Molecular Biology and Evolution, 2025, **42**, 1–18
https://doi.org/10.1093/molbev/msaf030
Advance access publication 3 February 2025
Discoveries

MBE OXFORD

Genomic Signatures of Adaptation to Stress Reveal Shared Evolutionary Trends Between *Tetrahymena utriculariae* and Its Algal Endosymbiont, *Micractinium tetrahymenae*

Joseph B. Kelly [1,*] **David E. Carlson** [2] **Manuela Reuter** [1] **Annette Sommershof**[3] **Lubomír Adamec** [4] **Lutz Becks** [1]

[1]Aquatic Ecology and Evolution, Department of Biology, University of Konstanz, Konstanz 78464, Germany
[2]Institute for Advanced Computational Science, Stony Brook University, Stony Brook, NY 11794, USA
[3]Department of Biology, University of Konstanz, Konstanz 78464, Germany
[4]Department of Experimental and Functional Morphology, Institute of Botany CAS, Třeboň CZ-379 01, Czech Republic

***Corresponding author:** E-mail: joseph.kelly@uni-konstanz.de.
Associate editor: Joanna Kelley

Abstract

The evolution of intracellular endosymbiosis marks a major transition in the biology of the host and endosymbiont. Yet, how adaptation manifests in the genomes of the participants remains relatively understudied. We investigated this question by sequencing the genomes of *Tetrahymena utriculariae*, a commensal of the aquatic carnivorous bladderwort *Utricularia reflexa*, and its intracellular algae, *Micractinium tetrahymenae*. We discovered an expansion in copy number and negative selection in a TLD domain-bearing gene family in the genome of *T. utriculariae*, identifying it as a candidate for being an adaptive response to oxidative stress resulting from the physiology of its endosymbionts. We found that the *M. tetrahymenae* genome is larger than those of other *Micractinium* and *Chlorella* and contains a greater number of rapidly expanding orthogroups. These were enriched for Gene Ontology terms relevant to the regulation of intracellular signal transduction and cellular responses to stress and stimulus. Single-exon tandem repeats were overrepresented in paralogs belonging to these rapidly expanding orthogroups, which implicates long terminal repeat retrotransposons (LTRs) as potential agents of adaptation. We additionally performed a comparative transcriptomic analysis of *M. tetrahymenae* in a free-living state and in endosymbiosis with *T. utriculariae* and discovered that the genes that are differentially expressed were enriched for pathways that evidence shifts in energy generation and storage and in cellular protection strategies. Together, our results elucidate the axes along which the participants must adapt in this young endosymbiosis and highlight evolutionary responses to stress as a shared trend.

Keywords: ciliate, green algae, endosymbiosis, tandem gene duplications, LTR, genome, stress

Introduction

The evolution of endosymbiosis between two unicellular organisms marks a substantial transition in biological organization. The most profound example is embodied in the evolution of chloroplasts and mitochondria from prokaryotic ancestors. This transition released the host cells from the energetic constraints imposed by their surface-to-volume ratio, resulting in a 200,000-fold increase in the number of genes expressed and, ultimately, the diversity of eukaryotic form and function (Lane and Martin 2010). A fundamental interest therefore lies in understanding the processes and conditions that shape the evolution of intracellular endosymbiosis (hereafter "endosymbiosis") and the genomic changes involved.

Apart from the "original" endosymbioses that ultimately created eukaryotic cells, there exist myriad endosymbiotic relationships across the tree of life that have molded the biology of the participants. Of particular interest are symbioses between heterotrophic hosts and autotrophic endosymbionts, such as those of ciliates (phylum Ciliophora) and green algae (phylum Chlorophyta). Endosymbiotic states have evolved multiple times independently in both phyla (Esteban et al. 2010; Pröschold et al. 2011), whereby multiple algal cells inhabit the interior of each individual host ciliate cell, bestowing mixotrophy to the ciliate via the algae's photosynthetic products. Examples of algae-bearing ciliates can be found from multiple classes such as *Stentor pyriformis* of Heterotrichea (Hoshina et al. 2021), *Paramecium chlorelligerum* of Oligohymenophorea (Lanzoni et al. 2016), and *Coleps viridis* of Prostomatea (Pröschold et al. 2021). The majority of chlorophyte algal symbionts of ciliates, on the other hand, belong to a single class, the Trebouxiophyceae (Pröschold et al. 2011). The associations are stable in that the algae are inherited by the daughter ciliate cells upon division of the host; however, transmission modes can also be mixed (Meier and Wiessner 1988). In addition to being phylogenetically widespread, chlorophyte-harboring ciliates can be found worldwide in a variety of wetlands and other freshwater bodies (Esteban et al. 2010). Given the prevalence and diversity of these ciliate-algal pairs and the predisposition for endosymbiosis to evolve between these phyla, they constitute ideal study systems in which to study how this major step in evolution happens.

The *Paramecium bursaria* species complex (Spanner et al. 2022) is the most widely-used study system among algae-bearing

Received: July 4, 2024. **Revised:** October 20, 2024. **Accepted:** January 17, 2025

ciliates. *Paramecium bursaria* is found primarily in association with the green algae *Chlorella variabilis* and *Micractinium conductrix*, although other endosymbiotic algal species have been reported, sometimes in co-residence (Hoshina and Fujiwara 2012). Their interactions are typically defined as mutualistic in that the host receives algal photosynthates and oxygen and the algae receive nitrogen supplementation, CO_2, and protection from viruses (reviewed in Fujishima and Kodama 2022). However, recent studies support an alternative model describing the benefits as being context-dependent and exploitative on behalf of the ciliate (Lowe et al. 2016; Horas et al. 2022). The latter view is consistent with the basic evolutionary tenet that each partner lineage evolves with regard to its own fitness independent of the detriment that could be inflicted on the other lineage. In line with this idea are the steep biotic and abiotic gradients that distinguish free-living and endosymbiotic states and which the participants must negotiate upon the establishment of their association. For example, the photosynthesis of the algae would be expected to elevate the concentration of reactive oxygen species inside the host cell, whereas the endosymbiont must deal with changes in oxygen concentrations, pH regimes, osmolarity, and host enzymes. Given the inherent selfishness of each participating lineage and the obstacles that arise from their partner's physiology, the fundamental question remains: how does each species evolve in the first place such that stability in the endosymbiosis is maintained?

Prior studies of algal endosymbionts that leveraged their genomic resources have elucidated processes that mediate their interactions with ciliates. These include evolution in nutritionally relevant algal genes such as the expansion of amino acid importers in *C. variabilis* (Blanc et al. 2010) that could function to aid in the uptake of host amino acids as a nitrogen source and the identification of a candidate pathway driving sugar excretion by *M. conductrix* (Arriola et al. 2018). RNA–RNA interactions have also been proposed as a stabilizing mechanism employed by *C. variabilis* in that this species possesses mRNAs with regions that are complementary to host transcripts. These mRNAs are released from the algae upon digestion by the host and are processed to short RNAs by the host's cellular machinery (Dicer). The short RNAs then function to knock-down host transcripts, resulting in an inhibition of host growth and are hence thought to impose a control on algal digestion rates (Jenkins et al. 2021).

Studies using *P. bursaria* genomes have also delimited molecular mechanisms and patterns of genome evolution that likely aid in stabilizing the endosymbiosis. RNAi experiments targeting the most differentially expressed host genes between endosymbiont-bearing and endosymbiont-free ciliates found that cultures with knocked-down *glnA* transcripts possessed fewer endosymbiotic algae per host cell (He et al. 2019), implicating glutamine as an amino acid that the host uses to control its algal load. Expansions in the number of genes bearing domains relevant to endosymbiosis have also been found in *P. bursaria* genomes, such as the TLD (TLDc) domain of oxidation resistance proteins (Oliver et al. 2011; Finelli and Oliver 2017), which could translate to an increased tolerance of the reactive oxygen species produced during algal photosynthesis (Cheng et al. 2020). Despite the wealth of knowledge procured from *P. bursaria* and its algae, these species are the only endosymbiotic ciliate-algal pairs with reference genomes to date (Blanc et al. 2010; Arriola et al. 2018; He et al. 2019; Cheng et al. 2020), thus limiting our understanding of how endosymbiosis evolves between these phyla.

Tetrahymena utriculariae is an intriguing example of ciliate-algal endosymbiosis. In stark contrast to the cosmopolitan *P. bursaria,* which has been isolated worldwide from a variety of freshwater sources, *T. utriculariae* is endemic to the southern part of Africa (Šimek et al. 2017; Sirová et al. 2020). It has only been isolated from the traps of the carnivorous aquatic plant *U. reflexa* (Pitsch et al. 2017), which constitute a harsh, inescapable environment characterized by the presence of digestive enzymes, a low pH (4.2 to 7.2) (Sirová et al. 2003), a cocktail of organic substances (Sirová et al. 2011), and anoxia (Adamec 2007). *Tetrahymena utriculariae* is also a specialist with regard to its endosymbiotic partnership and has only been found in association with the algae *Micractinium tetrahymenae* (Fig. 1a and b). The endosymbiosis is thought to be facultative in nature (Pitsch et al. 2017; Šimek et al. 2017) and, like *P. bursaria*, *T. utriculariae* derives its energy partially from its algal symbionts (Šimek et al. 2017). However, how these two species have adapted to endosymbiosis at the genomic level is still unknown.

To address this question, we performed a comparative genomic analysis of *T. utriculariae* and *M. tetrahymenae* against other members in their respective genera. Evidence of genomic expansions was discovered in both species that are relevant to cellular responses to stress and which were produced in large part by tandem gene duplication events. Differential expression analysis of free-living and endosymbiotic *M. tetrahymenae* also identified shifts in metabolic pathways that suggest transitions in primary metabolism and cellular maintenance strategies that mirror those of endosymbiotic *C. variabilis* and *M. conductrix*. Overall, these findings elucidate the adaptive changes that occur at the genomic level during co-evolution between these endosymbiotic partners, hence furthering our understanding of how such associations are maintained over evolutionary timescales.

Results

We used long-read sequencing data to produce genome assemblies of the green *T. utriculariae*, its resident algae *M. tetrahymenae* (Fig. 1a and b), and two close relatives of each species to identify signatures of adaptation in the context of endosymbiosis. As *Tetrahymena* have two nuclear genomes constituting diploid micronuclei (MIC) and a larger, transcriptionally active, polyploid macronucleus (MAC) that is derived from the MIC, we devised a genome assembly pipeline that leverages MIC-specific DNA motifs to decontaminate the sequencing reads and produce a MAC assembly (see Materials and Methods section). This pipeline was benchmarked using DNA sequencing reads produced from MAC-enriched and whole-cell samples for *T. malaccensis* 23b, a close relative of our focal species, *T. utriculariae*, for which we were unable to separate the nuclear fractions physically. The overall identity of the pairwise genome-wide alignments of the MAC-enriched and metagenomic (whole-cell extract) *T. malaccensis* 23b assemblies was >99.99%, with 100% of the contigs from the whole-cell sample aligning and >95.6% of the contigs from the MAC-enriched sample aligning (supplementary fig. S1, Supplementary Material online). Quality metrics such as the number of gene models inferred, intron and exon characteristics, and genome contiguity were comparable between the two assemblies (Table 1, supplementary table S1, Supplementary Material online). The *T. utriculariae* assembly was similar to its closest relatives, *T. malaccensis* and *T. thermophilia*, in terms of GC content, genome length, and the number of gene models inferred.

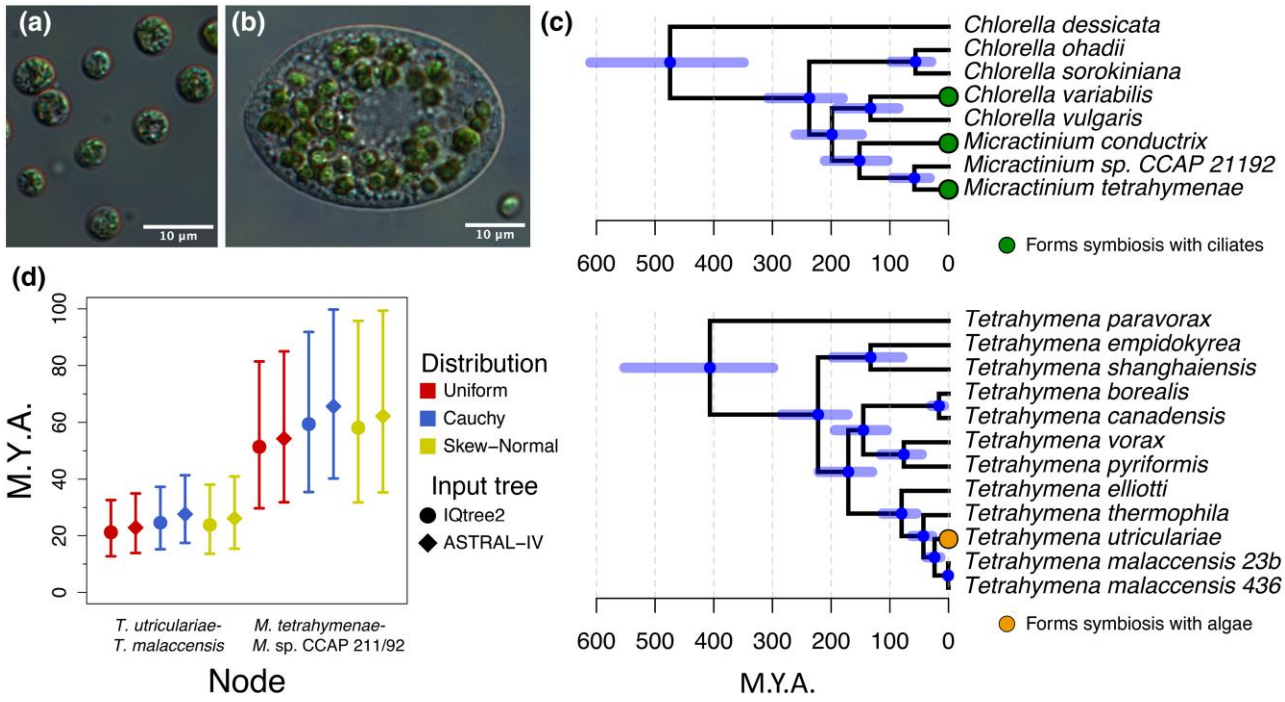

Fig. 1. The evolutionary history of the endosymbiosis on a geological timescale. a) Free-living *Micractinium tetrahymenae*. b) *Tetrahymena utricularlae* with endosymbiotic *M. tetrahymenae*. c) Subtrees extracted from the skew-normal MCMCtree output based on the IQtree2 topology for the *Chlorella-Micractinium* (top panel) and *Tetrahymena* (bottom panel) clades, with node bars reporting 95% CIs for date estimates. d) Mean posterior densities of dates for the most recent nodes producing the *T. utricularlae* and *M. tetrahymenae* lineages with 95% CIs, inferred using MCMCtree.

