Current Biology

Article

A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements

Highlights

- A cryptomonad hosts two distinct bacterial endosymbionts and a bacteriophage
- The bacteriophage infects the endosymbiont, *Megaira polyxenophila*
- Both bacterial endosymbionts and bacteriophage encode eukaryotic-like proteins
- Seven distinct genomes are present in the single-celled cryptomonad

Authors

Emma E. George, Dovilė Barcytė, Gordon Lax, ..., Julius Lukeš, Marek Eliáš, Patrick J. Keeling

Correspondence

3mma6eorg3@gmail.com

In brief

George et al. describe the genomic and metabolic complexity of two bacterial endosymbionts and an endosymbiont-infecting bacteriophage in the single-celled alga, *Cryptomonas gyropyrenoidosa*. This complex symbiosis involves seven genomes within a single eukaryotic cell and has been retained in culture for over 50 years.

George et al., 2023, Current Biology *33*, 1–15
May 22, 2023 © 2023 Elsevier Inc.
https://doi.org/10.1016/j.cub.2023.04.010

CellPress

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

Current Biology

CellPress

Article

A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements

Emma E. George,[1,6,7,*] Dovilė Barcytė,[2,3] Gordon Lax,[1] Sam Livingston,[1] Daria Tashyreva,[4] Filip Husnik,[3] Julius Lukeš,[4,5] Marek Eliáš,[2] and Patrick J. Keeling[1]

[1]University of British Columbia, Department of Botany, Vancouver V6T 1Z4, Canada
[2]University of Ostrava, Faculty of Science, Department of Biology and Ecology, 701 00 Ostrava, Czech Republic
[3]Okinawa Institute of Science and Technology, Okinawa, 904-0495, Japan
[4]Institute of Parasitology, Biology Center, Czech Academy of Sciences, 370 05 České Budějovice (Budweis), Czech Republic
[5]University of South Bohemia, Faculty of Sciences, 370 05 České Budějovice (Budweis), Czech Republic
[6]Present address: Scripps Institution of Oceanography, Integrative Oceanography Division, University of California, San Diego, La Jolla 92093, CA, USA
[7]Lead contact
*Correspondence: 3mma6eorg3@gmail.com
https://doi.org/10.1016/j.cub.2023.04.010

SUMMARY

Symbiosis between prokaryotes and microbial eukaryotes (protists) has broadly impacted both evolution and ecology. Endosymbiosis led to mitochondria and plastids, the latter spreading across the tree of eukaryotes by subsequent rounds of endosymbiosis. Present-day endosymbionts in protists remain both common and diverse, although what function they serve is often unknown. Here, we describe a highly complex community of endosymbionts and a bacteriophage (phage) within a single cryptomonad cell. Cryptomonads are a model for organelle evolution because their secondary plastid retains a relict endosymbiont nucleus, but only one previously unidentified *Cryptomonas* strain (SAG 25.80) is known to harbor bacterial endosymbionts. We carried out electron microscopy and FISH imaging as well as genomic sequencing on *Cryptomonas* SAG 25.80, which revealed a stable, complex community even after over 50 years in continuous cultivation. We identified the host strain as *Cryptomonas gyropyrenoidosa*, and sequenced genomes from its mitochondria, plastid, and nucleomorph (and partially its nucleus), as well as two symbionts, *Megaira polyxenophila* and *Grellia numerosa*, and one phage (MAnkyphage) infecting *M. polyxenophila*. Comparing closely related endosymbionts from other hosts revealed similar metabolic and genomic features, with the exception of abundant transposons and genome plasticity in *M. polyxenophila* from *Cryptomonas*. We found an abundance of eukaryote-interacting genes as well as many toxin-antitoxin systems, including in the MAnkyphage genome that also encodes several eukaryotic-like proteins. Overall, the *Cryptomonas* cell is an endosymbiotic conglomeration with seven distinct evolving genomes that all show evidence of inter-lineage conflict but nevertheless remain stable, even after more than 4,000 generations in culture.

INTRODUCTION

Symbioses between microbial eukaryotes (protists) and prokaryotes (bacteria and archaea) are extremely diverse in both taxonomy and metabolic functions, and the complex interactions involved in symbioses led to various evolutionary outcomes.[1] Two well-studied ancient endosymbioses resulted in stable, genetically integrated organelles (mitochondria and plastids), but the more common evolutionary outcomes involve the replacement of the endosymbiont in the host (i.e., extinction of the endosymbiont)[2,3] or short-term invasion by “professional” endosymbionts.[4,36]

Cryptomonad algae (cryptophytes) have been well studied primarily because they are a model for organelle evolution. The cryptomonad plastid was acquired through secondary endosymbiosis with a red alga, but unlike most other plastid acquisitions via secondary endosymbiosis, cryptomonads retain a highly reduced nucleus of the red algal symbiont called a nucleomorph, encoding approximately 500 genes.[5,6] Thus, most cryptomonads harbor four distinct genomes: nuclear, mitochondrial, plastid, and nucleomorph. These are already complex cells, but in one cryptomonad, endosymbiosis has gone even further. The freshwater *Cryptomonas* strain SAG 25.80 has been shown by microscopy to contain intracellular bacteria, some of which harbor virus-like particles (VLPs).[7,8] Endosymbiont and VLP abundance varies with the host growth phase, and the attempts to completely remove the endosymbionts with antibiotics have been unsuccessful, implying their tight integration and perhaps an essential role for the host.[8]

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

Table 1. Summary of sequenced genomes from *Cryptomonas gyropyrenoidosa* SAG 25.80

Genome	Taxonomy	Host	Genome size (bp)	%GC	Coverage	Plasmids
MAnkyphage	*Caudoviricetes*	*Megaira polyxenophila*	38,450	33.7	51,533×	–
Megaira polyxenophila	*Rickettsiales*; *Rickettsiaceae*	*Cryptomonas gyropyrenoidosa*	1,727,493	34	740×	2
Grellia numerosa	*Rickettsiales*; *Midichloriaceae*	*Cryptomonas gyropyrenoidosa*	1,448,196	30.7	541×	1
Plastid	Cryptomonad	*Cryptomonas gyropyrenoidosa*	128,773	33.8	1,746×	–
Mitochondrion	Cryptomonad	*Cryptomonas gyropyrenoidosa*	38,907	30.2	441×	–
Nucleomorph Chromosome 1	Cryptomonad	*Cryptomonas gyropyrenoidosa*	192,567	22.3	20×	–
Nucleomorph Chromosome 2	Cryptomonad	*Cryptomonas gyropyrenoidosa*	171,999	22.7	23×	–
Nucleomorph Chromosome 3	Cryptomonad	*Cryptomonas gyropyrenoidosa*	117,437	23.1	22×	–

See also Figures S3 and S5.

Tripartite systems with phage, endosymbionts, and eukaryotes are rarely observed, with most examples coming from animals.[9–11] However, the systems that have been studied suggest a complex network of interactions among all partners. The best characterized system is the arthropod-infecting endosymbiont *Wolbachia* and its phage WO, which encodes several eukaryotic proteins, including a spider toxin.[12] Endosymbiont-infecting phages can also provide protection to their bacterial and eukaryotic hosts. For example, phage-encoded toxins produced by the bacterial endosymbiont *Hamiltonella defensa* protect the aphid host against a parasitoid wasp,[10] and a phage-encoded protein enables bacterial symbionts in sponges to evade the host's immune system.[13] Very few endosymbiont-infecting phages are known from protists,[14–16] and most examples come from transmission electron microscopy (TEM) observations without genomic data to confirm the presence of phages.[7,17,18]

The *Cryptomonas* system is a particularly interesting case involving a protist host, because it also serves as an evolutionary microcosm experiment of sorts, since the culture has a long and unusually well-documented history. The *Cryptomonas* strain in question was incorporated into the Culture Collection of Algae at Göttingen University (SAG) in 1980, but the culture originated from the personal research collection of the famous naturalist and SAG founder, Ernst Georg Pringsheim, who isolated the strain pre-1970 (personal communication by Dr. Maike Lorenz, SAG Curator). The presence of bacterial endosymbionts and VLPs in *Cryptomonas* sp. SAG 25.80 was not discovered until 1988,[7] but the endosymbionts were certainly present when the culture was originally established because the culture was maintained in a lab setting with no exposure to environmental samples or other cultures. Additionally, *Cryptomonas* sp. SAG 25.80 has continuously grown via serial transfers since its isolation and has never undergone a "pause" in generations due to cryopreservation.

Here, we carry out genomic characterization of *Cryptomonas* sp. SAG 25.80, identified as *Cryptomonas gyropyrenoidosa*, and show that it harbors two different *Rickettsiales* endosymbionts and an endosymbiont-infecting phage. We characterize the genomes of the bacteria and phage, which altogether form a quadripartite symbiosis with complex viral-bacterial-eukaryotic interactions that have been stable for over 50 years in culture.

RESULTS

The complex community within *Cryptomonas gyropyrenoidosa* SAG 25.80 reveals seven distinct genomes

Hybrid genomic assemblies with long-read and short-read sequences produced high coverage organellar genomes from *Cryptomonas* sp. SAG 25.80, endosymbiotic bacterial genomes, and a single phage genome (Table 1). The 18S rRNA gene and the ITS2 region of *Cryptomonas* sp. SAG 25.80 shared high sequence similarity (>99%) with those previously reported from the authentic strain of *Cryptomonas gyropyrenoidosa* (GenBank: AJ421149.1 and AJ566154.1, respectively),[19] substantiating identification of the SAG strain as this species. An analysis of 16S rRNA genes present in the assembly revealed two bacterial endosymbionts. One exhibited 100% sequence identity to *Candidatus* Megaira polyxenophila (*Rickettsiaceae*) (Figure S1), and an additional analysis of the RNA polymerase β' subunit also confirmed the placement of the *Cryptomonas* endosymbiont within the *Ca.* Megaira polyxenophila clade (Figure S2). The other endosymbiont was assigned to *Ca.* Grellia numerosa (also known as *Ca.* Bandiella numerosa; *Midichloriaceae*), although it exhibited seven mismatches when compared with the type strain of the species (Figure S1). Both bacteria were confirmed to be endosymbionts of *C. gyropyrenoidosa* (see fluorescence *in situ* hybridization [FISH] results below), and the phage was confirmed to infect *Ca.* Megaira polyxenophila (see results below). The endosymbionts will be referred to without the *Candidatus* prefix from here on out.

