ORIGINAL PAPER

Examining the Relationship Between the Testate Amoeba *Hyalosphenia papilio* (Arcellinida, Amoebozoa) and its Associated Intracellular Microalgae Using Molecular and Microscopic Methods

Agnes K.M. Weiner[a,b,1], Billie Cullison[a,1], Shailesh V. Date[c], Tomáš Tyml[c,d], Jean-Marie Volland[c,d], Tanja Woyke[d], Laura A. Katz[a,e], and Robin S. Sleith[a,2,3]

[a]Smith College, Department of Biological Sciences, Northampton, Massachusetts, USA

[b]NORCE Climate and Environment, NORCE Norwegian Research Centre AS, Jahnebakken 5, 5007 Bergen, Norway

[c]Laboratory for Research in Complex Systems, Menlo Park, California, USA.

[d]DOE Joint Genome Institute, Berkeley, California, USA

[e]University of Massachusetts Amherst, Program in Organismic and Evolutionary Biology, Amherst, Massachusetts, USA

Submitted February 5, 2021; Accepted December 9, 2021

Monitoring Editor: Ross F. Waller

Running title: *Hyalosphenia papilio* and Microalgae

[1]These authors contributed equally and share first authorship.

[2]Current address: Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

[3]Corresponding author; e-mail rsleith@smith.edu (R. S. Sleith).

Symbiotic relationships between heterotrophic and phototrophic partners are common in microbial eukaryotes. Among Arcellinida (Amoebozoa) several species are associated with microalgae of the genus *Chlorella* (Archaeplastida). So far, these symbioses were assumed to be stable and mutualistic, yet details of the interactions are limited. Here, we analyzed 22 single-cell transcriptomes and 36 partially-sequenced genomes of the Arcellinida morphospecies *Hyalosphenia papilio*, which contains *Chlorella* algae, to shed light on the amoeba-algae association. By characterizing the genetic diversity of associated *Chlorella*, we detected two distinct clades that can be linked to host genetic diversity, yet at the same time show a biogeographic signal across sampling sites. Fluorescence and transmission electron microscopy showed the presence of intact algae cells within the amoeba cell. Yet analysis of transcriptome data suggested that the algal nuclei are inactive, implying that instead of a stable, mutualistic relationship, the algae may be temporarily exploited for photosynthetic activity before being digested. Differences in gene expression of *H. papilio* and *Hyalosphenia elegans* demonstrated increased expression of genes related to oxidative stress. Together, our analyses increase knowledge of this host-symbiont association and reveal 1) higher diversity of associated algae than previously characterized, 2) a transient association between *H. papilio* and *Chlorella* with unclear benefits for the algae, 3) algal-induced gene expression changes in the host.

Key words: Protists; *Chlorella*; symbiosis transcriptomics; genomics; fluorescence microscopy; TEM.

Introduction

Many lineages of microbial eukaryotes live in close association with intra- or extracellular symbionts that can either be other microbial eukaryotes or prokaryotes (reviewed in: Gast et al. 2007; Nowack and Melkonian 2010). The nature of these associations and their degree of closeness fall along a spectrum, reaching from obligate to facultative symbioses (e.g., Stoecker et al. 2009). While obligate symbionts are unable to survive without their host, facultative symbionts can also be found as free-living organisms (e.g., Fisher et al. 2017). Among the most common symbioses of microbial eukaryotes is the association between a heterotrophic organism and microalgae (Esteban et al. 2010; Nowack and Melkonian 2010). This relationship provides benefits to the heterotroph in the form of organic compounds produced by the algae through photosynthesis, while the algae may benefit from a refuge from predation and a controlled cytoplasmic environment (e.g., Johnson et al. 2007). Examples from across microbial eukaryotes for these types of associations include the green algal symbionts of *Paramecium bursaria* (e.g., Takahashi 2016) and multiple symbionts from across the tree of life in diverse Foraminifera species (e.g., *Minutocellus polymorphus, Navicula* sp., e.g. Schmidt et al. 2018).

Symbioses can be further classified based on the duration of the relationship between host and symbiont, from transient to stable, and eventually even organelle incorporation (e.g., Nowack and Melkonian 2010; Stoecker et al. 2009). In addition, symbioses can provide a variable degree of benefit to the partners from being mutualistic and beneficial for each to being antagonistic and thus only beneficial to one partner at the expense of the other (e.g., Bronstein 1994). Studying examples of symbioses from along these spectra is a useful way to gain an understanding of the evolution of photosynthetic organisms and how photosynthetic organelles may have been acquired throughout the history of life (e.g., Lara and Gomaa 2017).

Associations with microalgae are commonly found among shell-building amoebae of the order Arcellinida (Amoebozoa), that are the focus of this study. At least three morphospecies of Arcellinida are known to harbor photosynthetic organisms inside their cytoplasm (e.g., Gomaa et al. 2014; Lara and Gomaa 2017). These associations are assumed to be stable and mutualistic, as the amoeba species containing these symbionts are usually not found without them (e.g., Lara and Gomaa 2017). Genotyping of the microalgae in Arcellinida revealed a surprisingly low diversity among the *Chlorella* strains associated with these amoebae and identified them as strains that are common within a range of other microbial eukaryotes, including ciliates (Flemming et al. 2020; Gomaa et al. 2014; Zagata et al. 2016). Based on analysis of the plastid *rbcL* gene, Gomaa et al. (2014) even argued that a single clade of *Chlorella* is the dominant symbiont across diverse species of Arcellinida, including *Hyalosphenia papilio* and *Heleopera sphagni,* and Rhizaria: including *Archerella flavum and Placocista spinosa.* In addition, no variability among the algae was found to be associated with the high amounts of cryptic diversity observed in some of the host species (e.g., Gomaa et al. 2014; Singer et al. 2019). The lack of diversity in *Chlorella* strains within these diverse species has led to the postulation that while vertical acquisition of symbionts is possible, a great portion of these symbionts are acquired horizontally from the environment, though symbiont acquisition has never been directly observed in these uncultivable organisms (Lara and Gomaa 2017).

In this study, we examine the relationship between the Arcellinida morphospecies *Hyalosphenia papilio* and its associated algae *Chlorella* sp. (Archaeplastida) in detail by using sequencing and imaging approaches. *H. papilio* is one of the most abundant species of Arcellinida in low pH bogs, where it is found on *Sphagnum* moss (Gomaa et al. 2014; Heal 1962; Lahr et al. 2019; Ruggiero et al. 2020). This species appears bright green under the light microscope due to the high number of *Chlorella* cells contained within its cytoplasm. We investigate the diversity of *Chlorella* algae living within *H. papilio* samples from locations across New England, USA, to assess whether extensive sampling in one area would reveal additional

algal diversity than previously observed, and to assess whether symbiont diversity is related to host diversity and/or shows biogeographic signal across sampling sites. Further, we explore the nature of the relationship between host and algae to gain a better understanding of where it falls along the spectra regarding stability and degree of benefit to the partners. If, for example, the organisms have a stable, mutualistic relationship, we would expect evidence of a transcriptionally active *Chlorella* cell living inside *H. papilio*, whereas if the relationship is transient, we may see signs of degradation and/or transcriptional inactivity of the algal nucleus. To explore these relationships, we take advantage of the "bycatch" from single-cell whole genome and transcriptome amplifications targeting the Arcellinida, as symbiont nucleic acids are co-amplified. Consequently, these samples represent the microbiome present within the Arcellinida test (i.e. shell) at the time of DNA/RNA amplification. Here we focus on the presence of algal chloroplast and nuclear DNA/RNA in both whole genome amplifications (WGAs) and whole transcriptome amplifications (WTAs). We also carried out fluorescence microscopy on sections of resin embedded *H. papilio* harboring *Chlorella* cells as well as transmission electron microscopy (TEM) to shed further light on the state of the algae within its host.

In addition, we sought to understand the effect of symbiont-induced oxidative stress on host gene expression. Free oxygen radicals that are produced during photosynthesis by the symbiont can have detrimental effects on the host if not enough antioxidants are present (e.g., Betteridge 2000). This stress can lead to changes in gene expression in the host to maintain the integrity of cellular functioning (e.g., Johnson et al. 2007). To investigate the impact of photosynthesis by the algal symbionts in *H. papilio*, we analyzed differential gene expression between *H. papilio* and its congener *Hyalosphenia elegans*, which lacks photosynthetic symbionts. Together, our analyses contribute to a better understanding of the association between mixotrophic microbial eukaryotes and their photosynthetic symbionts.