Table 1 Comparison of *Tetrahymena* genome assemblies

Species	Contigs or scaffolds	Total length (Mb)	% GC	N50 (Kb)	Genes	Genome/proteome BUSCO			
						% Single-copy complete	% Complete duplicated	% Single-copy fragmented	% Missing
T. utricularlae A2B2	328	99.8	22.8	531.7	22,852	93.6/95.3	1.2/2.9	1.2/0	4.0/1.8
T. malaccensis 23b (MAC)	179	108	22.7	1046	23,102	97.7/97.1	0.6/2.3	0/0.6	1.7/0
T. malaccensis 23b (whole-cell)	171	107.7	22.7	1046.4	22,937	96.5/96.5	0.6/2.3	0/0.6	2.9/0.6
T. malaccensis 436	554	106.7	22.6	496.1	26,378	97.1/96.5	1.2/3.5	0/0	1.7/0
T. thermophila SB210	181	103.3	22.3	929.7	26,996	97.1/98.2	0.6/1.2	1.2/0	1.1/0.6
T. elliotti 4EA	331	90.8	22.9	1022.5	22,562	94.7/95.3	1.2/3.5	0.6/0.6	3.5/0.6
T. borealis X4H2	325	93.5	24	638.7	21,943	95.3/94.7	1.8/4.7	0.6/0.6	2.3/0
T. canadensis	1637	103.4	24.2	416	25,188	96.5/98.2	1.2/1.2	0/0	2.3/0.6
T. paravorax	448	108.4	28.6	574.2	25,551	94.7/97.7	1.2/0.6	1.2/0.6	2.9/1.1
T. shanghaiensis	2660	95.6	20.1	153.6	21,982	86.5/84.2	10.5/11.1	0.6/1.8	2.4/2.9
T. empidokyrea	558	84.9	25.7	447.1	20,847	77.8/74.3	20.5/25.7	0/0	1.7/0
T. pyriformis	367	116.1	28.2	610.2	26,866	95.3/95.3	1.2/2.9	2.3/1.8	1.2/0
T. vorax	512	114.8	28.4	658.3	25,238	94.2/95.9	4.1/3.5	0.6/0	1.1/0.6

BUSCO scores were calculated using alveolata_odb10 lineage. Source databases are listed in supplementary table S2, Supplementary Material online. Genomes and proteomes produced in the current study are underlined.

The nuclear assembly of *M. tetrahymenae* consisted of 18 contigs (Table 2), nine of which (ptg000003l, ptg000014l, ptg000005l, ptg000008l, ptg000012l, ptg000011l, ptg000002l, ptg000015l, ptg000013l) were capped by "TTTAGGG" canonical plant telomeric repeats on both ends and seven of which (ptg000010l, ptg000009l, ptg000004l, ptg000007l, ptg000006l, ptg000001l, ptg000109l) had telomeres on one end. The *M.* sp. CCAP 211/92 genome assembly was less contiguous although comparable in terms of completeness to the *M. tetrahymenae* genome based on BUSCO scores (Table 2).

Time-Calibrated Phylogeny With MCMCtree

We performed dating using MCMCtree with seven fossil calibrations spanning a 104-taxon phylogeny of Diaphoratickes, inferred via IQtree2 and ASTRAL-IV using 50 BUSCO peptide sequences that were identified using the eukaryota_odb10 database. The recovered phylogenies were well-supported as the mean bootstrap support value across nodes was 97.9% for IQtree2 (supplementary fig. S2a, Supplementary Material online) and the mean local posterior probability across nodes was 91.8% for ASTRAL-IV

Table 2 Comparison of *Chlorella* and *Micractinium* genome assemblies

Species	Contigs or scaffolds	Total length (Mb)	% GC	N50 (Kb)	Genes	Genome/proteome BUSCO			
						% Single-copy complete	% Complete duplicated	% Single-copy fragmented	% Missing
M. tetrahymenae UKON-001	18	66.9	65.3	4697	14,086	94.9/97.2	0.9/2.0	1.4/0.2	2.8/0.6
M. sp. CCAP 211/92	225	50.1	65.1	372.1	11,539	94.1/96.4	0.2/0.7	1.6/0.8	4.1/2.1
M. conductrix SAG 241-80	300	61	67.2	1210.5	9217	93.5/94.3	1.4/1.6	1.5/0.7	3.6/3.4
C. variabilis NC64A	414	46.2	67.1	1469.6	9780	90.7/91.5	0.7/1.1	3.7/2.9	4.9/4.5
C. vulgaris CCAP 211/11P	45	40.4	61.6	2825.1	10,808	96.4/98.3	0.2/0.7	1.2/0.2	2.2/0.8
C. dessicata UTEX 2437	29	20.7	45	1303.9	8418	94.0/90.4	4.2/3.9	0.6/0.8	1.2/4.9
C. ohadii strain 1	486	57.1	63.9	329	10,866	94.6/94.9	0.5/0.9	1.4/1.0	3.5/3.2
C. sorokiniana UTEX 1602	159	59.6	64	2593	9526	96.2/96.6	0.5/0.9	0.9/0.6	2.4/1.9

BUSCO scores were calculated using chlorophyta_odb10 database. Source databases are listed in supplementary table S2, Supplementary Material online. Genomes and proteomes produced in the current study are underlined.

(supplementary fig. S2b, Supplementary Material online). The topologies of the Trebouxiophyceae differed between the two phylogenies in that the *C. variabilis*-*C. vulgaris* clade was recovered as sister to *Micractinium* by IQtree2 and *C. ohadii*-*C. sorokiniana* as sister to *Micractinium* by ASTRAL-IV. The Oligohymenophorea clade encompassing *Paramecium* and *Tetrahymena* was identical between the ASTRAL-IV and IQtree2 phylogenies. *Tetrahymena utriculariae* and *M. tetrahymenae* were recovered as sister taxa to *T. malaccensis* and *M. sp*. CCAP 211/92, respectively, consistent with the 18S phylogenies that we inferred during strain selection (Fig. 1c, supplementary fig. S3 to S9, Supplementary Material online). The two phylogenies were overall highly similar (Robinson-Foulds distance = 0.13) and the clades descending from the fossil-calibrated nodes were monophyletic in both (supplementary fig. S4 to S9, Supplementary Material online).

We performed a sensitivity analysis for the MCMCtree pipeline evaluating three major parameters: the distribution specified for the node priors on the MCMCtree input tree, the method of tree construction used to infer the input tree (IQtree2 vs. ASTRAL-IV), and the fossil calibrations used. The age estimates were generally robust to the removal of fossils; however, older age estimates relative to those made by the full seven-fossil calibration scheme were produced by the removal of the streptophyte fossils using the IQtree2 input (supplementary fig. S10, Supplementary Material online) and similarly by the removal of the ciliophoran fossil for ASTRAL-IV, although with a mix of node ages being under- and overestimated for the skew-normal and uniform prior distributions (supplementary fig. S11, Supplementary Material online). The age estimates for bifurcations corresponding to the speciation events producing *T. utriculariae* and *M. tetrahymenae* were likewise stable across prior distributions, input phylogenies, and fossil calibration schemes, although older divergence time estimates and wider 95% CIs again resulted from the aforementioned fossil removals (supplementary fig. S12, Supplementary Material online). The majority of the mean divergence time estimates were between 21 and 30 M.Y.A. for the *T. malaccensis*-*T. utriculariae* split and between 51 and 70 M.Y.A. for the *M. tetrahymenae*-*M.* sp. CCAP 211/92 split (Fig. 1d, supplementary fig. S12), Supplementary Material online.

Characteristics of the *M. tetrahymenae* Genome

We performed a comparative genomic analysis to investigate the adaptations of *M. tetrahymenae* associated with an endosymbiotic lifestyle. The *M. tetrahymenae* genome was found to be larger than other *Chlorella* and *Micractinium* genomes in terms of overall length and in the number of genes it contains (Fig. 2a). An analysis with CAFE5 was conducted to infer orthogroups undergoing significant (i.e. rapid) changes in member size using a subtree constituting the *Chlorella-Micractinium* clade that was extracted from the time-calibrated phylogeny produced with the IQtree2 input tree, skew-normal node priors, and the full-fossil calibration scheme. This analysis identified *M. tetrahymenae* as containing the most orthogroups undergoing significant expansions and as having one of the lowest estimates for the number of orthogroups undergoing significant contractions among *Chlorella* and *Micractinium* (Fig. 2b). The most enriched Gene Ontology (GO) terms in genes belonging to rapidly expanding orthogroups in *M. tetrahymenae* were related to response to external stimulus, intracellular signaling, and stress responses, followed by several GO terms relevant to interactions with nucleic acid residues, and finally positive regulation of transport (Fig. 2c). The enrichment of the top three terms, "regulation of response to stimulus" (GO:0048583), "regulation of intracellular signal transduction" (GO:1902531), and "regulation of response to stress" (GO:0080134) were driven by expansions in three gene families as they collectively contained 84.2% of the paralogs belonging to rapidly expanding gene families bearing these three GO terms (Fig. 2d). Using the EGGnog-mapper annotations, these orthogroups were identified by their gene symbols as *CLA4* (OG0000408: GO:0048583, GO:0080134, GO:1902531), *AKAP7* (OG0000061: GO:0048583 and GO:1902531), and *OACYL* (OG0000152: GO:0048583 and GO:0080134). The genes belonging to these orthogroups were distributed across five contigs (Fig. 2e). Every copy from the *M. tetrahymenae* genome within these three orthogroups was found to be transcribed in our dataset, suggesting that they are functional genes.

Given the topological differences between the IQtree2 and ASTRAL-IV phylogenies for the *Chlorella-Micractinium* clade, we repeated orthogroup inference and the CAFE5 analysis using the corresponding MCMCtree output produced using the ASTRAL-IV input tree. We found highly similar results in that 84 significantly expanded *M. tetrahymenae* orthogroups were recovered for the ASTRAL-IV analysis compared with the 86 significantly expanded orthogroups recovered for the IQtree2-based analysis. The significantly expanded orthogroups between the two analyses were largely overlapping in gene content as 97.5% (596 of 611) of the genes in significantly expanding *M. tetrahymenae* orthogroups in the IQtree2 analysis were also present in the significantly expanding orthogroups of the ASTRAL-IV analysis and, conversely, 99.0% (596/602) of the

