO-terms corresponded mainly to carbohydrate metabolism and to cell cycle regulation and chromosome segregation.

We also examined the enrichment of KEGG pathways, Pfam domains and InterPro protein domains using STRING (Szklarczyk et al. 2015) (Material and Methods). The details are shown in Table S6. In line with the results obtained for GO enrichment analysis, there was no

significant enrichment of Pfam or InterPro protein domains in the upregulated genes dataset. Only a small group of genes (11 genes) could be annotated with a KEGG pathway category, but these were only assigned to the general KEGG pathway category "Metabolic pathways" (KEGG ID 1100), which is not informative. On the other hand, the results obtained for the downregulated genes dataset are more significant and meaningful. KEGG pathways enriched in this dataset included ABC mediated transport, terpenoid backbone biosynthesis, steroid biosynthesis, ribosome biogenesis, biosynthesis of secondary metabolites, and glutathione metabolism. Pfam domains enriched in this dataset were restricted to ABC transporters (PF00005). Finally, InterPro domains that showed enrichment included the EGF-like domain (IPR000742), Peptidase M8—leishmanolysin (IPR001577), ABC transporter-like (IPR003439), Furin-like repeat (IPR006212), EGF-like, conserved site (IPR013032), P-loop containing nucleoside triphosphate hydrolase (IPR027417), AAA+ ATPase domain (IPR003593), Glutathione S-transferase N-terminal (IPR004045), Glutathione S-transferase, C-terminal-like (IPR010987), and Isoprenoid synthase domain (IPR008949) (Table S6).
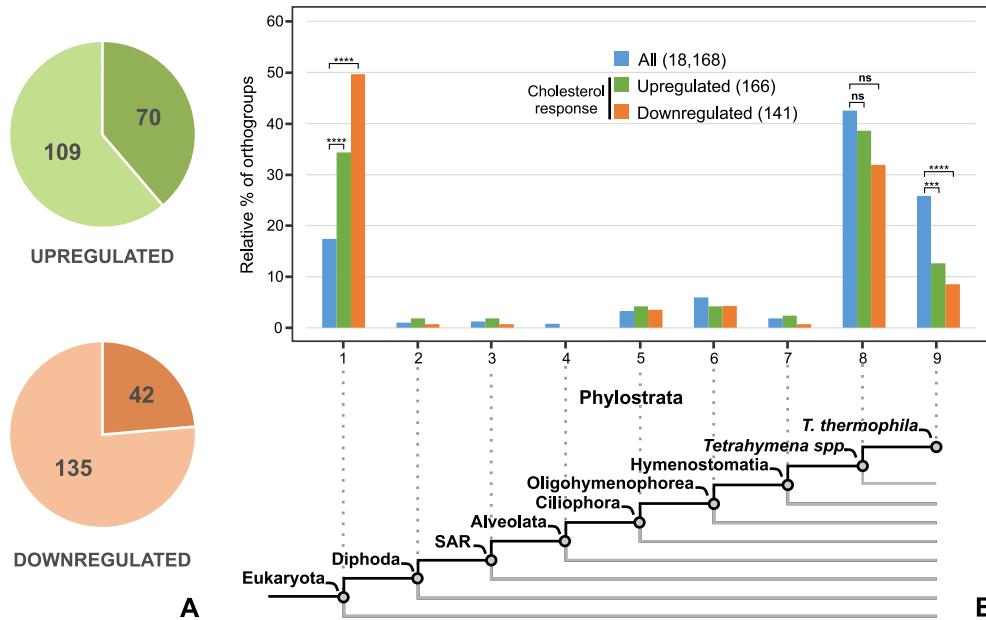
Figure 2 A considerable proportion of *Tetrahymena thermophila* sterol response genes is of unknown function. **A**. Pie charts depicting the number of DEGs with (light color) or without (dark color) hit in the eggNOG database. **B**. Phylostratigraphic distribution of the orthogroups that include at least one representative from *T. thermophila*'s proteome (*T. thermophila* OGs) in a consensus phylogeny that spans from the last eukaryotic common ancestor (LECA) to *T. thermophila*. The histogram above depicts the relative percentage of *T. thermophila* OGs that originated in each phylostrata. The total set of *T. thermophila* OGs is represented in blue. Green and orange bars specifically represent the *T. thermophila* OGs that include the upregulated and downregulated genes, respectively (****P-value < 0.0001, ***P-value < 0.001, ns = not significant).

Evolutionary origin of *Tetrahymena thermophila* cholesterol response

To estimate the age of the metabolic pathway involved in the response to exogenous sterols in *T. thermophila*, we carried out a phylostratigraphic analysis to determine at which ancestral point from the eukaryotic common ancestor to *T. thermophila* the DEGs had originated. For this purpose, we used OrthoFinder to cluster all protein sequences from *T. thermophila* and 30 other eukaryotic species into orthogroups (Table S1). Orthogroups (OGs) are constructed so that each one included sequences descended from a single ancestral gene. We used Dollo parsimony to infer the origin of each OG according to the distribution of species included in it. A total of 703,701 sequences from our eukaryotic taxon sampling were clustered into 306,738 OGs, from which 18,168 include at least one of the 26,996 sequences from *T. thermophila*'s predicted proteome (*T. thermophila* OGs). 3,160 of these *T. thermophila* OGs (17.39%) were presumably already present in the last eukaryotic common ancestor (LECA), 7,728 (42.54%) originated in the ancestor of the *Tetrahymena* genus, and 4,693 (25.83%) are exclusive of *T. thermophila*. Of the genes involved in the response to cholesterol, the 179 genes upregulated in our dataset are distributed into 166 OGs (upregulated OGs) that include 10,673 eukaryotic sequences, 1,109 from *T. thermophila*. Of these, 57 OGs (34.34%) were already present in the LECA, 64 (38.55%) are exclusively present in *Tetrahymena* genus, and 21 (12.65%) are specific to

T. thermophila. The 177 downregulated genes are distributed into 141 OGs (downregulated OGs) that include 9,850 eukaryotic sequences, 834 from *T. thermophila.* Of these, 70 OGs (49.65%) were already present in the LECA, 45 (31.91%) are found in *Tetrahymena* and 12 (8.51%) are specific to *T. thermophila*. Genes of LECA origin are significantly overrepresented in our set of DEGs. In contrast, genes specific to *T. thermophila* are underrepresented (one-tailed Fisher's exact test) (Fig. 2B).

Even though there is an overrepresentation of LECA OGs in the DEGs dataset, there is a considerable number of OGs with a more restricted taxonomic distribution. In particular, 170 *T. thermophila* OGs that include at least one DEG would have originated after the emergence of Ciliophora (103 upregulated OGs and 68 downregulated OGs, 1 OG include both upregulated and downregulated genes) (Table S7).