Results

Generation of Single-cell Genomes and Transcriptomes

For this study we analyzed a total of 36 single-cell whole genomes and 22 transcriptomes of the Arcellinida morphospecies *H. papilio*. In addition, we generated 12 transcriptomes for the congener species *H. elegans*. (Supplementary Material Table S1). All of the sequenced data are available on GenBank (BioProject number PRJNA761372).

Symbiont and Host Diversity and Phylogeny

We assessed the diversity of algal symbionts associated with *H. papilio* using the *rbcL* marker gene characterized from 17 partially-sequenced WGAs from which we recovered full-length *rbcL* sequences (Supplementary Material Table S1, File S1). We did not detect a full-length *rbcL* contig in the remaining 19 WGAs, and were unable to capture chloroplast transcripts from WTAs due to their lack of poly-A tails. The resulting multiple sequence alignment and phylogeny (Fig. 1; Supplementary Material Files S1 and S2) contained the 17 sequences from our samples as well as 208 sequences from GenBank and the maximum likelihood tree recovered the testate amoebae *Chlorella* symbiont (TACS) I clade reported in Gomaa et al. (2014). In addition to this previously known clade, we discovered a novel clade, TACS II, that represents samples of *Chlorella* associated with testate amoebae from across the northeastern USA (Fig. 1; Supplementary Material File S2). By applying a read mapping approach (see Methods and Supplementary Material Fig. S1) we were able to identify three additional *H. papilio* samples as containing TACS I despite their lack of full-length *rbcL* sequences. Overall, we found five samples that contain symbionts of the TACS I clade and 15 that contain TACS II symbionts (Supplementary Material Table S1).

In addition to the symbiont diversity, we also assessed the host genetic diversity by reconstructing a phylogenetic tree based on COI sequences extracted from the WGA samples for which we had *rbcL* sequences (Fig. 2; Supplementary Material Table S1, Files S3 and S4).

We identified three genetic clades within our samples (clade M (n=3), clade K (n=2), clade F (n=12); Fig. 2**)**, all of which had been described before (Gooma et al. 2014; Heger et al. 2013; Singer et al. 2019). In addition, there were three additional samples that did not have enough reads for assembly, but the reads were long enough to be compared to the alignment and to identify genotypes: two grouped with clade M and one with clade F. Linking host diversity to the encountered symbiont clades we observed that TACS I is exclusively associated with clade M hosts, whereas symbionts of TACS II are found within *H. papilio* clades K and F (Supplementary Material Table S1; Fig. 2). We did not find evidence for the occurrence of multiple symbionts of different clades within one host cell as the read mapping-based analysis did not detect polymorphisms within individual reference alignments as would be expected if multiple *Chlorella* clades were present in each amoeba (Fig. S1).

Examining the distribution of our samples, we found a biogeographic pattern in the occurrence of both the host clades as well as the two symbiont clades. The five *H. papilio* clade M samples that all contain TACS I algae only occurred in Harvard Forest, while *H. papilio* clade K containing TACS II was found in Acadia Bog and clade F with TACS II occurred in the open bogs at Hawley Bog and Orono Bog (Supplementary Material Table S1; Figs 1, 2).

Because sequences from three free-living *Chlorella* (GenBank accession numbers: KM514889, KM514890, KM514866), isolated from lakes in Jiangsu Province, China (Zou et al. 2016), are sister to the testate amoebae-associated clades (Fig. 1), we assessed the potential monophyly of TACSI and TACS II. Both AU and SH tests of alternative topologies reject the hypothesis of a single origin of *Chlorella* symbiont clades, either when we constrain the three clades (TACS I, TACSII, *Paramecium bursaria* symbiont clade; p-SH 0.0267, p-AU 0.0109) or only the *Chlorella* strains from testate amoebae (clades TACSI and TACSII; p-SH 0.0054, p-AU 0.0036), suggesting two independent origins of the symbiotic association between *Chlorella* and testate amoebae.

Assessing the Presence and Expression of *Chlorella* Nuclei

To determine where the *H. papilio* – *Chlorella* relationship falls on the spectrum spanning stable to transient symbioses, we first assessed whether the *Chlorella* cells in *H. papilio* are intact cells including nuclei or if, in the extreme case of kleptoplasty, only their chloroplasts are retained inside the amoebae. A substantial part of the granuloplasmic mass of *Hyalosphenia* cytoplasm was occupied by autofluorescence emitting (in a wide range of wavelengths, from blue to far-red) *Chlorella* cells (Fig. 3A, B). Both DNA staining on semi-thin sections and TEM analysis revealed the presence of seemingly intact *Chlorella* cells within *H. papilio*, at least some of which contain a nucleus (Fig. 3). The *Chlorella* cells measure on average 4.18 +/- 0.32 µm in diameter and their nucleus was on average 1.16 +/- 0.20 µm in diameter. While the sectioning plane did not always cut through a nucleus, the observation of six consecutive sections (0.5 µm each) stained with DAPI did not reveal a single anucleate algae. In addition to the algae cells found in the *H. papilio* cytoplasm, we also detected *Chlorella* cells in food vacuoles (Supplementary Material Fig. S2). Poorly preserved ultrastructure details (e.g., indistinguishable chloroplasts or nuclei; Supplementary Material Fig. S2) along with entirely missing autofluorescence signal (Fig. 3A, B, D) in these cells indicated that digestion was underway.

We then tested whether or not the *Chlorella* nuclei are actively transcribing by using a phylogenomic approach including a wide diversity of eukaryotic taxa to identify if genes in the transcriptome samples came from the *Chlorella* nucleus. Since we chose conserved eukaryotic genes with general housekeeping functions (Supplementary Material Table S2), they can be expected to be expressed in the amoebae and the algae cells, assuming the algae are complete and actively expressing their genes. If, on the other hand, the algae are deprived of their nuclei or the nuclei are in an inactive state, we should only obtain transcripts from the host and not the algae.

Analyzing a total of 150 gene trees, we observed only six cases in which sequences from the transcriptome data fell among Archaeplastida clades (Fig. 4; Supplementary Material

Table S2). These six sequences came from five different samples, whereas the remaining 17 *H. papilio* transcriptome samples never had sequences fall among Archaeplastida. This suggests that the algae may not be actively expressing their housekeeping genes, either because the nuclei – and over time the entire algae cells – are being degraded or that their metabolism is closely linked to the amoeba and general housekeeping functions are fulfilled by the host. For further investigation of the transcriptional activity of the *Chlorella* nuclei, we also searched for the presence of *Chlorella* genes related to photosynthesis. These genes would be expected to be present in a long-term, stable symbiosis, as nuclear products are necessary to maintain chloroplast function. However, this analysis revealed a similarly low level of nuclear signal from the algae as we only found a signal in seven out of 22 transcriptome samples and the highest number of genes expressed was four out of 27 in one sample (Fig. 4; Supplementary Material Table S2).

To further assess the stability of the relationship between the *Chlorella* and their amoebae hosts, we diluted 10 Arcellinida cells in filtered bog water, essentially removing free-living *Chlorella* as a food source. At the start of this experiment (day 0), and after 3, 5, 7, and 12 days, we assessed the presence of *Chlorella* plastid genome by partial sequencing of single-cell whole genome amplifications. Using average coverage of the chloroplast genome, we found highest coverage on day 0 with a rapid decline over time (Fig. 5), suggesting a gradual degradation of the *Chlorella* cells and chloroplasts. This is consistent with light micrographs taken during the starvation experiment in which the algae cells and chloroplasts appear increasingly degraded as time progresses (Fig. 5; Supplementary Material Fig. S3).

Differential Gene Expression

To measure the effect of the presence of photosynthetic *Chlorella* within the amoeba cell, we characterized genes that were expressed in *H. papilio* but absent from *H. elegans,* a species that does not have an association with algal cells. Using PhyloToL (Cerón-Romero et al. 2019)

as a tool for homology assessment, we found a total of 132 genes from 92 gene families expressed in *H. papilio* that were not expressed in the majority of the congener *H. elegans*. We used Blast2GO (Götz et al. 2008) to identify the functions of these gene families. The majority of gene families (86) fulfilled housekeeping functions, e.g. they contribute to metabolic processes and biosynthesis. However, we also found six gene families that are related to oxidation-reduction processes (Supplementary Material Table S3, File S5). When comparing the functional categories of these 92 differentially expressed gene families to a random set of 92 gene families that are shared between the two species, we found the same "housekeeping functions". However, oxidation-reduction processes as functional category is missing from the shared dataset, making it a unique category in *H. papilio* that could indicate a response to the presence of symbionts.