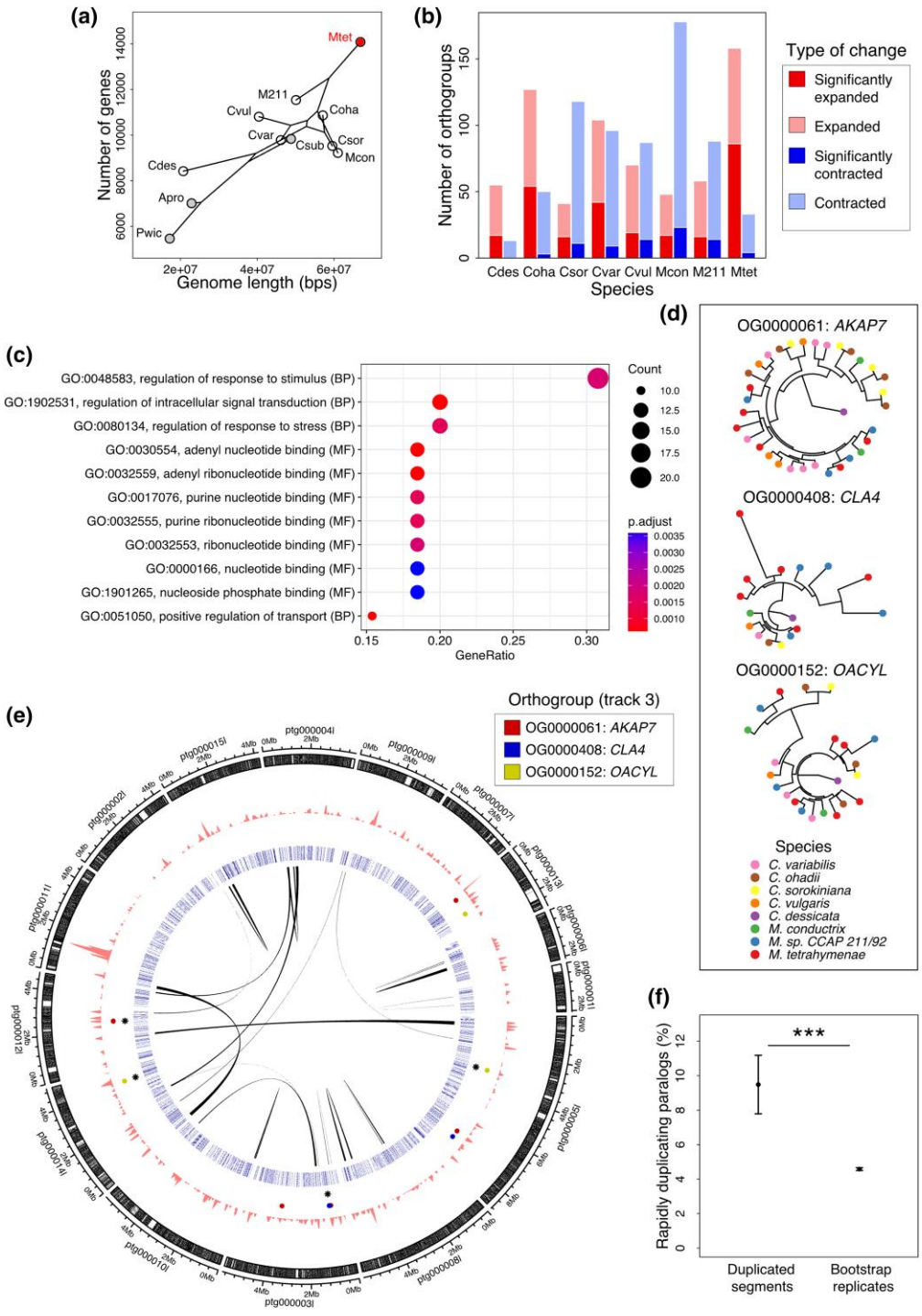

Fig. 2. The expanded genome of *Micractinium tetrahymenae*. a) Comparison of genome sizes and number of gene models of *Micractinium* and *Chlorella* species and three outgroup species, shaded in gray (*Coccomyxa subellipsoidea*, *Prototheca wickerhamii*, and *Auxenochlorella protothecoides*). Taxa are connected by their phylogenetic relatedness as inferred by IQtree2. b) Orthogroups identified as undergoing change in copy number by CAFE5 along each of the terminal branches in the *Chlorella* and *Micractinium* subtree. Abbreviations are made using the first letter of the genus and either the first three letters of the species or the first three numbers of the CCAP accession number. c) GO terms identified as being enriched in the genes belonging to rapidly expanding orthogroups. BP and MF refer to the "biological process" and "molecular function" GO categories. d) Gene trees of rapidly expanding orthogroups in *M. tetrahymenae* involved "regulation of response to stimulus" (GO:0048583), "regulation of intracellular signal transduction" (GO:1902531), and "regulation of response to stress" (GO:0080134). Trees were inferred in IQtree2 using AA alignments and are rooted on the outgroup copy from *C. dessicata*. e) Circos plot of the *M. tetrahymenae* genome assembly. Starting from the outside going inward, the tracks are (i) the *M. tetrahymenae* contigs (omitting two short contigs, ptg000109l at 44.7 kb containing only ten gene models and ptg000020l at 54.6 kb containing only six gene models) with genes represented by shaded segments, (ii) density of genes in rapidly expanding orthogroups (significant at $P < 0.01$) in 100 kb sliding windows, (iii) coordinates of *CLA4*, *AKAP7*, and *OACYL* copies, (iv) asterisks denoting tandem duplication events in *CLA4*, *AKAP7*, and *OACYL*, (v) coordinates of genes annotated as tandemly duplicated, and (vi) segmental duplications. f) A comparison of the percent of genes contained in duplicons belonging to rapidly expanding orthogroups against a distribution of 100k genome-wide bootstrap replicates, produced by sampling 80-kb segments of the genome (the mean size of a duplicon) and calculating the percent of genes in each bootstrap replicate belonging to rapidly expanding orthogroups. Significance was inferred using a permutational *t*-test in the R package Deducer v0.7-9 (Fellows 2012) specifying 10k permutations and a two-sided alternative hypothesis (Welsh *t*-statistic = 2.9, *P*-value = 2×10^{-4}).

genes in the significantly expanding orthogroups inferred in the ASTRAL-IV analysis were also found in the significantly expanding orthogroups from the IQtree2 analysis. The ASTRAL-IV-based analyses recovered the same 11 enriched GO terms from the IQtree2 analysis and two additional terms in the "biological process" category, "response to stress" (GO:0006950) and "response to oxygen levels" (GO:0070482). For the sake of stringency, we considered only the GO terms that were enriched in both analyses.

We explored the relative contributions of agents of genome expansions including those made by different types of gene duplication events and the activities of transposable elements (TEs). A total of 4.05% of the genome was detected as having undergone segmental duplications (Fig. 2e). The genes within the duplicons contained a significantly higher proportion of paralogs belonging to rapidly expanding orthogroups (Fig. 2f) although these did not include the three orthogroups representing *CLA4*, *AKAP7*, and *OACYL* (Fig. 2e). Instead, tandem duplications produced 67% of the paralogs in these orthogroups, which were present in four genomic clusters. *Micractinium tetrahymenae* had a higher percentage of tandemly duplicated genes (22.1%) relative to *M.* sp. CCAP 211/92 (17.4%) and *M. conductrix* (9.7%), although this comparison could be biased by differences in genome assembly contiguity. In total, 55.8% of *M. tetrahymenae* paralogs belonging to rapidly expanding orthogroups were tandem duplicates and 7.4% were contained in segmental duplications. *Micractinium tetrahymenae* was found to possess the highest per-bp content of long interspersed nuclear elements (LINEs), CACTA transposons, and long terminal repeat retrotransposons (LTRs) relative to other *Micractinium* and *Chlorella* (supplementary fig. S13, Supplementary Material online). We focused further analysis on LTRs given that they are not only a causative factor of tandem gene duplications, but also because the new gene copies that they generate are identifiable by their absence of introns due to switching to mRNA as a template during reverse transcription (reviewed in Krasileva 2019). Thus, we screened for single-exon gene models and found that they were present at a three-fold greater frequency (11.4%) in tandemly duplicated genes belonging to rapidly expanding orthogroups relative to non-tandemly duplicated genes (3.9%), and over a two-fold greater frequency relative to the complete set of tandemly duplicated genes (4.8%). The *CLA4* and *OACYL* orthogroups each contained two single-exon tandemly duplicated copies as well. However, only 12.3% of the basepairs masked as LTRs in *M. tetrahymenae* were assembled into intact LTRs by EDTA. Given this shortcoming of the TE annotation pipeline, we were unable to perform a comprehensive investigation into whether the single-exon genes were nested within intact LTRs.

Differential Expression Analysis of Free-Living and Symbiotic *M. tetrahymenae*

We sequenced the total pool of mRNA transcripts in free-living and endosymbiotic *M. tetrahymenae* to determine which pathways are recruited during endosymbiosis. The transcriptome shifted substantially in expression profiles (Fig. 3a). The differentially expressed (DE) genes were enriched for two pathways involving primary metabolism, KEGG pathway map00500 (starch and sucrose metabolism), and map00910 (nitrogen metabolism), and map04976 (bile secretion; Fig. 3b–d). The DE genes in map00500 form two chains through the pathway; the first is comprised of genes involved in the synthesis and breakdown of cellulose and the second of genes that catalyze the synthesis of starch and its subsequent hydrolysis to maltose (supplementary fig. S14, Supplementary Material online). Within map00910, the DE genes are carbonic anhydrase (*can*, *cah*) and genes involved in ammonium conversions, both in the capacity of primary metabolism via assimilatory nitrate reduction (ferredoxin-nitrite reductase, *nirA*) and in maintaining ammonia homeostasis (glutamate dehydrogenase, *gdhA*; supplementary fig. S15, Supplementary Material online). The enrichment of map04976, which constitutes a pathway likely absent in unicellular algae due to their lack of cholangiocytes and a digestive system, was a product of DE in six copies of *ABCG2* and one copy of each *AQP1* and *ATP1A* (supplementary fig. S16, Supplementary Material online).

Evolution in the *T. utriculariae* Genome

We performed a comparative genomic analysis for *T. utriculariae* that was similar to the one conducted for *M. tetrahymenae*. We discovered that the size of the *T. utriculariae* genome and the number of genes that it possesses were comparable to its closest relatives, *T. thermophilia* and *T. malaccensis* (Table 1). Three of the top five most enriched PFAM terms in the set of genes belonging to significantly expanding orthogroups in *T. utriculariae* are related to mobile genetic elements (Fig. 4a) and include the transposase family (HTH Tnp Tc3 2); DDE 3, a member of the DDE domain class whose members catalyze transposition and which were often located on *T. utriculariae* gene copies annotated as transposases (Fig. 4b); and DDE 1, which references the DDE superfamily of endonucleases. The third and fourth most enriched terms were REJ, a domain whose function is unknown but has been proposed to be active in maintaining homeostasis via ion transport (Moy et al. 1996), and TLD (TLDc), a domain that plays a role in protection against oxidative stress (Oliver et al. 2011; Finelli et al. 2016; Finelli and Oliver 2017). The enrichment of TLD was produced solely by expansions in one orthogroup (OG0000067), of which 22 of the 30 members in *T. utriculariae* had the domain (Fig. 4c). None of the genes with REJ or TLD domains had transposase or DDE annotations and the REJ and TLD-containing copies did not appear to be clustered with genes bearing transposase and DDE terms (Fig. 4b).

Given the relevance of the TLD domain in protecting the ciliate against the reactive oxygen species (ROS) that would be expected to accompany algal endosymbionts, we pursued an in-depth evolutionary analysis of the expanded orthogroup bearing the domain. Eighteen of the TLD-bearing copies and five of the non-TLD bearing copies were identified as tandem duplicates by i-Adhore and were often clustered in pairs (supplementary fig. S17, Supplementary Material online). Seventeen of the 22 TLD-bearing members were transcriptionally active (Fig. 4d), suggesting they encode functional proteins. Neither Interproscan nor eggNOG-mapper retrieved functional annotations for the remaining eight paralogs (represented as open circles on Fig. 4c and d), many of which were at the ends of long branches on the gene tree. Seven of these copies were transcriptionally active (Fig. 4d). TLD-bearing members of this orthogroup in *T. utriculariae* were found to be predominantly under negative purifying selection (84 of the 236 codons tested by FEL) whereas only two residues were inferred to be under positive selection (Fig. 4e, supplementary Data S1, Supplementary Material online). The majority of the testable sites (167 of 236 codons) were in the TLD domain due to a much higher degree of sequence variability in the N-terminus. Sixty-five codons (38.9% of the testable sites)

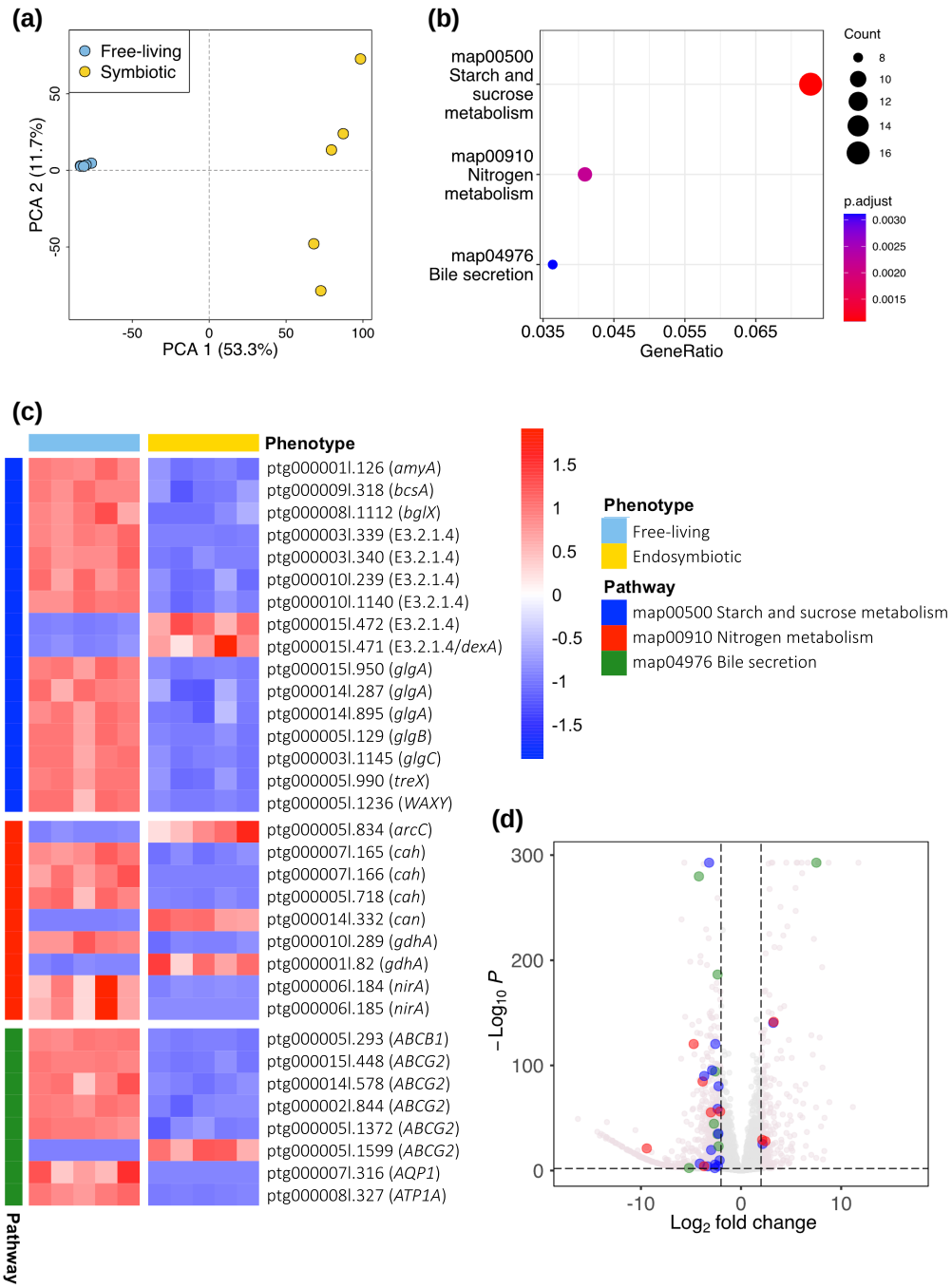

Fig. 3. Shift in expression profiles of *Micractinium tetrahymenae* between free-living and endosymbiotic states. a) PCA produced using the "prcomp" function in R with the normalized count matrix of transcriptomic reads for all genes. Each point corresponds to a separate sample. b) KEGG pathways enriched in the differentially expressed set of genes. c) DE genes belonging to the enriched pathways. Row names are the names of gene models in *M. tetrahymenae* with the gene symbols taken from the KO annotations made by eggNOG-mapper. Columns represent a given experimental replicate. The color gradient is scaled and centered based on the values within each row. d) Volcano plot highlighting DE genes in a free-living versus. endosymbiotic state of *M. tetrahymenae*. Enlarged points correspond to the genes within the three enriched pathways.

in the TLD domain were inferred to be under negative selection, consistent with the TLD domain's function and structure being highly sensitive to mutations (Finelli et al. 2016).

