1 **Article**

2 Limits to the cellular control of sequestered cryptophyte prey in the 3 marine ciliate *Mesodinium rubrum*

4

5 Andreas Altenburger[1,2*], Huimin Cai[3*], Qiye Li[4*], Kirstine Drumm[5], Miran Kim[5,6], Yuanzhen
6 Zhu[4], Lydia Garcia-Cuetos[2], Xiaoyu Zhan[4], Per Juel Hansen[5], Uwe John[7,8], Shuaicheng Li[3¤],
7 Nina Lundholm[2¤]

8

9 [1]The Arctic University Museum of Norway, UiT The Arctic University of Norway, 9037
10 Tromsø, Norway

11 [2]Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen,
12 Denmark

13 [3]Department of Computer Science, City University of Hong Kong, Hong Kong 999077, China

14 [4]BGI-Shenzhen, Shenzhen 518083, China

15 [5]Department of Biology, University of Copenhagen, 3000 Helsingør, Denmark

16 [6]Research Institute for Basic Sciences, Chonnam National University, Gwangju 61186,
17 Republic of Korea

18 [7]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, 27570
19 Bremerhaven, Germany

20 [8]Helmholtz Institute for Functional Marine Biodiversity, 23129 Oldenburg, Germany

21 *shared first authorship

22 ¤To whom correspondence may be addressed. Email: shuaicli@cityu.edu.hk or
23 nlundholm@snm.ku.dk.

24

25

26 The authors declare no competing financial interests

Abstract

The marine ciliate *Mesodinium rubrum* is famous for its ability to acquire and exploit chloroplasts and other cell organelles from some cryptophyte algal species. We sequenced genomes and transcriptomes of free-swimming *Teleaulax amphioxeia,* as well as well-fed and starved *M. rubrum* in order to understand cellular processes upon sequestration under different prey and light conditions. From its prey, the ciliate acquires the ability to photosynthesize as well as the potential to metabolize several essential compounds including lysine, glycan, and vitamins that elucidate its specific prey dependency. *M. rubrum* does not express photosynthesis related genes itself, but elicits considerable transcriptional control of the acquired cryptophyte organelles. This control is limited as light dependent transcriptional changes found in free-swimming *T. amphioxeia* got lost after sequestration. We found strong transcriptional rewiring of the cryptophyte nucleus upon sequestration, where 35% of the *T. amphioxeia* genes were significantly differentially expressed within well-fed *M. rubrum*. Qualitatively, 68% of all genes expressed within well-fed *M. rubrum* originated from *T. amphioxeia*. Quantitatively, these genes contributed up to 48% to the global transcriptome in well-fed *M. rubrum* and down to 11% in starved *M. rubrum*. This tertiary endosymbiosis system functions for several weeks, when deprived of prey. After this point in time, the ciliate dies if not supplied with fresh prey cells. *M. rubrum* represents one evolutionary way of acquiring photosystems from its algal prey, and might represent a step on the evolutionary way towards a permanent tertiary endosymbiosis.

Introduction

Endosymbiotic events have enabled eukaryotes to photosynthesize. More than a billion years ago, during a primary endosymbiosis event, a photosynthesizing

cyanobacterium was retained by a non-plastidic unicellular eukaryote. Since then, chloroplasts have spread throughout the eukaryotic tree of life by secondary and tertiary endosymbiosis.

Teleaulax amphioxeia is an ecologically important, phototrophic marine unicellular eukaryote (protist) with a worldwide distribution (1). It is 8 – 11 µm long and a member of the enigmatic group of cryptophytes, a group that is challenging to place in the evolutionary tree of life (2). Most cryptophytes have permanent chloroplasts, originating from a secondary endosymbiosis event between a red alga and a phylogenetically distinct, non-photosynthetic host (3, 4). Due to this origin, cryptophyte chloroplasts have a complex membrane topology with four membranes that enclose a nucleomorph between the outer two and the inner two membranes (5-7). The nucleomorph is a highly reduced remnant of the endosymbiotic red algal nucleus. Cryptophytes hence possess DNA of different origin: red algal nuclear DNA in the nucleomorph, chloroplast DNA, cryptophyte mitochondrial DNA, and cryptophyte nuclear DNA (8).

Being primary producers, phototrophic cryptophytes are at the base of the marine food web, and grazed upon by heterotrophic and mixotrophic protists alike (9). One of these grazers is *Mesodinium rubrum,* an abundant and ecologically important ciliate. *M. rubrum* is widely distributed in coastal ecosystems and known for causing non-toxic red tides (10-12). Acquisition of phototrophy by retaining a chloroplast that originated from a secondary endosymbiosis event is regarded as a tertiary endosymbiosis (13, 14). *M. rubrum* preys on cryptophytes belonging to the genera *Geminigera, Teleaulax* and *Plagioselmis. M. rubrum* cells keep around 20 chloroplasts from its cryptophyte prey, and usually a single enlarged prey nucleus located close to the nuclei of the ciliate (ciliates have two macronuclei and one micronucleus) (15-17). In order to sustain

its maximum growth rate of ~0.5 per day, *M. rubrum* has to ingest ~one cryptophyte per day (18, 19). *M. rubrum* covers typically > 98% of its carbon need via photosynthesis at natural prey concentrations, and can replicate the acquired chloroplasts approximately four times after prey deprivation. Eventually, the chloroplasts are degraded, and *M. rubrum* dies unless new cryptophyte prey cells are ingested (18-21). Thus, this tertiary endosymbiosis between a cryptophyte and *M. rubrum* is not permanent and stable, but species-specific (22, 23).

The regulation of cryptophyte genes within *M. rubrum* has previously been studied using RNA-seq, or Expressed Sequence Tags and microarray approaches (24, 25). These studies found a remarkable cellular and metabolic chimerism between host and prey, and showed that *M. rubrum* not only sequesters the organelle machinery of its prey, but also the anabolic potential of the sequestered organelles (25). Most cryptophyte genes involved in photosynthesis were up-regulated after sequestration of the cryptophyte nucleus and chloroplasts into the ciliate (24). However, previous studies had the challenge to distinguish between transcripts originating from *M. rubrum* and transcripts originating from the prey cryptophytes. We used genomic DNA (gDNA) data from free-swimming *T. amphioxeia* and prey-starved *M. rubrum* to overcome this problem. By screening for *k*-mers shared between gDNA reads and transcripts, we were able to assign transcripts to the right species by sequence signature. Using this approach, we could follow the transcriptional changes upon sequestration for cryptophyte and ciliate genes separately. We investigated changes in the level of *T. amphioxeia* genes expressed before and after ingestion by *M. rubrum* and compared those with starved *M. rubrum* cells that had lost the prey nucleus (Fig. 1). We explored changes in the regulation of the sequestered cryptophyte nuclei in response to changing light and time conditions (night, morning and day) corresponding to darkness,

20 minutes after turning on the light, and full light, and focused for the first time on transcriptional changes of ciliate genes under different light conditions and prey availabilities.

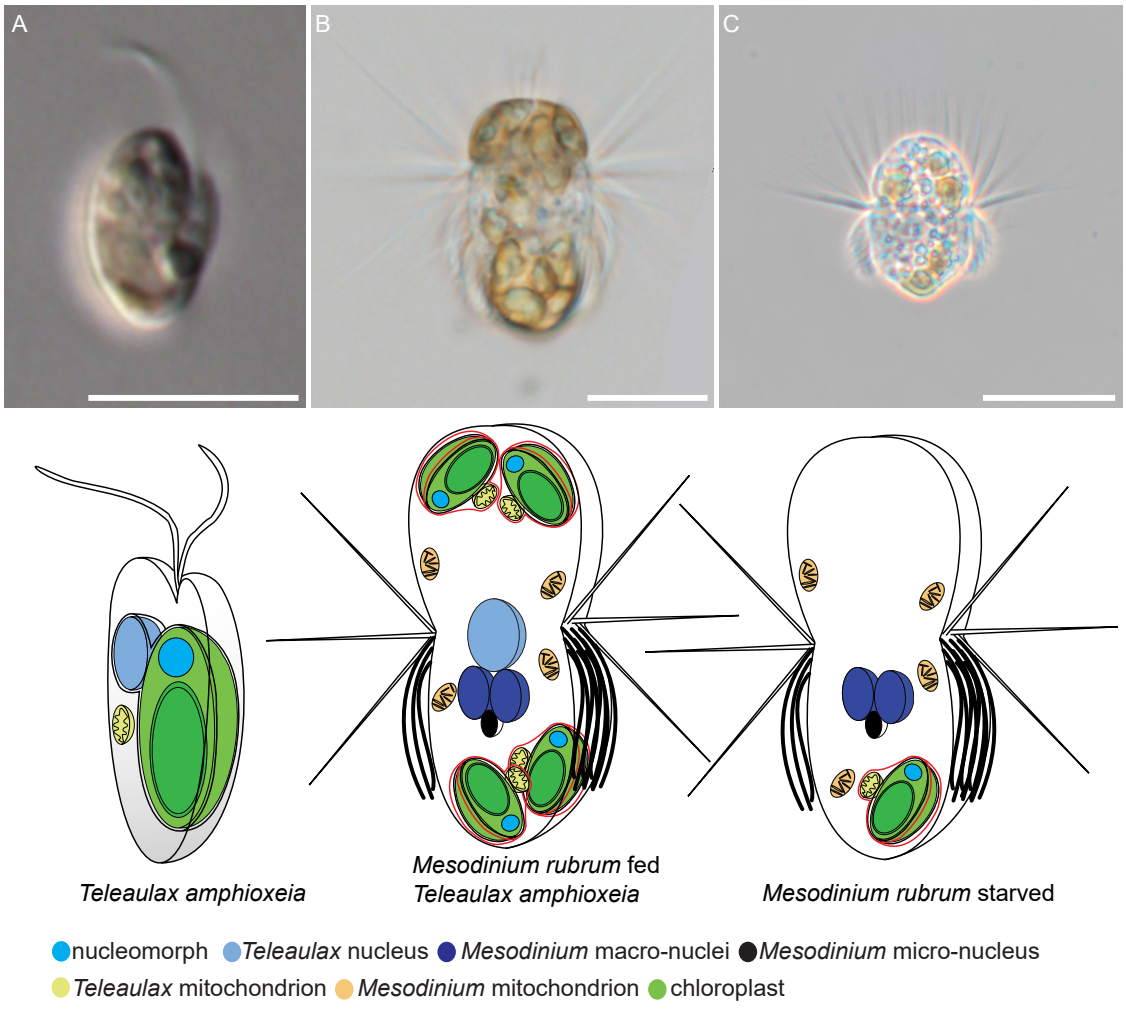


Fig. 1. Light micrographs of *Teleaulax amphioxeia* and *Mesodinium rubrum* with corresponding cartoons. (*A*) free swimming *T. amphioxeia* with chloroplast, nucleomorph, mitochondrion and nucleus. The outer membrane of the nucleus is connected to the outer membrane of the chloroplast. (*B*) well-fed *M. rubrum* with two macronuclei, one micronucleus, and one enlarged cryptophyte nucleus. *M. rubrum* contains its own mitochondria, cryptophyte mitochondria, and cryptophyte chloroplasts that are arranged along the periphery of the cell. (*C*) starved *M. rubrum* with two macronuclei, one micronucleus and ciliate mitochondria. Note: starved *M. rubrum* were defined as cultures where at least 90% of cells had lost the cryptophyte nucleus. Note also: Well-fed cells of *M. rubrum* have one enlarged cryptophyte nucleus, which is always located in the center of the cell, termed CPN (centered prey nucleus) (24). Well-fed cells might keep some extra prey nuclei in the periphery of the cell. Upon ciliate cell division, one of the two

115 daughter cells receives the CPN, while in the other, one of the extra prey nuclei migrate close to the
116 ciliate nuclei and enlarges (16). Scale bar equals 5 μm in (*A*), and 10μm in (*B*) and (*C*).