From the algal host cell, three out of the four *C. gyropyrenoidosa* genomes were assembled: mitochondrial, plastid, and nucleomorph. The gene content of the *C. gyropyrenoidosa* mitochondrial and plastid (Figure S3) genomes was similar to other published cryptomonad genomes.[20–22] The nucleomorph genome was assembled into the canonical three-chromosomal organization found in all other cryptomonad nucleomorphs,[23] including *Cryptomonas*.[24] Five chromosomal ends harbored a subtelomeric repeat containing the standard rRNA operon (18S-5.8S-28S rRNA genes) followed by the 5S rRNA gene, as is common for cryptomonad nucleomorph chromosomes, while the remaining end lacked the rRNA operon and included only the 5S rRNA gene, as has been found for two nucleomorph chromosome ends in the closely related species *C. paramecium*.[23] The nuclear genome of *C. gyropyrenoidosa* was also present at low coverage and was

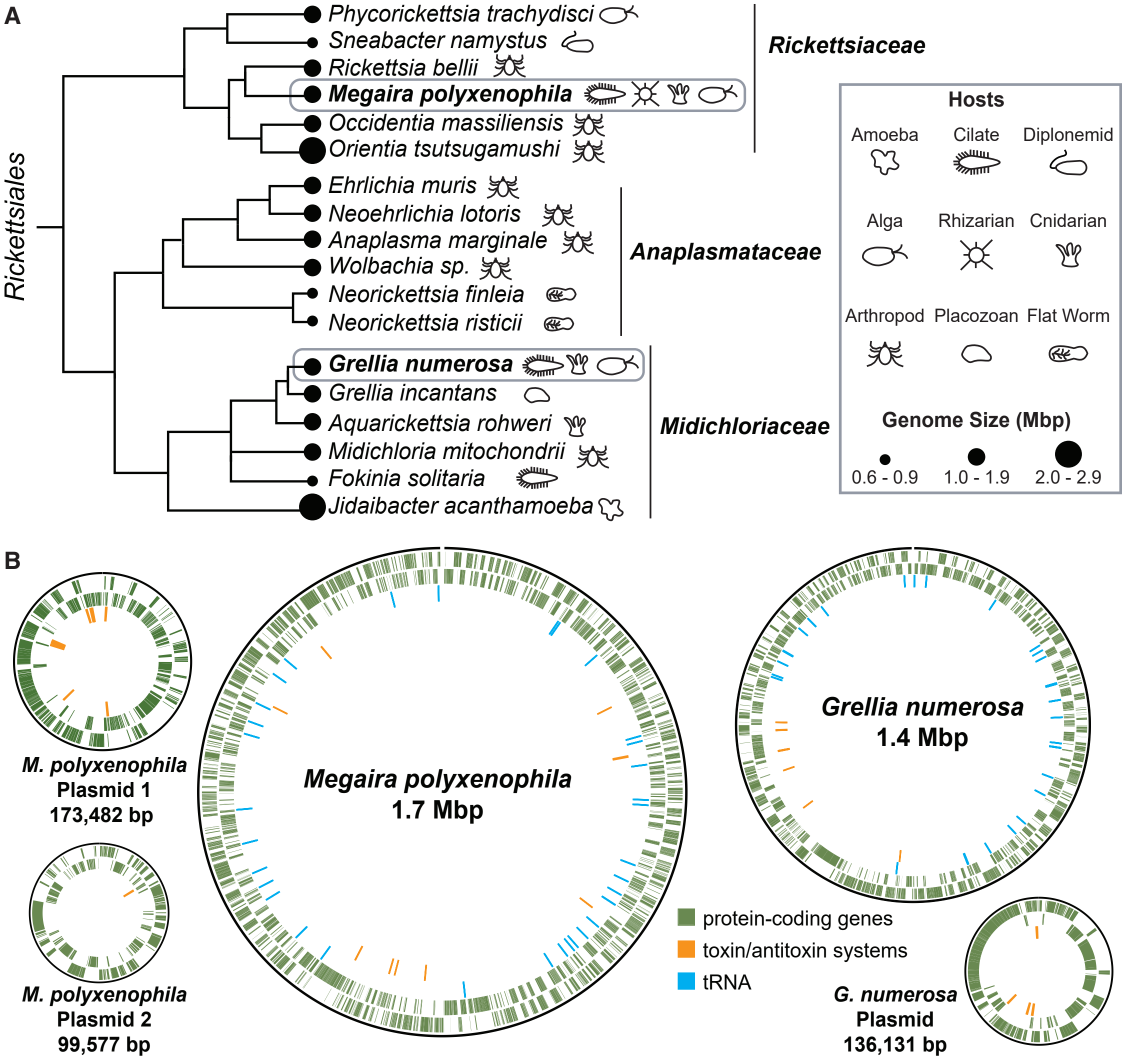

Figure 1. Genomic overview of bacterial endosymbionts in *Cryptomonas gyropyrenoidosa* SAG 25.80

(A) A 16S rDNA phylogeny-based schematic of the *C. gyropyrenoidosa* endosymbionts, *M. polyxenophila* (*Rickettsiaceae*) and *G. numerosa* (*Midichloriaceae*) and their closely related *Rickettsiales* taxa. Genome sizes are indicated by black circles, and icons depict the host of the endosymbiont.

(B) Characterization of *M. polyxenophila* and *G. numerosa* chromosomes and plasmids.

Genome plots (not drawn to the same scale) display protein-coding genes (green), tRNA (blue), toxin and antitoxin systems (orange). Note: eight TAs are present in the *G. numerosa* genome, but several TAs are encoded together and appear as a single TA on the genome plot. See also Figures S1 and S2 and Data S1A.

highly fragmented (>30,000 contigs), which is characteristic of other sequenced cryptomonad nuclear genomes.[25] The nuclear genome was not investigated further in this study.

The genomes of the two endosymbiotic bacteria were also assembled: the *M. polyxenophila* genome consisted of a 1.7-Mbp chromosome (98.6% completeness) and two plasmids (173,482 and 99,577 bp; BioSample SAMN30671969), while *G. numerosa* had a 1.4-Mbp chromosome (100% completeness) and one plasmid (136,132 bp; BioSample SAMN30671970) (Figure 1). The *M. polyxenophila* genome also contained a larger number of protein-coding genes (1,915) than that of *G. numerosa* (1,403), but in *M. polyxenophila*, transposable elements made up a greater proportion of coding sequences (>340 transposons, including insertion sequences; 16%), compared with *G. numerosa* (50 transposons; 4%). Both endosymbionts had multiple toxin-antitoxin (TA) systems that presumably serve as transcription and translation regulators, and seven out of eight TAs were located in a specific region of the *G. numerosa* chromosome, whereas the nine identified TAs in *M. polyxenophila* were spread throughout the chromosome (Figure 1; Data S1A). Over ten TAs were also present on the two *M. polyxenophila* plasmids.

Finally, a high coverage, complete phage genome (38,448 bp) belonging to *Caudoviricetes* (Figure 2) was also present in the *Cryptomonas* metagenome assembly. The GC content of phage and host genomes are typically similar,[26] and the *Caudoviricetes* GC content (34%) matched that of *M. polyxenophila* (34%) but not *G. numerosa* (31%) or any other assembled bacterial genomes from the culture. Additionally, a break in the *M. polyxenophila* genome assembly was the result of

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

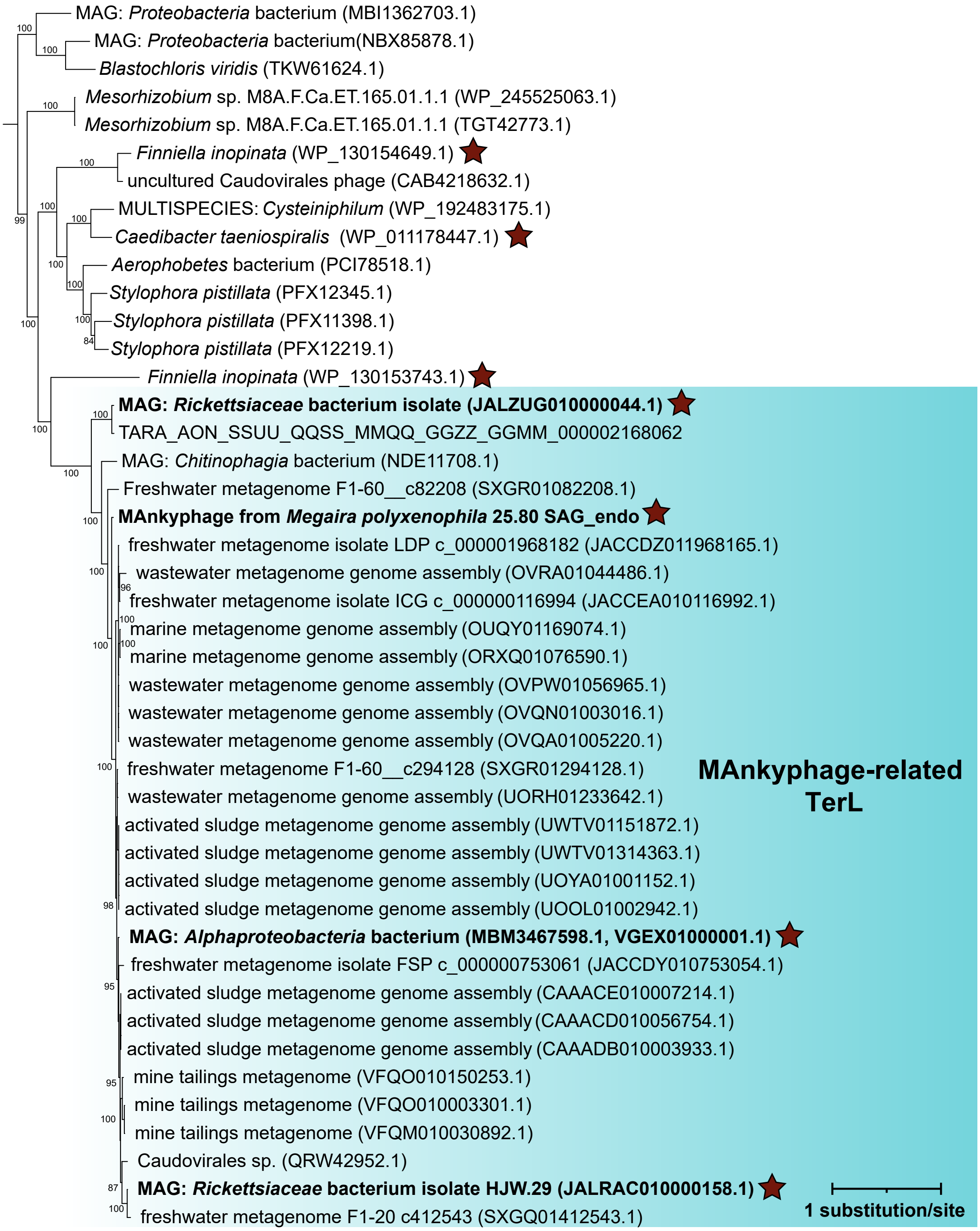

Figure 2. A maximum likelihood phylogenetic tree inferred from an alignment of phage large terminase subunit

The dataset used for the tree inference consisted of 5,256 sequences with 606 aligned positions. Star icons indicate phage sequences from known bacterial endosymbionts, and sequences referenced in the text are highlighted in bold. The dataset was assembled by combining MAnkyphage large terminase subunit (TerL), best hits to it in the NCBI nr protein sequence database, additional sequences highly similar to the MAnkyphage TerL obtained by conceptual translation of nucleotide sequences identified by TBLASTN searches against the whole-genome shotgun (WGS) NCBI database, and 5,130 reference TerL sequences from Benler et al.[66] The substitution model employed was LG + F + G4, support values represent 1,000 bootstrap pseudoreplicates. See also Data S3 for the full-length tree. See also Data S1H.