Discussion

The Age of the *T. uricularіae* and *M. tetrahymenae* Endosymbiosis

By studying the evolutionary processes that are at play in lineages that have recently made the transition to endosymbiosis, we can better understand how stability might be first achieved in its early stages. We found that the endosymbiotic phenotype of *T. uricularіae* is relatively new on geological timescales. Specifically, the majority of posterior means of divergence times for the *T. malaccensis*-*T. uricularіae* bifurcation were estimated between 21 and 30 M.Y.A. which, following the principal of maximum parsimony in that the symbiotic phenotype was derived only once and not lost in any lineages, provides an upper boundary for its age. The recency of the endosymbiosis matches with the observation that it is

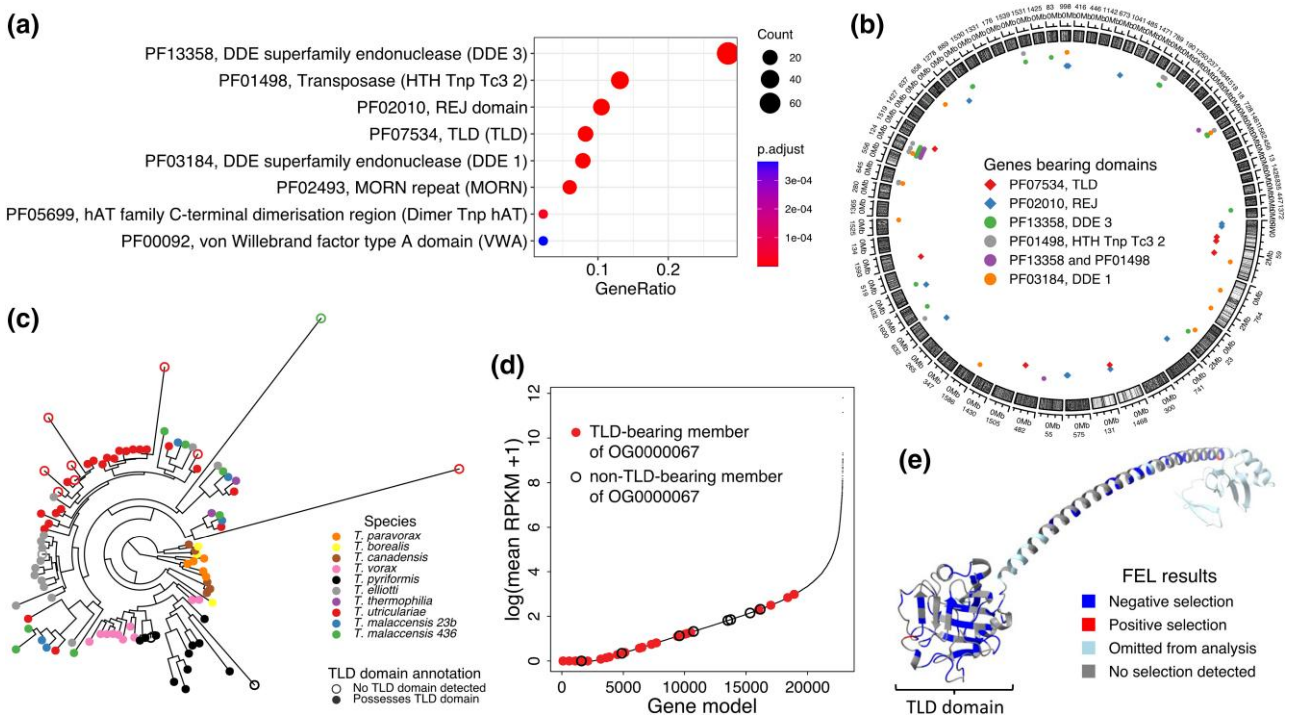

Fig. 4. Gene family expansions and residue-level evolution of paralogs with TLD domains in *Tetrahymena utriculariae*. a) Enriched PFAM terms in the set of genes belonging to significantly expanding orthogroups in *T. utriculariae*. b) Circos plot of the *T. utriculariae* genome assembly. The outermost track depicts the *T. utriculariae* contigs that are larger than 0.5 Mb with genes shaded in dark, and the inner track denotes the coordinates of the gene copies in rapidly expanding orthogroups that bear the top five most enriched PFAM terms. c) Gene tree of orthogroup OG0000067, characterized by genes containing the TLD domain. Tree inference was performed on an AA alignment in IQtree2. d) RPKM values of the OG0000067 members in *T. utriculariae*, calculated from the feature count matrix of transcripts mapped to the *T. utriculariae* genome assembly. e) Three-dimensional structure prediction of the *T. utriculariae* TLD-bearing copy with the highest RPKM value, produced using ColabFold and colored according to FEL analysis results. The error plot of this prediction is represented in supplementary fig. S18, Supplementary Material online.

facultative and the prediction that endosymbiosis evolves higher dependence over time, and from a facultative to an obligate interaction (Estrela et al. 2016). The young age of the phenotype could also help account for the observation that the genome expansion of *M. tetrahymenae* is a characteristic that evolved rapidly, at least in the context of the *Chlorella* and *Micractinium* clade, since its divergence from *M.* sp. CCAP 211/92. A previous study investigating trends in genome size evolution in the *Arsenophonus* endosymbionts of insects found evidence that more recent endosymbionts tended to have larger genomes (Siozios et al. 2024), a pattern which the authors argued is consistent with the rapid gain of new functions and evolutionary innovation. A hypothesis that warrants further investigation is that the endosymbiotic association between *M. tetrahymenae* and *T. utriculariae* is younger than those between *P. bursaria* and *M. conductrix* or *C. variabilis*, two algal species that do not have expanded genomes. Indeed, a genus-wide 18S rRNA phylogenetic analysis recovered *P. bursaria* as the earliest branching lineage within *Paramecium* (Fokin et al. 2004), leaving the possibility for an earlier establishment of algal endosymbiosis.

The Role of Genome Expansions as Agents of Adaptation in Endosymbiosis

Adaptation can occur within a lineage through the rapid expansion of orthogroups (i.e. gene families; Kondrashov 2012; Nakjang et al. 2013; Pan et al. 2013; Xu et al. 2016). We found that the *M. tetrahymenae* genome possessed a greater number of rapidly proliferating orthogroups relative to other *Micractinium* and *Chlorella*, consistent with its larger genome size. Both tandem and segmental duplications received support for being active in driving paralog generation within these orthogroups; however, tandem duplication was the more dominant process. The *M. tetrahymenae* genome was also distinguished from congeners by its high TE content with LTRs as one of the most abundant classes, which can produce tandem duplications that are identifiable by their lack of introns (Krasileva 2019). Indeed, single-exon genes were overrepresented in the tandem duplicates of rapidly expanding orthogroups, thus highlighting a potential role of LTRs in driving adaptive evolution in *M. tetrahymenae*.

We investigated the functional consequences of the genome expansion in *M. tetrahymenae* and discovered that the rapidly expanding orthogroups were enriched for genes related to stress responses and intracellular signaling, which was driven by expansions in three orthogroups, two of which contained single-exon copies as tandem repeats (*CLA4* and *OACYL*). The first was identified as the serine/threonine-protein kinase *CLA4*, which in *Saccharomyces cerevisiae* is required for polarized cell growth and proper timing of cytokinesis (Cvrčková et al. 1995) and is also involved in activating cellular signaling cascades in response to severe osmotic stress (Raitt 2000; Viéitez et al. 2020). The second is *AKAP7* (synonym of *AKAP18*, *AKAP15*), a gene that belongs to the family of A-kinase-anchoring proteins (AKAPs) that tether the ubiquitous cAMP-dependent protein kinase (PKA) and localize it to subcellular compartments, thus bestowing specificity to cAMP/PKA (Trotter et al. 1999; Dema et al. 2015). The various isoforms of *AKAP7* in humans and mice are involved in

regulating the activities of Ca^{2+} and Na^{+} channels at the membrane surface (AKAPα,β; Gray et al. 1998; Tibbs et al. 1998) and in mediating the increase of aquaporin-2 (*AQP2*) insertion into the apical surface of the cell membrane (AKAPδ), thus participating in the maintenance of cell water balance (Horner et al. 2012). The third orthogroup was identified as *OACYL*, a gene that lacks thorough investigation into its function. However, all members of this orthogroup were annotated as bearing domain similarity to DnaJ/Hsp40, a molecular chaperone that is highly conserved across prokaryotes and eukaryotes (Hennessy et al. 2000) and which acts in concert with DnaK/Hsp70 and GrpE to refold denatured or incorrectly folded proteins (Szabo et al. 1996; Winter and Jakob 2004; Celaya et al. 2016). The concerted activity of these enzymes helps ensure normal protein folding under oxidative stress in *Salmonella* (Kim et al. 2018, 2021), although numerous other functions have been proposed for DnaJ homologues including chaperoning of steroid receptors (Moffatt et al. 2008), regulation of the cell cycle (Tsai et al. 2006), and tolerance to heat and cellular toxins such as ethanol (Susin et al. 2006).

Micractinium tetrahymenae must perform in two distinct environments given the facultative symbiosis with *T. utricullariae*, inside the ciliate and in the free-living state, which likely require different adaptations that could be mediated by the different sets of paralogs within the *M. tetrahymenae* genome. This pattern of adaptation is supported by the differential expression of multiple members within a given gene family between free-living and endosymbiotic *M. tetrahymenae*. One such example is *ABCG2*, whose members are involved in suberin production and maintaining pH and osmotic balance in *Arabidopsis thaliana* (Yadav et al. 2014; de Silva et al. 2021) and are involved in the expulsion of a broad array of xenobiotics and endogenous compounds from the cell in humans and mice (Kusuhara and Sugiyama 2007). Two other gene families fitting this pattern are *gdhA*, coding for glutamate dehydrogenase, which performs the reversible conversion of glutamate to α-ketoglutarate and ammonia and is thus implicated in maintaining ammonia homeostasis in mammals (Spanaki and Plaitakis 2012) and submerged macrophytes (Xian et al. 2020); and endoglucanase E3.2.1.4, which can hydrolyze a broad range of polysaccharide substrates in rice (Yoshida and Komae 2006).

Physiological Changes in *M. tetrahymenae* Upon Transitioning to Endosymbiosis

The resources that are available to *M. tetrahymenae* likely shift upon switching to endosymbiosis. Acquisition of nitrogen by free-living chlorophytes can be performed via assimilatory nitrate reduction, whereby extracellular nitrate is imported into the cell and reduced to nitrite by nitrate reductase (*NR*), which is further reduced to ammonium by nitrite reductase (*nirA*; Guerrero et al. 1981). The downregulation of *nirA*, which produces the controlling enzyme of this pathway in *Arabidopsis* (Takahashi et al. 2001), suggests a departure from reliance on extracellular nitrate in endosymbiotic *M. tetrahymenae*. A potential parallel is seen with *P. bursaria* and *C. variabilis* in that endosymbiotic *C. variabilis* switch to using amino acids from *Paramecium* as a nitrogen source (Kato et al. 2006; Blanc et al. 2010; Kodama et al. 2014; He et al. 2019). Leveraging the algae's reliance on amino acids, *P. bursaria* appears to control its symbiont population by adjusting subsidies of glutamine (He et al. 2019). Such a transition in which nitrogen sources are obtained and how they are metabolized might also be present in *M. tetrahymenae*.

Plants and algae polymerize excess glucose into starch as a means for energy storage. In short, starch biosynthesis is a three-step process that is initiated by glucose-1-phosphate adenylyltransferase (EC:2.7.7.27, *glgC*), which takes glucose-1-phosphate and ATP as substrates to produce ADP-glucose. Next, starch synthases (EC:2.4.1.21, *glgA* and EC:2.4.1.242, *WAXY*) add ADP-glucose molecules to a growing amylose chain, which is then concatenated as a branch to a growing cluster of other such chains by starch branching enzymes (EC:2.4.1.18, *glgB*; Martin and Smith 1995). The starch molecules can then be hydrolyzed by isoamylase (EC:3.2.1.68, *treX*) to produce maltodextrin (Yokobayashi et al. 1970), which is then further hydrolyzed into maltose molecules by alpha-amylase (EC:3.2.1.1, *amyA*) (Janeček 1997). In endosymbiotic *M. tetrahymenae*, all the aforementioned genes leading to the production of starch and its subsequent hydrolysis are collectively downregulated. One possible reason for this expression pattern is that endosymbiotic *M. tetrahymenae* transfer excess sugars to the *Tetrahymena* host, a phenomenon that is well-documented between *C. variabilis* and *P. bursaria* (Brown and Nielsen 1974; Pardy et al. 1989) and is likely also performed by *M. conductrix* based on experimental evidence of maltose excretion (Arriola et al. 2018).

The Evolution of Oxidation Resistance Genes in *T. utricullariae*

We discovered that the genome of *T. utricullariae* was less differentiated than *M. tetrahymenae* from its congeners regarding its size, the number of genes it possesses, and in the number of rapidly expanding orthogroups. The best-supported signal of adaptation that we found was in the expansion of an orthogroup bearing TLD, a domain that characterizes the *OXR1* oxidation resistance gene family and is evolutionarily conserved across eukaryotes (Volkert et al. 2000; Durand et al. 2007). The TLD domain is alone sufficient to bestow proteins with protective properties against oxidative stress (Oliver et al. 2011) and has been demonstrated to play essential roles in oxidation resistance across distantly related organisms including mice (Oliver et al. 2011), *Anopheles gambiae* (Jaramillo-Gutierrez et al. 2010), and *Saccharomyces cerevisiae* (Elliott and Volkert 2004). This gene family also contains a high frequency of tandem duplicates in *T. utricullariae*, but given the high background rate of single-exon genes in *Tetrahymena* and the inference that LTRs only constitute 0.1% of the genome, we did not pursue LTRs as a causative factor. Tandem duplicates have previously been found to be concentrated in and near the centromeres of *T. thermophilia* MIC chromosomes and were hence proposed to be the consequence of a greater rate of illegitimate recombination events in this region (Hamilton et al. 2016). We therefore suggest that the next step towards identifying the mechanisms underlying the production of new copies within this gene family should involve the production of a MIC genome assembly to explore this possibility. This could perhaps be accomplished by using chromosomal conformation capture techniques as the MIC and MAC likely possess different chromatin geometries owing to their differences in ploidy and transcriptional profiles.