To better understand the evolutionary origin of these 170 *T. thermophila* OGs with DEGs of recent origin, we tested whether they have a resemblance to either prokaryotic or eukaryotic proteins outside Ciliophora (Material and Methods). Hundred and two of these OGs are likely to have originated de novo as none of them aligned with any protein outside Ciliophora. In contrast, 68 OGs have at least one member that aligned with a protein outside Ciliophora, from which 24 have the best scoring hit against prokaryotic proteins belonging to 16 bacterial and 2 archaeal species (Table S7). This suggests potential prokaryotic contributions to the origin of the *T. thermophila* cholesterol response pathway. Previous reports

have already suggested that some enzymes involved in the tetrahymanol biosynthesis pathway could have been horizontally acquired from prokaryotes (Tomazic et al. 2014).

Analysis of differentially expressed genes

Tetrahymanol biosynthetic pathway

Tetrahymanol is synthesized by direct cyclization of squalene, whose biosynthesis in *T. thermophila* follows the canonical mevalonate pathway (Conner et al. 1968; Raederstorff and Rohmer 1988). This pathway involves the conversion of mevalonate to mevalonate 5-phosphate by the mevalonate kinase and the successive action of phosphomevalonate kinase, mevalonate pyrophosphate decarboxylase, isopentenyl diphosphate isomerase, farnesyl (polyprenyl) diphosphate synthase, and squalene synthase (Fig. 3A). Genes for each of these enzymes were identified in the *T. thermophila* genome, together with the enzymes of the mevalonate synthesis pathway: thiolase, hydroxymethyl glutaryl-CoA (HMG-CoA) synthase, and HMG-CoA reductase (Tomazic et al. 2014). Interestingly, in this species, the last two enzymatic activities seem to be carried out by a single 1,364 aminoacid polypeptide encoded by the gene TTHERM_00691180 (Fig. 3C). Our results show the simultaneous downregulation of six out of the nine genes encoding these enzymes (Fig. 3B and Table S4). Squalene cyclase (also known as tetrahymanol synthase or THC1, TTHERM_01008630) and squalene synthase (SQS1, TTHERM_00382150) are, respectively, the first and second most downregulated genes in our dataset (Table S4). These results are in agreement with early biochemical data (Beedle et al. 1974; Warburg et al. 1982).

The exceptions are acetyl-CoA acyltransferase (thiolase, TTHERM_000091579), mevalonate kinase (TTHERM_00637680), and phosphomevalonate kinase (TTHERM_00371030), which are not significantly differentially expressed (Fig. 3B and Table S8). Our results indicate that mevalonate synthesis was mainly inhibited by transcriptional repression of the bifunctional enzyme HMG-CoA synthase/reductase (−4.483-fold). Repression of polyprenyl diphosphate synthase (−10.303-fold), isopentenyl diphosphate isomerase (−9.006-fold), and diphosphomevalonate decarboxylase (−3.908-fold) also appear to be important for the inhibition of squalene synthesis (Fig. 3B and Table S4). A similar inhibition effect of cholesterol on the mevalonate, squalene and/or polycyclic triterpenoid biosynthesis pathways has been reported in other organisms, including mammals (Maxwell et al. 2003), *Saccharomyces cerevisiae* (Bammert and Fostel 2000), and the filasterean amoeba *Capsaspora owczarzaki* (Najle et al. 2016).

Exogenous sterol modification pathway

As described above, the modifications carried out by *T. thermophila* on exogenous sterols include the de-ethylation of 29C phytosterols, desaturation at position C22(23) of sterol lateral chains and desaturations at C5(6) and C7(8) of the B-ring (Fig. 4A). All these activities are dependent on molecular oxygen and NAD(P)H, and with the exception of C7(8) desaturation, they require cytochrome *b5* as an electron donor (Nusblat et al. 2005). These biochemical properties are equivalent to those of the FAHD superfamily, which includes fatty acid, sphingolipid, and sterol hydroxylases and desaturases, as well as alkane monooxygenases (Uttaro 2006). These are integral
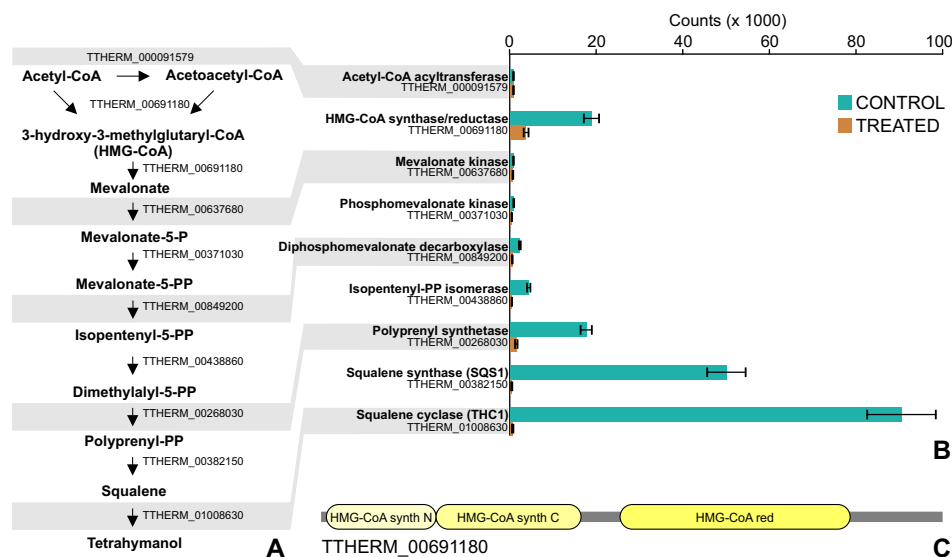