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

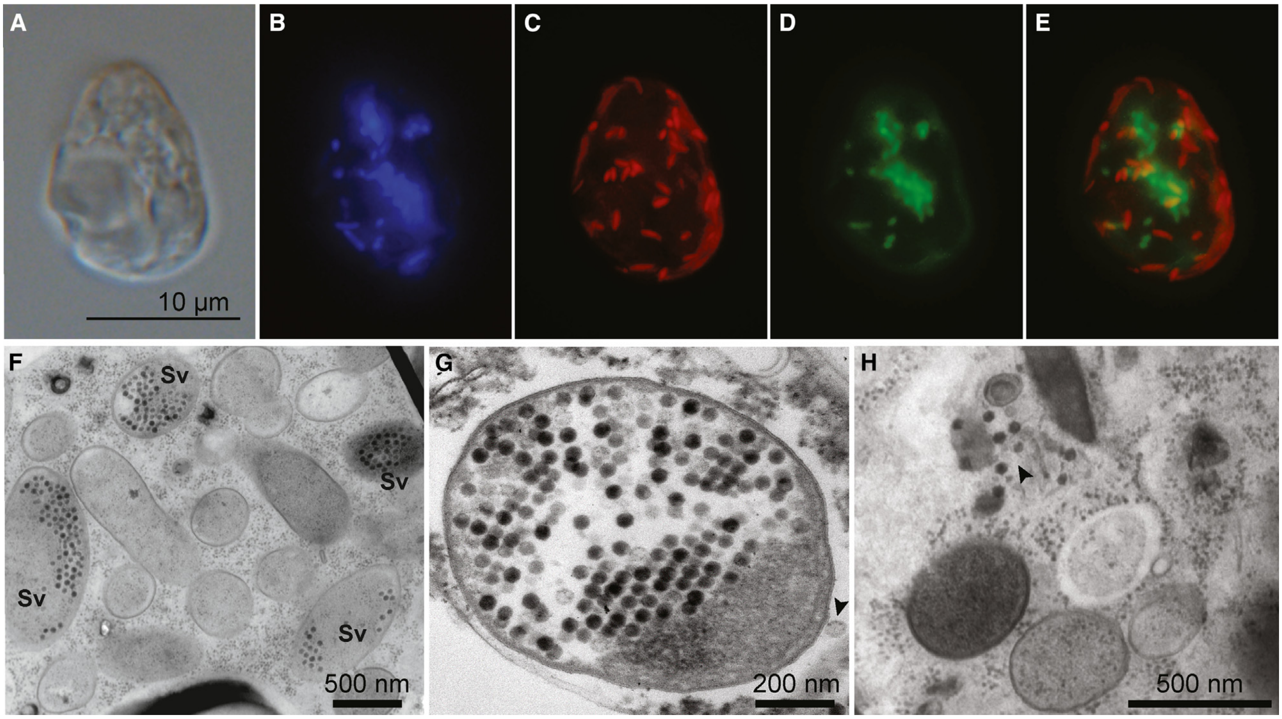

Figure 3. Microscopy of *Cryptomonas gyropyrenoidosa* SAG 25.80 with bacterial endosymbionts

(A) DIC; (B) DAPI; (C) FISH-*M. polyxenophila* probe; (D) FISH-*G. numerosa* probe; (E) overlay of (C) and (D); (F) endosymbionts clustered in the host cytoplasm, including endosymbionts with virus-like particles (Sv); (G) endosymbiont with virus-like particles within the bacterial cytoplasm and attached to the bacterial cell's surface (arrowhead); and (H) bacterial endosymbionts and a membrane-like structure (i.e., putative autolysosome vacuole) that potentially contains virus-like particles (arrowhead). See also Figures S4A and S4C and Table S2.

transposable elements and phage sequences that appeared to interfere with the assembly; these likely represent a prophage and transposons in subpopulations of *M. polyxenophila*. Scaffolds from an assembly with only Illumina reads also contained both *M. polyxenophila* and phage sequences (Table S1), providing additional evidence of the phage's integration into the *M. polyxenophila* genome. Owing to the presence of genes encoding ankyrin repeat (ANK) proteins (see below), we propose the name MAnkyphage for this *Megaira*-infecting phage. To check for sequence reads containing both MAnkyphage and *M. polyxenophila* regions, the Nanopore long reads were mapped to the MAnkyphage genome, and the mapped reads were then extracted and mapped to the *M. polyxenophila* genome. The MAnkyphage mapped reads aligned throughout the *M. polyxenophila* genome and generally aligned to transposon-rich regions (Data S2), confirming that the phage infects *M. polyxenophila*, but making the identification of any specific integration site of MAnkyphage into the *M. polyxenophila* genome extremely difficult.

TEM and FISH show differential morphology and relative abundance of endosymbionts along with evidence of a *Megaira polyxenophila* infecting bacteriophage

We confirmed the presence of *M. polyxenophila* and *G. numerosa* in *C. gyropyrenoidosa* host cells, using FISH and TEM. Both endosymbionts co-occurred in all host cells examined with FISH (Figures 3C–3E), and the abundance of *M. polyxenophila* was consistently higher than that of *G. numerosa* in all observed host cells. This is consistent with the higher read coverage of *M. polyxenophila* (740×) compared with *G. numerosa* (541×) that was also observed in genomic assemblies. In the TEM results, the majority of intracellular bacteria were located in the host cytoplasm (Figures 3F–3H), but a few bacterial cells were contained in putative peribacterial membranes (Figure S4A). Intracellular bacteria surrounded by peribacterial membranes in *C. gyropyrenoidosa* SAG 25.80 were previously observed[7] and likely correspond with *G. numerosa*, consistent with the fact that bacteria—identified here as *G. numerosa* (Figure S1)—enclosed by a membrane were reported from hydra cells.[27] However, we only observed bacteria within peribacterial membranes in the chemically fixed TEM samples and not samples preserved by high-pressure freezing (HPF), making the possibility that the membranes are fixation artifacts difficult to rule out.

We also confirmed the presence of VLPs with TEM, and VLPs were present in 17% ± 8% of endosymbionts in each *Cryptomonas* cell (Figures 3F–3H). VLPs were primarily observed within the bacterial cytoplasm, although they were occasionally attached to the outer membrane of the endosymbionts (Figure 3G) and in putative *Cryptomonas* autolysosome-like vacuoles (Figure 3H), as previously observed.[7] Additionally, VLPs were not present in bacteria within the putative peribacterial membranes (Figure S4A), consistent with the previous reports.[7,8] In *Cryptomonas* cells preserved with HPF, two bacterial subpopulations were distinguished based on different electron densities of their cytoplasm (Figure 3H), but whether this was due to species-specific properties or specific cellular processes (e.g., viral infection) remains unclear.

The large number of VLPs observed in the *Cryptomonas* TEM micrographs (Figures 3F–3H) and the high read coverage of the MAnkyphage genome (51,533×) suggested that this phage actively infects at least one of the *Cryptomonas* endosymbionts. The relationship between genome and capsid size is generally conserved in phages with similar capsid architecture due to the physical constraints of packaging DNA,[28] and a linear regression model using MAnkyphage and published phage data (Table S2) showed a significant correlation between genome size and capsid diameter ($p < 0.01$, $R^2 = 0.45$) (Figure S4C). Finally, the lack of VLPs in the bacteria surrounded by putative peribacterial membranes (Figure S4A), presumed to be *G. numerosa*, suggested that the cytoplasmic endosymbionts infected with VLPs are *M. polyxenophila*. This evidence, along with genes shared between the phage and *M. polyxenophila*, and the presence of related phages in publicly available metagenome-assembled genomes (MAGs) of *Megaira* spp. (see below) are consistent with *M. polyxenophila* as the MAnkyphage host, but the strongest evidence for this remains the sequence data for phage integration specifically in the *M. polyxenophila* genomic assemblies (Data S2; Table S1).

Endosymbiont metabolic pathways reveal a potential dependence of *M. polyxenophila* on *G. numerosa*

Both endosymbionts encode sets of genes relating to metabolism similar to what has been reported for other *G. numerosa* and *M. polyxenophila* strains.[29–31] This includes mostly intact cell membrane and wall biosynthesis but reduced carbon metabolism. For example, enzymes are present for TCA and pyruvate decarboxylation but not for glycolysis (Figure S5). Overall, *G. numerosa* has a broader biosynthetic capacity than *M. polyxenophila*, where *G. numerosa* encodes enzymes for gluconeogenesis and the biosynthesis of various co-factors (Figure S5). The endosymbionts rely on host metabolites and therefore harbor an array of transporters (Data S1B), including ATP/ADP translocases to import ATP or ADP from the eukaryotic host.

Interestingly, a possible metabolic dependency of *M. polyxenophila* on *G. numerosa* was predicted in the metabolic model analyses. *Grellia numerosa* encodes a complete queuosine biosynthesis pathway (Figure S5A) that produces the non-standard nucleoside queuosine found in certain tRNAs, but *M. polyxenophila* lacks genes for two essential enzymes, QueE and QueC (Figure S5B), catalyzing two consecutive internal steps of the pathway producing the queuosine precursor, 7-cyano-7-deazaguanine (also known as PreQ0). This is unexpected, since two different strains of *M. polyxenophila* sequenced from green algal hosts have an intact queuosine biosynthesis pathway.[30,31] To rule out that the missing queuosine biosynthesis genes were not due to a genome assembly artefact, the original metagenomic assembly from *C. gyropyrenoidosa* was checked, and no *Megaira*-like *queE* or *queC* sequences were present. Interestingly, *queE* and *queC* are encoded close to each other—separated by only a single gene—in the other two *M. polyxenophila* strains from green algae, and when these genomes were compared, a broader region spanning multiple genes, including *queE* and *queC*, is absent in *M. polyxenophila* from *C. gyropyrenoidosa*. The missing region coincides with a break in the co-linearity with the other *Megaira* genomes, so a genomic rearrangement event may have led to the loss of this region. Only bacteria have thus far been reported to possess the queuosine biosynthesis pathway,[32] so *C. gyropyrenoidosa* presumably cannot provide PreQ0 to the endosymbiont to compensate for the *queE* and *queC* loss. Our *C. gyropyrenoidosa* nuclear genome sequence is too fragmented to rule out the presence of queuosine biosynthesis genes in the alga, but a search of the transcriptome assemblies available for two other representatives of the genus *Cryptomonas*[33] did not reveal any candidates for *queE* and *queC* homologs. Hence, we propose that *M. polyxenophila* in *C. gyropyrenoidosa* acquires PreQ0 from the co-occurring *G. numerosa* endosymbiont, a notion supported also by the fact that both endosymbionts encode homologs of the YhhQ protein (Data S1B) known to mediate PreQ0 transport.[34] *Grellia numerosa* and another *Megaira* species co-infect a hydra host as well,[27] but it is currently unknown whether they also share this potential metabolic link since their genomes are unavailable.

Cryptomonas endosymbiont genomes encode a battery of proteins for putative host interactions

The *Cryptomonas* bacterial endosymbionts exhibited molecular machineries known to be involved in endosymbiosis, including type IV secretion systems (T4SSs; Data S1C). The T4SS in *Rickettsiales* is responsible for the secretion of proteins involved in eukaryotic host interactions,[35] but very few potential T4SS effectors were identified in the *Cryptomonas* endosymbionts. However, a large number of proteins with signal peptides were found in both endosymbionts (106 in *G. numerosa* and 130 in *M. polyxenophila*), and many of these proteins may be secreted into the host cytoplasm through general secretion mechanisms. Only *G. numerosa* encoded flagellar proteins (Data S1C), and despite having almost a complete set of flagellar genes, no visible flagella were present in TEM micrographs. The retention of flagellar genes is characteristic of *Midichloriaceae*,[29] and the loss of flagella is common in *Rickettsiaceae* where it has occurred independently multiple times.[36] Both endosymbionts also encode gene transfer agents (GTAs; Data S1D)—phage-like structures involved in horizontal gene transfer (HGT) in bacteria[37,38]—which are also commonly found in *Rickettsiales* genomes.[39–42]

Additional proteins involved in putative interactions with *Cryptomonas* were identified, including proteins with tetratricopeptide repeats (TPRs), ANKs, and leucine-rich repeats (LRRs) (Data S1E). ANKs are among the most common protein domains in eukaryotes, and a varying number of ANK-containing proteins (one up to >120) are present in different *Rickettsiales* bacteria.[43] Both *M. polyxenophila* and *G. numerosa* encoded several such proteins (seven and six, respectively; Data S1E). In some of the proteins, ANK domains were fused to outer membrane proteins, including an adhesion protein with a LPXTG motif in *M. polyxenophila* (gene *peg.48*) and two calcium-selective channel proteins in *G. numerosa* (*peg.773* and *peg.774*), suggesting direct protein-protein interactions between the bacteria and *Cryptomonas*. Additionally, an aminotransferase and a transcriptional regulator (MocR family) protein contained an ANK domain in *M. polyxenophila* (*peg.1028*).