Discussion

The three main insights from this study are: 1) two distinct clades of *Chlorella* are associated with the Arcellinida morphospecies *H. papilio* in our New England samples, a finding that contrasts with previous claims of a single world-wide partnership; 2) our analyses suggest that the relationship may be transient as we find no evidence of gene expression from the green algal nucleus, despite the presence of a nucleus as shown by TEM and DAPI; and 3) analyses of the host – *H. papilio* – transcriptomes suggest changes in gene expression consistent with oxidative stress.

Two Distinct Clades of *Chlorella* are Associated with Testate Amoebae in New England Bogs and Fens

Analysis of *Chlorella rbcL* sequences from a single morphospecies of testate amoebae sampled in New England refutes the “one alga to rule them all” hypothesis of a single *Chlorella* strain found across multiple morphospecies of testate amoebae (Gomaa et al. 2014). Instead, we find evidence for at least two *Chlorella* clades existing within *Hyalosphenia papilio* in New England. Phylogenetic analyses show that these two clades are non-monophyletic and are separated by free living strains (e.g. *Chlorella* sp. sens*u* Zou et al. 2016 clade) and symbiotic strains (e.g., *Chlorella variabilis-Paramecium bursaria* symbiont clade; Fig. 1). The observation of multiple clades of *Chlorella* is consistent with other systems that show a wide diversity of symbionts: a study of green algal symbionts of *Paramecium bursaria* found *Chlorella variablis*, *Chlorella vulgaris*, *Micractinium conductrix*, and *Choricystis parasitica* as endosymbionts (Flemming et al. 2020; Fujishima et al. 2012; Pröschold et al. 2011; Zagata et al. 2016).

The morphospecies *H. papilio* is known to contain a high amount of genetic diversity that has been assigned to ~12 clades (e.g., Heger et al. 2013; Singer et al. 2019), three of which are represented in our samples (Fig. 2). While we observe a link between symbiont and host diversity, with the TACS I clade exclusively being associated with clade M amoebae, this pattern is overlain by a biogeographic distribution pattern and can therefore not be considered as evidence for host-symbiont coevolution. Only once amoebae of different genotypes containing symbionts of different clades are found at the same sampling station can a case for genotype specific associations be made. To date, it appears more likely that testate amoebae pick up their symbionts from the environment instead of inheriting them vertically, as has already been argued by Gomaa et al. (2014). Future studies could test whether cross-infection experiments allowed *H. papilio* of different genetic types to accept symbiont clades from different geographic areas.

Considering the proximity of our sampling locations, encountering a biogeographic distribution pattern was rather surprising. We sampled amoebae from Harvard Forest and Hawley Bog, sites which are separated by only ~80 km with no major dispersal barriers; given

288 the small size of *Chlorella* we expected high dispersal rates and distances at the scale of the
289 Northeastern USA (Foissner 2006). However, none of the *Chlorella* genotypes found at Harvard
290 Forest (TACS I) were found at Hawley Bog or either of the Maine sampling sites (Figs 1, 2).
291 Indeed, only five of our *H. papilio* samples contained TACS I, with most amoebae (15) harboring
292 TACS II *Chlorella* symbionts. This is in contrast to the study of Gomaa et al. (2014), who
293 sampled testate amoebae from diverse Arcellinida clades across the world and predominantly
294 detected green algae from the TACS I clade with only a few samples harboring *Chlorella* from
295 other clades (e.g. 18/LC/10-KJ446811). The sampling of Gomaa et al. (2014) was limited to
296 nutrient rich and poor fens and bogs at high latitudes (>46° N) and it is possible that the TACS I
297 clade has a northern distribution that made it less likely for our study to detect. Alternatively, the
298 ecology of these sites could explain the distribution patterns of *Chlorella* clades. Hawley Bog,
299 Orono Bog, and Big Heath all represent low-nutrient bog habitats (Davis and Anderson 2001;
300 Kearsley 1999) while our sampling site at Harvard Forest is a rich fen impacted by damming in
301 the late 19th century (Swan and Gill 1970). Interestingly, the worldwide Holarctic sampling of
302 Gomaa et al. (2014) did not detect amoeba from clades F, K, or M. Clades K and M are only
303 known from Eastern North America (Singer et al. 2019) where Gomaa et al. (2014) did not
304 sample, while clade F has been collected in Alaska where Gomaa et al. (2014) collected only
305 samples from clades C and D. There remain three lineages (B, I, and L) from Northwest North
306 America from which no *Chlorella* have been genotyped, leaving the possibility that further
307 symbiont diversity exists. Additional sampling of *H. papilio* from diverse ecologies and latitudes
308 will help to elucidate this pattern.
309 Despite finding a higher symbiont diversity than previously reported, we still only
310 detected a single clade of symbiont within each amoeba individual. This is in agreement with the
311 results of cloning experiments carried out by Gomaa et al. (2014) and is either simply due to the
312 biogeographic distribution of the algae or it may suggest some type of preferential feeding by

313 the host and/or competition between *Chlorella* genotypes preventing colonization of *H. papilio*
314 by multiple genotypes.
315
316 **Relationship Between Host and Symbiont May be Transient**
317 In contrast to the expectation of a mutualistic symbiosis between Arcellinida and *Chlorella*, our
318 analyses suggest that only the chloroplast genome, and not the *Chlorella* nucleus, are active
319 within *H. papilio* (Fig. 4). Support for this includes our observation of no *Chlorella* nuclear
320 transcripts (either housekeeping or photosynthesis related) in the transcriptome samples, the
321 degradation of chloroplast DNA in genome samples as the amoebae are starved (Fig. 5), and
322 the presence of *Chlorella* cells in food vacuoles in TEM images (Supplementary Material Fig.
323 S2). It is possible that the algae are taken up by the host from the environment, their
324 photosynthetic activity is exploited while they are maintained inside the host cell and later they
325 are digested as additional food source, consistent with our starvation experiment. Considering
326 our observations, the *Chlorella – H. papilio* association may therefore fall closer to a transient
327 relationship on the spectrum of symbioses. This relationship resembles an “intermediate
328 mixotrophic mechanism” as defined by Esteban et al. (2010) as organisms that retain
329 photosynthetic cells for a period of time followed by digestion. Though the long-term functioning
330 of chloroplasts relies on nuclear signaling and nuclear encoded proteins (Pogson et al. 2008), in
331 several systems chloroplasts remain viable for days to weeks without nuclear products or
332 signals: in ciliate-algae relationships the kleptoplast survival period can last up to a month
333 (Johnson et al. 2007) while in Foraminifera-algae relationships the survival period is up to a
334 maximum of around 10 weeks (Correia and Lee 2002). The molecular and/or physiological
335 mechanisms limiting nuclear activity in the *H. papilio – Chlorella* association remain to be
336 elucidated, and coupling cultivation with techniques such as fluorescence in-situ hybridization
337 (FISH) will allow for fine grained analysis of host and symbiont identities (McManus and Katz
338 2009).

Oxidative Stress from *Chlorella* Influences Host Expression

The presence of a photosynthetic symbiont affecting the gene expression of a host has been documented in other lineages (Betteridge 2000; Kodama et al. 2014). Of the 92 gene families that are expressed in *H. papilio* and not *H. elegans*, six are related to reducing oxidative stress (Supplementary Material Table S3, with two gene families containing paralogs). As *H. papilio* has a photosynthetic symbiont, there is a likelihood that the increase of reactive oxygen species present can create changes within the host as oxidative stress leads to free radicals (reactive oxygen species (ROS)) building up in greater proportion to the production of antioxidants (Betteridge 2000), which may be processed by these enzymes (Supplementary Material Table S3). Similarly, in the case of the *Paramecium bursaria* - *Chlorella variabilis* relationship, differentially expressed genes in cells with symbionts include the down-regulation of oxidoreductase processes (Kodama et al. 2014). Changes in gene expression are also found in the sea anemone, *Anemonia viridis,* that has a photosynthetic protist living within it (Richier et al. 2005). As the protist photosynthesizes, the anemone upregulates the production of antioxidant enzymes to combat this increase of oxygen (Richier et al. 2005).

Taken together our results suggest that the Arcellinida morphospecies *H. papilio* harbors diverse *Chlorella* strains that influence host gene expression, and that the relationship between the amoebae and the microalgae *Chlorella* may not be as advantageous to both partners as was previously assumed. Further studies of the microbiomes of microbes will surely reveal additional nuances to symbiotic relationships.