ROS are reduced forms of oxygen that are byproducts of photosynthesis and respiration (Muller 2000; Han et al. 2001; Edreva 2005) that can also function as cellular signaling molecules (Laloi et al. 2004). ROS are damaging in excess

concentrations to organisms as they are more reactive to biomolecules than molecular oxygen (Lesser 2006; Khorobrykh et al. 2020). Extracellular ROS production is common among aquatic organisms (Hansel and Diaz 2021) and, in eukaryotic phytoplankton, extracellular H_2O_2 and O^{2-} production has been reported to span from 0.01 to 10^5 fmol cell^{-1} h^{-1} (Diaz and Plummer 2018). ROS production can be increased in response to cellular stress and can trigger ROS waves that function in cell-to-cell communication, a phenomenon originally described in vascular plants (Miller et al. 2009) that is evolutionarily conserved across eukaryotes (Fichman et al. 2023). Among the species that have been evaluated and in which stress-elicited ROS waves were detected in the form of H_2O_2 bursts were the chlorophytes *Chlamydomonas reinhardtii* and *Chlorella vulgaris*, the latter a close relative of the *Micractinium* and *Chlorella* symbionts of ciliates (Fichman et al. 2023). Thus, ROS concentrations may be elevated inside the ciliate cytoplasm due to stress-induced signaling by the algal symbionts and due to the algae's cellular respiration and photosynthesis.

In *P. bursaria*, ROS production by algae has been proposed as a destabilizing influence on the endosymbiosis (Kawano et al. 2004) and at least two mediation strategies have been identified. The first is behavioral and involves the expulsion of algae by *P. bursaria* when oxygen concentrations in the growth media are experimentally elevated, thus stimulating ROS production by the algae (Kawano et al. 2010). The second is the evolution of greater resistance towards ROS species, which has been demonstrated experimentally by a greater LD_{50} of H_2O_2 in *P. bursaria* in both aposymbiotic states (ciliates where the algal endosymbionts have been experimentally removed) and algal-bearing states compared with *Paramecium* species that do not establish endosymbiosis with algae (Kawano et al. 2004). Therefore, we postulate that the convergent expansions of TLD-bearing gene families in *P. bursaria* (Cheng et al. 2020) and *T. utriculariae* constitute a shared evolutionary response to oxidative stress caused by their algal endosymbionts.

Conclusion

Identifying how participating lineages evolve during the early stages of endosymbiosis is pivotal to understanding how stability is maintained in the partnership. To this end, we analyzed the genomes of *T. utriculariae* and *M. tetrahymenae*, a ciliate host and its endosymbiotic algae. Based on the principal of maximum parsimony and the recent speciation event producing the *T. utriculariae* lineage from a nonsymbiotic ancestor, the endosymbiosis can be reasoned to be relatively young on geological timescales. Rapid expansions in gene families were discovered in both organisms that appear to be evolutionary responses to stress, consistent with the idea that novel paralogs can facilitate the gain of new functions that help organisms adapt to the steep biotic and abiotic clines that would be expected in a nascent endosymbiosis. These include a TLD domain-bearing orthogroup in *T. utriculariae* that represents an example of convergent evolution between this ciliate and *P. bursaria*, and which is likely an adaptation to elevated ROS levels from the algal symbionts. In *M. tetrahymenae*, an overrepresentation of single-exon tandem repeats was also discovered in the rapidly expanding orthogroups, which highlights LTR retrotransposons as potential vehicles for adaptation in this species. Among the rapidly expanding gene families that contain single-exon tandem repeats were ones whose candidate functions involve stress responses, such as ensuring proper folding of proteins under oxidative stress and activating intracellular signaling in response to osmotic stress. Additionally, a differential expression analysis of mRNA transcripts of *M. tetrahymenae* in a free-living versus an endosymbiotic state revealed transitions in nitrogen sources, cell wall restructuring, and the discontinuation of storage of excess sugars. These expression patterns could underlie parallel physiological changes that are elicited upon endosymbiosis in *M. tetrahymenae* and the algal symbionts of *P. bursaria*, *C. variabilis* and *M. conductrix*. Thus, the *T. utriculariae* and *M. tetrahymenae* genomes chronicle adaptation and the nature of host-endosymbiont crosstalk in the early stages of this association and help elucidate how this major evolutionary step in biological organization occurs between ciliates and algae.

Materials and Methods

Strain Selection

A search for the closest relatives of our focal species was performed using the web-based version of BLASTn (Altschul et al. 1990) via the NCBI website. The publicly available *M. tetrahymenae* ITS1-5.8S rRNA-ITS2 sequence (NCBI accession: LT605003.1) and *T. utriculariae* 18S rRNA sequence (NCBI accession: LT605001.1) were queried against the complete nonredundant (nr/nt) database. The sequences for the top 100 HSPs of each search were filtered to remove duplicates and subjected to multiple sequence alignment with the corresponding query sequence in MAFFT v7.490 (Katoh and Standley 2013). The model of nucleotide substitution best fitting each alignment was then inferred using ModelFinder (Kalyaanamoorthy et al. 2017), identifying SYM + I + G for Chlorellaceae and GTR + I + G for *Tetrahymena* according to BIC scores. Phylogenetic tree inference was performed in MrBayes v3.2.6 (Ronquist et al. 2012) using two runs each with four MCMCs of a length of 50,000,000 generations. Consensus trees were calculated for each alignment following a 25% burn-in. The next closest relatives to *M. tetrahymena* and *T. utriculariae* were selected based on nearest phylogenetic distance and the criteria that the strain could be obtained from a culture collection (supplementary fig. S3a and b, Supplementary Material online). *Tetrahymena malaccensis* strain 23b was chosen as the next closest relative to *T. utriculariae* as a draft genome is not yet available for this strain and, given that it mates with *T. malaccensis* strain 436, a draft genome would help enable this species' further use as a model system. *Micractinium* sp. strain CCAP 211/92 was identified as the corresponding algal strain for sequencing. The sequence alignments and accompanying MrBayes code blocks are supplied as supplementary Data Files S2 and S3, Supplementary Material online.

Culture Source and Maintenance

A fresh culture of *T. malaccensis* strain 23b (Tetrahymena Stock Center Strain ID: SD01730) was obtained from the Tetrahymena Stock Center, Cornell University, and *M.* sp. strain CCAP 211/92 was obtained from the Culture Collection of Algae and Protozoa, The Scottish Association for Marine Science, Oban, Scotland. *Tetrahymena utriculariae* was isolated by picking individual single cells from a disrupted trap of *Utricularia reflexa* originating from Okavango Swamp, Botswana, by hand using a pipette under a dissecting microscope

and transferring each to rich wheat grass medium (doi:10.1101/pdb.rec12079) in 40 μL volumes. A single isolate was maintained as strain "A2B2". *Micractinium tetrahymenae* was isolated from *T. utricullariae* A2B2 by transferring individual cells of *T. utricullariae* through five sequential droplets of Woods Hole Culture Medium (WC) media (Guillard and Lorenzen 1972) to wash away extracellular algae. The rinsed ciliate cells were subsequently plated on WC agar, which resulted in the bursting of the ciliates, thus releasing the algae. Algal cells growing from this plate were established in liquid culture and then replated in serial dilutions to obtain a clonal lineage, designated strain "UKON-001". For standard culturing, *T. malaccensis* 23b was maintained in SPP media (Gorovsky et al. 1975), *T. utricullariae* in rich wheat grass medium, *M. tetrahymenae* in WC media, and *M.* sp. CCAP 211/92 in modified Bold's Basal Media (MBBM) media with ammonium replacing nitrate as the nitrogen source (Frickel et al. 2016) at 20 °C.

Nuclear Isolation of *T. malaccensis* 23b and HMW DNA Isolations

Two hundred milliliters of *T. malaccensis* 23b cells in early stationary phase were pelletized at 4 °C and 1800 × *g* (rotor break off) for 15 minutes. Cells were resuspended in 14 mL of nuclear isolation buffer (NIB: 292 mM sucrose, 210 μM $MgCl_2$, 0.1% spermidine [v/v], pH 6.75, with 0.125% NP-40 [v/v]) and incubated for five min on ice, vortexing every 15 seconds. The resulting suspension was filtered across a 10 μm nylon mesh and centrifuged at 4 °C and 1200 × *g* for 12 minutes. The pellet was rinsed in 15 mL NIB and repelletized at 4 °C and 1400 × *g* for 12 minutes. The pellet was resuspended in 1 mL Tris–EDTA buffer and stained with DAPI in a 3 μM final concentration. Micro (MIC) and macronuclear (MAC) fractions were then separated using a BD FACSAria™ III with a gating strategy using DAPI intensity, size (FSC), and granularity (SSC). Only MAC fractions with over 99% purity were used for DNA isolation. A similar pipeline was attempted for *T. utricullariae* but was unsuccessful.

DNA was isolated from five samples using the Wizard HMW DNA Isolation Kit (Promega catalogue number A2920). The first sample was highly enriched for DNA originating from the *T. malaccensis* 23b MAC and was produced from the sorted fraction. The second and third were produced from pelletized intact cells from 200 mL of *T. malaccensis* 23b culture and 400 mL of *T. utricullariae* culture, the latter of which was washed across a 10 μm mesh to deplete any bacterial or fungal cells present in the culture. The fourth and fifth samples were from 400 mL of each *M. tetrahymenae* and *M.* sp. CCAP 211/92 cultures in early stationary phase, which were pelletized and pulverized in liquid nitrogen prior to chemical lysis. The DNA isolations underwent library preparation with the SMRTbell Express Template Prep Kit 2.0 and sequencing on a PacBio Sequel II with one SMRT cell 8 M devoted to each sample.

RNAseq of Ciliates and Algae

We performed an experiment to identify genes that are differentially expressed in *M. tetrahymenae* in a free-living state versus a state of endosymbiosis within *T. utricullariae*. Five monocultures of *M. tetrahymenae* and five cultures of *T. utricullariae* bearing symbiotic *M. tetrahymenae* were grown to late log growth phase in WC media under 50 μmol $m^{-2}s^{-1}$ constant light intensity at 18 °C. Free-living *M. tetrahymenae* cells were separated from the *T. utricullariae* cells in the "symbiotic" treatment by filtering the culture across a cell strainer with a 10 μm mesh size and subsequently washing the cells with 30 mL of WC media, at which point a negligible density of free-living *M. tetrahymenae* was observed. Cells from all cultures were pelletized at 4 °C and 1800 × *g* (rotor break off) and immediately snap-frozen in liquid nitrogen. Cell pellets were resuspended in 350 μL of buffer RLT from the RNeasy Micro Kit (Qiagen, catalogue number: 74004) supplemented with 0.143 M β-mercaptoethanol, transferred to BeadBug™ prefilled tubes, 2.0 mL capacity (Merck, catalogue number Z763764) and disrupted on a MP Biomedicals FastPrep-24™ 5G using three consecutive 120 s runs at a speed of 6 m/s. RNA was isolated from the lysate using the RNeasy Micro Kit following the manufacturer's instructions. To aid in gene model inference, RNA was also isolated from cultures of *T. malaccensis* 23b and *M.* sp. CCAP 211/92 in late log growth phase, pelletized according to the above specifications for *T. utricullariae*. All RNA samples underwent library preparation that included the enrichment of eukaryotic mRNA via poly-A pulldown and were sequenced using 2 × 150 PE chemistry on an Illumina NovaSeq 6000 targeting 30 Gb of raw data for each the *T. malaccensis* 23b and *M.* sp. CCAP 211/92 samples, 9 Gb per monoculture sample of *M. tetrahymenae*, and 18 Gb per *T. utricullariae* sample.

Genome Assembly

We used the long-read sequencing data from the MAC-enriched and whole-cell gDNA samples of *T. malaccensis* 23b to devise and benchmark an assembly pipeline that we could then use to assemble the MAC genome of *T. utricullariae*. We built the pipeline around two features of *Tetrahymena* genomes that theoretically could be used to tease apart MAC and MIC DNA sequences. The first are the chromosomal breakage sequences (CBS), which are 15-bp-long motifs that are highly conserved among *Tetrahymena* (Hamilton et al. 2016). CBSs are specific to the MIC and mark the chromosomal coordinates that are cleaved during the production of the MAC genome (Yao et al. 1987). By removing reads bearing CBS from the raw data, the assembly of chimeric contigs would be prevented as such reads could bridge MAC-destined and MIC-specific genomic sequences. The resulting assembly would then be expected to contain MIC-specific contigs that represent stretches of chromosome between MAC-destined fragments and separate MAC contigs. The second biological characteristic of the nuclear dimorphism we leveraged is the difference in ploidy between the MAC and MIC genomes of *Tetrahymena*, in that the much larger MAC contains >20-fold greater DNA content relative to the MICs (Prescott 1994). Thus, the MIC-specific contigs could be identified on the basis of their much lower coverage and removed from the dataset, producing a MAC-only assembly.

We implemented this pipeline by first removing raw reads bearing CBS motifs from the whole-cell *T. malaccensis* 23b sequencing dataset, which were present in 0.13% of the reads. Given the short length of CBSs and the low error rate of PacBio HiFi CCS reads, we decided to remove reads only if they contained a 15-mer bearing 100% identity to a given CBS sequence or its reverse complement. As an additional quality control step to further validate the purity of the sorted MACs from which the MAC-enriched sequencing data were produced, we checked the MAC-enriched *T. malaccensis*

23b reads for CBS motifs, again requiring 100% identity to a CBS, and found none. Contigs were then assembled for both the MAC-enriched reads and the whole-cell read set with CBS-bearing sequences removed using Flye v2.9.1 (Kolmogorov et al. 2019). The raw reads were then mapped back to these preliminary assemblies using minimap v2.24 (Li 2018) to generate coverages of each contig. The contigs were then processed with PurgeHaplotigs v1.1.2 (Roach et al. 2018), which identifies assembly artefacts on the basis of redundancy and coverage. The degree of similarity between the MAC-enriched assembly and the MAC assembly produced from the whole-cell sample was evaluated by performing whole-genome alignment in nucmer v4.0.0 of the MUMmer4 package (Marçais et al. 2018). Assembly completeness was evaluated with BUSCO v5.4.3 (Manni et al. 2021) run on "genome" mode with the alveolata_odb10 lineage specified. Additional quality metrics were calculated using QUAST v5.0.2 (Mikheenko et al. 2018). We then repeated the assembly pipeline for *T. utricullariae*, in which 0.64% of the reads bore CBS and were removed. An additional decontamination step was performed since reads of at least algal and bacterial origin were expected to be present. This was accomplished using Kraken 2.0.7-beta (Wood et al. 2019) to classify the raw *T. utricullariae* reads by taxonomy. Only those annotated as being ciliophoran (or child terms) were retained for further processing

We screened the assemblies for mitochondrial sequences by querying the assemblies against three *Tetrahymena* mitochondrial genomes (NCBI accessions AF396436.1, DQ927303.1, and DQ927305.1) using BLASTn v2.10.0+ (Altschul et al. 1990) and found that the ~52 kb *T. utricullariae* and the ~54 kb *T. malaccensis 23b* mitochondrial genomes were recovered as single linear contigs, consistent with the observation of linear mitochondrial genomes of other *Tetrahymena* (Brunk 2003; Moradian et al. 2007).