Figure 3 Expression of tetrahymanol biosynthesis pathway genes. **A**. Sequential steps in the mevalonate pathway from Acetyl-CoA to tetrahymanol, indicating the name of the intermediate products and the identity of the genes encoding the responsible enzymes. **B**. Barplot of the expression values of each gene in the different conditions (CONTROL: without sterols added, TREATED: with the addition of 5 μg/ml cholesterol for 2 h) showing the downregulation of key enzymes in the pathway. Bars show standard error. **C**. Domain architecture of the *Tetrahymena thermophila* bifunctional HMG-CoA synthase/reductase.
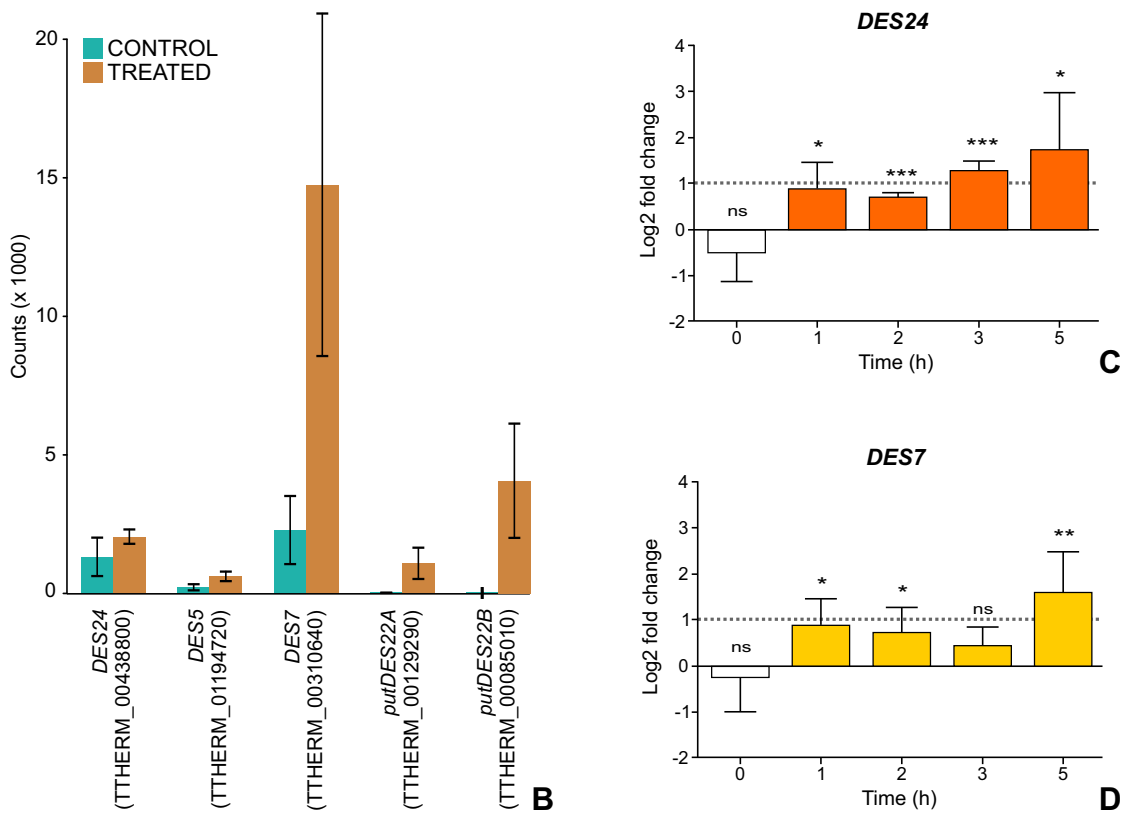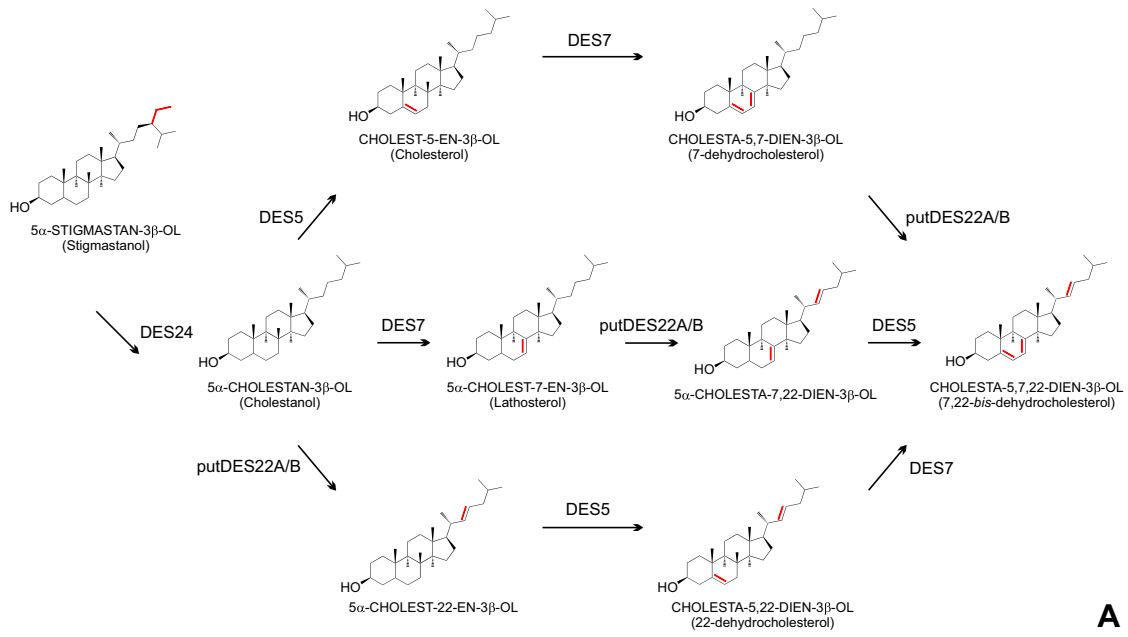
Figure 4 Expression of exogenous sterols modification pathway genes. **A**. Pathway of conversion of the saturated C29 sterol stigmastanol into 7,22-*bis*-dehydrocholesterol in *Tetrahymena thermophila* based on (Tomazic et al. 2011). **B**. Barplot of the expression values of known genes involved in the pathway (*DES24*, *DES5,* and *DES7*) and the two putative *DES22* genes (*putDES22A* and *putDES22B*, see text for details), in the different conditions (CONTROL: without sterols added, TREATED: with the addition of 5 µg/ml cholesterol for 2 h). Bars show standard error. **C** and **D**. Time-course RT-qPCR measurement of *DES24* and *DES7* expression upon addition of 5 µg/ml β-sitosterol. Grey lines indicate the threshold of $\log_2$-fold change = 1 (2-fold change) to consider genes as significantly upregulated. Statistical analysis was performed by *T*-test (unpaired, one-tailed), against a control without β-sitosterol (*P* values, *0.5, **0.01, ***0.001, ns = not significant; *n* = 4).

membrane, nonheme, di-iron containing enzymes that share three characteristic histidine boxes putatively involved in chelating the di-iron center (Uttaro 2006). Two subfamilies of FAHD can be established based on the length of spacing between the first and third histidine boxes. Short spacing desaturases subfamily includes sterol desaturases, as well as sphingolipid and steroid hydroxylases, whereas long spacing desaturases include sphingolipid and fatty acid desaturases and alkane monooxygenases (Sperling et al. 2003).