Methods

Sample collection and preparation: We collected Arcellinida samples at four different bogs and fens in New England: “Hawley Bog” (42.576774, -72.890266) and “Harvard Forest “(42.531728, -72.189973) in

Massachusetts and “Orono Bog” (44.870752, -68.723785) and “Big Heath” in Acadia National Park (44.335780, -68.274809) in Maine (Supplementary Material Table S1). At each site we collected a handful of *Sphagnum* moss from different sampling points. Back in the lab we washed off the Arcellinida from the moss by putting about 10 strands of moss into 50 ml conical tubes with 20 ml filtered (2 µm filter) water from the sampling sites and shaking the tubes. The moss and water were then filtered over a 300 µm mesh into a Petri dish to eliminate large particles. We then hand-picked individual healthy-looking Arcellinida cells from the filtrate under the microscope. We only chose *H. papilio* cells that were bright green, indicating the presence of healthy, undigested microalgae. Each individual was photo-documented and cleaned in filtered (2 µm filter) bog water before being transferred in 1 µl bog water to a sterile 0.2 ml tube and stored at -80 °C before either genome or transcriptome amplification, which took place within 1-3 days of collection.

Single-cell transcriptomics and genomics: For the amplification of the transcriptomes of individual *H. papilio* and *H. elegans* cells we first added 1.4 µl nuclease-free water to the tubes with the isolated amoebae and then 0.25 µl of the lysis buffer contained within the SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing (TaKaRa Bio USA, Inc., Mountain View, CA, USA). The subsequent transcriptome amplification and reverse transcription we conducted according to the manufacturer’s protocol, yet in quarter reactions. The SMART-Seq® v4 Ultra® Low Input RNA Kit was selected for this study as it has been applied successfully across diverse microbial eukaryotes, including the difficult to lyse Foraminifera (e.g., Weiner et al. 2020).

For the generation of single-cell genomes we used the REPLI-g Single-Cell Kit (Qiagen, Germantown, MD, USA). We added 1 µl of single cell water and 1.5 µl DLB buffer (both are part of the kit) to the picked cell and then followed the protocol according to the manufacturer’s instructions.

After transcriptome/genome amplification we purified all samples using the AMPure XP PCR Purification system (Beckman Coulter Life Sciences, Indianapolis, IN, USA) and quantified the amplified nucleic acid content using a Qubit 3.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA). Barcoded sequencing libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) and the samples were sequenced on a HiSeq 4000 platform at the Institute for Genome Sciences at the University of Maryland.

Transcriptome/genome assembly and post-assembly processing: We used FastQC (Andrews, 2010) for quality checking of the raw sequencing reads, trimmed adapters using the BBMap toolkit (Bushnell 2014) and then assembled the reads using SPAdes (for genomes) or rnaSPAdes (for transcriptomes; Bankevich et al. 2012). After assembly, we processed the transcriptomes using our phylogenomic pipeline PhyloToL (Cerón-Romero et al. 2019). This post-assembly processing includes the removal of prokaryotic sequences and ribosomal DNA, assignment of the assembled sequences to gene families as defined by OrthoMCL (Li et al. 2003) using USearch (Edgar 2010), translation into amino acid sequences and removal of short, highly identical sequences.

Algal diversity and phylogeny: In each of the 36 trimmed genome assemblies (Supplementary Material Table S1) we searched for contigs containing the chloroplast-encoded *rbcL* sequence using BLASTn version 2.9.0 (Boratyn et al. 2012) with a database from the *Chlorella rbcL* reference sequence (KJ446796.1) from Gomaa et al. (2014). We added 208 representative *Chlorella rbcL* sequences (limited in length to 700-5,000bp) from GenBank (including sequences from Gomaa et al. (2014)) to our dataset and aligned all sequences using Mafft version 7.419 (Katoh and Standley 2013). We trimmed the alignment to match the length of the *rbcL* sequences from Gomaa et al. (2014) and used jModelTest2 version 2.1.10 to select the best substitution model using the CIPRES Science Gateway (Miller et al. 2010). We then built a maximum likelihood phylogeny using RAxML version 8.2.12 (Stamatakis 2014), with GTRGAMMAI specified as the substitution model and 100 bootstraps.

To test for evidence of multiple engulfing events of *Chlorella* algae by distinct protist lineages (i.e., amoebae, ciliates) we performed AU and SH tests with 10,000 re-samplings using the RELL method in IQ-TREE version 1.6.12 (Nguyen et al. 2015). We tested three hypotheses: 1) the best tree from the RAxML analysis above, 2) a constrained tree requiring monophyletic clades of amoebae- and ciliate-associated *Chlorella*, and 3) a constrained tree requiring a monophyletic clade of amoebae-associated *Chlorella*.

To evaluate the results of the *rbcL* phylogeny, and to assess evidence for both multiple *Chlorella* strains per amoeba cell and low coverage samples (samples without enough reads for successful *de-novo* assembly, but enough reads to determine *rbcL* type) we assembled the trimmed reads to the

Chlorella rbcL reference sequence KJ446796.1 using BBMap version 37.56 (Bushnell 2014). Reference alignments were checked by eye using Geneious version 2019.0.4 (Kearse et al. 2012).

Host diversity and phylogeny: To assess the genetic diversity of the hosts and the potential link with algal diversity we extracted mitochondrial cytochrome oxidase (COI) gene sequences from the *H. papilio* genome samples from which we obtained *rbcL* sequence data. We applied a read mapping approach using the BBMap toolkit (Bushnell 2014) to identify reads from the genome datasets that match *H. papilio* COI sequences from GenBank representing the diversity of genetic lineages encountered within this morphospecies to date (e.g., Gomaa et al. 2014; Heger et al. 2013; Singer et al. 2019). We then assembled the mapped reads using SPAdes (Bankevich et al. 2012) and selected the longest contigs for phylogenetic analysis. We produced a multiple sequence alignment with the inferred COI sequences, the GenBank *H. papilio* sequences and two sequences of *Nebela* sp. as outgroups using MAFFT version 7.419 (Katoh and Standley 2013). We trimmed the MSA to adjust the sequences to equal length and built a phylogenetic tree using RAxML version 8.2.12 (Stamatakis 2014) with rapid bootstrapping using a GTRCAT model and automatic halting of bootstrapping under the autoMRE criterion. All phylogenetic programs were accessed *via* the CIPRES Science Gateway (Miller et al. 2010). We visualized the tree in FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and rooted it with the outgroup sequences.

Fluorescence and transmission electron microscopy: To assess the state of the algae cells within *H. papilio*, we conducted fluorescence and transmission electron microscopy. For fluorescence microscopy, hand-picked *H. papilio* cells were preserved in 100% ethanol and subsequently infiltrated and embedded in medium grade LR-White resin. Semithin sections (500 nm) were mounted on glass slides and stained with 4′,6-diamidino-2-phenylindole (DAPI) to reveal the presence of nuclei. The stained sections were observed using an inverted epifluorescence microscope (Axio Observer.D1, Carl Zeiss, Jena, Germany) equipped with a monochrome high-resolution camera (AxioCam MRm, Carl Zeiss, Jena, Germany). The green autofluorescence was imaged and subtracted to the blue fluorescence in order to reveal the DNA signal. Autofluorescence subtraction and overlays were done using the GIMP® software.

For transmission electron microscopy (TEM) we again hand-picked individual *H. papilio* cells from environmental samples. We transferred the cells from the original petri dish to a drop of freshly-filtered (2

448 µm filter) in situ water and repeated this washing step until all obvious contamination on the outside of the
449 cells was removed. We then fixed the cells overnight in 3% glutaraldehyde (GTA) in 0.1 M PHEM buffer
450 (60 mM PIPES, 25 mM HEPES, 2 mM $MgCl_2$, 10 mM EGTA, pH 6.9).
451 After washing off the fixative, the cells were post-fixed for 1 hour using 1% aqueous osmium
452 tetroxide solution and subsequently washed three times in water before dehydration through an acetone
453 series. Cells were finally infiltrated and flat-embedded in Epon-Araldite resin as described in Müller-
454 Reichert et al. (2003). Thin sections (100 nm) were mounted on formvar coated slot grids and stained for
455 7 min in 2% uranyl acetate followed by 10 min in Reynolds Lead Citrate. The cells were observed and
456 imaged with a Jeol® 1400 TEM.
457 **Assessing algal nuclear activity in transcriptome samples:** If the algae residing within *H.*
458 *papilio* cells are actively transcribing their nuclear DNA, we expect their transcriptomes to be present in
459 our sequencing reads, as the nucleic acids of organisms living within the shell or cytoplasm of the
460 amoeba will be co-amplified. In order to search for an algae signal in our transcriptome data, we used
461 PhyloToL (Cerón-Romero et al. 2019) to produce multiple sequence alignments and gene trees for 150
462 conserved eukaryotic gene families (Supplementary Material Table S2). These gene families were
463 selected based on their presence in at least four of the five major eukaryotic clades (SAR, Amoebozoa,
464 Archaeplastida, Excavata and Opisthokonta). We included a total of 278 species of all major eukaryotic
465 clades, bacteria and archaea in the analysis in addition to our 22 *H. papilio* and 12 *H. elegans*
466 transcriptomes. If both the amoeba and algae transcriptomes were present in a sample, we would expect
467 to find sequences of one sample appearing in Amoebozoa and Archaeplastida in the gene trees. We
468 used custom Python scripts to determine the sister branch of each sequence of our transcriptome
469 samples in each gene tree (Supplementary Material Table S2). In addition, we calculated the branch
470 length to avoid long-branch artifacts. All identified cases were additionally screened by eye.
471 To further investigate active gene expression by the algae we also assessed the number of times
472 that *H. papilio* samples appeared among green algae in the gene trees of 27 photosynthesis related gene
473 families (Supplementary Material Table S2). We used the KEGG (Kanehisa and Goto 2000) pathway for
474 *Chlorella* photosynthesis to identify nuclear encoded genes associated with photosynthesis. These genes
475 were then analyzed using PhyloToL following the methods described above.