117

Results and Discussion

Transcriptomic profiles and reference gene set constructions for *Teleaulax amphioxeia* and *Mesodinium rubrum*

We performed RNA-seq on cultures of free-swimming *T. amphioxeia*, *M. rubrum* well-fed, and *M. rubrum* prey-starved for more than 4 weeks (i.e. more than 90% of the cells in the *M. rubrum* culture had lost the central cryptophyte nucleus) (Fig. 1). Each culture was sampled at three time points during the light dark cycle: night (6 hours after the light was switched off), morning (20 minutes after the light was switched on), and day (7 hours after the light was switched on) (Fig. 2A). Biological triplicates were collected for each condition, and an average of 184 million reads were generated for each biological replicate (Supplementary Table 1). To accurately discriminate the species-origin of each assembled transcript, we also performed genome sequencing for DNA extracted from free-swimming *T. amphioxeia* and starved *M. rubrum*, respectively (Fig. 2A; Supplementary Table 1).

As no reference nuclear genomes were available for *T. amphioxeia* and *M. rubrum,* we *de novo* assembled the transcriptome of each sample separately in a first step, followed by stepwise combining the transcripts assembled from each sample. The species identity of each transcript was determined by screening the *k*-mers shared between gDNA reads and transcript sequences (Fig. 2B; Supplementary Table 2; see methods for details). This allowed us to identify 72,061 and 22,250 non-redundant transcripts (i.e. genes) as *T. amphioxeia*- and *M. rubrum*-origin, respectively (Fig. 2C). To access the representativeness of the reference gene sets, we aligned the RNA-seq reads from free-swimming *T. amphioxeia* samples to the 72,061 *T. amphioxeia* genes, and aligned the reads from *M. rubrum* samples to the collection of 72,061 *T. amphioxeia* and 22,250 *M. rubrum* genes (Note: *M. rubrum* samples transcribed genes

143 from both the host and prey genomes). On average, 97.7% of the reads (ranging 96.2%
144 - 98.6%) could be mapped back to the reference gene sets, 95.7% (ranging 93.1% -
145 97.2%) were aligned in proper pairs, and 88.0% (ranging 84.9% - 89.6%) had mapping
146 quality ≥ 30 (Supplementary Table 3), demonstrating that most sequences in the
147 transcriptomes are present uniquely in the two reference gene sets. We also aligned
148 the RNA-seq reads from free-swimming *T. amphioxeia* samples to the collection of *T.*
149 *amphioxeia* and *M. rubrum* genes, and observed less than 0.02% of the aligned reads
150 being mistakenly mapped to *M. rubrum* genes, highlighting the reliability of our DNA-
151 based species assignment process.

152 We annotated 62.9% of *T. amphioxeia* and 50.6% of *M. rubrum* genes by searching
153 against different functional databases (Fig. 2C). Interestingly, the GC content of *T.*
154 *amphioxeia* genes was around 59%, thus considerably higher than the GC content of
155 *M. rubrum* with 35% (Fig. 2C; Fig. 2D). This GC deviation further supports that the
156 genes were assigned to the right species.

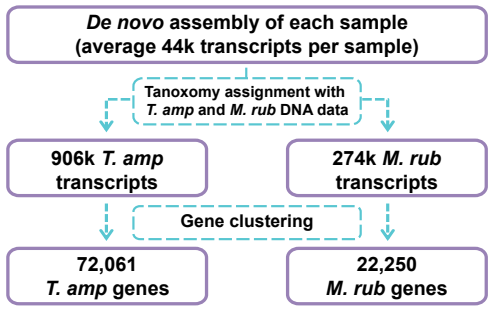
157

A

Number of samples collected

	DNA	RNA		
		Night (Dark 6h)	Morning (Light 20m)	Day (Light 7h)
T. amp (Free-swimming)	1	3	3	3
M. rub (Well-fed)	-	3	3	3
M. rub (Starved)	2	3	3	3

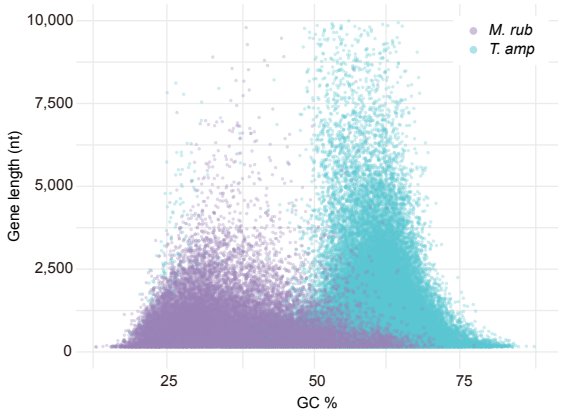
B



C

	T. amphioxeia	*M. rubrum*
Assembly		
Number of genes	72,061	22,250
Total length (nt)	78,195,118	19,532,541
Mean length (nt)	1,085	878
N50 (nt)	2,052	1,259
GC content	58.6%	34.5%
Genes with ORF [≥ 50 aa]	52,544 (72.9%)	21,120 (94.9%)
Annotation		
KEGG	37,675	8,877
GO	22,099	6,895
UniProtKB/Swiss-Prot and TrEMBL	44,786	11,159
Annotated genes	45,297 (62.9%)	11,250 (50.6%)

D



158

159 **Fig 2.** Workflow and transcriptome features of *Teleaulax amphioxeia* and *Mesodinium rubrum*. (*A*)
160 Sampling strategy. (*B*) Analysis workflow. (*C*) Summary of the non-redundant reference gene sets
161 constructed from the *de novo* transcriptome assembly. Abbreviations: nt, nucleotides; aa, amino acids;
162 ORF, open reading frame; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology.
163 (*D*). comparison of gene length and GC content for *M. rubrum* and *T. amphioxeia* genes respectively.

164

165

166 ## *M. rubrum* keeps all the genetic material and transcribes most genes from the
167 ## acquired cryptophyte nuclei

168 By searching for the *T. amphioxeia* genes in the *M. rubrum* gDNA sequence reads, we
169 retrieved almost all (97.3% - 99.9%) of the 72,061 *T. amphioxeia* genes in the two
170 starved *M. rubrum* DNA samples (Fig. 3A; see methods), suggesting that *M. rubrum*
171 keeps all the genetic material from the acquired cryptophyte nuclei. Next, we examined
172 the transcriptional activity of the *T. amphioxeia* nuclei upon sequestration by *M.*

rubrum. Gene expression measurement indicated that on average 63% of the *T. amphioxeia* genes were actively transcribed (TPM ≥ 1) inside *M. rubrum* at some time-point during the sampling cycle, which comprised 82% of the *T. amphioxeia* genes, when considering all sampling points together (Fig. 3B). Even though these ratios were lower than those observed in the free-swimming *T. amphioxeia* samples (77% in average and 96% in combination), they did indicate that the majority of *T. amphioxeia* genes (82%) were actively transcribed inside the *M. rubrum* cells (Fig. 3B). At the same time, an average of 90% of the *M. rubrum* genes were actively transcribed regardless of light and prey availability (Supplementary Fig. S1).

Up to 68.4 ± 1.2% of the genes transcribed within the well-fed *M. rubrum* cells originated from *T. amphioxeia*. This proportion was maintained at 50.4 ± 1.3% for the starved *M. rubrum* samples (Fig. 3C). The contribution of cryptophyte genes to the global transcriptome of well-fed *M. rubrum* in the present study (68.4 ± 1.2%), is higher than previous estimates (13.5% in (24), 58-62% in (25)). However, when taking the transcriptional abundance of each gene into account, the contribution of *T. amphioxeia* transcripts to the global *M. rubrum* transcriptomes was much lower, ranging from 47.5 ± 1.3% (well fed morning) to 10.2 ± 0.9% (starved night) (Fig. 3D). Thus, despite the fact that most *T. amphioxeia* genes were transcribed inside *M. rubrum*, the gene products from *M. rubrum* dominated the mRNA pools of the host cells even in well-fed cells. In a well-integrated endosymbiotic system, one would expect to find a lower qualitative expression of endosymbiont genes: only genes that are beneficial to the host will be expressed, while genes not needed by the host will suffer depletion. Given that 82% of all *T. amphioxeia* genes were expressed at some time point within *M. rubrum*, it is likely that many of the *T. amphioxeia* transcripts are not photosynthesis related and rather by-products. Their functional benefit for *M. rubrum* is not obvious.