LRR domains are also involved in eukaryotic protein interactions, and LRR-containing proteins are common in bacterial

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

endosymbionts.[36,44–46] Six LRR-containing proteins were present in *M. polyxenophila* (Data S1E), and although the majority of these proteins had unknown functions, one protein (*peg.67*) included a cupin-like domain 8 (cupin_8; Pfam: PF13621; InterPro: IPR041667). The LRR/cupin_8 protein is encoded on plasmid 1, upstream of an ANK protein and two acetyltransferase (GNAT) proteins. The LRR/cupin_8 protein shared sequence identity with KDM8 (JMJD5) proteins in eukaryotes (25%–29%; E-value 2E−9) and cupin-like domain-containing proteins in *Bacteroidota* (26%–33%; E-value 7E−7). The eukaryotic KDM8 proteins carry the JmjC domain (Pfam: PF08007; InterPro: IPR003347), which belongs to the cupin superfamily and also shared similarity with cupin_8. Proteins in the cupin superfamily have extremely diverse functions,[47] and many eukaryotic JmjC-containing proteins are involved in histone modification.[48] These proteins can also act as ribosomal oxygenases in both bacteria and eukaryotes.[49]

To determine if the LRR/cupin_8 protein in *M. polyxenophila* was derived from HGT from *Bacteroidota* or eukaryotes, we first compared the protein domain architecture of the BLAST results. The LRR domain found in *peg.67* was absent in all *Bacteroidota* cupin-like proteins but present in several eukaryotic KDM8 proteins. Therefore, the domain architecture of the *M. polyxenophila* protein was more similar to eukaryotes than to bacteria. We then inferred maximum likelihood (ML) trees of the full-length protein and cupin_8/JmjC domain alignments from the top bacterial and eukaryotic BLAST results, plus cupin-like proteins from *Rickettsiales* (Figure S6). In both trees, the *M. polyxenophila* cupin_8 protein fell in a highly supported clade with sequences from *Bacteroidota* and not with other *Rickettsiales* or eukaryotes. Although the placement suggested HGT from *Bacteroidota* to *M. polyxenophila*, the LRR/cupin_8 protein had a very long branch in both the full-length protein (Figure S6) and domain-only trees (data not shown), so both its evolution and function remain unclear.

In *G. numerosa*, a protein with a SET domain was identified (*peg.693*; Data S1E). SET domain proteins are involved in the modification of eukaryotic histones by bacterial pathogens including *Legionella pneumophila*,[50,51] *Chlamydia trachomatis*,[52] and *Bacillus anthracis*.[53] However, many free-living bacteria also encode SET-containing proteins with widespread functions, including lysine methyltransferases.[54] In pathogenic bacteria, SET domain proteins contain nuclear localization signals (NLSs) to target the host nucleus, but no NLSs were identified in either the SET-containing protein in *G. numerosa* or the LRR/cupin_8 protein in *M. polyxenophila*, suggesting that these proteins are not involved in histone modification. However, three putative NLSs were predicted in an ANK protein from *G. numerosa* (*peg.589*; Data S1E).

Megaira and *Grellia* endosymbionts have similar functional diversity

To examine the genomic and metabolic diversity of the endosymbionts with broad host ranges, we conducted a comparison of all available *Megaira* and *Grellia* genomes (Figure 4). Two additional genomes from each group are currently available[29,30,55]: *M. polyxenophila* from the green algae *Mesostigma viride* (1.5 Mbp; GenBank: GCA_020410825.1) and *Carteria cerasiformis* (1.3 Mbp; GenBank: GCA_913698045.1), *G. numerosa* from the ciliate *Euplotes* (1.1 Mbp; GenBank: GCA_021811875.1), and *G. incantans* from the placozoan *Trichoplax* (1.3 Mbp; GenBank: PRJEB30343). The genomes primarily varied in the number of pseudogenes and mobile elements (e.g., transposons), and *M. polyxenophila* in *Cryptomonas* had the greatest number of both (Figure 4A). Despite differences in genome size, pseudogene and mobile element abundance, and hosts, all *M. polyxenophila* strains had high average nucleotide identities (97%–99% ANI), which support a strain-level designation for the three endosymbionts.[56,57] As expected for *Grellia* spp., the ANI between the two *G. numerosa* strains was much higher (94%) than between each *G. numerosa* strain and *G. incantans* (81% ANI).

The functional diversity of the endosymbionts with broad host ranges was surprisingly similar (Figure 4B). An OrthoFinder analysis of the three *M. polyxenophila* strains and three *Grellia* spp. genomes revealed that 556 orthologous genes were shared between the six endosymbionts. A large number of orthologs (162) were also found in some but not all *M. polyxenophila* and *Grellia* spp. genomes. The *M. polyxenophila* strains shared an additional 364 orthologs with one another, whereas 209 orthologs were specific to *Grellia* spp. We classified the proteins into clusters of orthologous groups (COGs) (Figure S4B), using WebMGA and the National Center for Biotechnology Information' (NCBI's) COG database, and found that the endosymbionts had similar abundances of COGs in all functional categories except two: COGs involved in motility due to the presence of flagella in *Grellia* spp. and COGs involved in replication, recombination, and repair caused by the large number of transposon-related proteins in *M. polyxenophila* from *Cryptomonas*.

The *Megaira polyxenophila* infecting bacteriophage, MAnkyphage, encodes eukaryotic-like proteins

The *Megaira* phage, or MAnkyphage, encodes 51 genes including both core genes (e.g., structure, replication, and transcriptional regulators) and accessory genes (Figure 5; Data S1F). Some structural elements like baseplate components were not identified, but several genes encoding hypothetical proteins may have structural functions (Figure 5A). The phage harbors a patatin-like phospholipase (*peg.6*), a protein also found in the unrelated *Wolbachia* phage WO where it is proposed to have lytic activity.[58] Bacterial endosymbionts and pathogens also use patatin-like proteins to disrupt eukaryotic host membranes,[59,60] and we found that both *Cryptomonas* bacterial endosymbionts encode patatin-like phospholipases. The endosymbiont proteins shared low sequence similarity with MAnkyphage patatin-like phospholipase (<28%), likely due to different targets for lysis (e.g., eukaryotic vs. bacterial membranes). Additional eukaryotic-like proteins encoded by MAnkyphage included two ANK-carrying proteins (Figure 5A; Data S1F). ANK-containing proteins are commonly found in other phages of bacterial symbionts from eukaryotes (e.g., *Wolbachia* phage WO and sponge-symbiont phages), and their functions range from reproductive manipulation of eukaryotic hosts to evasion of eukaryotic host immune systems by bacterial symbionts.[12,13]

The functional role of many MAnkyphage proteins were unclear, including a CCDC90-like protein and several proteins with domains of unknown function (e.g., DUF1064 and DUF5681) (Figure 5A; Data S1F). MAnkyphage also encodes an incomplete DUF3685 protein domain (*peg.13*), which is a

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

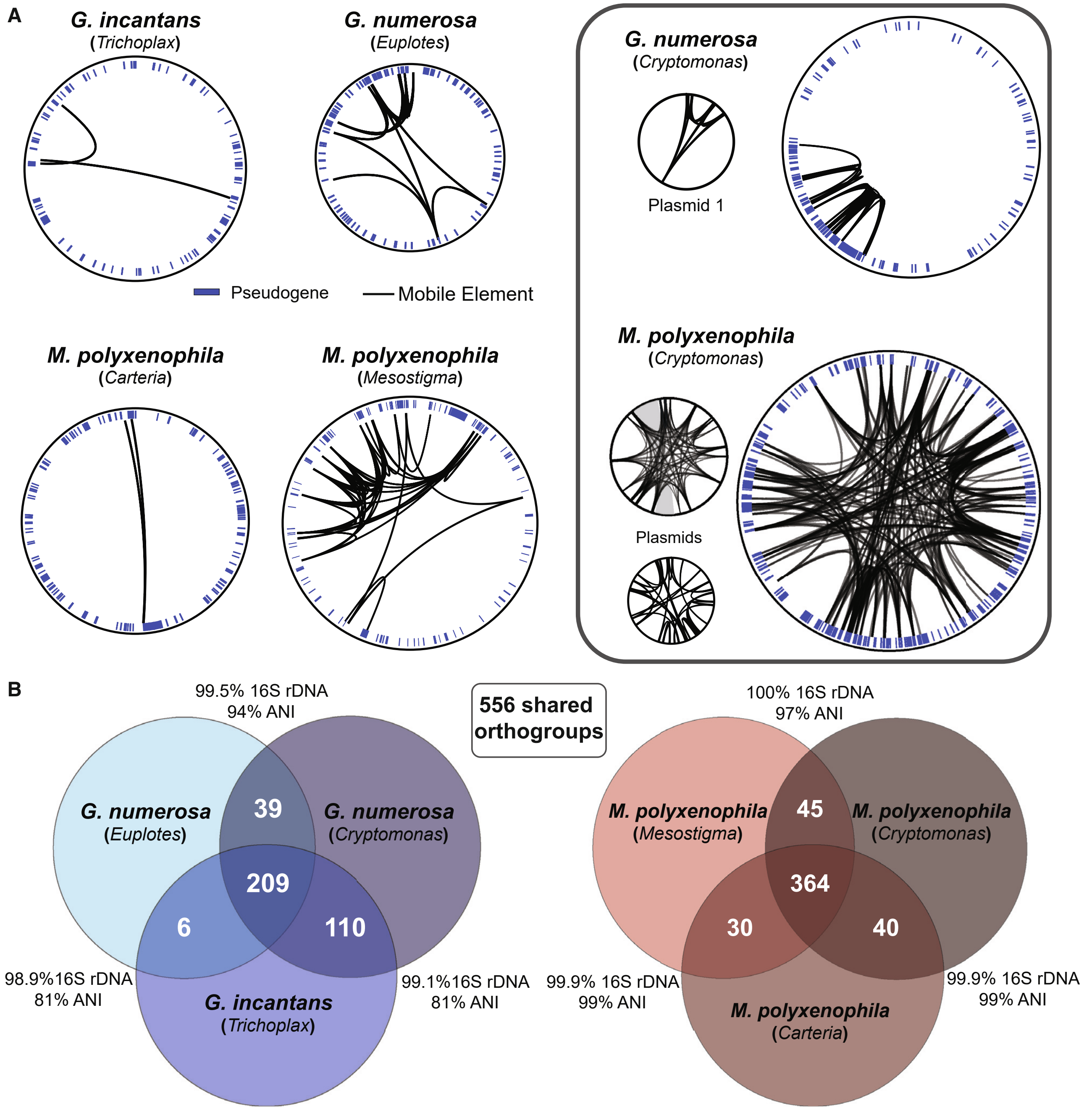

Figure 4. Genome comparisons of *Grellia* and *Megaira* species/strains

The available *Megaira* genomes include *M. polyxenophila* in *Mesostigma* (green alga), *Carteria* (green alga), and *Cryptomonas* (cryptomonad—this study). The available *Grellia* genomes include *G. incantans* from *Trichoplax* (placozoan) and *G. numerosa* from *Euplotes* (ciliate) and *Cryptomonas* (cryptomonad—this study).

(A) Genome plots with pseudogenes (blue rectangles) and mobile elements with homologous pairs connected by black lines.