Starvation experiment: In order to investigate the stability of the amoeba/algae relationship over time, we maintained living *H. papilio* in the lab for two weeks. These cells were obtained from the environment as described above, isolated from the surrounding substrate by hand-picking from the petri-dish and placed in a new dish with freshly-filtered (0.2 µm filter) in situ water, which did not contain free-living *Chlorella*, *Sphagnum* or any other food sources. We kept the dishes in an incubator at 20 °C under a 12h light/dark cycle. We then removed two cells each on day 0, 3, 5, 7, and 12, photo-documented them and froze them in RLT buffer for subsequent single-cell genome analysis. The amplified genomes were sequenced and processed and the assemblies then scanned for chloroplast contigs using BLASTn. All procedures were conducted as described above. We used GeSeq (Tillich et al. 2017) for annotation of the contigs and to determine if they derive from chloroplasts and then mapped reads to the longest identified chloroplast contig using BBMap version 37.56 (Bushnell 2014) with default settings. The average coverage of each alignment was determined using bedtools genomecov (Quinlan and Hall 2010).

Differential gene expression: To assess whether the presence of the algae within the *H. papilio* cells has an influence on the gene expression of the amoeba we compared the composition of the transcriptomes of 16 of our *H. papilio* samples to our 12 samples of *H. elegans* (Supplementary Material Table S1), a closely related species without associated algae. After counting all gene families present in each transcriptome sample using custom Python scripts, we identified gene families differentially present in the two morphospecies. We then used PhyloToL (Cerón-Romero et al. 2019) to construct multiple sequence alignments and gene trees to assess homology prior to functional analysis in Blast2GO (Götz et al. 2008). To compare the functional categories of the differentially expressed genes, we analyzed a set of equivalent size that was made up of randomly chosen gene families shared among both amoeba species.

Acknowledgements

This study was financially supported by grants from the National Institute of Health (grant number R15HG010409) and the National Science Foundation (grant numbers OCE-1924570,

DEB-1651908) to LAK. JM, TT, and SVD were supported by the Gordon and Betty Moore Foundation (grant number GBMF7617). We thank the Donner Cryo-EM resource at the Lawrence Berkeley National Laboratory for providing access to the TEM. Work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. We thank Aaron Ellison for discussions on bog ecology and gratefully acknowledge the following Smith College individuals: Naomi Ostriker for conducting the starvation experiment, Olivia Dufour for the amoeba and *Chlorella* drawings and Auden Cote-L'Heureux for help with Python scripts. We also thank other members of the Katzlab for helpful comments on the manuscript and extend our thanks to the sequencing center at the Institute for Genome Sciences at the University of Maryland.

Data Availability

All sequenced transcriptomes are available on GenBank under the SRA BioProject PRJNA761372.

Figures

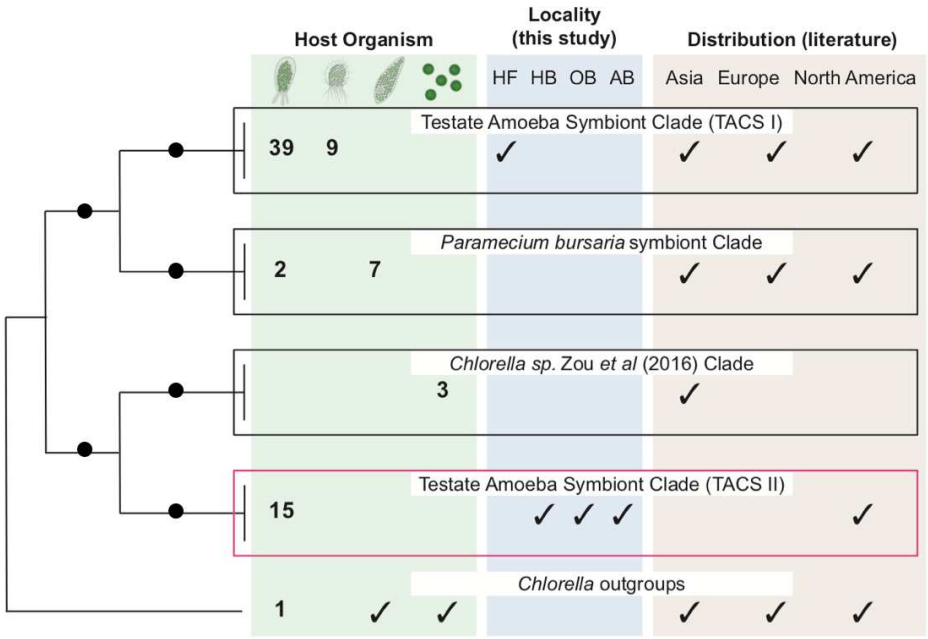


Figure 1. Genetic diversity and phylogeny of *Chlorella* algae. A maximum likelihood phylogeny of a ~700 bp portion of the large subunit of the ribulose-bisphosphate carboxylase (rbcL) chloroplast gene indicates that *Hyalosphenia papilio* samples from New England harbor two non-monophyletic *Chlorella* lineages (TACS I and II). The host organisms depicted with line drawings are *H. papilio*, *Placocista spinosa* (representing all other testate amoebae from Gomaa et al, 2014), *Paramecium bursaria,* as well as free-living *Chlorella*. The two TACS clades are separated by *Chlorella* clades that are either primarily associated with *Paramecium* or free-living (Zou et al. 2016 clade). Our work across the northeastern USA located lineage TACS II at three sites (Hawley Bog (HB), Orono Bog (OB) and Big Heath (AB)) and TACS I at one site (Harvard Forest (HF)), while previous world-wide work in testate amoebae predominantly recovered TACS I (Gooma et al. 2014). Numbers indicate the number of sequences in the alignment corresponding to a given host, check marks indicate geographic distribution, except in the case of the outgroups where they represent the 149 outgroup sequences split between free living and *Paramecium* associated. Bootstrap support values greater than 70 are indicated with filled circles.