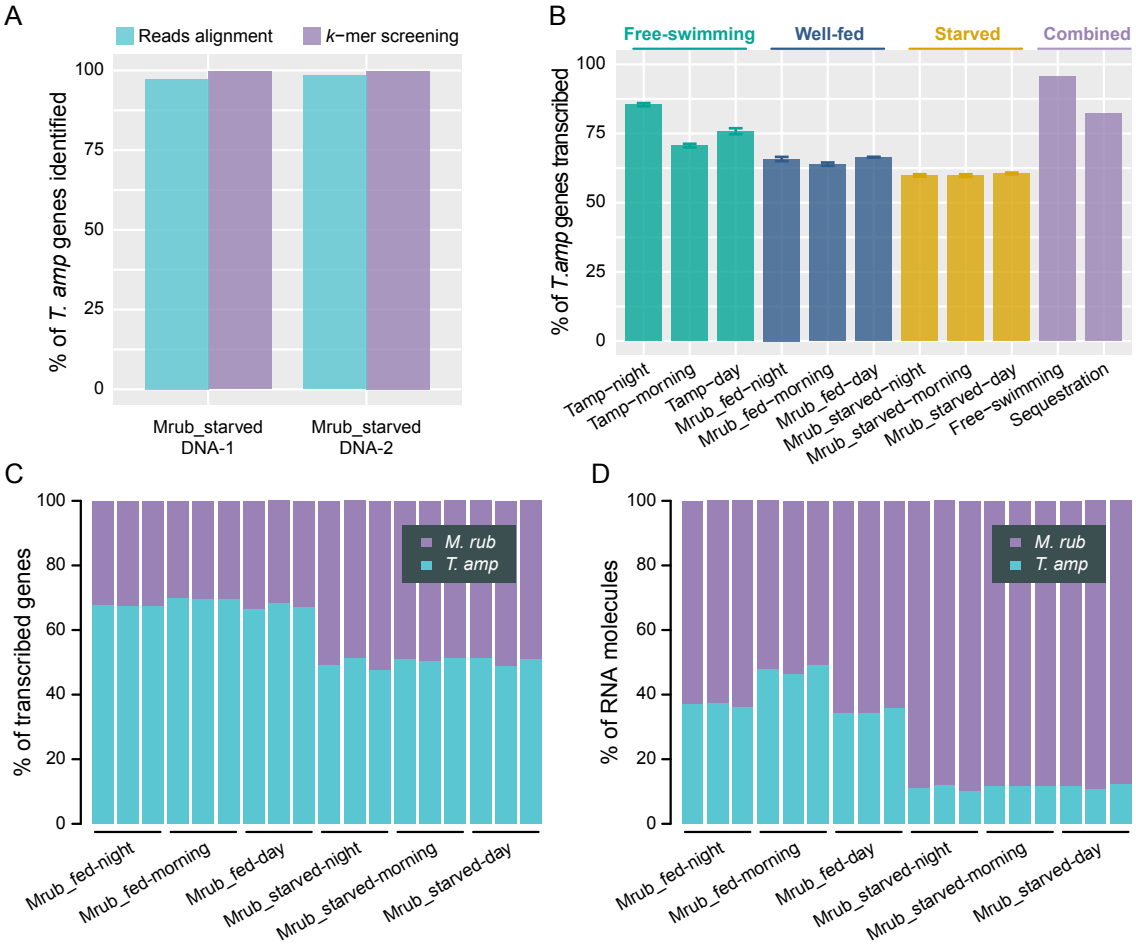


Fig. 3. Global transcriptome features of *M. rubrum* (*A*) percentage of *T. amphioxeia* genes identified in the starved *M. rubrum* DNA data by read alignment and *k*-mer screening methods. (*B*) proportion of actively transcribed *T. amphioxeia* genes before and after sequestration. (*C*) global transcriptome of *M. rubrum* with proportion of contributing *T. amphioxeia* and *M. rubrum* genes. (*D*) global transcriptome of *M. rubrum* with proportion of transcript abundance originating from *T. amphioxeia* or *M. rubrum*.

Cryptophyte nuclei present dramatic transcriptional rewiring upon sequestration

Principal component analysis (PCA) with the *T. amphioxeia* gene expression matrix separated all the 27 samples into three distinct clusters of free-swimming *T. amphioxeia,* well-fed *M. rubrum* and prey-starved *M. rubrum* (Fig. 4A), with the

distance separating free-swimming *T. amphioxeia* samples from all *M. rubrum* samples being larger than the distance separating well-fed and starved *M. rubrum* groups. This suggests that the condition of sequestration alone induced an overwhelming amount of transcriptional changes when compared with other experimental conditions for the *T. amphioxeia* genes. Consistently, this was supported also by weighted gene correlation network analyses (WGCNA) using the same matrix after filtration for lowly expressed genes (44,241 *T. amphioxeia* genes with mean normalized count ≥ 10) identified six modules. The first two modules comprised 81% of the input genes and enriched *T. amphioxeia* genes that were prevailingly down- and up-regulated after sequestration by *M. rubrum*, respectively (Supplementary Fig. S2). More specifically, differential gene expression analyses between the three sample groups revealed by PCA showed that 34.8% and 31.9% of *T. amphioxeia* genes were significantly differentially expressed ($|\log_2 FC| > 1.5$ and FDR < 0.01) upon sequestration in well-fed and prey-starved *M. rubrum*, respectively (Fig. 4B). All these results consistently emphasize that a profound transcriptional rewiring occurs in the *T. amphioxeia* nuclei after sequestration by *M. rubrum*, as reported previously (24, 25).

Functional enrichment analyses for the sequestration-induced differentially expressed genes (DEGs) revealed that *T. amphioxeia* genes related to ion transmembrane transport, signal transduction, cell motility, and regulation of metabolic processes were down-regulated. On the other hand, up-regulation after sequestration was observed for genes involved in photosynthesis, RNA processing, DNA replication and repair, lipid and protein metabolism, and metabolism of diverse compounds (e.g. nucleic acid, carbohydrate, amino acid, carboxylic acid and pigment) (Fig. 4C, Supplementary Table 6, Supplementary Table 7). These results were generally consistent with previous observations by Kim *et al.* (24) and Lasek-Nesselquist *et al.* (25). Interestingly, we also

found that the up-regulated DEGs were enriched in DNA replication, repair and recombination, and cell cycle. These comprised *T. amphioxeia* genes encoding cyclins (e.g. CycA, CycH), cyclin-dependent kinases (e.g. CDK2, CDK7), cell division control proteins (e.g. Cdc6, Cdc7, Cdc45) and almost all the genes involved in the eukaryotic replication complex (Fig. 4D, Supplementary Fig. S3). This suggests that the sequestered nuclei are able to replicate their DNA. Besides, in contrast to Lasek-Nesselquist *et al.* (25), we did not observe downregulation of genes involved in protein processing pathways (Supplementary Fig. S4). On the contrary, many genes involved in endoplasmic reticulum membrane and mRNA surveillance pathway were up-regulated after sequestration (Supplementary Fig. S4 and S5). This implies that the sequestered prey nuclei play an active regulatory role in transcription, translation and also in transportation of *T. amphioxeia* gene products.

Of note, only few *T. amphioxeia* genes (927) were identified as DEGs between well-fed and prey-starved *M. rubrum* samples (Fig. 4B), and the majority of DEGs were shared between free-swimming-vs-inside well-fed *M. rubrum* and free-swimming-vs-inside starved *M. rubrum* (supplementary Fig. S6). This demonstrates that the global transcriptional patterns of the *T. amphioxeia* nuclei inside well-fed *M. rubrum* cells were highly similar with those inside starved *M. rubrum* cells.

These findings are unexpected as most of the starved cells had lost their prey nuclei; less than 10% of the starved *M. rubrum* cells had preserved the enlarged centered prey nucleus (CPN). A typical well-fed *M. rubrum* cell has about 20 chloroplasts and a single enlarged CPN, that is located at more or less the same position anterior to the two macronuclei within *M. rubrum* (16). With each chloroplast, *M. rubrum* takes up one cryptophyte nucleus. Well-fed *M. rubrum* cells can contain multiple prey nuclei, i.e. the CPN and some extra prey nuclei that are kept in the periphery of the cell (16). The

finding that well-fed and starved cells have similar gene expression patterns, suggests that only the CPN is actively transcribed inside *M. rubrum*. Otherwise, the multiple periphery nuclei have to perform a somehow concerted gene transcription with the CPN inside well-fed *M. rubrum* cells.

Experimental evidence suggests that chloroplasts can divide within *M. rubrum* without the presence of cryptophyte nuclei (16, 18). It is also known that photosynthesis in *M. rubrum* is related to the percentage of cells with a CPN, not to the number of chloroplasts (16). Starved *M. rubrum* cells that have lost the CPN, will usually have some chloroplasts remaining in the cell. A reason why those chloroplasts can survive within *M. rubrum* might be that they are particularly robust, with a comparatively large gene set (26, 27). Likely, the nucleomorph plays a crucial role in enabling the *T. amphioxeia* chloroplasts to divide within *M. rubrum* and renders it a favored prey in comparison to chloroplasts with smaller gene sets (28).