Algal genome assemblies were produced using hifiasm v0.16.1 (Cheng et al. 2021) and curated using PurgeHaplotigs. Assembly metrics were calculated using BUSCO with the chlorophyta_odb10 lineage specified and with QUAST. Circular organellar assemblies were produced by rerunning hifiasm using only reads larger than 21 kb. The organellar origin of the assemblies was determined by querying the contigs against the publicly available *C. variabilis* and *C. sorokiniana* plastid and mitochondrial assemblies (NCBI accessions: CM009642.1, CM009641.1, KP271968.1, KP271969.1) using BLASTn. Contaminating organellar contigs were identified and removed from the original hifiasm assemblies by querying each contig against the circular organellar contigs using BLASTn, producing a nuclear-only assembly for each alga. All subsequent comparative genomic analyses were performed on the nuclear genome assemblies for both *Tetrahymena* and the algae. Additionally, prior to further comparative genomic analysis, organellar contigs were removed from the publicly available *Tetrahymena* and trebouxiophyte genome assemblies and any corresponding gene models. For the sake of resource generation, the *Tetrahymena* and *Micractinium* organellar assemblies are submitted alongside the nuclear assemblies.

Gene Model Inference and Functional Annotations

The nuclear contigs of *M. tetrahymenae*, *M.* sp. CCAP 211/92, *T. utricullariae*, and the two *T. malaccensis* 23b assemblies were soft-masked with RepeatMasker v4.0.7 (Smit & Green, RepeatMasker at http://www.repeatmasker.org), using a de novo inferred library from RepeatModeler v2.0.1 (Flynn et al. 2020) and the existing *Tetrahymena* and *Chlorella* repeat libraries (specified using "-species"). Raw RNA sequencing reads were filtered and trimmed using fastp v0.19.6 (Chen 2023) with the settings "poly_g_min_len 10" and "poly_x_min_len 10". De novo transcriptomes were assembled in Trinity v2.5.1 (Grabherr et al. 2011). The resulting assemblies were then used to construct spliced transcript alignments with PASA v2.5.2 (Haas 2003). The spliced alignments with complete ORFs were then used to train the gene prediction software Augustus v3.2.2 (Stanke et al. 2008). Publicly available proteomes from close relatives of our focal species (supplementary table S2, Supplementary Material online) were aligned to the contigs using miniprot v0.12 (Li 2023). The miniprot protein alignments and PASA alignments were then supplied as hints for the final round of gene prediction in Augustus. The miniprot alignments were also supplied to GeneMark-EP, packaged in the GeneMark-ES/ET/EP Suite v4.71 (Brůna et al. 2020), to generate a second source of ab initio gene models. EVidenceModeler v2.1.0 (Haas et al. 2008) was then used to generate consensus gene models from the PASA alignments, miniprot alignments, Augustus and GeneMark gene models. Summary statistics of the gene models were generated using AGAT v1.0.0 (https://www.doi.org/10.5281/zenodo.3552717) and by calculating the BUSCO scores of the proteomes with the appropriate lineage-specific database specified (alveolata_odb10 lineage for *Tetrahymena* and chlorophyta_odb10 for *Chlorella-Micractinium*). The full pipeline of EDTA v2.0.0 (Ou et al. 2019) was also run to provide a more comprehensive annotation and classification of the transposable elements. Functional annotations of the gene models were inferred for the publicly available *Tetrahymena* and algal genome assemblies and for our four genome assemblies using eggNOG-mapper v2.1.12 (Cantalapiedra et al. 2021) for KEGG pathways, preferred gene names, and GO terms and Interproscan v5.39-77.0 (Jones et al. 2014) for PFAM terms.

Phylogenetic Inference and Fossil Dating

One hundred and one publicly-available proteomes representing a broad taxonomic sampling of the eukaryotic Diaphoratickes superclade (supplementary table S2, Supplementary Material online) were combined with the proteomes generated in the current study, producing an initial dataset of 105 taxa. Protein sequences corresponding to eukaryotic BUSCOs were identified by running BUSCO on "proteome" mode on each proteome specifying the eukaryota_odb10 lineage. A representative copy was chosen at random if the BUSCO was complete but duplicated. Amino acid residues were aligned within each BUSCO across taxa using Muscle v5.1 (Edgar 2004). Alignments were filtered using Gblocks v0.91b (Castresana 2000) with default parameters, except "-b = h". BUSCOs that were present in <25% of the taxa or that had filtered alignment lengths of <50 positions were removed from the dataset. Additionally, *Euplotes vannus* was removed as it had >75% missing data.

To investigate the sensitivity of the comparative genomic pipeline (below) to the method of phylogenetic inference used, we performed tree construction using two independent methods, IQtree v2.2.3 (Minh et al. 2020) and ASTRAL-IV v1.19.4.5 (Zhang and Mirarab 2022). For the IQtree2 analysis, the model of protein evolution best fitting each alignment was inferred using ModelFinder (Kalyaanamoorthy et al. 2017), implemented in IQtree v2.2.3 (Minh et al. 2020). Any alignments that failed the composition chi^2 test were

removed, resulting in a final set of 50 BUSCO alignments for the 104 taxa. Phylogenetic tree inference was performed using IQtree2 on the concatenated alignment of 50 BUSCOs, which had a final length of 7,181 positions, using a partitioned analysis that specified the models of AA evolution identified by ModelFinder for each individual BUSCO and by running 10,000 ultrafast bootstrap replicates (Hoang et al. 2018). Gene trees were inferred to provide as input for the ASTRAL-IV analysis for the same 50 BUSCOs using RAXML v8.2.9 (Stamatakis 2014), for which 20 independent runs on distinct starting trees, a gamma model of rate heterogeneity, and automatic protein substitution model selection were specified. ASTRAL-IV was then run using 32 initial rounds of placement and 32 subsampling rounds per exploration step (-r 32 -s 32). The final alignment and partition are supplied in supplementary Data Files S4 and S5, Supplementary Material online.

The phylogenies were time-calibrated using MCMCtree in PAML v4.9 (Yang 2007) using seven fossil calibrations following the age specifications of Morris et al. 2018 and Strassert et al. 2021 (supplementary table S3, Supplementary Material online). We performed a sensitivity analysis by repeating the MCMCtree pipeline specifying either the skew-normal, uniform, or Cauchy prior node distribution using both the IQtree2 and ASTRAL-IV as input trees and four fossil calibration schemes. The first fossil calibration scheme consisted of the full set of seven calibration points. The second and third involved the removal of either the Ciliophora or Chlorophyta fossils (the fossils "closest" to the focal taxa of this paper on the phylogenies), and the fourth involved the removal of the embryophyte, angiosperm-gymnosperm, and angiosperm fossils, as they represented a bias within streptophytes by virtue of having more fossil calibrations relative to other clades. Input files were prepared using MCMCtreeR (Puttick 2019). MCMCtree was performed with each of the aforementioned combinations of parameters using two independent runs, each with a length of 20,000,000 generations, sampling every 100 generations, after a 200,000 generation burn-in. Each run was initialized from a different seed that was randomly generated from a uniform distribution set to range from 1 to 1,000,000 using the "runif" function in R. For each run we assumed independent rates model for the molecular clock, birth-death priors, $\alpha = 2$ and $\beta = 14.21$ for the Dirichlet-gamma prior for the mean substitution rate, and $\alpha = 2$ and $\beta = 2$ for the Dirichlet-gamma prior for the rate drift parameter. Convergence and sufficiently large ESS (>200) of parameter estimates was confirmed using Tracer v1.7.2 (Rambaut et al. 2018).

Inference of Rapidly Contracting and Expanding Gene Families

Orthogroups were inferred using the subtrees representing the *Chlorella-Micractinium* and *Tetrahymena* clades, extracted from the skew-normal MCMCtree output tree, and the corresponding taxa's proteomes using OrthoFinder v2.5.4 (Emms and Kelly 2019) specifying BLASTp for homolog detection. CAFE5 (Mendes et al. 2021) was then run using the orthogroup count file and the aforementioned subtrees, restricted to taxa with a minimum single-copy complete proteome BUSCO score of 90% (alveolata_odb10 lineage for *Tetrahymena* and chlorophyta_odb10 for *Chlorella-Micractinium*) to mitigate differences in genome completeness. The optimal number of gamma categories was determined for each tree by allowing K to vary from 2 to 5 and by repeating each model five times to evaluate convergence and -lnl values. Additionally, a Poisson distribution was specified at the root. A *P*-value cutoff of 0.01 was implemented to designate an orthogroup as undergoing significant change in copy number along a branch. Given that the branching order of the *Micractinium-Chlorella* clade differed between the IQtee2 and ASTRAL-IV analyses in that *C. vulgaris-C. variabilis* were sister to *Micractinium* in the IQtree2 phylogeny and *C. sorokiniana-C. ohadii* were sister to *Micractinium* in the ASTRAL-IV phylogeny, we repeated this pipeline using the MCMCtree time-calibrated phylogeny inferred using the ASTRAL-IV input tree starting from the OrthoFinder step to determine if the topological incongruency affected the results.

Gene Duplication Analysis

Segmental and tandem duplications were identified with iAdhore v3.0 (Proost et al. 2012) using the OrthoFinder-inferred orthogroups as the input for the homology table. The analysis was run on collinear mode with a gap size of 30, q-value cutoff of 0.75, FDR multiple hypothesis correction, *P*-value cutoff of 0.01, three anchor points, and using the gg2 alignment method.

Comparative Transcriptomic Analysis of *M. Tetrahymenae*

Raw RNA sequencing reads were trimmed and filtered with fastp as described above. Splice-aware alignments were conducted against the *M. tetrahymenae* and *T. utriculariae* nuclear genome assemblies with STAR v2.7.3a (Dobin et al. 2013) using default mapping parameters. A read count matrix of transcripts was compiled with featureCounts v2.0.0 (Liao et al. 2014) that discarded multi-mapping, multi-overlapping, and chimeric reads, counted only primary alignments, and required the mapping of both ends to the transcript. Differentially expressed genes were identified between the endosymbiotic and free-living *M. tetrahymenae* samples using the R package DEseq2 v1.36.0 (Love et al. 2014) specifying the "apeglm" shrinkage estimator (Zhu et al. 2019).

Functional Enrichment Analysis

An enrichment analysis was performed for the differentially expressed genes in *M. tetrahymenae* in a symbiotic versus free-living state and the genes contained in significantly expanding orthogroups in *M. tetrahymenae* and *T. utriculariae* using the "enricher" function of the R package clusterProfiler v4.4.4 (Wu et al. 2021). A *P*-value cutoff of 0.01 was used with FDR *P*-value corrections, requiring a minimum gene set size of 20 members and default settings for other parameters.

Analysis of TLD Domain-bearing Novel Protein Family in *Tetrahymena*

FEL (Kosakovsky Pond and Frost 2005), implemented in HyPhy v2.5.60 (Kosakovsky Pond et al. 2020), was used to identify positions in the codon-aware multiple sequence alignment under positive and negative selection for the TLD-bearing members of the *Tetrahymena* orthogroup OG0000067 in *T. utriculariae* using a phylogeny-wide analysis. As a three-dimensional alternative to an amino acid alignment, we used ColabFold (Jumper et al. 2021; Mirdita et al. 2022) implemented in ChimeraX (Pettersen et al. 2021) specifying PDB templates and energy-minimization to predict the structure of a representative TLD-bearing

copy of this orthogroup in *T. utricularia*e, selected as it had the highest RPKM value. The codon-aware alignment of these copies is also provided as supplementary Data File S1, Supplementary Material online.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

This material is based upon work supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2109477, awarded to J.B.K. and is further funded by a grant made by the Gordon and Betty Moore Foundation to LB (grant no. 9196; https://doi.org/10.37807/GBMF9196) and by the long-term research development project of the Czech Academy of Sciences (RVO 67985939) awarded to L.A. We acknowledge the flow cytometry center (FlowKon) at the University of Konstanz for the use of flow cytometers and the expert support in instrument usage and data analysis. We also thank Sandra Hamm of the Bioimaging Center (BIC) at the University of Konstanz for help in obtaining microscopy images of the study organisms. The authors would like to thank the Stony Brook Research Computing and Cyberinfrastructure and the Institute for Advanced Computational Science at Stony Brook University for access to the high-performance SeaWulf computing system, which was made possible by $1.85M in grants from the National Science Foundation (awards 1531492 and 2215987) and matching funds from the Empire State Development's Division of Science, Technology and Innovation (NYSTAR) program (contract C210148).

Data Availability

The sequencing data and genome assemblies produced during this study are available under NCBI BioProject PRJNA 1120015. Strains *Micractinium tetrahymenae* UKON-001 and *Tetrahymena utricularia*e A2B2 have been deposited in the Culture Collection of Algae and Protozoa under accession numbers CCAP 248/22 and CCAP 1630/23.

References

Adamec L. Oxygen concentrations inside the traps of the carnivorous plants *Utricularia* and *Genlisea* (Lentibulariaceae). *Ann Bot*. 2007:100(4):849–856. https://doi.org/10.1093/aob/mcm182.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990:215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Arriola MB, Velmurugan N, Zhang Y, Plunkett MH, Hondzo H, Barney BM. Genome sequences of *Chlorella sorokiniana* UTEX 1602 and *Micractinium conductrix* SAG 241.80: implications to maltose excretion by a green alga. *Plant J*. 2018:93(3):566–586. https://doi.org/10.1111/tpj.13789.

Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, Lindquist E, Lucas S, Pangilinan J, Polle J, *et al*. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell*. 2010:22(9):2943–2955. https://doi.org/10.1105/tpc.110.076406.

Brown JA, Nielsen PJ. Transfer of photosynthetically produced carbohydrate from endosymbiotic Chlorellae to *Paramecium bursaria*. *J Protozool*. 1974:21(4):569–570. https://doi.org/10.1111/j.1550-7408.1974.tb03702.x.

Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform*. 2020:2(2):lqaa026. https://doi.org/10.1093/nargab/lqaa026.

Brunk CF. Complete sequence of the mitochondrial genome of *Tetrahymena thermophila* and comparative methods for identifying highly divergent genes. *Nucleic Acids Res*. 2003:31(6):1673–1682. https://doi.org/10.1093/nar/gkg270.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. EggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021:38(12):5825–5829. https://doi.org/10.1093/molbev/msab293.

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000:17(4):540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Celaya G, Fernández-Higuero JA, Martin I, Rivas G, Moro F, Muga A. Crowding modulates the conformation, affinity, and activity of the components of the bacterial disaggregase machinery. *J Mol Biol*. 2016:428(11):2474–2487. https://doi.org/10.1016/j.jmb.2016.04.027.

Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*. 2023:2(2):e107. https://doi.org/10.1002/imt2.107.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021:18(2):170–175. https://doi.org/10.1038/s41592-020-01056-5.

Cheng Y-H, Liu C-FJ, Yu Y-H, Jhou Y-T, Fujishima M, Tsai IJ, Leu J-Y. Genome plasticity in *Paramecium bursaria* revealed by population genomics. *BMC Biol*. 2020:18(1):180. https://doi.org/10.1186/s12915-020-00912-2.

Cvrčková F, De Virgilio C, Manser E, Pringle JR, Nasmyth K. Ste20-like protein kinases are required for normal localization of cell growth and for cytokinesis in budding yeast. *Genes Dev*. 1995:9(15):1817–1830. https://doi.org/10.1101/gad.9.15.1817.

Dema A, Perets E, Schulz MS, Deák VA, Klussmann E. Pharmacological targeting of AKAP-directed compartmentalized cAMP signalling. *Cell Signal*. 2015:27(12):2474–2487. https://doi.org/10.1016/j.cellsig.2015.09.008.

de Silva NDG, Murmu J, Chabot D, Hubbard K, Ryser P, Molina I, Rowland O. Root suberin plays important roles in reducing water loss and sodium uptake in *Arabidopsis thaliana*. *Metabolites*. 2021:11(11):735. https://doi.org/10.3390/metabo11110735.

Diaz JM, Plummer S. Production of extracellular reactive oxygen species by phytoplankton: past and future directions. *J Plankton Res*. 2018:40(6):655–666. https://doi.org/10.1093/plankt/fby039.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics*. 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

Durand M, Kolpak A, Farrell T, Elliott NA, Shao W, Brown M, Volkert MR. The OXR domain defines a conserved family of eukaryotic oxidation resistance proteins. *BMC Cell Biol*. 2007:8(1):13. https://doi.org/10.1186/1471-2121-8-13.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004:32(5):1792–1797. https://doi.org/10.1093/nar/gkh340.

Edreva A. Generation and scavenging of reactive oxygen species in chloroplasts: a submolecular approach. *Agric Ecosyst Environ*. 2005:106(2-3):119–133. https://doi.org/10.1016/j.agee.2004.10.022.

Elliott NA, Volkert MR. Stress induction and mitochondrial localization of Oxr1 proteins in yeast and humans. *Mol Cell Biol*. 2004:24(8):3180–3187. https://doi.org/10.1128/MCB.24.8.3180-3187.2004.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y.

Esteban GF, Fenchel T, Finlay BJ. Mixotrophy in ciliates. *Protist*. 2010:161(5):621–641. https://doi.org/10.1016/j.protis.2010.08.002.

Estrela S, Kerr B, Morris JJ. Transitions in individuality through symbiosis. *Curr Opin Microbiol*. 2016:31:191–198. https://doi.org/10.1016/j.mib.2016.04.007.

Fellows I. Deducer: a data analysis GUI for *R*. *J Stat Softw*. 2012:49(8):1–15. https://doi.org/10.18637/jss.v049.i08.

Fichman Y, Rowland L, Oliver MJ, Mittler R. ROS are evolutionary conserved cell-to-cell stress signals. *Proc Natl Acad Sci U S A*. 2023:120(31):e2305496120. https://doi.org/10.1073/pnas.2305496120.

Finelli MJ, Oliver PL. TLDc proteins: new players in the oxidative stress response and neurological disease. *Mamm Genome*. 2017:28(9-10):395–406. https://doi.org/10.1007/s00335-017-9706-7.

Finelli MJ, Sanchez-Pulido L, Liu KX, Davies KE, Oliver PL. The evolutionarily conserved Tre2/Bub2/Cdc16 (TBC), lysin motif (LysM), domain catalytic (TLDc) domain is neuroprotective against oxidative stress. *J Biol Chem*. 2016:291(6):2751–2763. https://doi.org/10.1074/jbc.M115.685222.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020:117(17):9451–9457. https://doi.org/10.1073/pnas.1921046117.

Fokin SI, Przyboś E, Chivilev SM, Beier CL, Horn M, Skotarczak B, Wodecka B, Fujishima M. Morphological and molecular investigations of *Paramecium schewiakoffi* sp. nov. (Ciliophora, Oligohymenophorea) and current status of distribution and taxonomy of *Paramecium* spp. *Eur J Protistol*. 2004:40(3):225–243. https://doi.org/10.1016/j.ejop.2004.02.001.

Frickel J, Sieber M, Becks L. Eco-evolutionary dynamics in a coevolving host–virus system. *Ecol Lett*. 2016:19(4):450–459. https://doi.org/10.1111/ele.12580.

Fujishima M, Kodama Y. Mechanisms for establishing primary and secondary endosymbiosis in *Paramecium*. *J Eukaryot Microbiol*. 2022:69(5):e12901. https://doi.org/10.1111/jeu.12901.

Gorovsky MA, Yao M-C, Keevert JB, Pleger GL. Isolation of micro- and macronuclei of *Tetrahymena pyriformis*. *Methods Cell Biol*. 1975:9:311–327. https://doi.org/10.1016/S0091-679X(08)60080-1.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011:29(7):644–652. https://doi.org/10.1038/nbt.1883.

Gray PC, Johnson BD, Westenbroek RE, Hays LG, Yates JR, Scheuer T, Catterall WA, Murphy BJ. Primary structure and function of an A kinase anchoring protein associated with calcium channels. *Neuron*. 1998:20(5):1017–1026. https://doi.org/10.1016/S0896-6273(00)80482-1.

Guerrero MG, Vega JM, Losada M. The assimilatory nitrate-reducing system and its regulation. *Annu Rev Plant Physiol*. 1981:32(1):169–204. https://doi.org/10.1146/annurev.pp.32.060181.001125.

Guillard RRL, Lorenzen CJ. Yellow-green algae with chlorophyllide C1,2. *J Phycol*. 1972:8:10–14. https://doi.org/10.1111/j.1529-8817.1972.tb03995.x.

Haas BJ. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*. 2003:31(19):5654–5666. https://doi.org/10.1093/nar/gkg770.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol*. 2008:9(1):R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Hamilton EP, Kapusta A, Huvos PE, Bidwell SL, Zafar N, Tang H, Hadjithomas M, Krishnakumar V, Badger JH, Caler E V, *et al*. Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *Elife*. 2016:5:e19090. https://doi.org/10.7554/eLife.19090.

Han D, Williams E, Cadenas E. Mitochondrial respiratory chain-dependent generation of superoxide anion and its release into the intermembrane space. *Biochem J*. 2001:353(2):411–416. https://doi.org/10.1042/bj3530411.

Hansel CM, Diaz JM. Production of extracellular reactive oxygen species by marine biota. *Ann Rev Mar Sci*. 2021:13(1):177–200. https://doi.org/10.1146/annurev-marine-041320-102550.

He M, Wang J, Fan X, Liu X, Shi W, Huang N, Zhao F, Miao M. Genetic basis for the establishment of endosymbiosis in *Paramecium*. *ISME J*. 2019:13(5):1360–1369. https://doi.org/10.1038/s41396-018-0341-4.

Hennessy F, Cheetham ME, Dirr HW, Blatch GL. Analysis of the levels of conservation of the J domain among the various types of DnaJ-like proteins. *Cell Stress Chaperones*. 2000:5(4):347–358. https://doi.org/10.1379/1466-1268(2000)005<0347:AOTLOC>2.0.CO;2.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018:35(2):518–522. https://doi.org/10.1093/molbev/msx281.

Horas EL, Metzger SM, Platzer B, Kelly JB, Becks L. Context-dependent costs and benefits of endosymbiotic interactions in a ciliate–algae system. *Environ Microbiol*. 2022:24(12):5924–5935. https://doi.org/10.1111/1462-2920.16112.

Horner A, Goetz F, Tampé R, Klussmann E, Pohl P. Mechanism for targeting the A-kinase anchoring protein AKAP18δ to the membrane. *J Biol Chem*. 2012:287(51):42495–42501. https://doi.org/10.1074/jbc.M112.414946.

Hoshina R, Fujiwara Y. Photobiont flexibility in *Paramecium bursaria*: double and triple photobiont co-habitation. *Adv Microbiol*. 2012:2(03):227–233. https://doi.org/10.4236/aim.2012.23027.

Hoshina R, Tsukii Y, Harumoto T, Suzaki T. Characterization of a green *Stentor* with symbiotic algae growing in an extremely oligotrophic environment and storing large amounts of starch granules in its cytoplasm. *Sci Rep*. 2021:11(1):2865. https://doi.org/10.1038/s41598-021-82416-9.

Janeček S. α-amylase family: molecular biology and evolution. *Prog Biophys Mol Biol*. 1997:67(1):67–97. https://doi.org/10.1016/S0079-6107(97)00015-1.

Jaramillo-Gutierrez G, Molina-Cruz A, Kumar S, Barillas-Mury C. The *Anopheles gambiae* oxidation resistance 1 (OXR1) gene regulates expression of enzymes that detoxify reactive oxygen species. *PLoS One*. 2010:5(6):e11168. https://doi.org/10.1371/journal.pone.0011168.

Jenkins BH, Maguire F, Leonard G, Eaton JD, West S, Housden BE, Milner DS, Richards TA. Emergent RNA–RNA interactions can promote stability in a facultative phototrophic endosymbiosis. *Proc Natl Acad Sci U S A*. 2021:118(38):e2108874118. https://doi.org/10.1073/pnas.2108874118.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014:30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021:596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017:14(6):587–589. https://doi.org/10.1038/nmeth.4285.

Kato Y, Ueno S, Imamura N. Studies on the nitrogen utilization of endosymbiotic algae isolated from Japanese *Paramecium bursaria*. *Plant Sci*. 2006:170(3):481–486. https://doi.org/10.1016/j.plantsci.2005.09.018.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013:30(4):772–780. https://doi.org/10.1093/molbev/mst010.

Kawano T, Irie K, Kadono T. Oxidative stress-mediated development of symbiosis in green paramecia. In: Seckbach J, Grube M, editors. *Symbioses and stress*. Dordrecht, Netherlands: Springer; 2010. p. 177–195.

Kawano T, Kadono T, Kosaka T, Hosoya H. Green paramecia as an evolutionary winner of oxidative symbiosis: a hypothesis and