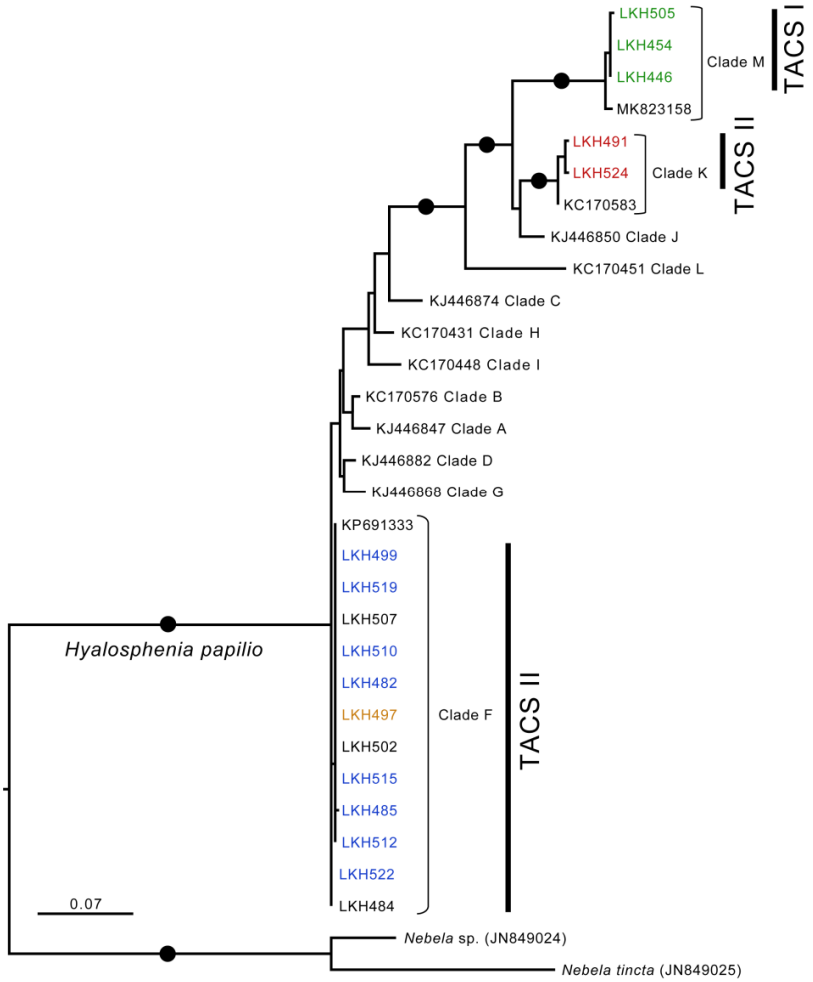
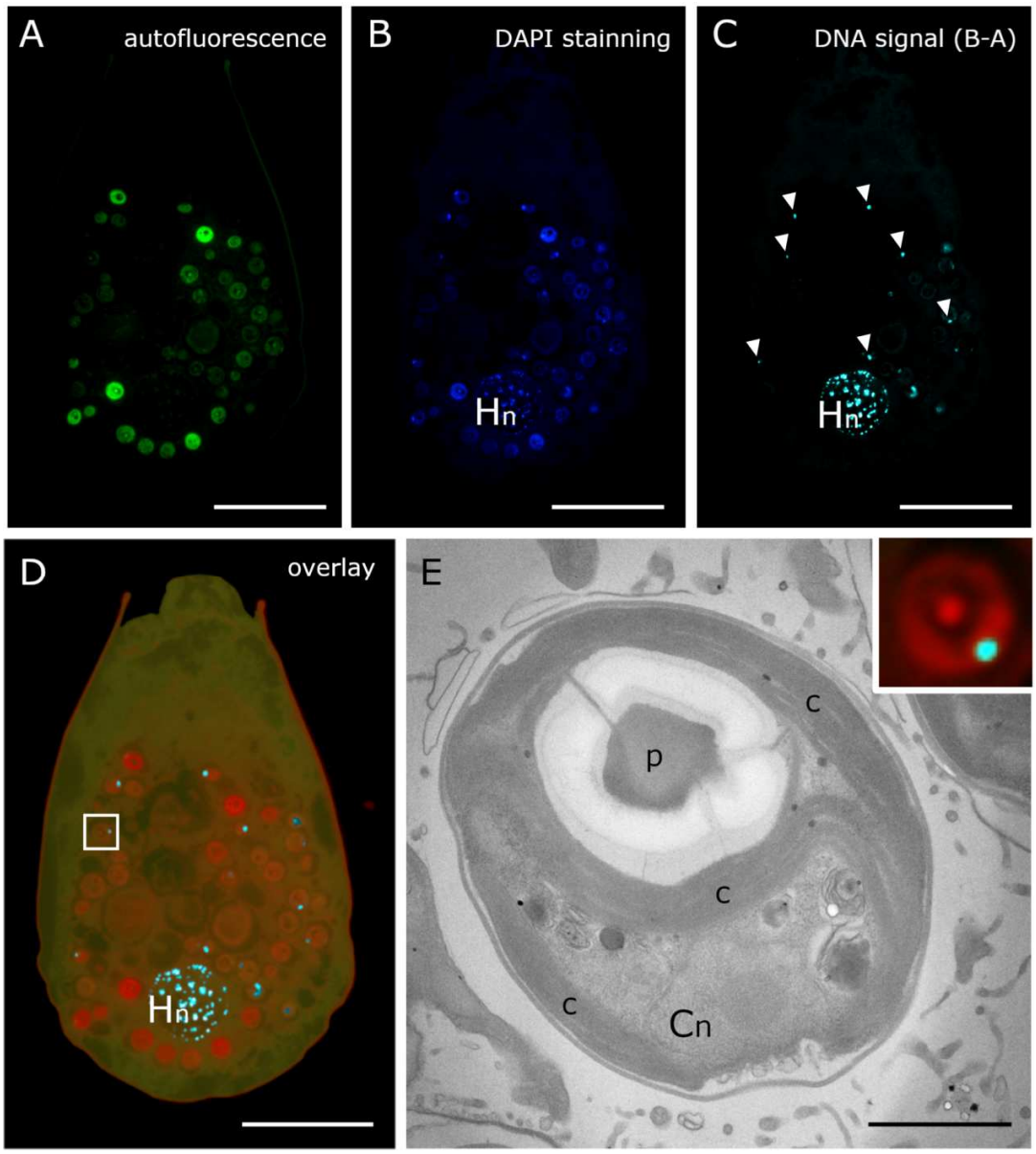


Figure 2. Genetic diversity within *Hyalosphenia papilio*. A maximum likelihood phylogeny of the genetic diversity within the morphospecies *H. papilio* based on a ~430 bp fragment of the mitochondrial cytochrome oxidase (COI) gene. Included in the tree are 17 COI sequences extracted from our genome samples, *H. papilio* sequences from GenBank representative for all known clades (Gomaa et al. 2014; Heger et al. 2013; Singer et al. 2019) and *Nebela* sp. sequences as outgroups. Our samples group among clades M, K and F. Samples in clade M are associated with symbionts of the TACS I clade and clades K and F with symbionts of TACS II. Colored in green are samples collected at Harvard Forest, in blue are samples from Hawley Bog, red from Acadia National Park and orange from Orono Bog. Bootstrap support values greater than 70 are indicated with filled circles.



555
556
557 **Figure 3.** *Hyalosphenia papilio* harboring *Chlorella* symbionts. **A-D.** Semi-thin sections of a
558 single *Hyalosphenia papilio* cell inspected by fluorescence microscopy, Hn *Hyalosphenia*
559 nucleus. **A.** Green autofluorescence of the *Chlorella* chloroplasts; **B.** Blue autofluorescence
560 along with DNA (DAPI staining) signal. **C.** DNA signal (teal color, the DAPI-stained structures
561 yielded by subtraction of the green autofluorescence (A) from the blue fluorescence (B)),
562 *Chlorella* nuclei (arrowheads); **D.** Overlay of fine structure, DAPI signal, and autofluorescence
563 across blue, green, red and far-red channels. **E.** Fine structure of *Chlorella* symbiont showing its
564 nucleus (Cn), chloroplast (c), and pyrenoid (p); insert displays detail of fine structure and

fluorescence overlay (D, area delimited by white square; nucleus = teal color, chloroplast and pyrenoid autofluorescence = red). Scale bar A-D = 30 µm; scale bar E = 1 µm (insert = ×5.3).