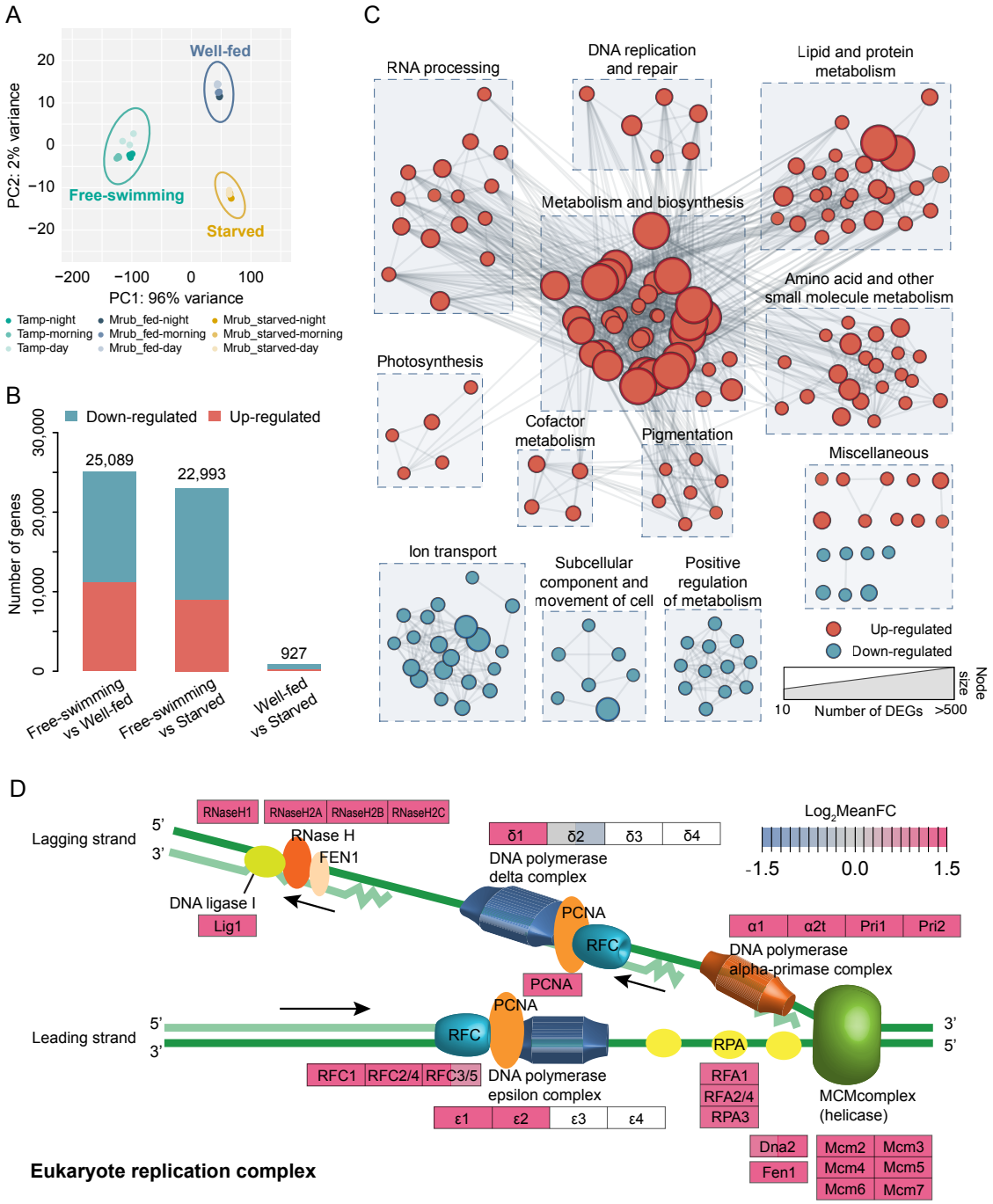


Fig. 4. Changes in gene expression of *Teleaulax amphioxeia* genes in response to sequestration. (*A*) principal component analysis of *T. amphioxeia* genes show a clear segregation between free-swimming, well-fed, and starved samples. (*B*) amount of significantly differentially expressed genes ($|\log_2 FC| > 1.5$ and FDR < 0.01) upon sequestration. (*C*) GO enrichment results for *T. amphioxeia* genes up-/down-regulated after sequestration in well-fed samples visualized as an enrichment map. Nodes represent enriched gene-sets and edges represent mutual overlap between gene-sets, thus clustering highly redundant gene-sets. (*D*) changes in *T. amphioxeia* gene expression in the eukaryote replication

complex pathway. Left part of each box shows log_2 fold change in gene expression for free-swimming vs. well-fed samples. Right part of each box shows log_2 fold change in gene expression for free-swimming vs. starved samples.

Cryptophyte responses to light- and time-changes get lost upon sequestration by *M. rubrum*

Free-swimming *T. amphioxeia* is expected to adjust its gene expression pattern according to light and time changes during night, morning and day as any other photosynthetic organism with permanent chloroplasts. This was confirmed by the PCA result (Fig. 4A). To get a closer look at the light- and time-dependent transcriptional responses of the cryptophyte genes before and after sequestration, we conducted pairwise correlation analyses of the free-swimming *T. amphioxeia* samples and the *M. rubrum* samples, respectively. While relatively low correlations were observed among *T. amphioxeia* samples from different light and time conditions (Fig. 5A), we found all the *M. rubrum* samples showing consistently high pairwise correlations (Fig. 5B), implying that the cryptophyte nuclei had lost the ability to adjust gene expression according to light and time changes upon sequestration. This conclusion is further confirmed by the overwhelming amount of dark/light-responding DEGs (10,828; $|log_2FC| > 1.5$ and FDR < 0.01) identified in free-swimming *T. amphioxeia* samples in comparison to those identified in *M. rubrum* samples (157; Fig. 5C).

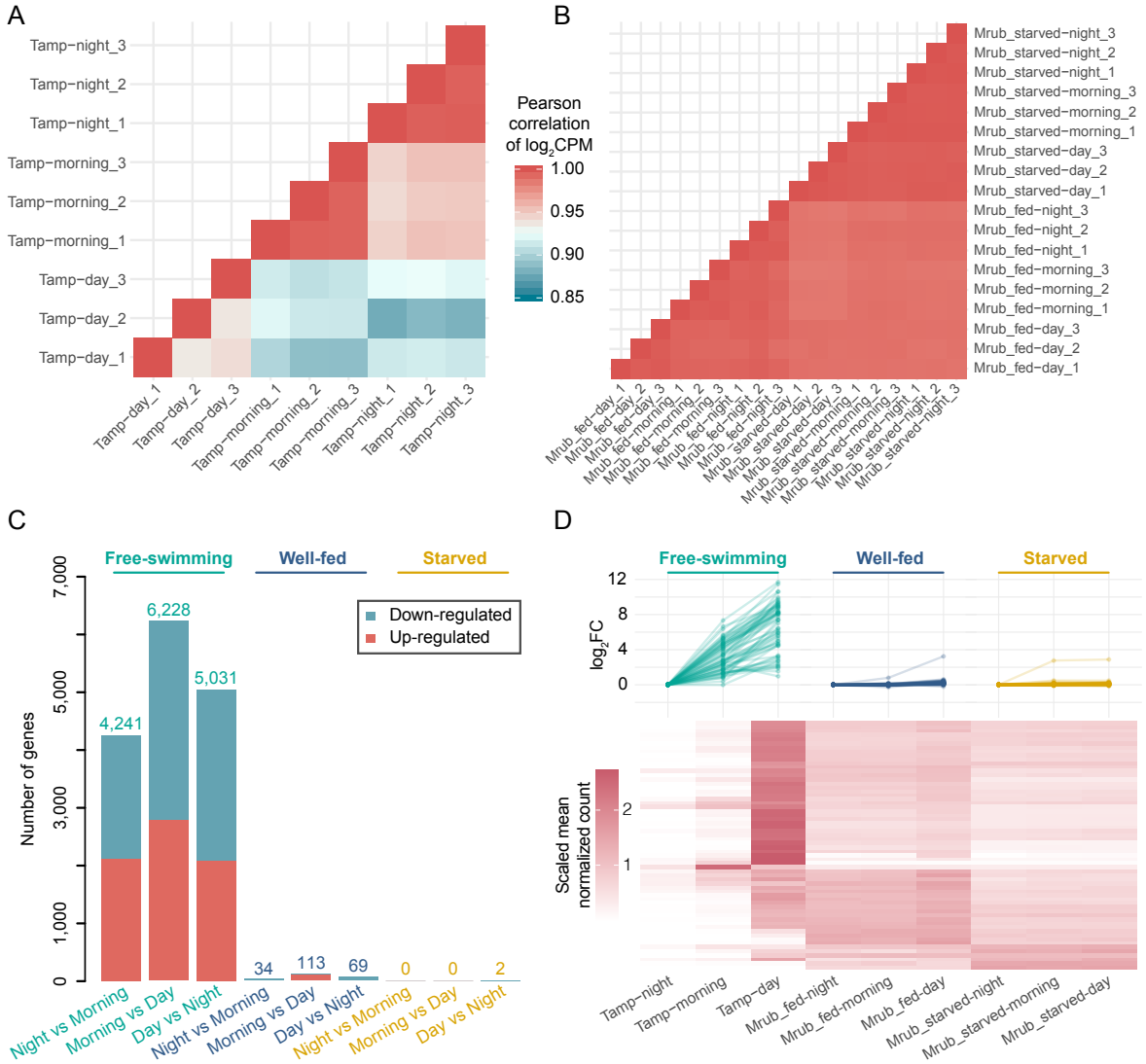
T. amphioxeia genes, significant differentially expressed according to light and time changes, were functionally enriched in photosynthesis, oxidative phosphorylation, glycolysis and circadian entrainment related pathways (Supplementary Table 8), consistent with the expectation for a free-living photosynthetic organism. Of note, *T. amphioxeia* genes involved in circadian entrainment were generally down-regulated upon sequestration (Supplementary Table 7). This might partly account for the loss of time/light response of the cryptophyte nuclei upon sequestration. The expression of

309 many genes in free-swimming *T. amphioxeia* responded to light (i.e. DEGs up-
310 regulated in morning and day versus night). After sequestration, these light responding
311 genes maintained high expression levels at night. Including DEGs encoding for light-
312 harvesting complex and light reaction of photosynthesis, which are responsible for
313 harvesting and transferring light energy and obviously not needed at night (Fig. 5D,
314 Supplementary Table 9). The loss of the dark/light response together with the over-
315 expression of potentially undesired genes strongly suggests that *M. rubrum* can elicit
316 only one expression pattern out of its acquired cryptophyte nucleus regardless of light
317 condition and prey availability (i.e. the number of acquired prey nuclei).
318 Interestingly, in a different system, the Antarctic Ross Sea dinoflagellate acquires
319 transient chloroplasts from haptophyte prey, and the expression of kleptoplast-targeted
320 genes is also unaffected by environmental parameters such as light (29).

321 During evolution, foreign chloroplasts have ended up in other protists in many different
322 ways (13). In some protists, intact endosymbionts are well integrated into host cells
323 (30). Other protists reduce ingested algal cells, and keep prey nuclei as well as other
324 cell organelles beside the chloroplasts (like *M.* rubrum). Yet, other protists retain
325 exclusively the chloroplasts for shorter, (i.e. the ciliate *Strombidium*) or longer time (i.e.
326 the dinoflagellate *Dinophysis*) (31, 32). This can be interpreted as evolutionary steps
327 towards permanent endosymbiosis. In a first step, a prey cell is taken up by a host and
328 not digested. In a second step, the host gets some control over the gene expression
329 of the acquired cell via the ingested prey nuclei – that is where *M. rubrum* is right now.
330 In a third step, only the chloroplasts are retained, but need to be replaced with time (ie.
331 *Dinophysis* (33)). In the final step, the genes from the host and the acquired cell (or
332 organelles such as chloroplasts and nuclei) need to align in order to fine-tune the gene
333 expression according to environmental conditions. Whether or not *M. rubrum* is on its

334 way towards a permanent tertiary endosymbiosis is speculative. Such a step will
335 depend on the ability of *M. rubrum* to divide and keep the sequestered prey nuclei
336 permanently, or rely on gene transfer from the algal prey to the ciliate nuclei.