The *T. thermophila* genome encodes nine putative genes encoding short spacing desaturases. Eight of them are orthologs of sterol desaturases or sterol C4 methyl oxidases and one is an orthologue of sphingolipid hydroxylases (Nusblat et al. 2009). Two of these genes have been identified and characterized as phytosterol C24 de-ethylase (*DES24*) and sterol C5-desaturase (*DES5*), encoded respectively by genes TTHERM_00438800 and TTHERM_01194720 (Nusblat et al. 2009; Tomazic et al. 2011). Under our experimental conditions, *DES24* was significantly differentially expressed but remained below the 2-fold change threshold that we used (Table S9). However, *DES5* was included in the upregulated genes dataset (2.320 fold) (Fig. 4B and Tables S4). The four putative sterol C4-methyl oxidase orthologues showed no clear sterol modification activity on the substrates tested (Najle et al. 2013), and, appropriately, none of them was differentially expressed in our dataset (Table S9).

The fact that *DES24* was not significantly induced in our study suggests that the inducer, if any, could be its natural substrate, a 29C phytosterol-like β-sitosterol, rather than cholesterol. To test this hypothesis, we carried out an RT-qPCR quantification of *DES24* and *DES7* transcripts in RNA samples obtained from *T. thermophila* grown in the presence or absence of β-sitosterol. *DES24* expression was below the 2-fold change threshold after 2 h of treatment with β-sitosterol. However, *DES24* expression exceeded that threshold three hours posttreatment, and induction continued up to 5 h (Fig. 4C). The maximal expression of *DES7* in β-sitosterol-supplemented medium was also delayed when compared to the expression in cholesterol-containing medium, displaying an expression peak at 5 h of treatment (compare Fig. 1A and 4D). This delay in *DES7* induction could be explained by the action of DES24 on β-sitosterol, supported by the basal level of expression of *DES24* in the absence of sterols (see Fig. 4B). Upon incorporation, β-sitosterol can be first converted into cholesterol by the action of DES24, and we hypothesize that cholesterol might trigger the induction of *DES7* at a later time.

The rapid induction of *DES7* by cholesterol was confirmed by our RNA-seq analysis (6.026-fold, Fig. 4B and Table S4). DES7 belongs to a different class of oxygenases, phylogenetically or structurally unrelated to FAHD. They have a Rieske type [2Fe-2S] cluster and a nonheme iron-binding domain. During the reaction, a ferredoxin reductase receives electrons from NAD(P)H and transfers them to the Rieske domain of the oxygenase, after which they are transferred to the iron-binding domain. In three-component systems, a small ferredoxin carries the electrons from the reductase to the oxygenase (Ferraro et al. 2005; Petrusma et al. 2014). We were unable to identify orthologues of ferredoxin reductases but detected three putative ferredoxins (TTHERM_00161840, TTHERM_00256970, and TTHERM_00590010), although none of them appeared differentially regulated.

Candidate genes for sterol C22 desaturases

C22 desaturation is one of the most robust sterol desaturating activities found in *T. thermophila* and was shown to be induced by cholesterol (Nusblat et al. 2005). In fungi and plants, this activity is involved in the synthesis of ergosterol and phytosterols and is carried out by cytochrome P450 oxygenases. The cytochrome P450 oxygenase superfamily of enzymes is ubiquitously distributed in all domains of life and unrelated to FAHD or to Rieske-oxygenases. They are involved in the metabolism of a wide variety of substrates, such as steroids and other endogenous molecules, as well as xenobiotic compounds. A previous detailed analysis of the cytochrome P450 oxygenases present in *T. thermophila* (Fu et al. 2009) revealed 44 putative genes belonging to 13 families and 21 subfamilies, but none of them corresponding to the sterol C22 desaturase group. In our transcriptomic analysis, we did not find differential expression of any P450 oxygenase gene. Taken together, both observations suggest that the sterol C22 desaturase is likely to be a different type of oxygenase. With this in mind, we searched for putative oxygenases among the DEGs in our RNA-seq dataset. *T. thermophila*'s genome encodes 16 putative genes for oxygenases belonging to the long spacing subfamily of FAHD (section 3.5.2) (Cid et al. 2017). Five of them were differentially expressed (Table S4) in the presence of cholesterol: (1) TTHERM_00334320 was downregulated (−5.056 fold) and showed similarity to sphingolipid desaturases; (2) TTHERM_00052620 was induced 2.490-fold and is most probably a stearoyl-CoA desaturase, sharing 30% identity with *Saccharomyces cerevisiae* OLE1; (3) TTHERM_00339850 was induced 5.137-fold and shares 58% identity with a paralog (TTHERM_00138530) recently characterized as Δ6 linoleoyl desaturase (Dahmen et al. 2013); and (4) the last two genes, TTHERM_00085010 and TTHERM_00129290 have 56% identity, they have no clear homology to described enzymes and are the second (75.412-fold) and third (70.078-fold) most upregulated genes in our dataset, respectively. In previous studies, these two sequences were grouped together with other two paralogous genes that are not differentially expressed in our dataset (TTHERM_01133950 and TTHERM_00331050), and with fungal acylamide delta-3 desaturases (Cid et al. 2017). These enzymes, only characterized in fungi, introduce a double bond at the C-3 position of the hydroxyl-fatty acyl moiety of sphingolipids (Cid et al. 2017). To date, no delta-3 unsaturated compound, neither as the free fatty acid nor in the fatty acyl moiety of sphingolipids or triglycerides, has been found in *Tetrahymena* (Cid et al.

Najle et al.

2017), casting doubt on the possible function of the proteins encoded by these genes. Based on their notable induction by cholesterol, the apparent absence of delta-3 desaturase activity in the ciliate and the lack of sequence identity to other known desaturases, we analyzed both genes as possible candidates to encode sterol C22 desaturases: TTHERM_00129290 (putative sterol C22 desaturase A, *putDES22A*) and TTHERM_00085010 (putative sterol C22 desaturase B, *putDES22B*) (Fig. 4).

We knocked down the expression of both candidate genes by RNA interference (RNAi) by feeding. To this end, we used the method employed previously to identify *DES7* (Najle et al. 2013) (see Materials and Methods). As shown in Fig. 5, the knockdown of either of the putative *DES22* genes led to a significant reduction in 22-dehydrocholesterol production, compared with the control. These results strongly suggest that both genes encode sterol C22 desaturases.