(B) *M. polyxenophila* (*Rickettsiaceae*) and *Grellia* spp. (*Midichloriaceae*) share 556 orthologous genes. *Grellia* spp. share an additional 209 orthologs, and *M. polyxenophila* strains share an additional 364 orthologs. Average nucleotide identity (ANI) and 16S rRNA gene (16S rDNA) sequence similarities are shown between each species/strain. See also Figure S4B.

homolog of the cyanobacterial CheY-like effector and the plastid-encoded eukaryotic protein, Ycf55 (Figure 5D). These Ycf55/CheY-like response regulators harbor a receiver domain and a hypothetical effector domain, and although the receiver domain varies, the DUF3685 effector is highly conserved in cyanobacteria, plants, and various groups of algae including cryptomonads.[61] However, no DUF3685 genes were detected in the complete assembly of the *C. gyropyrenoidosa* plastid genome (Figure S3), and the gene appears to be expendable for plastid function in other cryptomonads.[22] A protein alignment

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

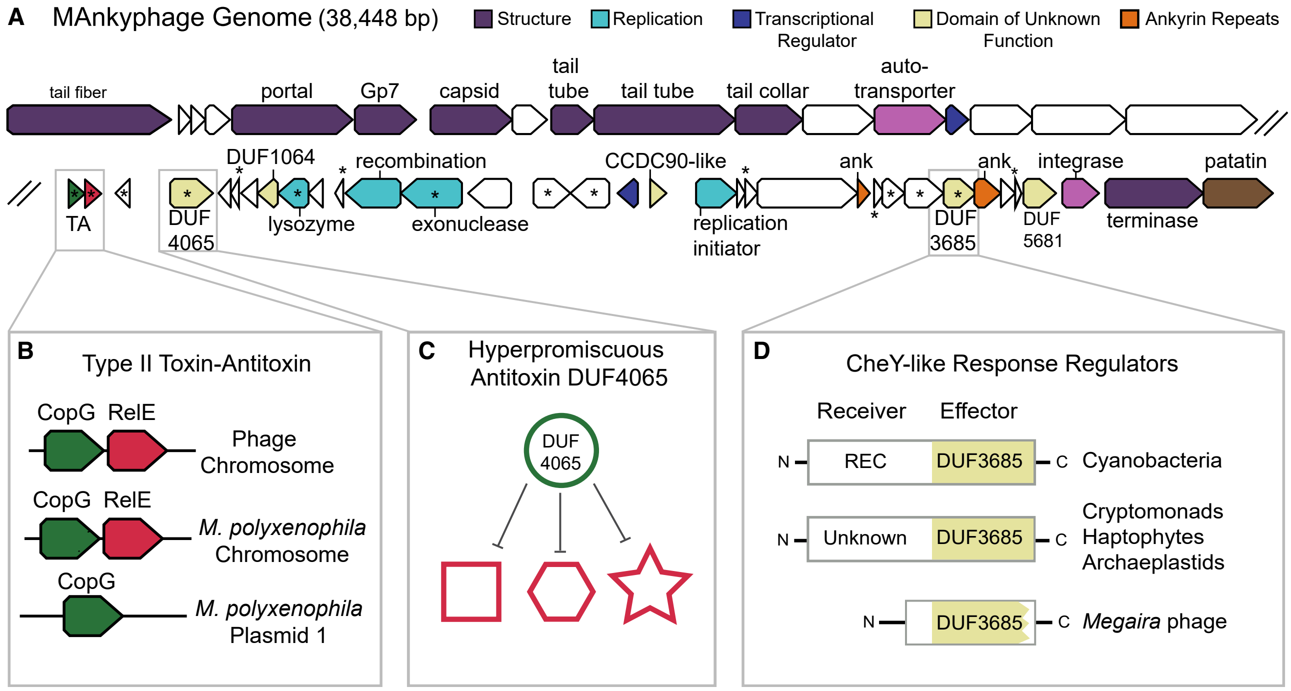

Figure 5. The *Megaira* phage (MAnkyphage) genome with core and accessory genes

(A) Core genes involved in structure, replication, and regulation of host transcription were identified using BLAST, HMMER, and HHpred searches. Putative accessory genes include proteins with domains of unknown functions (DUFs) and ankyrin repeat domains, along with toxin-antitoxin systems. Asterisks indicate genes absent in MAnkyphage from a *Megaira* metagenome-assembled genome sampled from a freshwater lake in Tanzania (GenBank: VGEX01000001.1).

(B) Type II toxin-antitoxin system found in MAnkyphage and *M. polyxenophila* genomes.

(C) Proteins with the hyperpromiscuous antitoxin domain, DUF4065, inhibit a diverse set of toxins.

(D) MAnkyphage protein with incomplete DUF3685. The effector domain, DUF3685, of the CheY-like response regulators is found in cyanobacteria and several photosynthetic eukaryotes, including cryptomonads. See also Figures S4D and S7 and Data S1 and S2.

of the DUF3685 domain from cyanobacteria, plants, green and red algae, haptophytes, and cryptomonads showed that the *Megaira* phage DUF3685 homolog falls within the cryptomonads, although the clade has low support (bootstrap value = 70%) (Figure S7) and the branch is long, so both its evolution and function remain uncertain.

Finally, MAnkyphage harbors the TA, RelE/CopG (*peg.38* and *peg.39*), which is also present in the *M. polyxenophila* chromosome (69% and 63% sequence similarity, respectively) and plasmid (only CopG—68% sequence similarity) (Figure 5B). Additionally, MAnkyphage encodes the hyperpromiscuous antitoxin protein domain DUF4065, which is known to neutralize diverse toxins,[62] hence the phage-encoded antitoxin likely regulates various TA systems in *M. polyxenophila* (Figure 5C). Putative viral defense mechanisms in *M. polyxenophila* and *G. numerosa* were also examined, and a type II restriction modification system was identified in *M. polyxenophila* (Data S1G), which uses endonucleases to cleave viral DNA.[63] Another potential viral defense protein, dGTPase (*peg.818*), was present in both *M. polyxenophila* and *G. numerosa*. In this defense system, dGTPases remove deoxynucleotides, specifically dGTP, during phage infection, which starves the phage of DNA components,[64] but dGTPases are common in *Rickettsiales*[36] and may have non-viral defense functions as well. TA systems can also serve as phage defense,[65] but none of the seven identified TAs in *M. polyxenophila* have known phage-defense functions.

The environmental distribution of MAnkyphage is widespread

To determine the distribution of MAnkyphage in the environment and its potential host range, publicly available sequence data, including metagenomes and MAGs, were searched for close homologs of MAnkyphage large terminase subunit (TerL), a hallmark phage gene commonly employed as a phylogenetic marker.[66] We identified 29 complete or partial sequences that formed a fully supported (bootstrap value = 100) tight clade sister to the TerL sequence derived from the genome of a bacterial endosymbiont belonging to *Holosporaceae* (*Finniella inopinata*). This clade included sequences with ≥70% amino acid identity to TerL (*peg.7*) from MAnkyphage (Figure 2). MAnkyphage-related TerL sequences were identified in freshwater metagenomes sampled around the world, along with a few brackish or marine metagenomes (Data S1H). One of the MAnkyphage-related TerL sequences came from a MAG (GenBank: VGEX01000001.1) from a freshwater lake in Tanzania, representing a close relative of *M. polyxenophila*,[31] which additionally included most of the other MAnkyphage genes (35 out of 51) (Figure 5; Data S1F), and the encoded proteins had relatively high sequence similarities (>80%) to that of MAnkyphage from *Cryptomonas*. The phage sequences from the Tanzanian lake *Megaira* MAG also showed synteny with the MAnkyphage genome from *Cryptomonas* (Figure S4D). Another MAnkyphage relative, albeit represented by a much less complete set of

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

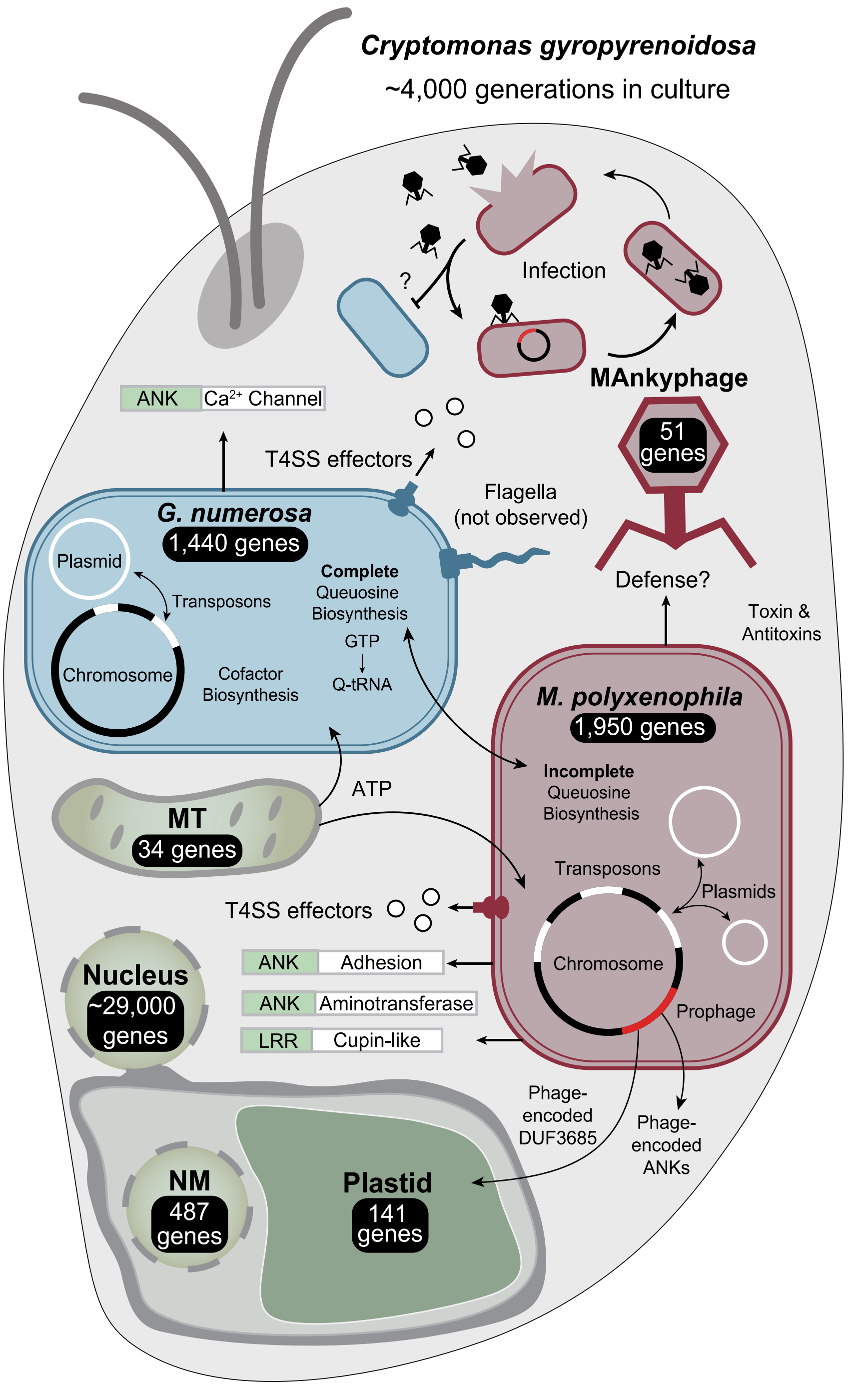

Figure 6. Putative interactions between the MAnkyphage, two bacterial endosymbionts (*M. polyxenophila* and *G. numerosa*), and the cryptomonad host

Cryptomonas gyropyrenoidosa harbors 4 genomes (in green): nuclear, mitochondrial (MT), nucleomorph (NM), and plastid. Mobile elements are found in the endosymbionts' chromosomes and plasmids, particularly in *M. polyxenophila*. Only *G. numerosa* encodes a flagellum, but both endosymbionts have type IV secretion systems (T4SS) and ATP/ADP translocases. The endosymbionts and phage also encode eukaryotic-like protein domains like ankyrin repeats (ANKs) and leucine-rich repeats (LRRs). Other putative eukaryotic interactions include the MAnkyphage-encoded protein (DUF3685) with homology to response regulators in plastids and cyanobacteria. MAnkyphage harbors a type II toxin-antitoxin system present in *M. polyxenophila*. The tally of genomes in this single-celled cryptomonad comes to 7 plus 3 bacterial plasmids. See also Figures S3 and S5–S7 and Data S1.

sequences, was encountered in a MAG (GenBank: JALRAC000000000.1) retrieved from an estuary in south China, and the MAG belonged to *M. polyxenophila* based on the analysis of RNA polymerase β' subunit sequences (Figure S2). Finally, a potentially complete phage genome somewhat more distantly related to MAnkyphage (GenBank: JALZUG010000044.1; 42,689 bp with sequence identity at both termini suggesting a circular-mapping molecule) was found in a MAG from a marine kelp-associated metagenome; the MAG corresponds to a *Megaira*-related bacterium, based on phylogenetic evidence (Figure S2). Overall, this suggests that phages of the MAnkyphage clade infect *Megaira* spp. from diverse freshwater, brackish, and marine environments.