337



338

339 **Fig. 5.** Changes in light and time controlled gene expression of free-swimming *T. amphioxeia* and after
340 sequestration by *M. rubrum*. (*A*) Pearson correlation analysis of *T. amphioxeia* genes among different
341 samples show differences according to time and light condition. (*B*) Pearson correlation analysis of *T.*
342 *amphioxeia* genes after sequestration by *M. rubrum* reveals an expression pattern that is independent
343 of light and prey availability. (*C*) amount of *T. amphioxeia* genes that were differentially expressed
344 according to time and light condition in free-swimming cells and after sequestration by *M. rubrum.* (*D*)

small panels show expression fold change of light dependent *T. amphioxeia* DEGs at night, night-versus-morning, and night-versus-day in free-swimming, inside well-fed *M. rubrum* and inside starved *M. rubrum* condition. The heat map shows *T. amphioxeia* genes that got differentially expressed according to time and light condition before sequestration by *M. rubrum* but maintained at high expression levels at night after sequestration.

M. rubrum **fine-tunes gene expression in response to prey availability and up-regulates genes involved in transport when well fed**

The problem that *M. rubrum* faces is that it has to deal with different genetic codes. Ciliates show deviations in the genetic code and it has been suggested that these deviations have occurred multiple times independently (34). *M. rubrum* uses a genetic code that is different from cryptophytes and other eukaryotes, for instance it translates UAA and UAG into tyrosine and not into STOP codons (35). By retaining organelles from its cryptophyte prey, *M. rubrum* can use the prey nucleus to serve the chloroplast gene products using the standard code. Given this is possible for several cryptophytes such as *Teleaulax amphioxeia, T. acuta* and *Geminigera cryophila* (TPG clade), the question remains as to why only these taxa and no other cryptophytes apparently can be exploited (15). *M. rubrum* is known to feed on cryptophyte species belonging to different clades, but cannot utilize them for growth and photosynthesis, with the exception of the TPG clade (15, 36).

The construction of reference gene sets separately for *M. rubrum* and *T. amphioxeia* allowed us to compare the gene compositions of the host and its prey in this endosymbiotic system. By mapping the *M. rubrum* and *T. amphioxeia* genes to KEGG pathways, we found that the majority of pathways were present in both species, whereas *M. rubrum* lacked genes involved in photosynthesis, antenna proteins for photosynthesis and carotenoid biosynthesis (Fig. 6A), although we could not

completely rule out the possibility that these genes were presented in the *M. rubrum* genome but not expressed. Interestingly, we also found that the pathways related to the biosynthesis or metabolism of some essential compounds such as lysine, glycan and several kinds of vitamins (pantothenate, riboflavin and biotin) were absent or non-expressed in *M. rubrum* (Fig. 6A), indicating that *M. rubrum* has to obtain the metabolic potential for these compounds from the prey. This explains its dependency on grazing. Of note, most of these *M. rubrum*-absent pathways were actually up-regulated in *T. amphioxeia* after sequestration (Fig. 6A), implicating that *M. rubrum* is able to obtain these nutrients from its cryptophyte prey without digesting the acquired organelles, which is a critical step towards a permanent endosymbiosis. It can be speculated that, in addition to the nucleomorph of the chloroplast, these micronutrients are the reason why *M. rubrum* sequesters exclusively species of the TPG clade.

Next, we investigated the transcriptional changes of *M. rubrum* genes in response to different light conditions and prey availability (well-fed or prey-starved). PCA with the *M. rubrum* gene expression matrix revealed that the *M. rubrum* samples were clustered according to light condition as well as according to prey availability (Fig. 6B). The response to the supply of prey was stronger than the response to light (Fig. 6B). This finding was confirmed by a WGCNA analysis, which uncovered 10 co-expression modules. The two largest modules comprised up to 35% of the input genes and enriched *M. rubrum* genes that were prevailingly down- and up-regulated after starvation (Supplementary Fig. S7). However, DEG analyses using the same cutoff as the *T. amphioxeia* genes ($|\log_2 FC| > 1.5$ and FDR < 0.01), or even lower cutoff ($|\log_2 FC| > 1$ and FDR < 0.01), only identified a small number of genes as DEGs between well-fed and prey-starved *M. rubrum* samples (Fig. 6C). Actually, the negligible number of DEGs identified between samples from different light conditions suggests that *M.*

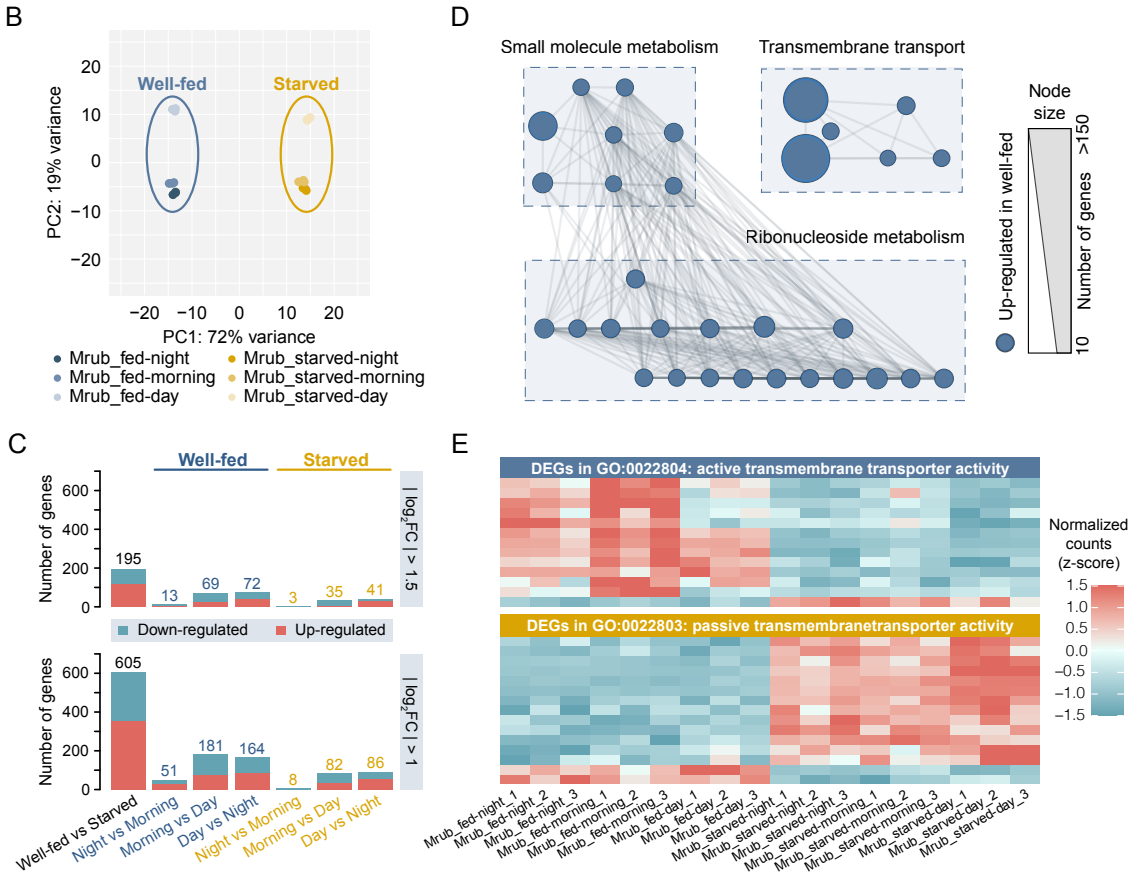
396 *rubrum* is not sensitive to light changes, and that the few responses to light take
397 comparatively long time (Fig. 6B, C).

398 To uncover the functional preference of *M. rubrum* genes in response to prey
399 availability, we conducted functional enrichment analysis for genes clustered in the two
400 largest co-expression modules of the WGCNA analysis. Module 1 comprised 3,285
401 genes that got downregulated upon prey starvation (Supplementary Fig. S7). Module
402 2 contained 2,387 *M. rubrum* genes that got upregulated upon starvation
403 (Supplementary Fig. S7). Genes downregulated upon prey starvation (upregulated in
404 well-fed condition) were enriched in small molecule metabolism, ribonucleoside
405 metabolism and transmembrane transport (Fig. 6D). Interestingly, genes related to
406 active transmembrane transporter activity (GO:0022804; adjusted p = 0.001) were
407 enriched in module 1 (Supplementary Table 10). This indicates that *M. rubrum*-derived
408 active transmembrane transporters play an important role in well-fed *M. rubrum* (Fig.
409 6E). In contrast, passive transmembrane transporter activity is more prominent in prey-
410 starved *M. rubrum* (Fig. 6E), indicating that well-fed *M. rubrum* is transporting
411 molecules among different cell compartments and actively coordinating biological
412 processes of itself and the cryptophyte prey within the cell.

413

A

Pathway	Number of assigned KO identifiers				Down/Not/Up-regulated in well-fed
	M. rubrum	*P. tetraurelia*	*T. thermophila*	*T. amphioxeia*	
Photosynthesis - antenna proteins [ko00196]	0	0	0	9	
Photosynthesis [ko00195]	0	0	0	24	
Pantothenate and CoA biosynthesis [ko00770]	0	5	5	14	
Carotenoid biosynthesis [ko00906]	0	0	0	12	
Lysine biosynthesis [ko00300]	0	1	1	12	
Other types of O-glycan biosynthesis [ko00514]	1	0	1	12	
Various types of N-glycan biosynthesis [ko00513]	1	12	13	21	
Riboflavin metabolism [ko00740]	1	4	2	14	
Other glycan degradation [ko00511]	1	4	4	10	
Biotin metabolism [ko00780]	1	1	1	10	
Histidine metabolism [ko00340]	2	8	5	14	



414

415 **Fig. 6.** Transcriptional changes in *M. rubrum* upon sequestration in different light conditions. (*A*)
416 comparison of the presence of genes from selected pathways in the ciliates *M. rubrum, Parameciu*m
417 *tetraurelia,* and *Tetrahymena thermophile* and the cryptophyte *T. amphioxeia* with differential expression
418 of *T. amphioxeia* genes upon sequestration (free-swimming versus well-fed), showing services provided
419 by *T. amphioxeia* to *M. rubrum*. (*B*) principal component analysis of *M. rubrum* genes. (*C*) barplot
420 showing the amount of significantly differentially expressed *M. rubrum* genes (|log$_2$FC| > 1.5 or |log$_2$FC|
421 > 1) according to prey and light conditions. (*D*) GO enrichment analysis of *M. rubrum* genes upregulated

in the well-fed samples visualized as an enrichment map. (*E*) heat map showing the differential expression (up/down fold change > 1.5) of active and passive transmembrane transporters in well-fed and starved *M. rubrum* cells.