supportive data. *Zeitschrift für Naturforschung C*. 2004:59(7-8): 538–542. https://doi.org/10.1515/znc-2004-7-816.
Khorobrykh S, Havurinne V, Mattila H, Tyystjärvi E. Oxygen and ROS in photosynthesis. *Plants*. 2020:9(1):91. https://doi.org/10.3390/plants9010091.
Kim J-S, Liu L, Fitzsimmons LF, Wang Y, Crawford MA, Mastrogiovanni M, Trujillo M, Till JKA, Radi R, Dai S, *et al*. Dksa–DnaJ redox interactions provide a signal for the activation of bacterial RNA polymerase. *Proc Natl Acad Sci U S A*. 2018:115(50):E11780–E11789. https://doi.org/10.1073/pnas.1813572115.
Kim J-S, Liu L, Vázquez-Torres A. The DnaK/DnaJ chaperone system enables RNA polymerase-DksA complex formation in *Salmonella* experiencing oxidative stress. *mBio*. 2021:12(3):e03443–e03420. https://doi.org/10.1128/mBio.03443-20.
Kodama Y, Suzuki H, Dohra H, Sugii M, Kitazume T, Yamaguchi K, Shigenobu S, Fujishima M. Comparison of gene expression of *Paramecium bursaria* with and without *Chlorella variabilis* symbionts. *BMC Genomics*. 2014:15(1):183. https://doi.org/10.1186/1471-2164-15-183.
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019:37(5): 540–546. https://doi.org/10.1038/s41587-019-0072-8.
Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc Lond B Biol Sci*. 2012:279(1749):5048–5057. https://doi.org/10.1098/rspb.2012.1108.
Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*. 2005:22(5):1208–1222. https://doi.org/10.1093/molbev/msi105.
Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, *et al*. Hyphy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol*. 2020:37(1):295–299. https://doi.org/10.1093/molbev/msz197.
Krasileva KV. The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes. *Curr Opin Plant Biol*. 2019:48:18–25. https://doi.org/10.1016/j.pbi.2019.01.004.
Kusuhara H, Sugiyama Y. ATP-binding cassette, subfamily G (ABCG family). *Pflugers Arch*. 2007:453(5):735–744. https://doi.org/10.1007/s00424-006-0134-x.
Laloi C, Apel K, Danon A. Reactive oxygen signalling: the latest news. *Curr Opin Plant Biol*. 2004:7(3):323–328. https://doi.org/10.1016/j.pbi.2004.03.005.
Lane N, Martin W. The energetics of genome complexity. *Nature*. 2010:467(7318):929–934. https://doi.org/10.1038/nature09486.
Lanzoni O, Fokin SI, Lebedeva N, Migunova A, Petroni G, Potekhin A. Rare freshwater ciliate *Paramecium chlorelligerum* Kahl, 1935 and its macronuclear symbiotic bacterium "*Candidatus* Holospora parva.". *PLoS One*. 2016:11(12):e0167928. 10.1371/journal.pone.0167928.
Lesser MP. Oxidative stress in marine environments: biochemistry and physiological ecology. *Annu Rev Physiol*. 2006:68(1):253–278. https://doi.org/10.1146/annurev.physiol.68.040104.110001.
Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018:34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
Li H. Protein-to-genome alignment with miniprot. *Bioinformatics*. 2023:39(1):btad014. https://doi.org/10.1093/bioinformatics/btad014.
Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014:30(7):923–930. https://doi.org/10.1093/bioinformatics/btt656.
Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol*. 2014:15(12):550. https://doi.org/10.1186/s13059-014-0550-8.
Lowe CD, Minter EJ, Cameron DD, Brockhurst MA. Shining a light on exploitative host control in a photosynthetic endosymbiosis. *Curr Biol*. 2016:26(2):207–211. https://doi.org/10.1016/j.cub.2015.11.052.
Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021:38(10):4647–4654. https://doi.org/10.1093/molbev/msab199.
Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol*. 2018:14(1):e1005944. https://doi.org/10.1371/journal.pcbi.1005944.
Martin C, Smith AM. Starch biosynthesis. *Plant Cell*. 1995:7(7): 971–985. https://doi.org/10.1105/tpc.7.7.971.
Meier R, Wiessner W. Infection of algae-free *Paramecium bursaria* with symbiotic *Chlorella* sp. Isolated from green paramecia. *Eur J Protistol*. 1988:24(1):69–74. https://doi.org/10.1016/S0932-4739(88)80011-7.
Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*. 2021:36(22-23):5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022.
Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*. 2018:34(13):i142–i150. https://doi.org/10.1093/bioinformatics/bty266.
Miller G, Schlauch K, Tam R, Cortes D, Torres MA, Shulaev V, Dangl JL, Mittler R. The plant NADPH oxidase RBOHD mediates rapid systemic signaling in response to diverse stimuli. *Sci Signal*. 2009:2(84):ra45–ra45. https://doi.org/10.1126/scisignal.2000448.
Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.
Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022:19(6):679–682. https://doi.org/10.1038/s41592-022-01488-1.
Moffatt NSC, Bruinsma E, Uhl C, Obermann WMJ, Toft D. Role of the cochaperone tpr2 in Hsp90 chaperoning. *Biochemistry*. 2008:47(31):8203–8213. https://doi.org/10.1021/bi800770g.
Moradian MM, Beglaryan D, Skozylas JM, Kerikorian V. Complete mitochondrial genome sequence of three *Tetrahymena* species reveals mutation hot spots and accelerated nonsynonymous substitutions in Ymf Genes. *PLoS One*. 2007:2(7):e650. https://doi.org/10.1371/journal.pone.0000650.
Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. The timescale of early land plant evolution. *Proc Natl Acad Sci U S A*. 2018:115(10): E2274–E2283. https://doi.org/10.1073/pnas.1719588115.
Moy GW, Mendoza LM, Schulz JR, Swanson WJ, Glabe CG, Vacquier VD. The sea urchin sperm receptor for egg jelly is a modular protein with extensive homology to the human polycystic kidney disease protein, PKD1. *J Cell Biol*. 1996:133(4):809–817. https://doi.org/10.1083/jcb.133.4.809.
Muller F. The nature and mechanism of superoxide production by the electron transport chain: its relevance to aging. *J Am Aging Assoc*. 2000:23(4):227–253. https://doi.org/10.1007/s11357-000-0022-9.
Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM, Heaps SE, Hirt RP, Martin Embley T. Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. *Genome Biol Evol*. 2013:5(12):2285–2303. https://doi.org/10.1093/gbe/evt184.
Oliver PL, Finelli MJ, Edwards B, Bitoun E, Butts DL, Becker EBE, Cheeseman MT, Davies B, Davies KE. Oxr1 is essential for protection against oxidative stress-induced neurodegeneration. *PLoS Genet*. 2011:7(10):e1002338. https://doi.org/10.1371/journal.pgen.1002338.
Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, *et al*. Benchmarking transposable element annotation methods for creation of a streamlined,

comprehensive pipeline. *Genome Biol*. 2019:20(1):275. https://doi.org/10.1186/s13059-019-1905-y.
Pan G, Xu J, Li T, Xia Q, Liu S-L, Zhang G, Li S, Li C, Liu H, Yang L, *et al*. Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. *BMC Genomics*. 2013:14(1):186. https://doi.org/10.1186/1471-2164-14-186.
Pardy RL, Spargo B, Crowe JH. Release of trehalose by symbiotic algae. *Symbiosis*. 1989:7:149–158.
Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci*. 2021:30(1):70–82. https://doi.org/10.1002/pro.3943.
Pitsch G, Adamec L, Dirren S, Nitsche F, Šimek K, Sirová D, Posch T. The green *Tetrahymena utricolariae* n. sp. (Ciliophora, Oligohymenophorea) with its endosymbiotic algae (*Micractinium* sp.), living in traps of a carnivorous aquatic plant. *J Eukaryot Microbiol*. 2017:64(3):322–335. https://doi.org/10.1111/jeu.12369.
Prescott DM. The DNA of ciliated protozoa. *Microbiol Rev*. 1994:58(2):233–267. https://doi.org/10.1128/mr.58.2.233-267.1994.
Proost S, Fostier J, De Witte D, Dhoedt B, Demeester P, Van de Peer Y, Vandepoele K. . i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res*. 2012:40(2):e11–e11. https://doi.org/10.1093/nar/gkr955.
Pröschold T, Darienko T, Silva PC, Reisser W, Krienitz L. The systematics of *Zoochlorella* revisited employing an integrative approach. *Environ Microbiol*. 2011:13(2):350–364. https://doi.org/10.1111/j.1462-2920.2010.02333.x.
Pröschold T, Rieser D, Darienko T, Nachbaur L, Kammerlander B, Qian K, Pitsch G, Bruni EP, Qu Z, Forster D, *et al*. An integrative approach sheds new light onto the systematics and ecology of the widespread ciliate genus *Coleps* (Ciliophora, Prostomatea). *Sci Rep*. 2021:11(1):5916. https://doi.org/10.1038/s41598-021-84265-y.
Puttick MN. MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics*. 2019:35(24):5321–5322. https://doi.org/10.1093/bioinformatics/btz554.
Raitt DC. Yeast Cdc42 GTPase and Ste20 PAK-like kinase regulate Sho1-dependent activation of the Hog1 MAPK pathway. *EMBO J*. 2000:19(17):4623–4631. https://doi.org/10.1093/emboj/19.17.4623.
Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 2018:67(5):901–904. https://doi.org/10.1093/sysbio/syy032.
Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 2018:19(1):460. https://doi.org/10.1186/s12859-018-2485-7.
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012:61(3):539–542. https://doi.org/10.1093/sysbio/sys029.
Šimek K, Pitsch G, Salcher MM, Sirová D, Shabarova T, Adamec L, Posch T. Ecological traits of the algae-bearing *Tetrahymena utricolariae* (Ciliophora) from traps of the aquatic carnivorous plant *Utricularia reflexa*. *J Eukaryot Microbiol*. 2017:64(3):336–348. https://doi.org/10.1111/jeu.12368.
Siozios S, Jimenez PN, Azagi T, Sprong H, Frost CL, Parratt SR, Taylor G, Brettell L, Liew KC, Croft L, *et al*. Genome dynamics across the evolutionary transition to endosymbiosis. *Curr Biol* 2024:34(24):5659–5670.e7. https://doi.org/10.1016/j.cub.2024.10.044.
Sirová D, Adamec L, Vrba J. Enzymatic activities in traps of four aquatic species of the carnivorous genus *Utricularia*. *New Phytol*. 2003:159(3):669–675. https://doi.org/10.1046/j.1469-8137.2003.00834.x.
Sirová D, Borovec J, Picek T, Adamec L, Nedbalová L, Vrba J. Ecological implications of organic carbon dynamics in the traps of aquatic carnivorous *Utricularia* plants. *Funct Plant Biol*. 2011:38(7):583–593. https://doi.org/10.1071/FP11023.
Sirová D, Kreidlová V, Adamec L, Vrba J. The ability of *Tetrahymena utricolariae* (Ciliophora, Oligohymenophorea) to colonize traps of different species of aquatic carnivorous *Utricularia*. *J Eukaryot Microbiol*. 2020:67(5):608–611. https://doi.org/10.1111/jeu.12812.
Spanaki C, Plaitakis A. The role of glutamate dehydrogenase in mammalian ammonia metabolism. *Neurotox Res*. 2012:21(1):117–127. https://doi.org/10.1007/s12640-011-9285-4.
Spanner C, Darienko T, Filker S, Sonntag B, Pröschold T. Morphological diversity and molecular phylogeny of five *Paramecium bursaria* (Alveolata, Ciliophora, Oligohymenophorea) syngens and the identification of their green algal endosymbionts. *Sci Rep*. 2022:12(1):18089. https://doi.org/10.1038/s41598-022-22284-z.
Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014:30(9):1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*. 2008:24(5):637–644. https://doi.org/10.1093/bioinformatics/btn013.
Strassert JFH, Irisarri I, Williams TA, Burki F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat Commun*. 2021:12(1):1879. https://doi.org/10.1038/s41467-021-22044-z.
Susin MF, Baldini RL, Gueiros-Filho F, Gomes SL. GroES/GroEL and DnaK/DnaJ have distinct roles in stress responses and during cell cycle progression in *Caulobacter crescentus*. *J Bacteriol*. 2006:188(23):8044–8053. https://doi.org/10.1128/JB.00824-06.
Szabo A, Korszun R, Hartl FU, Flanagan J. A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. *EMBO J*. 1996:15(2):408–417. https://doi.org/10.1002/j.1460-2075.1996.tb00371.x.
Takahashi M, Sasaki Y, Ida S, Morikawa H. Nitrite reductase gene enrichment improves assimilation of NO_2 in *Arabidopsis*. *Plant Physiol*. 2001:126(2):731–741. https://doi.org/10.1104/pp.126.2.731.
Tibbs VC, Gray PC, Catterall WA, Murphy BJ. AKAP15 Anchors cAMP-dependent protein kinase to brain sodium channels. *J Biol Chem*. 1998:273(40):25783–25788. https://doi.org/10.1074/jbc.273.40.25783.
Trotter KW, Fraser IDC, Scott GK, Stutts MJ, Scott JD, Milgram SL. Alternative splicing regulates the subcellular localization of a-kinase anchoring protein 18 Isoforms. *J Cell Biol*. 1999:147(7):1481–1492. https://doi.org/10.1083/jcb.147.7.1481.
Tsai M-F, Wang C-C, Chang G-C, Chen C-Y, Chen H-Y, Cheng C-L, Yang Y-P, Wu C-Y, Shih F-Y, Liu C-C, *et al*. A new tumor suppressor DnaJ-like heat shock protein, HLJ1, and survival of patients with non-small-cell lung carcinoma. *J Natl Cancer Inst*. 2006:98(12):825–838. https://doi.org/10.1093/jnci/djj229.
Viéitez C, Martínez-Cebrián G, Solé C, Böttcher R, Potel CM, Savitski MM, Onnebo S, Fabregat M, Shilatifard A, Posas F, *et al*. A genetic analysis reveals novel histone residues required for transcriptional reprogramming upon stress. *Nucleic Acids Res*. 2020:48(7):3455–3475. https://doi.org/10.1093/nar/gkaa081.
Volkert MR, Elliott NA, Housman DE. Functional genomics reveals a family of eukaryotic oxidation protection genes. *Proc Natl Acad Sci U S A*. 2000:97(26):14530–14535. https://doi.org/10.1073/pnas.260495897.
Winter J, Jakob U. Beyond transcription—new mechanisms for the regulation of molecular chaperones. *Crit Rev Biochem Mol Biol*. 2004:39(5-6):297–317. https://doi.org/10.1080/10409230490900658.
Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019:20(1):257. https://doi.org/10.1186/s13059-019-1891-0.
Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, *et al*. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*. 2021:2(3):100141. https://doi.org/10.1016/j.xinn.2021.100141.
Xian L, Zhang Y, Cao Y, Wan T, Gong Y, Dai C, Ochieng WA, Nasimiyu AT, Li W, Liu F. Glutamate dehydrogenase plays an important role in ammonium detoxification by submerged macrophytes. *Sci Total*

Environ. 2020:722:137859. https://doi.org/10.1016/j.scitotenv.2020.137859.

Xu Y, Zhu Y, Wang Y, Chang Y-F, Zhang Y, Jiang X, Zhuang X, ZhuY, Zhang J, Zeng L, *et al*. Whole genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic *Leptospira*. *Sci Rep*. 2016:6(1):20020. https://doi.org/10.1038/srep20020.

Yadav V, Molina I, Ranathunge K, Castillo IQ, Rothstein SJ, Reed JW. ABCG transporters are required for suberin and pollen wall extracellular barriers in *Arabidopsis*. *Plant Cell*. 2014:26(9):3569–3588. https://doi.org/10.1105/tpc.114.129049.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007:24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

Yao M-C, Zheng K, Yao C-H. A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in *Tetrahymena*. *Cell*. 1987:48(5):779–788. https://doi.org/10.1016/0092-8674(87)90075-4.

Yokobayashi K, Misaki A, Harada T. Purification and properties of *Pseudomonas* isoamylase. *Biochim Biophys Acta*. 1970:212(3):458–469. https://doi.org/10.1016/0005-2744(70)90252-4.

Yoshida K, Komae K. A rice family 9 glycoside hydrolase isozyme with broad substrate specificity for hemicelluloses in type II cell walls. *Plant Cell Physiol*. 2006:47(11):1541–1554. https://doi.org/10.1093/pcp/pcl020.

Zhang C, Mirarab S. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. *Mol Biol Evol*. 2022:39(12):msac215. https://doi.org/10.1093/molbev/msac215.

Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*. 2019:35(12):2084–2092. https://doi.org/10.1093/bioinformatics/bty895.