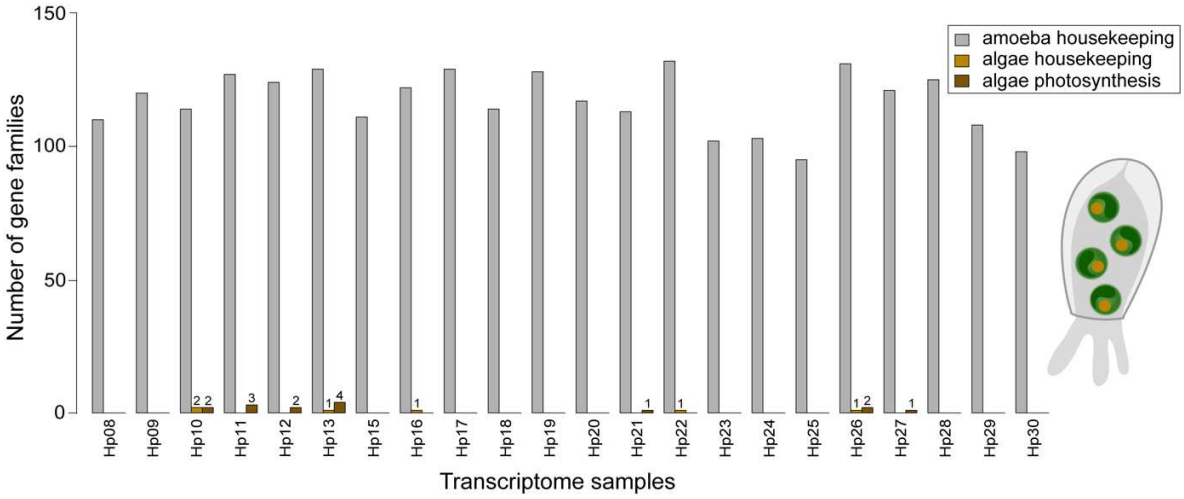
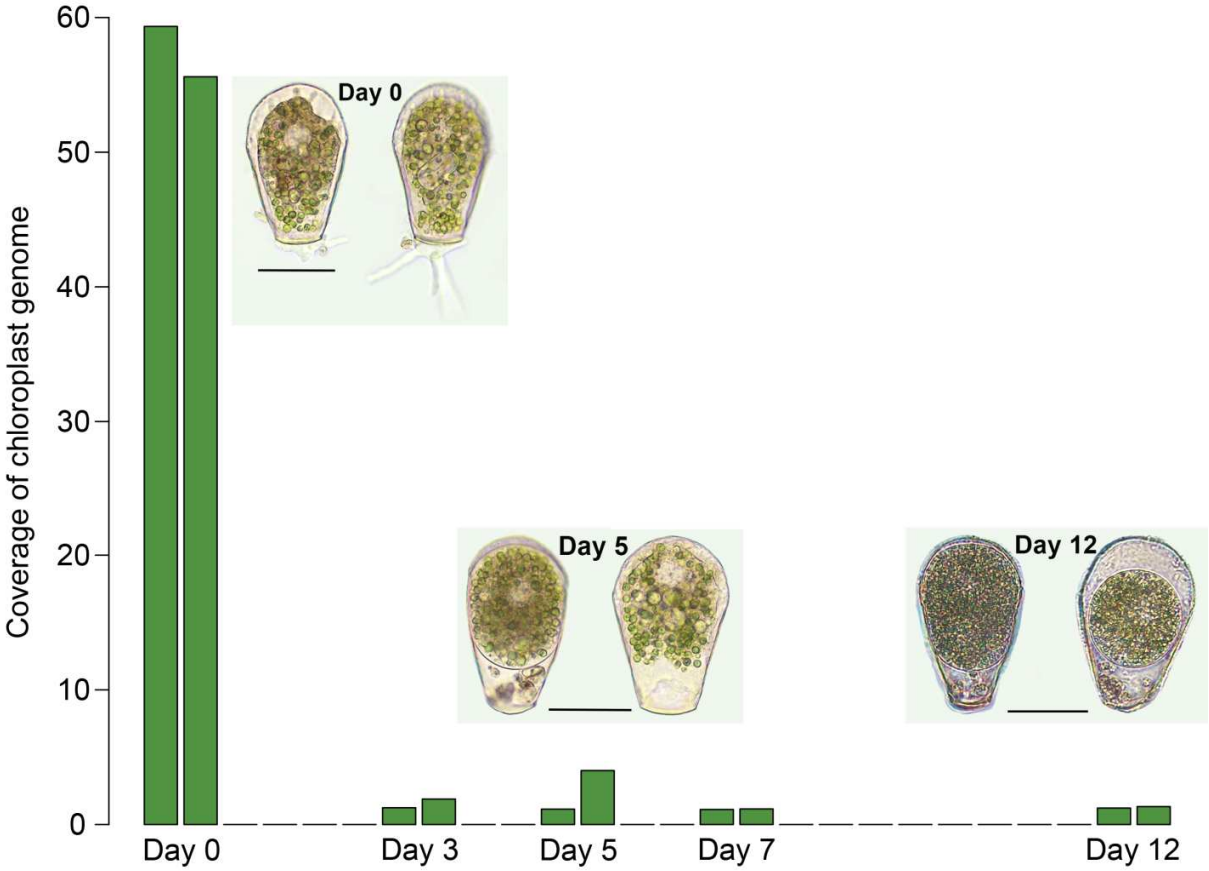


Figure 4. Presence of expressed algal nuclear genes (housekeeping and photosynthesis related) in *Hyalosphenia papilio* transcriptome samples. The bar chart indicates the number of gene families – out of 150 conserved housekeeping gene families and 27 photosynthesis related gene families – that are expressed by either *H. papilio* (grey) and/or the *Chlorella* symbionts (yellow, brown) in each of the samples from our transcriptome dataset. Exact numbers of gene families expressed by the algae are shown above the bars. PhyloToL was used to produce gene trees and assess the position of sequences among either Amoebozoa or Viridiplantae. In the housekeeping gene families very few contained sequences were classified as Viridiplantae, which suggests that *Chlorella* housekeeping genes are not being actively transcribed. There were also very few nuclear encoded photosynthesis genes recovered from the transcriptomes (Supplementary Material Table S2).



582
583
584 **Figure 5.** Starvation experiment. Evidence of a transient relationship seen in the reduction of
585 *Chlorella* chloroplast genome during a starvation experiment, indicated as average coverage
586 (average depth of reads per reference base). High genome coverage on Day 0 indicates healthy
587 chloroplasts, however, they quickly degrade in micrographs and genome coverage as the
588 amoeba is deprived of food over a period of 12 days. Scale bars in micrographs are 50 µm.
589

Supplementary Material

Table S1. Sampling table indicating details of all single-cell genomes and transcriptomes obtained for this study.

Table S2. Presence/absence of *H. papilio* sequences in the gene trees of 150 selected housekeeping gene families and 27 photosynthesis related gene families generated using PhyloToL. Further indicated are the sister branches to each sequence and the number of times we detect *Chlorella* signal in the *H. papilio* samples.

Table S3. Functional characterization using Blast2GO on genes expressed in *H. papilio*, but not *H. elegans* cells suggest that the algal symbiont is causing oxidative stress. We find eight genes involved in oxidation-reduction processes from six gene families.

Figure S1. Reads from two samples mapped to the TACSI *rbcL* reference from Gomaa et al. (2014; KJ446796.1). Differences from the reference are colored in the alignment. **A.** LKH454 from TACSI has no fixed differences to reference, and no consistent polymorphisms, though a few reads that differ at the 3' end. **B.** LKH484 from the TACSII lineage has many fixed differences from the reference but few polymorphisms.

Figure S2. Food vacuole containing *Chlorella* cells. In contrast with the intact *Chlorella* cells located in the *Hyalosphenia* cytoplasm (Fig. 2E), the partially digested cells present in the food vacuole lack a distinguishable nucleus, chloroplast, or pyrenoid.

Figure S3. Light microscope images of *H. papilio* cells from the starvation experiment sampled after 0, 3, 5, 7 and 12 days showing the gradual digestion of the *Chlorella* algae. Scale bars indicate 50 µm.

File S1. FASTA file of a MAFFT alignment of the rbcL sequences used for phylogenetic inference of *Chlorella* diversity.

File S2. RAxML tree showing *Chlorella* diversity. This tree formed the basis for the cartoon tree shown in Figure 1.

File S3. FASTA file of an alignment of the COI sequences used for inference of the genetic diversity within the *H. papilio* samples.

File S4. RAxML tree showing the *H. papilio* genetic diversity and separation into different genetic lineages. This tree includes sequences from all currently known genetic lineages of *H. papilio* as published in Gomaa et al. (2014) and Heger et al. (2013).

File S5. FASTA file of representative sequences that are differentially expressed in *H. papilio* compared to *H. elegans* and that may play a role in oxidative processes invoked by photosynthesis (Supplementary Material Table S3).

References

Andrews S (2010) Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [WWW Document]. URL https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19**:455–477

Betteridge DJ (2000) What is oxidative stress? Metabolism, Advances in Oxidative Stress. Proceedings of an "Expert Session" held on the Occasion of the Annual Meeting of the European Association for the study of Diabetes **49**:3–8. https://doi.org/10.1016/S0026-0495(00)80077-3

Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL (2012) Domain enhanced lookup time accelerated BLAST. Biol Direct **7**:12

Bronstein JL (1994) Conditional outcomes in mutualistic interactions. Trends Ecol Evol **9**:214–217

Bushnell B (2014) BBMap: A Fast, Accurate, Splice-Aware Aligner (No. LBNL-7065E). Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States)

Cerón-Romero MA, Maurer-Alcalá XX, Grattepanche J-D, Yan Y, Fonseca MM, Katz LA (2019) PhyloToL: A taxon/gene-Rich phylogenomic pipeline to explore genome evolution of diverse eukaryotes. Mol Biol Evol **36**:1831–1842

Correia MJ, Lee JJ (2002) How long do the plastids retained by *Elphidium excavatum* (Terquem) last in their host? Symbiosis **32**:27–38

Davis RB, Anderson DS (2001) Classification and distribution of freshwater peatlands in Maine. Northeastern Natural **8**:1–50