Conclusions

We found very strong transcriptional changes of *T. amphioxeia* genes after sequestration by *M. rubrum.* Upregulated prey genes were related to photosynthesis and metabolism, as well as biosynthesis of lysine and glycan, several kinds of vitamins and gene replication. These processes provide a gain for the host, *M. rubrum* and demonstrates its prey dependency. Light dependent transcriptional regulation of *T. amphioxeia* genes found in free-swimming condition got lost upon sequestration. The transcriptional pattern of *T. amphioxeia* genes in well-fed and prey-starved *M. rubrum* was highly similar, indicating that *M. rubrum* can only induce the expression of one particular pattern out of the acquired prey nucleus. *M. rubrum* shows only very few adjustments in its gene expression in response to different light conditions. Noticeable is the upregulation of active transmembrane transporters in well-fed *M. rubrum* and the role of passive transmembrane transporters in starved *M. rubrum.*

Materials and Methods

Cultures

Cultures were established from single-cell isolates of *Teleaulax amphioxeia* (SCCAP K-1837, collected in Elsinore Harbor, Denmark), and *Mesodinium rubrum* (MBL-DK2009 collected in September 2009 in Elsinore Harbor, Denmark). Cultures (*T. amphioxeia, M. rubrum* fed *T. amphioxeia*) were kept in triplicates and grown in glass

446 bottles in F/2 medium at 15 °C in a light/dark cycle of 16/8h with a light intensity of 100
447 µmol photons $m^{-2}s^{-1}$. During the exponential phase of growth, the ciliates were
448 transferred to new media when cell concentrations reached 5000 ml^{-1} or more.

449 **RNA extraction**

450 For RNA extraction, cultures were harvested in full light (7 hours into the light cycle),
451 in darkness (6 hours into dark cycle) and in the transition between dark and light (20
452 minutes into the light cycle). Cells of *M. rubrum* were harvested in a well-fed and a
453 starved stage.

454 For the well-fed condition, we checked before extraction that no free cryptophyte cells
455 remained in the medium and that at least 90% of all *M. rubrum* cells contained a
456 cryptophyte nucleus. This was done by staining the nuclei with Hoechst reagent
457 (#33342, Thermo Fisher Scientific, Waltham, USA), and checking 20 stained cells
458 under a fluorescent microscope. Harvesting of starved cells was done approximately
459 four weeks after the last cryptophytes had been seen in the culture. We confirmed the
460 loss of cryptophyte nuclei by staining with Hoechst reagent and checking for prey nuclei
461 under a fluorescence microscope. Cells were harvested after at least 90% of all *M.*
462 *rubrum* cells had lost their cryptophyte nucleus. Cells were harvested by centrifugation
463 in 10ml glass tubes at 3220 rcf for 10 minutes (see Supplementary Table 11 for cell
464 numbers in each harvest). Pellets were transferred to 1.5 mL LoBind Eppendorf tubes
465 and liquid nitrogen was directly added onto the pellets. The Eppendorf tubes were
466 stored on ice without allowing the pellets to thaw until the lysis buffer was added. RNA
467 was extracted using the column based Exiqon Cell and Plant RNA Isolation Kit
468 (#300110, Exiqon, Vedbæk, Denmark) following the 'plant' protocol. In addition, a
469 separate round of harvest has been transferred to hot Trizol and stored at -80 °C as
470 backup. Two samples (10 and 11) from this backup have been used for RNA extraction

using the Trizol method. Extracted RNA was stored at -80 ˚C until library preparation for sequencing.

DNA extraction

For DNA extraction *T. amphioxeia* cells as well as starved *M. rubrum* (fed *T. acuta)* cells were harvested as described above and DNA extracted using a KingFisher Duo Prime System (#5400110, Thermo Fisher Scientific, Waltham, USA) using the Plant DNA Kit and following the manufacturers recommendations.

Library construction and sequencing

The RNA-seq libraries were mainly prepared using the MGIEasy RNA Library Prep Set (V1.0, MGI Tech) with 1 ug total RNA as input and sequenced on the BGISEQ-500RS platform using the PE100 chemistry according to the standard protocols provided by MGI Tech Co., Ltd (Shenzhen, China). The only exception was one of the three biological replicates of Tamp-day, of which the amount of total RNA was less than 1 ug and failed to meet the requirement of the MGI kit. The RNA-seq library of this sample was prepared using the TruSeq Stranded mRNA LT Sample Prep kit (RS-122-2101, Illumina) with 500 ng total RNA as input, and sequenced on the Illumina HiSeq 4000 platform using the PE100 chemistry, according to the standard Illumina protocols (San Diego, CA, USA).

The DNA sequencing libraries of free-swimming *T. amphioxeia* and starved *M. rubrum* were prepared using the MGIEasy DNA Library Prep Kit (V1.1, MGI Tech) with 1 μg genomic DNA as input, and sequenced on the BGISEQ-500RS platform using the PE100 chemistry according to the standard protocols provided by MGI Tech Co., Ltd (Shenzhen, China).

Quality control of raw sequencing data

Prior to subsequent analyses, all the DNA- and RNA-seq raw reads were passed to SOAPnuke (v1.5.3) (37) for quality control by removal of adapter-contaminated reads and low-quality reads. Specifically, all the DNA-seq data were filtered by SOAPnuke with parameters *-n 0.02 -l 20 -q 0.3 -Q 2 -G -d*. For the RNA-seq data, we generated two versions of clean data. The first one was used for de novo transcriptome assembly and generated by parameters *-n 0.02 -l 20 -q 0.3 -p 1 -t 15,0,15,0 -Q 2 -G -d* (i.e. removing adapter-contaminated, low-quality and duplicated reads). The second version was used for gene expression measurement and generated by parameters *-n 0.02 -l 20 -q 0.3 -p 1 -t 15,0,15,0 -Q 2 -G* (i.e. removing adapter-contaminated and low-quality reads but keeping duplicated reads).

Construction of reference gene sets for *T. amphioxeia* and *M. rubrum*

A hierarchical strategy was employed to construct the reference gene sets for *T. amphioxeia* and *M. rubrum* with the RNA-seq and DNA-seq clean data.

(i) De novo transcriptome assembly for each RNA sample

The clean RNA reads (without duplicates) of each sample were first assembled using Trinity (v2.4.0) (38, 39) with parameters *--min_contig_length 150 --min_kmer_cov 2 --min_glue 3*. Then the highly similar sequences were clustered, and the redundant transcripts were removed from each transcriptome assembly using cd-hit-est (v4.6.8) (40, 41) with a sequence identity threshold of 0.95. Finally, the clean RNA reads (with duplicates) were aligned to each assembly to quantify the abundance of each gene defined by Trinity (i.e. a transcript cluster) using Salmon (v0.13.1) (42) with parameters *--validateMappings -l IU –allowDovetail*. The lowly expressed genes with TPM < 1 were removed from each transcriptome assembly, then only the longest transcript of each gene was kept. Detailed statistics for the transcriptome assembly for each RNA sample is available in Supplementary Table 2.

(ii) Assignment of transcripts to *T. amphioxeia* and *M. rubrum*

Species assignment was conducted based on the average nucleotide identity (ANI) of each transcript in relative to the DNA sequences of *T. amphioxeia* and *M. rubrum* estimated by Mash (v2.1) (43, 44). To eliminate potential prokaryotic contamination in the transcriptomes, we also built a data set containing 10,243,458 prokaryotic nucleotide sequences by extracting all Archaea and Bacteria sequences from the NCBI nt database (release 20190315). Specifically, we sketched at most 5,000 non-redundant 21-mers from each transcript, and compared them with the non-redundant 21-mers generated from all of the *T. amphioxeia* DNA reads, *M. rubrum* DNA reads and the prokaryotic nucleotide sequences, respectively, to estimate ANI of *T. amphioxeia* (ANI_{tamp}), *M. rubrum* (ANI_{mrub}) and prokaryotic (ANI_{prok}) for each transcript. Of note, even so, more than 90% of the cells in the starved *M. rubrum* cultures had lost the cryptophyte nucleus, genomic sequences of *T. amphioxeia* was still detectable in the starved *M. rubrum* DNA reads (although occurring at low abundances). Considering this, a transcript was assigned to: (1) *T. amphioxeia* if $ANI_{tamp} \geq 0.95$ and $ANI_{prok} < 0.95$; (2) *M. rubrum* if $ANI_{tamp} < 0.95$, $ANI_{mrub} \geq 0.95$ and $ANI_{prok} < 0.95$; (3) unknown sequences for other conditions (Supplementary Table 2). The unknown sequences were discarded from subsequent analyses.

(iii) Hierarchical removal of redundant transcripts

The reference gene sets of *T. amphioxeia* and *M. rubrum* were generated by combining *T. amphioxeia* and *M. rubrum* transcripts from all the samples, respectively. The highly similar sequences were clustered, and the redundant transcripts for the same species were removed using cd-hit-est (v4.6.8) (40, 41) with a sequence identity threshold of 0.95. To further remove redundant transcripts that failed to be clustered by cd-hit-est (caused by alternative splicing such as exon skipping, intron retention, etc.), two

545 rounds of Minimap2 spliced alignment and MCL clustering processes were performed.
546 Specifically, pairwise spliced alignment for all transcripts was generated using the all-
547 vs-all mode of Minimap2 (v2.10-r764) (45) with parameters *-aX -x splice*. Then, a graph
548 was built with the best alignment of each query that satisfied identity > 0.95 and
549 coverage > 70% against the shorter sequence. The vertices of the graph were the
550 transcripts and the edges were weighted by identity x coverage. Next, the graph was
551 inputted to Markov Clustering algorithm (mcl v14-137) (46) to cluster similar sequences
552 with the default power and inflation setting, and the longest transcript of each cluster
553 was kept as the representative. The Minimap-MCL process ran iteratively based on
554 the cluster representatives generated by the last iteration until no more pairwise
555 alignments satisfying the threshold were found. Then a second round of the Minimap-
556 MCL process was performed with the threshold identity > 0.98 and coverage of the
557 shorter sequence > 50% to generate the final non-redundant *T. amphioxeia* and *M.*
558 *rubrum* reference gene sets. The longest ORF for each gene was detected by
559 TransDecoder (v5.5.0) (39) with parameters *-m 50 --genetic_code universal* for the *T.*
560 *amphioxeia* genes and *-m 50 --genetic_code Mesodinium* for the *M. rubrum* genes.