Sterol transport and signaling

Eukaryotic cells acquire sterols via specialized transporters or lipoprotein receptors. In *Saccharomyces cerevisiae*, Pdr11p and Aus1p, two ATP-binding cassette (ABC) transporters located in the plasma membrane (PM) mediate sterols import during anaerobiosis (Li and Prinz 2004). Mammalian cells take up cholesterol mainly as cholesteryl esters, carried by low-density lipoproteins (LDL), using LDL receptors; while ABC transporters (i.e. ABCA1, ABCG1) are involved in cholesterol efflux. LDL receptors and their cargo are internalized via clathrin-mediated endocytosis. Cholesteryl esters are hydrolyzed in endosomes, and free cholesterol is attached to the luminal protein Niemann-Pick Type C Disease 2 (NPC2). In late endosomes/lysosomes, NPC2 transfers its cargo to the membrane protein NPC1, making cholesterol available for transfer to sterol binding proteins, which in turn distribute cholesterol to the PM and other organelles. Enterocytes and hepatocytes have an NPC1-related protein (NPC1L1)
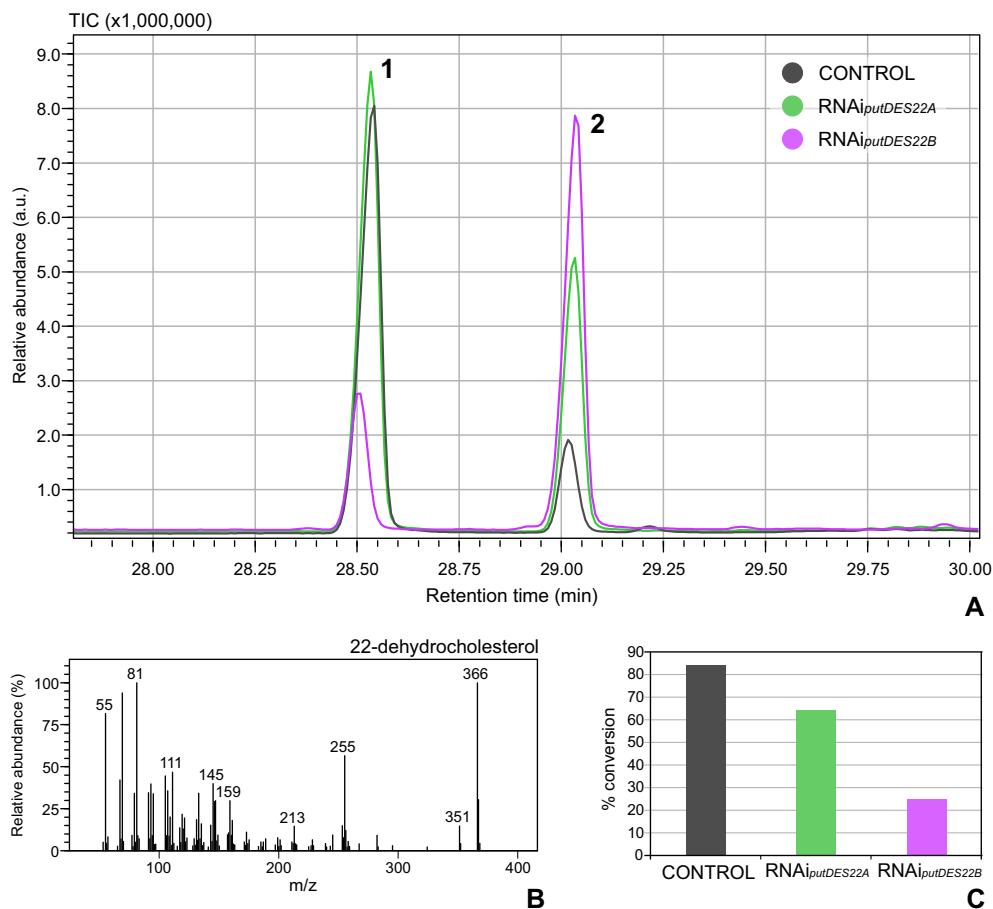

Figure 5 Inhibition of sterol C22-desaturase activity by RNA interference. **A.** GC-MS total ion chromatograms of acetylated sterols obtained from *Tetrahymena thermophila* KO640 strain (Najle et al. 2013) cultured for 48 h in SPP medium supplemented with 10 µg/ml cholesterol. The ciliates were feed with bacteria harboring an empty vector (CONTROL, dark grey) or vectors driving the production of dsRNA homologous to genes TTHERM_00129290 (*putDES22A*, green) or TTHERM_00085010 (*putDES22B*, purple). Feeding with dsRNA-producing bacteria was started 48 h before cholesterol addition. The product of cholesterol (peak **2**) bioconversion (22-dehydrocholesterol, peak **1**) was identified by its mass spectrum (**B**). **C.** Percentages of bioconversion were calculated based on the integrated peak areas from each chromatogram, as [Area$_{22\text{-dehydrocholesterol}}$/(Area$_{22\text{-dehydrocholesterol}}$ + Area$_{cholesterol}$) × 100]. One representative of three independent experiments is shown.

localized at the PM, involved in the clathrin-mediated uptake of free cholesterol (Wang and Song 2012).

Tetrahymena thermophila appears to internalize sterols mainly by phagocytosis (Elguero et al. 2018), but there is no information about the molecular mechanism underlying this process. The ciliate has a protein (TTHERM_00672270) that is 50% and 45% identical with human NPC1 and NPC1L1, respectively. However, TTHERM_00672270 was not differentially expressed in our RNA-seq (Table S3). NPC2 are small proteins containing an MD-2-related lipid-recognition domain (ML-domain, Pfam: PF02221), which mediates the direct binding of lipids (Inohara and Nuez 2002). *Tetrahymena thermophila* does not have recognizable orthologues of NPC2, but we have identified three genes encoding putative ML-domain-containing proteins (TTHERM_00354680, TTHERM_00030420, and TTHERM_00353379). Whereas TTHERM_00354680 was not differentially expressed (*ML_NR*, Table S3), TTHERM_00 030420 (*ML_UP*) and TTHERM_00353379 (*ML_DWN*) were, respectively, the most upregulated gene (147.65-fold) and the third most downregulated gene (−27.17-fold) in our RNA-seq experiments (Table S4).

The *T. thermophila* genome encodes 165 putative ABC transporters which were previously classified into eight subfamilies (Xiong et al. 2010). In our dataset, fifteen of these were differentially expressed (3 upregulated and 12 downregulated, see Table S4), and all showed a comparable percentage of sequence identity (50–46%) to known sterol transporters, such as *S. cerevisiae* Pdr11p and mammalian ABCA1. However, their involvement in sterol uptake or efflux cannot be predicted based solely on their primary structure. ABC transporters were also found to be differentially regulated in cholesterol-fed mice (Maxwell et al. 2003).