DISCUSSION

Despite being a single-celled organism, *Cryptomonas gyropyrenoidosa* SAG 25.80 represents a remarkably complex consortium of genomes united through endosymbiosis. This alga is at one level a quadripartite system with phage (MAnkyphage), two bacterial endosymbionts (*Grellia numerosa* and *Megaira polyxenophila*), and the host eukaryote, but the host is itself an ancient conglomeration consisting of four genome-containing compartments built through several rounds of endosymbiotic associations with bacteria (mitochondria and plastids), a red alga (the nucleomorph), and the host nucleocytoplasm (Figure 6).

Grellia and *Megaira* species infect a wide range of eukaryotes, from green algae and ciliates to corals and placozoans,[29–31,55,67–70] and these endosymbionts also coexist in hydra.[27,71] In *Cryptomonas*, both *G. numerosa* and *M. polyxenophila* harbor a set of eukaryote-interacting proteins, including many with LRR and ANK domains (Figure 6), and this arsenal likely contributes to the success of the endosymbionts in invading and persisting in eukaryotic hosts.[51] Other eukaryote-interacting proteins are also encoded by MAnkyphage (ANKs and DUF3685 plastid response regulator), which provides *M. polyxenophila* additional mechanisms for host interactions (Figures 5 and 6). Similar sets of proteins were found in *M. polyxenophila* and *G. numerosa* from other hosts (Figure 4) and represent general mechanisms

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

for host-endosymbiont interactions that contribute to the success of these endosymbionts with broad host ranges.

Given the diverse eukaryotic host range of *M. polyxenophila*,[30,31,67] its phage, MAnkyphage, may provide interesting comparisons between bacteria and phage interactions in single-celled versus multicellular hosts (e.g., protists vs. metazoans). However, the bacterial host specificity is unknown for MAnkyphage from *Cryptomonas*, and additional studies will be needed to determine whether it can infect other *M. polyxenophila* strains, or even other *Megaira* species. The existence of a MAnkyphage-related clade of viruses (Figure 2) identified from metagenomes and MAGs suggests that these related phages infect *M. polyxenophila* and likely other *Megaira* species from diverse freshwater, brackish, and marine environments (Data S1H). The geographical distribution of this MAnkyphage group is also widespread and includes locations from Tanzania, China, and the USA (Data S1H).

The dynamics of MAnkyphage infection are also of interest, since the population dynamics of its host are so odd. Several factors differentiate phages of intracellular versus free-living bacteria, including the medium for infection (eukaryotic cytoplasm), encounter rates with bacterial hosts (which are likely lower in the environment vs. infected eukaryotic cells), and cell wall/membrane barriers (bacterial versus bacterial and eukaryotic). MAnkyphage appears successful at infecting *M. polyxenophila* in *Cryptomonas* host cells: VLPs were present in 17% ± 8% of endosymbiont cells (Figure 3), similar to the previously reported 16%–23%.[8] However, the proportion of *Megaira* cells infected with the MAnkyphage prophage remains unknown since the two endosymbionts are hardly distinguishable with TEM, and temperate phage infections go unobserved. The only putative viral defense mechanisms identified in the *M. polyxenophila* genome was a restriction modification system that uses endonucleases to cleave viral DNA strands and a dGTPase that may help deplete available DNA components. Some TA systems also have phage-defense functions along with cellular regulation properties, so one of the many TAs present in *M. polyxenophila* may also provide a phage-defense function (Figure 1). Interestingly, MAnkyphage encodes a TA system (RelE/CopG) also present in *M. polyxenophila* (Figure 5), which likely regulates transcription of the host's toxin, or alternatively, the TA regulates MAnkyphage production and immunity, a function of other phage-encoded TAs.[65]

The complexity of this system, along with the potential conflict between the bacteria, phage, and host, seems to suggest that the system has convoluted layers of selection and may even perhaps be unstable over evolutionary time. However, the culture has been maintained for over 50 years in stable conditions, and with the known 5-day generation time of *C. gyropyrenoidosa* SAG 25.80,[7,8] we can estimate that host strain has maintained the endosymbionts and phage for close to 4,000 generations. This represents a remarkable degree of stability compared with other protist-symbiont systems, where endosymbionts can be rapidly lost in culture.[1] The longevity of MAnkyphage and its *M. polyxenophila* host in culture also suggests that a balance is maintained between the lytic and lysogenic life cycles, since the infection and lysis of all viable endosymbionts would quickly lead to the extinction of MAnkyphage in *Cryptomonas* cells. Whether and how MAnkyphage may be transferred between *Megaira*-harboring *Cryptomonas* cells is another interesting question. The long-term retention of this intracellular community also raises questions about the selection pressures on the system, especially the differing selection pressures on the bacterial-bacterial, bacterial-eukaryotic, viral-bacterial, and viral-eukaryotic interactions. This phage-bacteria-protist system is convoluted, but its persistence in culture suggests all these pressures are somehow balanced.

Overall, we demonstrated the genomic and metabolic complexity of two bacterial endosymbionts and a phage in the single-celled alga, *Cryptomonas gyropyrenoidosa*. This is the first quadripartite system with a phage, bacterial endosymbionts, and eukaryotes described from algae at the molecular level, and the *C. gyropyrenoidosa* conglomeration has evolved through numerous symbiotic events, resulting in seven genomes within a single eukaryotic cell: the host nucleus, mitochondria, plastid and nucleomorph along with the endosymbionts, *Grellia numerosa* and *Megaira polyxenophila*, and a *Megaira*-infecting MAnkyphage. Our work adds additional layers of symbiotic complexity to the already intricate cryptomonad model of endosymbiosis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - ○ DNA isolation and sequencing
  - ○ Genome assemblies and annotation
  - ○ Fluorescence in situ hybridization (*FISH*)
  - ○ Transmission electron microscopy
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.04.010.

ACKNOWLEDGMENTS

We would like to thank the SAG curator, Maike Lorenz, for the history about the *Cryptomonas* SAG 25.80 culture. We also want to thank the University of British Columbia Culture Collection curator, Donna Dinh, and the UBC Bioimaging Facility (RRID: SCR_021304) for helping maintain and image the culture. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2014-03994) and from the Gordon and Betty Moore Foundation (https://doi.org/10.37807/GBMF9201) to P.J.K. Further support came from the Czech Science Foundation, specifically the projects 20-27648S to M.E. and 22-14356S and 22-01026S to J.L. E.E.G was also supported by the Simons Foundation Postdoctoral Fellowship in Marine Microbial Ecology (award ID: 993200), and D.B. was also supported by the FY2022 JSPS Postdoctoral Fellowship for Research in Japan (short-term).

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

AUTHOR CONTRIBUTIONS

Conceptualization, E.E.G., P.J.K., M.E., and D.B.; methodology, E.E.G., D.B., D.T., M.E., and P.J.K.; investigation, E.E.G., D.B., D.T., G.L., S.L., and F.H.; writing – original draft, E.E.G. and P.J.K.; writing – review & editing, E.E.G., D.B., M.E., P.J.K., D.T., G.L., S.L., and J.L.; funding acquisition, P.J.K., M.E., and J.L.; resources, P.J.K., M.E., and J.L.; supervision, P.J.K. and M.E.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: October 31, 2022
Revised: January 20, 2023
Accepted: April 6, 2023
Published: April 27, 2023

REFERENCES

1. Husnik, F., Tashyreva, D., Boscaro, V., George, E.E., Lukeš, J., and Keeling, P.J. (2021). Bacterial and archaeal symbioses with protists. Curr. Biol. *31*, R862–R877. https://doi.org/10.1016/J.CUB.2021.05.049.
2. Boscaro, V., Kolisko, M., Felletti, M., Vannini, C., Lynn, D.H., and Keeling, P.J. (2017). Parallel genome reduction in symbionts descended from closely related free-living bacteria. Nat. Ecol. Evol. *1*, 1160–1167. https://doi.org/10.1038/s41559-017-0237-0.
3. Boscaro, V., Syberg-Olsen, M.J., Irwin, N.A.T., George, E.E., Vannini, C., Husnik, F., and Keeling, P.J. (2022). All essential endosymbionts of the ciliate *Euplotes* are cyclically replaced. Curr. Biol. *32*, R826–R827. https://doi.org/10.1016/J.CUB.2022.06.052.
4. Husnik, F., and Keeling, P.J. (2019). The fate of obligate endosymbionts: reduction, integration, or extinction. Curr. Opin. Genet. Dev. *58–59*, 1–8. https://doi.org/10.1016/J.GDE.2019.07.014.
5. Curtis, B.A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M.C., Ball, S.G., Gile, G.H., Hirakawa, Y., et al. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature *492*, 59–65. https://doi.org/10.1038/nature11681.
6. Archibald, J.M. (2007). Nucleomorph genomes: structure, function, origin and evolution. BioEssays *29*, 392–402. https://doi.org/10.1002/BIES.20551.
7. Schnepf, E., and Melkonian, M. (1990). Bacteriophage-like particles in endocytic bacteria of *Cryptomonas* (Cryptophyceae). Phycologia *29*, 338–343. https://doi.org/10.2216/i0031-8884-29-3-338.1.
8. Schnepf, E., and Feith, R. (1992). Experimental studies to modify the number of endocytic bacteria in *Cryptomonas* strain SAG 2580 (Cryptophyceae) and their lysis by bacteriophages. Arch. Protistenkd. *142*, 95–100. https://doi.org/10.1016/S0003-9365(11)80074-6.
9. Yuksel, S.A., Thompson, K.D., Ellis, A.E., and Adams, A. (2001). Purification of *Piscirickettsia* salmonis and associated phage particles. Dis. Aquat. Organ. *44*, 231–235. https://doi.org/10.3354/dao044231.
10. Degnan, P.H., and Moran, N.A. (2008). Diverse phage-encoded toxins in a protective insect endosymbiont. Appl. Environ. Microbiol. *74*, 6782–6791. https://doi.org/10.1128/AEM.01285-08.
11. Friedman, C.S., and Crosson, L.M. (2012). Putative phage hyperparasite in the Rickettsial pathogen of abalone, "*Candidatus* Xenohaliotis californiensis". Microb. Ecol. *64*, 1064–1072. https://doi.org/10.1007/s00248-012-0080-4.
12. Bordenstein, S.R., and Bordenstein, S.R. (2016). Eukaryotic association module in phage WO genomes from *Wolbachia*. Nat. Commun. *7*, 13155. https://doi.org/10.1038/ncomms13155.
13. Jahn, M.T., Arkhipova, K., Markert, S.M., Stigloher, C., Lachnit, T., Pita, L., Kupczok, A., Ribes, M., Stengel, S.T., Rosenstiel, P., et al. (2019). A phage protein aids bacterial symbionts in eukaryote immune evasion. Cell Host Microbe *26*, 542–550.e5. https://doi.org/10.1016/J.CHOM.2019.08.019.
14. Corsaro, D., Walochnik, J., Venditti, D., Müller, K.D., and Michel, R. (2013). Molecular identification of a phage-infected *Protochlamydia* strain naturally harboured by non-encysting *Naegleria*. Acta Protozool. *52*, 273–281. https://doi.org/10.4467/16890027AP.13.024.1316.
15. Pramono, A.K., Kuwahara, H., Itoh, T., Toyoda, A., Yamada, A., and Hongoh, Y. (2017). Discovery and complete genome sequence of a bacteriophage from an obligate intracellular symbiont of a cellulolytic protist in the termite gut. Microbes Environ. *32*, 112–117. https://doi.org/10.1264/jsme2.ME16175.
16. Tikhe, C.V., and Husseneder, C. (2017). Metavirome sequencing of the termite gut reveals the presence of an unexplored bacteriophage community. Front. Microbiol. *8*, 2548. https://doi.org/10.3389/fmicb.2017.02548.
17. Preer, J.R., and Jurand, A. (1968). The relation between virus-like particles and R bodies of *Paramecium aurelia*. Genet. Res. *12*, 331–340. https://doi.org/10.1017/S0016672300011915.
18. Vannini, C., Boscaro, V., Ferrantini, F., Benken, K.A., Mironov, T.I., Schweikert, M., Görtz, H.D., Fokin, S.I., Sabaneyeva, E.V., and Petroni, G. (2014). Flagellar movement in two bacteria of the family *Rickettsiaceae*: A re-evaluation of motility in an evolutionary perspective. PLoS One *9*, e87718. https://doi.org/10.1371/journal.pone.0087718.
19. Hoef-Emden, K., and Melkonian, M. (2003). Revision of the genus *Cryptomonas* (Cryptophyceae): a combination of molecular phylogeny and morphology provides insights into a long-hidden dimorphism. Protist *154*, 371–409. https://doi.org/10.1078/143446103322454130.
20. Kim, J.I., Moore, C.E., Archibald, J.M., Bhattacharya, D., Yi, G., Yoon, H.S., and Shin, W. (2017). Evolutionary dynamics of cryptophyte plastid genomes. Genome Biol. Evol. *9*, 1859–1872. https://doi.org/10.1093/GBE/EVX123.
21. Kim, J.I., Yoon, H.S., Yi, G., Shin, W., and Archibald, J.M. (2018). Comparative mitochondrial genomics of cryptophyte algae: gene shuffling and dynamic mobile genetic elements. BMC Genomics *19*, 275. https://doi.org/10.1186/S12864-018-4626-9.
22. Tanifuji, G., Kamikawa, R., Moore, C.E., Mills, T., Onodera, N.T., Kashiyama, Y., Archibald, J.M., Inagaki, Y., and Hashimoto, T. (2020). Comparative plastid genomics of *Cryptomonas* species reveals fine-scale genomic responses to loss of photosynthesis. Genome Biol. Evol. *12*, 3926–3937. https://doi.org/10.1093/GBE/EVAA001.
23. Tanifuji, G., Onodera, N.T., Wheeler, T.J., Dlutek, M., Donaher, N., and Archibald, J.M. (2011). Complete nucleomorph genome sequence of the nonphotosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. Genome Biol. Evol. *3*, 44–54. https://doi.org/10.1093/GBE/EVQ082.
24. Kim, J.I., Tanifuji, G., Jeong, M., Shin, W., and Archibald, J.M. (2022). Gene loss, pseudogenization, and independent genome reduction in non-photosynthetic species of *Cryptomonas* (Cryptophyceae) revealed by comparative nucleomorph genomics. BMC Biol. *20*, 227. https://doi.org/10.1186/S12915-022-01429-6.
25. Cenci, U., Sibbald, S.J., Curtis, B.A., Kamikawa, R., Eme, L., Moog, D., Henrissat, B., Maréchal, E., Chabi, M., Djemiel, C., et al. (2018). Nuclear genome sequence of the plastid-lacking cryptomonad *Goniomonas avonlea* provides insights into the evolution of secondary plastids. BMC Biol. *16*, 137. https://doi.org/10.1186/s12915-018-0593-5.
26. Almpanis, A., Swain, M., Gatherer, D., and McEwan, N. (2018). Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. Microb. Genomics *4*, e000168. https://doi.org/10.1099/mgen.0.000168.
27. Fraune, S., and Bosch, T.C.G. (2007). Long-term maintenance of species-specific bacterial microbiota in the basal metazoan *Hydra*. Proc.