561 (iv) Estimation of the representativeness of the reference gene sets

562 To evaluate the representativeness of the reference gene sets, we aligned the RNA
563 clean reads (with duplicates) from each of the *T. amphioxeia* samples to the *T.*
564 *amphioxeia* genes, and aligned the reads from each of the *M. rubrum* samples to the
565 collection of *T. amphioxeia* and *M. rubrum* genes using BWA-MEM (v0.7.16) (47) with
566 default parameters. Then, the numbers and ratios of reads being aligned, being
567 uniquely aligned (as defined by mapping quality ≥ 30), and being aligned in proper
568 pairs, in relative to the total numbers of inputted reads, were counted by samtools
569 flagstat (SAMtools v1.7) (48, 49). The completeness of the gene sets was assessed

by the Benchmarking Universal Single-Copy Orthologs (BUSCO v3) (50) with the eukaryota odb9 database (Supplementary Table 12).

Identification of *T. amphioxeia* genes present in prey-starved *M. rubrum* DNA data

To determine whether the whole prey nuclei were kept inside *M. rubrum* (Fig. 3A), DNA read alignment and alignment-free k-mer screening methods were used to identify *T. amphioxeia* genes in the two starved *M. rubrum* DNA samples. For the alignment-based method, we aligned the DNA reads of the two starved *M. rubrum* DNA samples to the *T. amphioxeia* gene set using BWA-MEM (v0.7.16) (47) and calculate the coverage of each gene using genomeCoverageBed (v2.26.0) (51) with default parameters. *T. amphioxeia* genes with coverage larger than 70% were considered present in the starved *M. rubrum* cells. The alignment-free k-mer screening method was described in the above section “Assignment of transcripts to *T. amphioxeia* and *M. rubrum*”. The *T. amphioxeia* genes present in the starved *M. rubrum* DNA samples were defined as $ANI_{tamp} \geq 0.95$ and $ANI_{prok} < 0.95$.

Expression level quantification

We aligned the RNA clean reads of each sample to the database containing all *T. amphioxeia* and *M. rubrum* genes and quantified the abundance of each gene using Salmon (v0.13.1) (42) with parameters *--validateMappings -l IU --allowDovetail*. Of note, more than 99.98% of the aligned reads from the free-swimming *T. amphioxeia* samples were mapped to the *T. amphioxeia* genes by Salmon, highlighting the reliability of our species assignment process described above. To eliminate the big disparity in the data amount originated from *T. amphioxeia* across different sample groups (Fig. 3D) for subsequent analyses, we downsampled the clean data of each sample to adjust the number of *T. amphioxeia* aligned reads in each sample to at most

595 ~40M. The raw counts of the *T. amphioxeia* genes in all the 27 samples were collected
596 (Supplementary Table 4). Then, we balanced the data amount originated from *M.*
597 *rubrum* across all *M. rubrum* samples in the same way and generated the raw counts
598 of the *M. rubrum* genes in all the 18 *M. rubrum* samples (Supplementary Table 4).

599 **Transcriptome features of *M. rubrum***

600 With the conclusion that *M. rubrum* keeps all the genetic material from the prey (Fig.
601 3A), we further investigated how actively the *T. amphioxeia* genes were transcribed in
602 the *M. rubrum* samples (Fig. 3B) and the component features of the global *M. rubrum*
603 transcriptomes (Fig. 3C and 3D). Specifically, three matrices were first collected from
604 the Salmon quantification output: EffectiveLength of each *T. amphioxeia* gene in each
605 sample, Counts of each *T. amphioxeia* gene in each sample and TPM of each *T.*
606 *amphioxeia* and *M. rubrum* gene in each *M. rubrum* sample. Next, the *T. amphioxeia*
607 TPM matrix were then re-calculated with the EffectiveLength and Counts matrices
608 according to formula: TPM_i = 10^6 * ($Counts_i$ / $EffectiveLength_i$) / [Σ ($Counts_i$ /
609 $EffectiveLength_i$)]. The genes with TMP ≥ 1 were deemed to be actively transcribed.
610 Finally, we calculated the proportion of actively transcribed *T. amphioxeia* genes in all
611 samples with the re-calculated *T. amphioxeia* TPM matrix, the ratio of actively
612 transcribed *T. amphioxeia* and *M. rubrum* gene numbers (
613 $NumGenes_{Tamp}$/($NumGenes_{Tamp}$+ $NumGenes_{Mrub}$) vs.
614 $NumGenes_{Mrub}$/($NumGenes_{Tamp}$+ $NumGenes_{Mrub}$)) and accumulated the abundance
615 of actively transcribed *T. amphioxeia* and *M. rubrum* genes (Σ TPM_{Tamp} / (Σ TPM_{Tamp}+
616 Σ TPM_{Mrub}) vs. Σ TPM_{Mrub} / (Σ TPM_{Tamp}+ Σ TPM_{Mrub})) in each *M. rubrum* sample with
617 the *T. amphioxeia* and *M. rubrum* TPM matrix. The proportion of actively transcribed
618 *M. rubrum* genes at different prey availabilities (Fig. S1) was counted in the same
619 manner as Fig. 3B.

620 **Identification of differentially expressed genes (DEGs)**

621 The DEG candidates between sample groups were detected by DESeq2 (v1.22.0)
622 (52). First, the library sizes across samples were normalized using the default median
623 ratio method, taking the raw counts as input. Next, to obtain dispersion estimates, the
624 type of fitting of dispersions to the mean intensity was set to be parametric. Then, we
625 used Wald significance tests (nbinomWaldTest) for model fitting and test statistics. The
626 Benjamini-Hochberg False Discovery Rate (FDR) was employed to adjust p-value for
627 multiple test correction. Finally, the significant DEGs were defined with the criteria of
628 basemean ≥ 20, adjusted p-value < 0.01 and $|\log_2$ (fold change)$|$ > 1.5 ($|\log_2$ (fold
629 change)$|$ > 1.5 and $|\log_2$ (fold change)$|$ > 1 were both used to detect significant *M.*
630 *rubrum* DEGs).

631 **Principal component analysis (PCA) clustering**

632 We used the normalized counts as described in the above DEG section to perform
633 PCA clustering. We filtered the gene counts with narrow variance (standard deviation
634 of normalized count < 10 across all samples) and then vst (variance stabilizing
635 transformation provided by DESeq2) transformed the normalized counts with
636 parameter *blind=FALSE*. Finally, we generated the PCA plots with the vst transformed
637 matrices using plotPCA function provided by DESeq2.

638 **Identification of co-expression modules**

639 Co-expression network analyses were performed using the weighted gene correlation
640 network analyses R package (WGCNA v1.68) (53). After filtering lowly expressed
641 genes (mean normalized count < 10 across all samples), the normalized read count
642 matrices were vst transformed with parameter *blind=FALSE*. Then the vst transformed
643 matrice were passed to the blockwiseModules function implemented in the WGCNA

644 package for identification of the signed co-expression modules with parameters
645 *maxBlockSize = 50000, networkType = signed, minModuleSize = 100, minKMEtoStay*
646 *= 0.6, minCoreKME = 0.5, mergeCutHeight = 0.15, numericLabels = TRUE,*
647 *pamRespectsDendro = FALSE* and *power = 16* for the *T. amphioxeia* genes and *power*
648 *= 18* for the *M. rubrum* genes.

649 **Functional annotation and enrichment analysis**

650 We aligned the gene sequences to the UniProtKB/Swiss-Prot database (release
651 20190408) with BLASTX (blast-2.2.26) using parameters *-F F -e 1e-5*. The best hit of
652 each query was retained based on the BLASTX bit score. The GO annotation of the
653 best aligned UniProt protein was then assigned to the query gene. To determine what
654 pathways the genes might be involved in, gene sequences were searched against the
655 KEGG database (v89.1)(54) with BLASTX (blast-2.2.26) using parameters *-F F -e 1e-*
656 *5*. The best hit of each query was retained based on the BLASTX bit score.

657 A set of better quality and reliable functional annotations was generated (the BLASTX
658 hits with e-value lower than 1e-10) for the functional enrichment analyses.
659 Hypergeometric tests were employed to examine whether a list of DEGs was enriched
660 in a specific GO term in relation to background genes as previously described(55), by
661 comparing the number of target DEGs annotated to this GO term, the number of target
662 DEGs not annotated to this GO term, the number of background genes (i.e. all the *T.*
663 *amphioxeia* or *M. rubrum* genes excluding the target DEGs) annotated to this GO term,
664 and the number of background genes not annotated to this GO term. P-values were
665 adjusted for multiple testing by applying FDR (56), and enriched GO terms were
666 considered significant for adjusted p-value < 0.05. The GO enrichment results were
667 visualized with EnrichmentMap (v3.2.1) (57) in Cytoscape (v3.7.2) (58).

KEGG enrichment analysis was done using the same principle as the GO enrichment. The regulation of gene expression involved in the enriched KEGG pathways was visualized Pathview (v1.23.3)(59).

Comparison of the presence of genes involved in the pathways in the ciliates and the cryptophyte

To compare the biological process compositions of the host *M. rubrum* and its prey *T. amphioxeia* in this endosymbiotic system, we first extracted and counted the KO (KEGG Ontology) terms and pathways present in the reference ciliates *Paramecium tetraurelia* and *Tetrahymena thermophile* directly from the KEGG database (v89.1). For the host *M. rubrum* and its prey *T. amphioxeia*, we aligned the gene sequences to the KEGG database (v89.1) with BLASTX (blast-2.2.26) using three sets of parameters: (a) *-F F -e 1e-5*, (b) *-F F -e 1e-10* and (c) *-F F -e 1e-20*. The KO terms associated with the pathways of the best aligned KEGG protein were then assigned to the query genes. The pathways involving less than 5 genes with BLASTX hits (e-value < 1e-5) in *M. rubrum* and at least 20 genes with BLASTX hits (e-value < 1e-20) in *T. amphioxeia* were considered absent or non-expressed in the host *M. rubrum* and provided by the prey *T. amphioxeia* in the endosymbiotic system. The number of present KO terms in the selected pathways in *M. rubrum* and *T. amphioxeia* shown in Fig. 6A were counted with the BLASTX hits (e-value < 1e-10).