Other sterol binding protein families include oxysterol binding proteins (OSBP), OSBP-related proteins (ORP) (Olkkonen 2013), and some members of the START domain-containing proteins (Alpy and Tomasetto 2014). These proteins are involved in nonvesicular trafficking of sterols and in signaling processes. *Tetrahymena thermophila* contains 14 putative ORP and 27 START domain-containing proteins. Our dataset shows that one ORP (TTHERM_01085530) and four START domain-containing proteins (TTHERM_000374979, TTHERM_00624850, TTHERM_00006380, and TTHERM_00939010) were significantly downregulated (Table S4). Whether they are involved in processes related to sterol trafficking or signaling in the ciliate remains to be elucidated.

Putative components of signaling pathways that were differentially regulated include 14 protein kinase domain-, 2 GTP binding domain-, 2 cAMP binding domain-containing proteins, 4 putative voltage-gated channels, and 1 protein phosphatase (Table S4).

Validation of differentially expressed genes by RT-qPCR

We carried out RT-qPCR on genes involved either in tetrahymanol synthesis or in cholesterol transport or modification. We selected some significantly upregulated
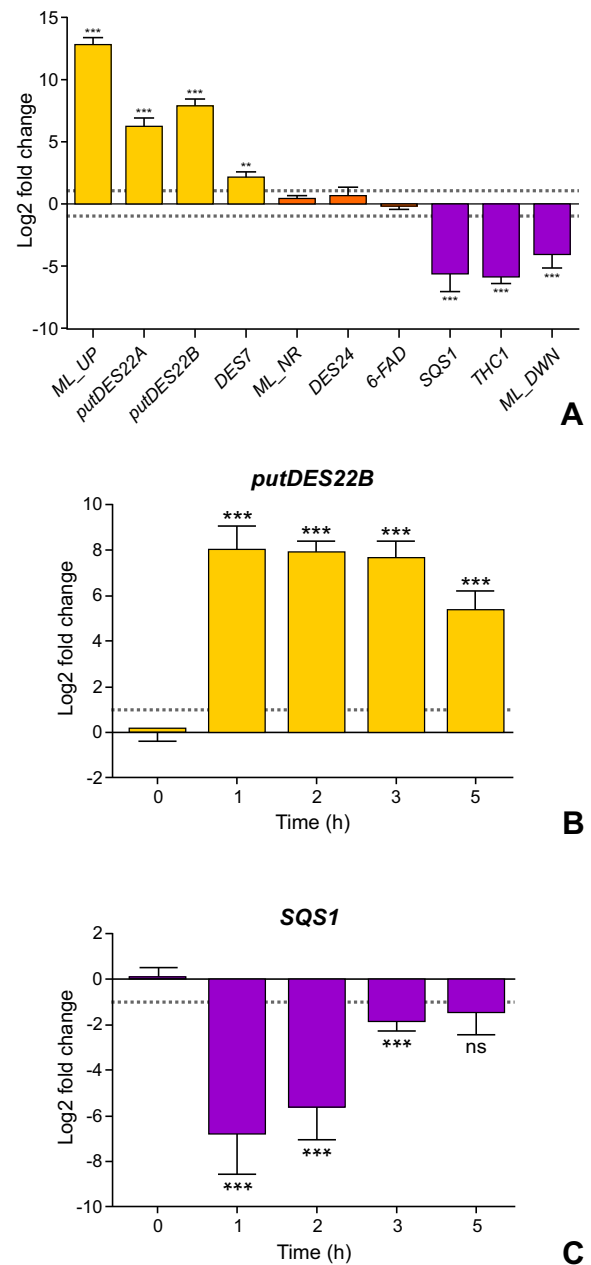
Figure 6 Validation of RNA-seq results by RT-qPCR. **A**. Cells were treated with 5 μg/ml cholesterol for 2 h prior to RNA extraction. RT-qPCR measurement of ten genes representative of the upregulated, downregulated, and not differentially expressed genes in the RNA-seq. **B–C**. Time-course RT-qPCR measurement of *putDES22B* and *SQS* genes in response to 5 μg/ml cholesterol. Grey lines indicate a threshold of absolute $\log_2$-fold change = 1 (absolute 2-fold change) to consider genes as significantly differentially expressed between conditions. Statistical analyses were performed by *T*-test (unpaired, one-tailed), against a control without cholesterol (*P* values, *0.5, **0.01, ***0.001, ns = not significant; *n* = 4).

(*ML_UP*, *putDES22A*, *putDES22B*, and *DES7*) and downregulated (*SQS1*, *THC1*, and *ML_DWN*) genes, together with some that show no significant variation in our

analysis (*ML_NR*, *DES24,* and *Δ6-FAD*). As shown in Fig. 6A, the expression patterns obtained by RT-qPCR were consistent with the results of our RNA-seq analysis.

Finally, we also analyzed the expression over time of two highly regulated genes, TTHERM_00085010 (*put-DES22B,* upregulated) and THERM_00382150 (*SQS1,* downregulated), after cholesterol treatment. As shown in Fig. 6B, although the maximal response was reached after one hour in both cases, it was maintained at a high level after two hours of stimulation. This result validates the choice of two hours' postcholesterol treatment as an appropriate time for our RNA-seq analysis.

DISCUSSION

Changes in the transcriptome of *T. thermophila* in response to exogenous cholesterol involve the differential regulation of 356 genes, most of which belong to orthogroups that are either ancestral to all eukaryotes or specific to the *Tetrahymena* genus.

The presence of cholesterol induces the expression of sterol C5 and C7 desaturases (DES5 and DES7, respectively), but does not significantly affect the expression of the phytosterol C24 de-ethylase (DES24). Interestingly, the second and third most upregulated genes in our dataset were identified as putative sterol C22 desaturases (*put-DES22A/B*). Knocking down the expression of each gene using RNAi by feeding led to a considerable reduction of total C22 desaturase activity. The sterol C22 desaturases known so far, which are widely distributed in eukaryotes, are cytochrome P450 oxygenases. However, the *put-DES22A/B* genes belong to the FAHD superfamily of enzymes. Thus, our findings provide the first evidence for a new family of sterol C22 desaturases, completely unrelated to cytochrome P450s. The two *T. thermophila* paralogs might have originated from other lipid-modifying FAHD by convergent evolution and gene duplication. The evolutionary constraints that could have led to this functional redundancy are unknown, although we can speculate that the two putDES22 paralogs can have different substrate preferences. We have previously shown that the deletion of *DES24* is lethal when a *T. thermophila* *ΔDES24* mutant strain is grown in 29C phytosterol-supplemented media (Tomazic et al. 2011). On the other hand, *T. thermophila* mutant strains lacking either *DES5* or *DES7* showed no apparent growth defects when cultivated in the presence of diverse sterol substrates (Nusblat et al. 2009; Najle et al. 2013 and Najle, S. R. & Uttaro, A. D., unpublished observations). Based on these observations, we also speculate that the capacity to carry out specific modifications on the lateral chain of sterols (namely C24 de-ethylation and C22 desaturation) might be essential for the ciliate. Whether this is the case for C22 desaturation remains to be investigated.