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

Natl. Acad. Sci. USA *104*, 13146–13151. https://doi.org/10.1073/PNAS.0703375104.

28. Lee, D.Y., Bartels, C., McNair, K., Edwards, R.A., Swairjo, M.A., and Luque, A. (2022). Predicting the capsid architecture of phages from metagenomic data. Comput. Struct. Biotechnol. J. *20*, 721–732. https://doi.org/10.1016/J.CSBJ.2021.12.032.
29. Giannotti, D., Boscaro, V., Husnik, F., Vannini, C., and Keeling, P.J. (2022). The "Other" *Rickettsiales*: an overview of the family "*Candidatus* Midichloriaceae.". Appl. Environ. Microbiol. *88*, e0243221. https://doi.org/10.1128/AEM.02432-21.
30. Davison, H.R., Pilgrim, J., Wybouw, N., Parker, J., Pirro, S., Hunter-Barnett, S., Campbell, P.M., Blow, F., Darby, A.C., Hurst, G.D.D., and Siozios, S. (2022). Genomic diversity across the *Rickettsia* and '*Candidatus* Megaira' genera and proposal of genus status for the Torix group. Nat. Commun. *13*, 2630. https://doi.org/10.1038/s41467-022-30385-6.
31. Davison, H.R., Hurst, G.D.D., and Siozios, S. (2023). '*Candidatus* Megaira' are diverse symbionts of algae and ciliates with the potential for defensive symbiosis. Microb. Genom. *9*, 000950. https://doi.org/10.1099/mgen.0.000950.
32. Zallot, R., Brochier-Armanet, C., Gaston, K.W., Forouhar, F., Limbach, P.A., Hunt, J.F., and De Crécy-Lagard, V. (2014). Plant, animal, and fungal micronutrient queuosine is salvaged by members of the DUF2419 protein family. ACS Chem. Biol. *9*, 1812–1825. https://doi.org/10.1021/cb500278k.
33. Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. *12*, e1001889. https://doi.org/10.1371/JOURNAL.PBIO.1001889.
34. Zallot, R., Yuan, Y., and De Crécy-Lagard, V. (2017). The *Escherichia coli* COG1738 member YhhQ is involved in 7-cyanodeazaguanine (preQ$_0$) transport. Biomolecules *7*, 12. https://doi.org/10.3390/BIOM7010012.
35. Driscoll, T.P., Verhoeve, V.I., Guillotte, M.L., Lehman, S.S., Rennoll, S.A., Beier-Sexton, M., Rahman, M.S., Azad, A.F., and Gillespie, J.J. (2017). Wholly *Rickettsia*! Reconstructed metabolic profile of the quintessential bacterial parasite of eukaryotic cells. mBio *8*, e00859–17. https://doi.org/10.1128/mBio.00859-17.
36. George, E.E., Husnik, F., Tashyreva, D., Prokopchuk, G., Horák, A., Kwong, W.K., Lukeš, J., and Keeling, P.J. (2020). Highly reduced genomes of protist endosymbionts show evolutionary convergence. Curr. Biol. *30*, 925–933.e3. https://doi.org/10.1016/j.cub.2019.12.070.
37. Lang, A.S., Zhaxybayeva, O., and Beatty, J.T. (2012). Gene transfer agents: phage-like elements of genetic exchange. Nat. Rev. Microbiol. *10*, 472–482. https://doi.org/10.1038/nrmicro2802.
38. Redfield, R.J., and Soucy, S.M. (2018). Evolution of bacterial gene transfer agents. Front. Microbiol. *9*, 2527. https://doi.org/10.3389/fmicb.2018.02527.
39. Lang, A.S., and Beatty, J.T. (2007). Importance of widespread gene transfer agent genes in α-proteobacteria. Trends Microbiol. *15*, 54–62. https://doi.org/10.1016/J.TIM.2006.12.001.
40. Shakya, M., Soucy, S.M., and Zhaxybayeva, O. (2017). Insights into origin and evolution of α-proteobacterial gene transfer agents. Virus Evol. *3*, vex036. https://doi.org/10.1093/VE/VEX036.
41. Christensen, S., and Serbus, L.R. (2020). Gene transfer agents in symbiotic microbes. In Symbiosis: Cellular, Molecular, Medical and Evolutionary Aspects, M. Kolc, ed (Springer), pp. 25–76. https://doi.org/10.1007/978-3-030-51849-3_2.
42. George, E.E., Tashyreva, D., Kwong, W.K., Okamoto, N., Horák, A., Husnik, F., Lukeš, J., and Keeling, P.J. (2022). Gene transfer agents in bacterial endosymbionts of microbial eukaryotes. Genome Biol. Evol. *14*, evc099. https://doi.org/10.1093/GBE/EVAC099.
43. Yurchenko, T., Ševčíková, T., Přibyl, P., El Karkouri, K., Klimeš, V., Amaral, R., Zbránková, V., Kim, E., Raoult, D., Santos, L.M.A., and Eliáš, M. (2018). A gene transfer event suggests a long-term partnership between eustigmatophyte algae and a novel lineage of endosymbiotic bacteria. ISME J. *12*, 2163–2175. https://doi.org/10.1038/s41396-018-0177-y.
44. Miyashita, H., Kuroki, Y., and Matsushima, N. (2014). Novel leucine rich repeat domains in proteins from unicellular eukaryotes and bacteria. Protein Pept. Lett. *21*, 292–305. https://doi.org/10.2174/09298665113206660112.
45. Zhou, J.M., and Chai, J. (2008). Plant pathogenic bacterial type III effectors subdue host responses. Curr. Opin. Microbiol. *11*, 179–185. https://doi.org/10.1016/J.MIB.2008.02.004.
46. Ishida, K., Sekizuka, T., Hayashida, K., Matsuo, J., Takeuchi, F., Kuroda, M., Nakamura, S., Yamazaki, T., Yoshida, M., Takahashi, K., et al. (2014). Amoebal endosymbiont *Neochlamydia* genome sequence illuminates the bacterial role in the defense of the host amoebae against *Legionella pneumophila*. PLoS One *9*, e95166. https://doi.org/10.1371/journal.pone.0095166.
47. Dunwell, J.M., Purvis, A., and Khuri, S. (2004). Cupins: the most functionally diverse protein superfamily? Phytochemistry *65*, 7–17. https://doi.org/10.1016/J.PHYTOCHEM.2003.08.016.
48. Tsukada, Y.I., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811–816. https://doi.org/10.1038/nature04433.
49. Chowdhury, R., Sekirnik, R., Brissett, N.C., Krojer, T., Ho, C.H., Ng, S.S., Clifton, I.J., Ge, W., Kershaw, N.J., Fox, G.C., et al. (2014). Ribosomal oxygenases are structurally conserved from prokaryotes to humans. Nature *510*, 422–426. https://doi.org/10.1038/NATURE13263.
50. Li, T., Lu, Q., Wang, G., Xu, H., Huang, H., Cai, T., Kan, B., Ge, J., and Shao, F. (2013). SET-domain bacterial effectors target heterochromatin protein 1 to activate host rDNA transcription. EMBO Rep. *14*, 733–740. https://doi.org/10.1038/EMBOR.2013.86.
51. Mondino, S., Schmidt, S., and Buchrieser, C. (2020). Molecular mimicry: a paradigm of host-microbe coevolution illustrated by *Legionella*. mBio *11*, e01201-20. https://doi.org/10.1128/mBio.01201-20.
52. Pennini, M.E., Perrinet, S., Dautry-Varsat, A., and Subtil, A. (2010). Histone methylation by NUE, a novel nuclear effector of the intracellular pathogen *Chlamydia trachomatis*. PLoS Pathog. *6*, e1000995. https://doi.org/10.1371/JOURNAL.PPAT.1000995.
53. Mujtaba, S., Winer, B.Y., Jaganathan, A., Patel, J., Sgobba, M., Schuch, R., Gupta, Y.K., Haider, S., Wang, R., and Fischetti, V.A. (2013). Anthrax SET protein: a potential virulence determinant that epigenetically represses NF-κB activation in infected macrophages. J. Biol. Chem. *288*, 23458–23472. https://doi.org/10.1074/jbc.M113.467696.
54. Alvarez-Venegas, R. (2014). Bacterial SET domain proteins and their role in eukaryotic chromatin modification. Front. Genet. *5*, 65. https://doi.org/10.3389/fgene.2014.00065.
55. Gruber-Vodicka, H.R., Leisch, N., Kleiner, M., Hinzke, T., Liebeke, M., McFall-Ngai, M., Hadfield, M.G., and Dubilier, N. (2019). Two intracellular and cell type-specific bacterial symbionts in the placozoan *Trichoplax* H2. Nat. Microbiol. *4*, 1465–1474. https://doi.org/10.1038/s41564-019-0475-9.
56. Konstantinidis, K.T., Ramette, A., and Tiedje, J.M. (2006). The bacterial species definition in the genomic era. Philos. Trans. R. Soc. Lond. B Biol. Sci. *361*, 1929–1940. https://doi.org/10.1098/RSTB.2006.1920.
57. Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. USA *106*, 19126–19131. https://doi.org/10.1073/PNAS.0906412106.
58. Bordenstein, S.R., and Bordenstein, S.R. (2022). Widespread phages of endosymbionts: phage WO genomics and the proposed taxonomic classification of Symbioviridae. PLoS Genet. *18*, e1010227. https://doi.org/10.1371/JOURNAL.PGEN.1010227.