Data availability

The raw sequencing reads produced in this study are deposited in the CNGB Nucleotide Sequence Archive (CNSA) with accession number CNP0000925 (https://db.cngb.org/cnsa/). The nucleotide sequences and functional annotations of the reference gene sets for *Teleaulax amphioxeia* and *Mesodinium rubrum* are deposited in the figshare repository.

693 ## Acknowledgements

694 The Danish Council for Independent Research funded this project (grant number 4181-
695 00484 to PJH). QL acknowledges the Major scientific and technological projects of
696 Hainan Province (ZDKJ2019011). LGC was supported by the Carlsberg Foundation
697 (2012_01_0515).

698 ## Author contributions

699 PJH, NL, UJ and AA conceived the work. AA, KD, MK, LGC did the culturing; AA and
700 LGC extracted RNA; QL, YZ and XZ did the sequencing; HC, AA and QL analyzed the
701 data under the supervision of SL and NL; AA, HC and QL drafted the manuscript. All
702 authors contributed to writing of the final version of the manuscript.

703 ## Conflict of Interest

704 The authors declare no conflict of interest.

705

706 ## References

707

708 1. J. I. Kim *et al.*, The Plastid Genome of the Cryptomonad *Teleaulax amphioxeia*. *PLoS One* **10**,
709 e0129284 (2015).
710 2. F. Burki *et al.*, Untangling the early diversification of eukaryotes: a phylogenomic study of the
711 evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B* **283** (2016).
712 3. T. Cavalier-Smith, Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* **5**,
713 174-182 (2000).
714 4. S. E. Douglas, C. A. Murphy, D. F. Spencer, M. W. Gray, Cryptomonad algae are evolutionary
715 chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* **350**, 148-151
716 (1991).
717 5. T. Cavalier-Smith, Principles of protein and lipid targeting in secondary symbiogenesis:
718 euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J.*
719 *Eukaryot. Microbiol.* **46**, 347-366 (1999).
720 6. B. A. Curtis *et al.*, Algal genomes reveal evolutionary mosaicism and the fate of
721 nucleomorphs. *Nature* **492**, 59-65 (2012).

7. S. Douglas *et al.*, The highly reduced genome of an enslaved algal nucleus. *Nature* **410**, 1091-1096 (2001).
8. K. Hoef-Emden, J. M. Archibald, "Cryptophyta (Cryptomonads)" in Handbook of the Protists, J. M. Archibald, A. G. B. Simpson, C. H. Slamovits, Eds. (Springer International Publishing, Cham, 2017), 10.1007/978-3-319-28149-0_35 chap. 24, pp. 851-891.
9. B. A. Ward, M. J. Follows, Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2958-2963 (2016).
10. L. Herfort, T. D. Peterson, V. Campbell, S. Futrell, P. Zuber, *Myrionecta rubra* (*Mesodinium rubrum*) bloom initiation in the Columbia River estuary. *Estuar. Coast. Shelf Sci.* **95**, 440-446 (2011).
11. M. D. Johnson *et al.*, The Genetic Diversity of *Mesodinium* and Associated Cryptophytes. *Front. Microbiol.* **7**, 2017 (2016).
12. T. Lindholm, P. Lindroos, A. C. Mork, Ultrastructure of the photosynthetic ciliate *Mesodinium rubrum*. *BioSyst.* **21**, 141-149 (1988).
13. E. C. Nowack, M. Melkonian, Endosymbiotic associations within protists. *Phil. Trans. R. Soc. B* **365**, 699-712 (2010).
14. M. D. Johnson, D. Oldach, C. F. Delwiche, D. K. Stoecker, Retention of transcriptionally active cryptophyte nuclei by the ciliate *Myrionecta rubra*. *Nature* **445**, 426-428 (2007).
15. P. J. Hansen, M. Moldrup, W. Tarangkoon, L. Garcia-Cuetos, O. Moestrup, Direct evidence for symbiont sequestration in the marine red tide ciliate *Mesodinium rubrum*. *Aquat. Microb. Ecol.* **66**, 63-75 (2012).
16. M. Kim, K. Drumm, N. Daugbjerg, P. J. Hansen, Dynamics of Sequestered Cryptophyte Nuclei in *Mesodinium rubrum* during Starvation and Refeeding. *Front. Microbiol.* **8**, 1-14 (2017).
17. S. W. Nam, J. W. Park, W. Yih, M. G. Park, W. Shin, The fate of cryptophyte cell organelles in the ciliate *Mesodinium* cf. *rubrum* subjected to starvation. *Harmful Algae* **59**, 19-30 (2016).
18. P. Juel Hansen, T. Fenchel, The bloom-forming ciliate *Mesodinium rubrum* harbours a single permanent endosymbiont. *Mar. Biol. Res.* **2**, 169-177 (2006).
19. M. Smith, P. J. Hansen, Interaction between *Mesodinium rubrum* and its prey: importance of prey concentration, irradiance and pH. *Mar. Ecol. Prog. Ser.* **338**, 61-70 (2007).
20. D. J. Matthew, K. S. Diane, Role of feeding in growth and photophysiology of *Myrionecta rubra*. *Aquat. Microb. Ecol.* **39**, 303-312 (2005).
21. T. Fenchel, P. J. Hansen, Motile behaviour of the bloom-forming ciliate Mesodinium rubrum. *Mar. Biol. Res.* **2**, 33-40 (2006).
22. D. E. Gustafson, D. K. Stoecker, M. D. Johnson, W. F. Van Heukelem, K. Sneider, Cryptophyte algae are robbed of their organelles by the marine ciliate *Mesodinium rubrum*. *Nature* **405**, 1049-1052 (2000).
23. E. Peltomaa, M. Johnson, *Mesodinium rubrum* exhibits genus-level but not species-level cryptophyte prey selection. *Aquat. Microb. Ecol.* **78**, 147-159 (2017).
24. G. H. Kim *et al.*, Cryptophyte gene regulation in the kleptoplastidic, karyokleptic ciliate *Mesodinium rubrum*. *Harmful Algae* **52**, 23-33 (2016).
25. E. Lasek-Nesselquist, J. H. Wisecaver, J. D. Hackett, M. D. Johnson, Insights into transcriptional changes that accompany organelle sequestration from the stolen nucleus of *Mesodinium rubrum*. *BMC Genomics* **16**, 805 (2015).
26. J. F. Allen, W. B. M. de Paula, S. Puthiyaveetil, J. Nield, A structural phylogenetic map for chloroplast photosynthesis. *Trends Plant Sci.* **16**, 645-655 (2011).
27. H. Qiu, M. Lee Jun, S. Yoon Hwan, D. Bhattacharya, Hypothesis: Gene-rich plastid genomes in red algae may be an outcome of nuclear genome reduction. *J. Phycol.* **53**, 715-719 (2017).
28. D. Grzebyk, O. Schofield, C. Vetriani, P. G. Falkowski, The Mesozoic Radiation of Eukaryotic Algae: The Portable Plastid Hypothesis. *J. Phycol.* **39**, 259-267 (2003).
29. E. Hehenberger, R. J. Gast, P. J. Keeling, A kleptoplastidic dinoflagellate and the tipping point between transient and fully integrated plastid endosymbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1910121116, 1-9 (2019).

30. G. I. McFadden, Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **6**, a016105 (2014).
31. M. G. Park, M. Kim, S. Kim, The Acquisition of Plastids/Phototrophy in Heterotrophic Dinoflagellates. *Acta Protozool.* **53**, 39-50 (2014).
32. M. D. Johnson, D. J. Beaudoin, The genetic diversity of plastids associated with mixotrophic oligotrich ciliates. *Limnol. Oceanogr.* **64**, 2187-2201 (2019).
33. M. Kim, S. Kim, W. Yih, M. G. Park, The marine dinoflagellate genus Dinophysis can retain plastids of multiple algal origins at the same time. *Harmful Algae* **13**, 105-111 (2012).
34. A. B. Tourancheau, N. Tsao, L. A. Klobutcher, R. E. Pearlman, A. Adoutte, Genetic code deviations in the ciliates: evidence for multiple and independent events. *EMBO J.* **14**, 3262-3267 (1995).
35. S. M. Heaphy, M. Mariotti, V. N. Gladyshev, J. F. Atkins, P. V. Baranov, Novel Ciliate Genetic Code Variants Including the Reassignment of All Three Stop Codons to Sense Codons in Condylostoma magnum. *Mol. Biol. Evol.* **33**, 2885-2889 (2016).
36. M. D. Johnson, D. J. Beaudoin, M. J. Frada, E. F. Brownlee, D. K. Stoecker, High grazing rates on cryptophyte algae in Chesapeake Bay. *Front. Mar. Sci.* **5**, 1-13 (2018).
37. Y. Chen *et al.*, SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1-6 (2018).
38. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652 (2011).
39. B. J. Haas *et al.*, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494-1512 (2013).
40. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).
41. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).
42. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417-419 (2017).
43. B. D. Ondov *et al.*, Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
44. B. D. Ondov *et al.*, Mash Screen: high-throughput sequence containment estimation for genome discovery. *Genome Biol.* **20**, 232 (2019).
45. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
46. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575-1584 (2002).
47. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
48. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
49. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).
50. F. A. Simao, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
51. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
52. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

825 53. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis.
826 *BMC Bioinformatics* **9**, 559 (2008).
827 54. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*
828 **28**, 27-30 (2000).
829 55. E. I. Boyle *et al.*, GO::TermFinder--open source software for accessing Gene Ontology
830 information and finding significantly enriched Gene Ontology terms associated with a list of
831 genes. *Bioinformatics* **20**, 3710-3715 (2004).
832 56. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful
833 Approach to Multiple Testing. *J. Roy. Stat. Soc. Ser. B. (Stat. Method.)* **57**, 289-300 (1995).
834 57. D. Merico, R. Isserlin, O. Stueker, A. Emili, G. D. Bader, Enrichment Map: A Network-Based
835 Method for Gene-Set Enrichment Visualization and Interpretation. *PLOS ONE* **5**, e13984
836 (2010).
837 58. P. Shannon *et al.*, Cytoscape: A Software Environment for Integrated Models of Biomolecular
838 Interaction Networks. *Genome Res.* **13**, 2498-2504 (2003).
839 59. W. Luo, C. Brouwer, Pathview: an R/Bioconductor package for pathway-based data
840 integration and visualization. *Bioinformatics* **29**, 1830-1831 (2013).

841