Repression of tetrahymanol biosynthesis by cholesterol occurs at the transcriptional level, affecting both the mevalonate pathway for squalene synthesis and squalene cyclization. This repression is due to the rapid and strong downregulation of genes encoding bifunctional HMG-CoA synthase/reductase, diphosphomevalonate decarboxylase, Isopentenyl-PP isomerase, polyprenyl synthetase, squalene synthase (SQS1), and squalene cyclase (THC1).

In summary, we experimentally determined and analyzed the transcriptional response to cholesterol in *T. thermophila*. We identified new candidate genes potentially involved in sterol transport and signaling, including two uncharacterized ML-domain-containing proteins with opposite transcriptional regulation. This work opens new avenues for research toward understanding the physiology and evolution of eukaryotic sterol metabolism.

LITERATURE CITED

Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Gall, L., Le Lynn, D. H., McManus, H., Mitchell, E. A. D., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A. & Spiegel, F. W. 2012. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.*, 59:429–493.

Alpy, F. & Tomasetto, C. 2014. START ships lipids across interorganelle space. *Biochimie*, 96:85–95.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215:403–410.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.*, 25:25.

Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Camara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.-M., Kissmehl, R., Klotz, C., Koll, F., Le Mouel, A., Lepere, G., Malinsky, S., Nowacki, M., Nowak, J. K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Betermier, M., Weissenbach, J., Scarpelli, C., Schachter, V., Sperling, L., Meyer, E., Cohen, J. & Wincker, P. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, 444:171–178.

Bammert, G. F. & Fostel, J. M. 2000. Genome-wide expression patterns in *Saccharomyces cerevisiae*: Comparison of drug

treatments and genetic alterations affecting biosynthesis of ergosterol. *Antimicrob. Agents Chemother.*, 44:1255–1265.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C. & Eddy, S. R. 2004. The Pfam protein families database. *Nucleic Acids Res.*, 32:D138–D141.

Beedle, A. S., Munday, K. A. & Wilton, D. C. 1974. Studies on the biosynthesis of tetrahymanol in *Tetrahymena pyriformis*. The mechanism of inhibition by cholesterol. *Biochem. J.*, 142:57–64.

Bligh, E. G. & Dyer, W. J. 1959. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.*, 37:911–917.

Bloch, K. 1965. The biological synthesis of cholesterol. *Science*, 150:19–28.

Bolger, A. M., Lohse, M. & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30:2114–2120.

Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, 34:525–527.

Buchfink, B., Xie, C. & Huson, D. H. 2014. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, 12:59.

Chang, T.-Y., Chang, C. C. Y., Ohgami, N. & Yamauchi, Y. 2006. Cholesterol sensing, trafficking, and esterification. *Annu. Rev. Cell Dev. Biol.*, 22:129–157.

Cid, N. G., Sanchez Granel, M. L., Montes, M. G., Elguero, M. E., Nudel, C. B. & Nusblat, A. D. 2017. Phylogenomic analysis of integral diiron membrane histidine motif-containing enzymes in ciliates provides insights into their function and evolutionary relationships. *Mol. Phylogenet. Evol.*, 114:1–13.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M. & Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21:3674–3676.

Conner, R. L., Landrey, J. R., Burns, C. H. & Mallory, F. B. 1968. Cholesterol inhibition of pentacyclic triterpenoid biosynthesis in *Tetrahymena pyriformis*. *J. Protozool.*, 15:600–605.

Conner, R. L., Mallory, F. B., Landrey, J. R. & Iyengar, C. W. 1969. The conversion of cholesterol to delta-5,7,22-cholestatrien-3-beta-ol by *Tetrahymena pyriformis*. *J. Biol. Chem.*, 244:2325–2333.

Conner, R. L., Mellory, F. B., Landrey, J. R., Ferguson, K. A., Kaneshiro, E. S. & Ray, E. 1971. Ergosterol replacement of tetrahymanol in Tetrahymena membranes. *Biochem. Biophys. Res. Commun.*, 44:995–1000.

Conner, R. L. & Ungar, F. 1964. The accumulation of cholesterol by *Tetrahymena pyriformis*. *Exp. Cell Res.*, 36:134–144.

Csűrös, M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics*, 26:1910–1912.

Dahmen, J. L., Olsen, R., Fahy, D., Wallis, J. G. & Browse, J. 2013. Cytochrome b5 coexpression increases *Tetrahymena thermophila* Δ6 fatty acid desaturase activity in *Saccharomyces cerevisiae*. *Eukaryot. Cell*, 12:923–931.

Derelle, R., Torruella, G., Klimeš, V., Brinkmann, H., Kim, E., Vlček, Č., Lang, B. F. & Eliáš, M. 2015. Bacterial proteins pinpoint a single eukaryotic root. *Proc. Natl Acad. Sci.*, 112:E693–E699.

Desmond, E. & Gribaldo, S. 2009. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol.*, 1:364–381.

Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M.,

Tallon, L. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Farzad, M., Carlton, J. M., Smith, R. K. J., Garg, J., Pearlman, R. E., Karrer, K. M., Sun, L., Manning, G., Elde, N. C., Turkewitz, A. P., Asai, D. J., Wilkes, D. E., Wang, Y., Cai, H., Collins, K., Stewart, B. A., Lee, S. R., Wilamowska, K., Weinberg, Z., Ruzzo, W. L., Wloga, D., Gaertig, J., Frankel, J., Tsao, C.-C., Gorovsky, M. A., Keeling, P. J., Waller, R. F., Patron, N. J., Cherry, J. M., Stover, N. A., Krieger, C. J., del Toro, C., Ryder, H. F., Williamson, S. C., Barbeau, R. A., Hamilton, E. P. & Orias, E. 2006. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.*, 4:e286.

Elguero, M. E., Sanchez Granel, M. L., Montes, M. G., Cid, N. G., Favale, N. O., Nudel, C. B. & Nusblat, A. D. 2018. Uptake of cholesterol by *Tetrahymena thermophila* is mainly due to phagocytosis. *Rev. Argent. Microbiol.*, 50:105–107.

Emms, D. M. & Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, 16:157.

Ewels, P., Magnusson, M., Lundin, S. & Käller, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32:3047–3048.

Ferraro, D. J., Gakhar, L. & Ramaswamy, S. 2005. Rieske business: structure-function of Rieske non-heme oxygenases. *Biochem. Biophys. Res. Commun.*, 338:175–190.

Fu, C., Xiong, J. & Miao, W. 2009. Genome-wide identification and characterization of cytochrome P450 monooxygenase genes in the ciliate *Tetrahymena thermophila*. *BMC Genom.*, 10:208.