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26**:2460–2461

Esteban GF, Fenchel T, Finlay BJ (2010) Mixotrophy in Ciliates. Protist **161**:621–641

Fields SD, Rhodes RG (1991) Ingestion and Retention of *Chroomonas* spp. (Cryptophyceae) by *Gymnodinium acidotum* (Dinophyceae). J Phycol **27**:525–529

Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA (2017) The evolution of host-symbiont dependence. Nat Commun **8**:15973

672 **Flemming FE, Potekhin A, Pröschold T, Schrallhammer M** (2020). Algal diversity in
673 *Paramecium bursaria*: Species identification, detection of *Choricystis parasitica*, and
674 assessment of the interaction specificity. Diversity **12**:287
675 **Foissner W** (2007) Protist Diversity and Distribution: Some Basic Considerations. In Foissner
676 W, Hawksworth DL (eds) Protist Diversity and Geographical Distribution. Topics in
677 Biodiversity and Conservation, Vol 8, Springer, Dordrecht, pp 1-8
678 **Fujishima M, Kodama Y** (2012) Endosymbionts in *Paramecium*. Europ J Protistol **48**:124-137
679 **Gast RJ, Moran DM, Dennett MR, Caron DA** (2007) Kleptoplasty in an Antarctic dinoflagellate:
680 caught in evolutionary transition? Environ Microbiol **9**:39–45
681 **Gomaa F, Kosakyan A, Heger TJ, Corsaro D, Mitchell EAD, Lara E** (2014) One alga to rule
682 them all: Unrelated mixotrophic testate amoebae (Amoebozoa, Rhizaria and
683 Stramenopiles) share the same symbiont (Trebouxiophyceae). Protist **165**:161–176
684 **Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón**
685 **M, Dopazo J, Conesa A** (2008) High-throughput functional annotation and data mining
686 with the Blast2GO suite. Nucleic Acids Res **36**:3420–3435
687 **Heal OW** (1962) The abundance and micro-distribution of testate amoebae
688 (Rhizopoda:Testacea) in *Sphagnum*. Oikos **13**:35–47
689 **Heger TJ, Mitchell EAD, Leander BS** (2013) Holarctic phylogeography of the testate amoeba
690 *Hyalosphenia papilio* (Amoebozoa: Arcellinida) reveals extensive genetic diversity
691 explained more by environment than dispersal limitation. Mol Ecol **22**:5172-5184
692 **Johnson MD, Oldach D, Delwiche CF, Stoecker DK** (2007) Retention of transcriptionally
693 active cryptophyte nuclei by the ciliate *Myrionecta rubra*. Nature **445**:426–428
694 **Katoh K, Standley DM** (2013) MAFFT Multiple Sequence Alignment Software Version 7:
695 Improvements in performance and usability. Mol Biol Evol **30**:772–780
696 **Kanehisa M, Goto S** (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids
697 Res **28**:27–30
698 **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,**
699 **Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A** (2012)
700 Geneious Basic: An integrated and extendable desktop software platform for the
701 organization and analysis of sequence data. Bioinformatics **28**:1647–1649
702 **Kearsley J** (1999) Non-forested Acidic Peatlands of Massachusetts: A Statewide Inventory and
703 Vegetation Classification. Massachusetts Natural Heritage and Endangered Species
704 Program, Westborough, MA, 30

Kodama Y, Suzuki H, Dohra H, Sugii M, Kitazume T, Yamaguchi K, Shigenobu S, Fujishima M (2014) Comparison of gene expression of *Paramecium bursaria* with and without *Chlorella variabilis* symbionts. BMC Genomics **15**:183

Lahr DJG, Kosakyan A, Lara E, Mitchell EAD, Morais L, Porfirio-Sousa AL, Ribeiro GM, Tice AK, Pánek T, Kang S, Brown MW (2019) Phylogenomics and morphological reconstruction of Arcellinida testate amoebae highlight diversity of microbial eukaryotes in the Neoproterozoic. Curr Biol **29**:991-1001.e3.

Lara E, Gomaa F (2017) Symbiosis Between Testate Amoebae and Photosynthetic Organisms. In Grube M, Seckbach J, Muggia L (eds) Algal and Cyanobacteria Symbioses. World Scientific Europe, pp 399–419

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res **13**:2178–2189

McManus GB, Katz LA (2009) Molecular and morphological methods for identifying plankton: what makes a successful marriage? J Plankton Res **31**:1119–1129

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In 2010 Gateway Computing Environments Workshop (GCE). Ieee, pp 1–8

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

Nowack ECM, Melkonian M (2010) Endosymbiotic associations within protists. Philos Trans R Soc Lond B Biol Sci **365**:699–712

Pogson BJ, Woo NS, Förster B, Small ID (2008) Plastid signalling to the nucleus and beyond. Trends Plant Sci **13**:602–609

Pröschold T, Darienko T, Silva PC, Reisser W, Krienitz L (2011) The systematics of *Zoochlorella* revisited employing an integrative approach. Environ Microbiol **13**:350-364

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**:841–842

Rauch C, Vries J de, Rommel S, Rose LE, Woehle C, Christa G, Laetz EM, Wägele H, Tielens AGM, Nickelsen J, Schumann T, Jahns P, Gould SB (2015) Why it Is time to look beyond algal genes in photosynthetic slugs. Genome Biol Evol **7**:2602–2607

Richier S, Furla P, Plantivaux A, Merle P-L, Allemand D (2005) Symbiosis-induced adaptation to oxidative stress. J Exp Biol **208**:277–285

738 **Ruggiero A, Grattepanche J-D, Weiner AKM, Katz LA** (2020) High diversity of testate
739 amoebae (Amoebozoa, Arcellinida) detected by HTS analyses in a New England fen
740 using newly designed taxon-specific primers. J Eukaryot Microbiol **67**:450–462
741 **Rumpho ME, Pelletreau KN, Moustafa A, Bhattacharya D** (2011) The making of a
742 photosynthetic animal. J Exp Biol 214:303–311
743 **Schmidt C, Morard R, Romero O, Kucera M** (2018) Diverse internal symbiont community in
744 the endosymbiotic foraminifera *Pararotalia calcariformata*: Implications for symbiont
745 shuffling under thermal stress. Front Microbiol **9**:2018
746 **Singer D, Mitchell EAD, Payne RJ, Blandenier Q, Duckert C, Fernández LD, Fournier B,**
747 **Hernández CE, Granath G, Rydin H, Bragazza L, Koronatova NG, Goia I, Harris LI,**
748 **Kajukało K, Kosakyan A, Lamentowicz M, Kosykh NP, Vellak K, Lara E** (2019)
749 Dispersal limitations and historical factors determine the biogeography of specialized
750 terrestrial protists. Mol Ecol **28**:3089–3100
751 **Stamatakis A** (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of
752 large phylogenies. Bioinformatics **30**:1312–1313
753 **Stoecker D, Johnson M, deVargas C, Not F** (2009) Acquired phototrophy in aquatic protists.
754 Aquat Microb Ecol **57**:279–310
755 **Swan JMA, Gill AM** (1970) The origins, spread, and consolidation of a floating bog in Harvard
756 Pond, Petersham, Massachusetts. Ecology **51**:829–840
757 **Takahashi T** (2016) Simultaneous evaluation of life cycle dynamics between a host
758 *Paramecium* and the endosymbionts of *Paramecium bursaria* using capillary flow
759 cytometry. Sci Rep **6**:31638
760 **Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S** (2017)
761 GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res
762 **45**:W6–W11
763 **Tsuchiya M, Miyawaki S, Oguri K, Toyofuku T, Tame A, Uematsu K, Takeda K, Sakai Y,**
764 **Miyake H, Maruyama T** (2020) Acquisition, maintenance, and ecological roles of
765 kleptoplasts in *Planoglabratella opercularis* (Foraminifera, Rhizaria). Front Mar Sci
766 **7**:2020
767 **Weiner AKM, Cerón-Romero MA, Yan Y, Katz LA** (2020) Phylogenomics of the epigenetic
768 toolkit reveals punctate retention of genes across eukaryotes. Genome Biol Evol
769 **12**:2196–2210

Zagata P, Greczek-Stachura M, Tarcz S, Rautian M (2016) The evolutionary relationships between endosymbiotic green algae of *Paramecium bursaria* syngens originating from different geographical locations. Folia Biol (krakow) **6**4:47–54

Zou S, Fei C, Song J, Bao Y, He M, Wang C (2016) Combining and comparing coalescent, distance and character-based approaches for barcoding microalgae: A test with *Chlorella*-like species (Chlorophyta). PloS ONE **11**:e0153833