Gao, F., Warren, A., Zhang, Q., Gong, J., Miao, M., Sun, P., Xu, D., Huang, J., Yi, Z. & Song, W. 2016. The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). *Sci. Rep.*, 6:24874.

He, D., Sierra, R., Pawlowski, J. & Baldauf, S. L. 2016. Reducing long-branch effects in multi-protein data uncovers a close relationship between Alveolata and Rhizaria. *Mol. Phylogenet. Evol.*, 101:1–7.

Horner, M. A., Pardee, K., Liu, S., King-Jones, K., Lajoie, G., Edwards, A., Krause, H. M. & Thummel, C. S. 2009. The Drosophila DHR96 nuclear receptor binds cholesterol and regulates cholesterol homeostasis. *Genes Dev.*, 23:2711–2716.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C. & Bork, P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.*, 34:2115–2122.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., Mcanulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A. F., Selengut, J. D., Sigrist, C. J. A. A., Thimma, M., Thomas, P. D., Valentin, F., Wilson, D., Wu, C. H. & Yeats, C. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res.*, 37:D211–D215.

Inohara, N. & Nuez, G. 2002. ML – A conserved domain involved in innate immunity and lipid metabolism. *Trends Biochem. Sci.*, 27:219–221.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, 40:109–114.

Li, Y. & Prinz, W. A. 2004. ATP-binding cassette (ABC) transporters mediate nonvesicular, raft-modulated sterol movement from the plasma membrane to the endoplasmic reticulum. *J. Biol. Chem.*, 279:45226–45234.

Love, M. I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15:550.

Mallory, F. B. & Conner, R. L. 1971. Dehydrogenation and dealkylation of various sterols by *Tetrahymena pyriformis*. *Lipids*, 6:149–153.

Mallory, F. B., Conner, R. L., Landrey, J. R. & Iyengar, C. W. L. 1968. The biosynthesis of 7,22-bisdehydrocholesterol from cholesterol. *Tetrahedron Lett.*, 9:6103–6106.

Mallory, F. B., Gordon, J. T. & Conner, R. L. 1963. The isolation of a pentacyclic triterpenoid alcohol from a protozoan. *J. Am. Chem. Soc.*, 85:1362–1363.

Masek, T., Vopalensky, V., Suchomelova, P. & Pospisek, M. 2005. Denaturing RNA electrophoresis in TAE agarose gels. *Anal. Biochem.*, 336:46–50.

Maxwell, K. N., Soccio, R. E., Duncan, E. M., Sehayek, E. & Breslow, J. L. 2003. Novel putative SREBP and LXR target genes identified by microarray analysis in liver of cholesterol-fed mice. *J. Lipid Res.*, 44:2109–2119.

Najle, S. R., Molina, M. C., Ruiz-Trillo, I. & Uttaro, A. D. 2016. Sterol metabolism in the filasterean *Capsaspora owczarzaki* has features that resemble both fungi and animals. *Open Biol.*, 6:160029.

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. *Mol. Biol. Evol.*, 30:1630–1643.

Nusblat, A. D., Muñoz, L., Valcarce, G. A. & Nudel, C. B. 2005. Characterization and properties of cholesterol desaturases from the ciliate *Tetrahymena thermophila*. *J. Eukaryot. Microbiol.*, 52:61–67.

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. *Eukaryot. Cell*, 8:1287–1297.

Olkkonen, V. M. 2013. OSBP-related proteins: liganding by glycerophospholipids opens new insight into their function. *Molecules*, 18:13666–13679.

Petrusma, M., Van Der Geize, R. & Dijkhuizen, L. 2014. 3-Ketosteroid 9α-hydroxylase enzymes: rieske non-heme monooxygenases essential for bacterial steroid degradation. *Antonie Van Leeuwenhoek*, 106:157–172.

Pfaffl, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, 29:e45.

Raederstorff, D. & Rohmer, M. 1988. Polyterpenoids as cholesterol and tetrahymanol surrogates in the ciliate *Tetrahymena pyriformis*. *Biochim. Biophys. Acta*, 960:190–199.

Soneson, C., Love, M. I. & Robinson, M. D. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.*, 4:1521.

Sperling, P., Ternes, P., Zank, T. K. & Heinz, E. 2003. The evolution of desaturases. *Prostaglandins Leukot. Essent. Fatty Acids*, 68:73–95.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., Von Mering, C. & von Mering, C. 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, 43:D447–D452.

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

Timmons, L. & Fire, A. 1998. Specific interference by ingested dsRNA. *Nature*, 395:854.

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*. *Eukaryot. Cell*, 10:423–434.

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. *Mol. Phylogenet. Evol.*, 74:122–134.

Uttaro, A. D. 2006. Critical review biosynthesis of polyunsaturated fatty acids in lower eukaryotes. *IUBMB Life*, 58:563–571.

Wang, L.-J. & Song, B.-L. 2012. Niemann-Pick C1-Like 1 and cholesterol uptake. *Biochim. Biophys. Acta*, 1821:964–972.

Warburg, C. F., Wakeel, M. & Wilton, D. C. 1982. The role of squalene synthetase in the inhibition of tetrahymanol biosynthesis by cholesterol in *Tetrahymena pyriformis*. *Lipids*, 17:230–234.

Xiong, J., Feng, L., Yuan, D., Fu, C. & Miao, W. 2010. Genome-wide identification and evolution of ATP-binding cassette transporters in the ciliate *Tetrahymena thermophila*: a case of functional divergence in a multigene family. *BMC Evol. Biol.*, 10:330.

Xiong, J., Lu, X., Zhou, Z. Z., Chang, Y., Yuan, D., Tian, M., Zhou, Z. Z., Wang, L., Fu, C., Orias, E. & Miao, W. 2012. Transcriptome analysis of the model protozoan, *Tetrahymena thermophila*, using deep RNA sequencing. *PLoS ONE*, 7:e30630.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Quality control of RNA-seq analysis.
Figure S2. GO-terms enriched among the downregulated genes.
Table S1. Taxon sampling used for phylostratigraphic analysis.
Table S2. Primers used in this study.
Table S3. Output of DESeq2 analysis showing all differentially expressed genes with a *P*-value ≤ 0.05.
Table S4. Differentially expressed genes showing absolute fold change ≥ 2, with eggNOG annotation.
Table S5. Gene ontology analysis for the downregulated genes dataset.
Table S6. Output of STRING results, including KEGG pathways, Pfam domain, and InterPro domain enrichment analyses.
Table S7. Orthogroups defined for phylostratigraphic analysis.
Table S8. Differential expression values of tetrahymanol biosynthesis pathway genes.
Table S9. Differential expression values of genes for the sterols modification pathway.