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

59. Banerji, S., and Flieger, A. (2004). Patatin-like proteins: a new family of lipolytic enzymes present in bacteria? Microbiology (Reading) *150*, 522–525. https://doi.org/10.1099/MIC.0.26957-0.

60. Rahman, M.S., Ammerman, N.C., Sears, K.T., Ceraul, S.M., and Azad, A.F. (2010). Functional characterization of a phospholipase A2 homolog from *Rickettsia typhi*. J. Bacteriol. *192*, 3294–3303. https://doi.org/10.1128/JB.00155-10.

61. Hovde, B.T., Starkenburg, S.R., Hunsperger, H.M., Mercer, L.D., Deodato, C.R., Jha, R.K., Chertkov, O., Monnat, R.J., and Cattolico, R.A. (2014). The mitochondrial and chloroplast genomes of the haptophyte *Chrysochromulina tobin* contain unique repeat structures and gene profiles. BMC Genomics *15*, 604. https://doi.org/10.1186/1471-2164-15-604.

62. Kurata, T., Saha, C.K., Buttress, J.A., Mets, T., Brodiazhenko, T., Turnbull, K.J., Awoyomi, O.F., Oliveira, S.R.A., Jimmy, S., Ernits, K., et al. (2022). A hyperpromiscuous antitoxin protein domain for the neutralization of diverse toxin domains. Proc. Natl. Acad. Sci. USA *119*. e2102212119. https://doi.org/10.1073/PNAS.2102212119.

63. Stern, A., and Sorek, R. (2011). The phage-host arms race: shaping the evolution of microbes. BioEssays *33*, 43–51. https://doi.org/10.1002/bies.201000071.

64. Tal, N., Millman, A., Stokar-Avihail, A., Fedorenko, T., Leavitt, A., Melamed, S., Yirmiya, E., Avraham, C., Brandis, A., Mehlman, T., et al. (2022). Bacteria deplete deoxynucleotides to defend against bacteriophage infection. Nat. Microbiol. *7*, 1200–1209. https://doi.org/10.1038/s41564-022-01158-0.

65. Li, Y., Liu, X., Tang, K., Wang, W., Guo, Y., and Wang, X. (2020). Prophage encoding toxin/antitoxin system PfiT/PfiA inhibits Pf4 production in *Pseudomonas aeruginosa*. Microb. Biotechnol. *13*, 1132–1144. https://doi.org/10.1111/1751-7915.13570.

66. Benler, S., Yutin, N., Antipov, D., Rayko, M., Shmakov, S., Gussow, A.B., Pevzner, P., and Koonin, E.V. (2021). Thousands of previously unknown phages discovered in whole-community human gut metagenomes. Microbiome *9*, 78. https://doi.org/10.1186/S40168-021-01017-W.

67. Lanzoni, O., Sabaneyeva, E., Modeo, L., Castelli, M., Lebedeva, N., Verni, F., Schrallhammer, M., Potekhin, A., and Petroni, G. (2019). Diversity and environmental distribution of the cosmopolitan endosymbiont “*Candidatus* Megaira.”. Sci. Rep. *9*, 1179. https://doi.org/10.1038/s41598-018-37629-w.

68. Zaila, K.E., Doak, T.G., Ellerbrock, H., Tung, C.H., Martins, M.L., Kolbin, D., Yao, M.C., Cassidy-Hanley, D.M., Clark, T.G., and Chang, W.J. (2017). Diversity and universality of endosymbiotic *Rickettsia* in the fish parasite *Ichthyophthirius multifiliis*. Front. Microbiol. *8*, 189. https://doi.org/10.3389/fmicb.2017.00189.

69. Yang, A., Narechania, A., and Kim, E. (2016). Rickettsial endosymbiont in the “early-diverging” streptophyte green alga *Mesostigma viride*. J. Phycol. *52*, 219–229. https://doi.org/10.1111/jpy.12385.

70. Boscaro, V., Husnik, F., Vannini, C., and Keeling, P.J. (2019). Symbionts of the ciliate *Euplotes*: diversity, patterns and potential as models for bacteria–eukaryote endosymbioses. Proc. Biol. Sci. *286*, 20190693. https://doi.org/10.1098/rspb.2019.0693.

71. Castelli, M., Sassera, D., and Petroni, G. (2016). Biodiversity of “non-model” *Rickettsiales* and their association with aquatic organisms. In Rickettsiales: Biology, Molecular Biology, Epidemiology, and Vaccine Development (Springer), pp. 59–91. https://doi.org/10.1007/978-3-319-46859-4_3.

72. Dellaporta, S.L., Wood, J., and Hicks, J.B. (1983). A plant DNA minipreparation: Version II. Plant Mol. Biol. Rep. *1*, 19–21. https://doi.org/10.1007/BF02712670.

73. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477. https://doi.org/10.1089/cmb.2012.0021.

74. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: interrogation of genome assemblies. F1000Res *6*, 1287. https://doi.org/10.12688/f1000research.12232.1.

75. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

76. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

77. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. *13*, e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

78. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. *25*, 1043–1055. https://doi.org/10.1101/GR.186072.114.

79. Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., et al. (2018). KBase: the United States Department of Energy systems biology KnowledgeBase. Nat. Biotechnol. *36*, 566–569. https://doi.org/10.1038/nbt.4163.

80. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

81. Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J. Mol. Biol. *430*, 2237–2243. https://doi.org/10.1016/J.JMB.2017.12.007.

82. Gabler, F., Nam, S.Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A.N., and Alva, V. (2020). Protein sequence analysis using the MPI bioinformatics toolkit. Curr. Protoc. Bioinform. *72*, e108. https://doi.org/10.1002/CPBI.108.

83. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. Nucleic Acids Res. *46*, W200–W204. https://doi.org/10.1093/NAR/GKY448.

84. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: the protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/NAR/GKAA913.

85. Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. *45*, W6–W11. https://doi.org/10.1093/NAR/GKX391.

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

87. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

88. Kempen, M. van, Kim, S.S., Tumescheit, C., Mirdita, M., Gilchrist, C.L.M., Söding, J., and Steinegger, M. (2022). Foldseek: fast and accurate protein structure search. https://doi.org/10.1101/2022.02.07.479398.

89. Karp, P.D., Latendresse, M., Paley, S.M., Ong, M.K.Q., Billington, R., Kothari, A., Weaver, D., Lee, T., Subhraveti, P., Spaulding, A., et al. (2016). Pathway Tools version 19.0: integrated software for pathway/genome informatics and systems biology. Brief. Bioinform. *17*, 887–890.

90. Abby, S.S., Néron, B., Ménager, H., Touchon, M., and Rocha, E.P.C. (2014). MacSyFinder: a program to mine genomes for molecular systems

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

with an application to CRISPR-Cas systems. PLoS One *9*, e110726. https://doi.org/10.1371/journal.pone.0110726.

91. Käll, L., Krogh, A., and Sonnhammer, E.L. (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. *338*, 1027–1036. https://doi.org/10.1016/J.JMB.2004.03.016.
92. Syberg-Olsen, M.J., Garber, A.I., Keeling, P.J., McCutcheon, J.P., and Husnik, F. (2022). Pseudofinder: detection of pseudogenes in prokaryotic genomes. Mol. Biol. Evol. *39*, msac153. https://doi.org/10.1093/MOLBEV/MSAC153.
93. Wu, S., Zhu, Z., Fu, L., Niu, B., and Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics *12*, 444. https://doi.org/10.1186/1471-2164-12-444.
94. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.
95. Nair, R., Carter, P., and Rost, B. (2003). NLSdb: database of nuclear localization signals. Nucleic Acids Res. *31*, 397–399. https://doi.org/10.1093/NAR/GKG001.
96. Bernhofer, M., Goldberg, T., Wolf, S., Ahmed, M., Zaugg, J., Boden, M., and Rost, B. (2018). NLSdb—major update for database of nuclear localization signals and nuclear export signals. Nucleic Acids Res. *46*, D503–D508. https://doi.org/10.1093/NAR/GKX1021.
97. Nguyen Ba, A.N., Pogoutse, A., Provart, N., and Moses, A.M. (2009). NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. BMC Bioinformatics *10*, 202. https://doi.org/10.1186/1471-2105-10-202.
98. Hugenholtz, P., Tyson, G.W., Webb, R.I., Wagner, A.M., and Blackall, L.L. (2001). Investigation of candidate division TM7, a recently recognized major lineage of the domain bacteria, with no known pure-culture representatives. Appl. Environ. Microbiol. *67*, 411–419. https://doi.org/10.1128/AEM.67.1.411-419.2001.
99. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/NAR/GKH340.
100. Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics *30*, 3276–3278. https://doi.org/10.1093/BIOINFORMATICS/BTU531.
101. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.
102. Rodriguez-R, L.M., and Konstantinidis, K.T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ *4*, e1900v1. https://doi.org/10.7287/peerj.preprints.1900v1.
103. Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., et al. (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genomics *9*, 75. https://doi.org/10.1186/1471-2164-9-75.
104. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Cryptomonas gyropyrenoidosa SAG 25.80	Culture Collection of Algae (SAG)	SAG 25.80
Chemicals, peptides, and recombinant proteins		
ProLong Gold antifade reagent with DAPI	Life Technologies	Cat# P36931
Megenus_487: GCCGGGGCTTTTTCTGTTGGT	Lanzoni et al.[67]	N/A
BanNum_173: CCTCTCGGCAATATACAGTA	Boscaro et al.[70]	N/A
Spurr's resin	Electron Microscopy Sciences	Cat# 14300
4% Glutaraldehyde	Electron Microscopy Sciences	Cat# 16220
Osmium tetroxide (OsO_4)	Ted Pella	Cat# 18459
Critical commercial assays		
DNAeasy PowerBiofilm kit	QIAGEN	Cat# 24000
QIAamp DNA minikit	QIAGEN	Cat# 51304
Invisorb® Spin Plant Mini Kit	STRATEC Molecular GmbH	Cat#1037100300
DNA minipreparation protocol	Dellaporta et al.[72]	N/A
Illumina DNA prepkit	Illumina	Cat# 20018705
TruSeq Nano DNA Kit	Illumina	Cat#15041110 Rev. D
MinION 1D ligation library	Oxford Nanopore	Cat# SQK-LSK109
SMRTbell library construction kit	PacBio	Cat#100-938-900
Deposited data		
Megaira polyxenophila SAG 25.80_endo genome	This paper	SAMN30671969
Grellia numerosa SAG 25.80_endo genome	This paper	SAMN30671970
MAnkyphage_25.80 genome	This paper	GenBank OP381185
Cryptomonas gyropyrenoidosa SAG 25.80 mitochondrialgenome	This paper	GenBank OQ603491
Cryptomonas gyropyrenoidosa SAG 25.80 plastidgenome	This paper	GenBank OQ612619
Cryptomonas gyropyrenoidosa SAG 25.80 nucleomorphgenome (3 chromosomes)	This paper	GenBank OQ709067, OQ709068, OQ709069
Software and algorithms		
SPAdes v3.11.1	Bankevich et al.[73]	http://cab.spbu.ru/files/release3.11.1/manual.html
Unicycler v0.4.7	Wick et al.[77]	https://github.com/rrwick/Unicycler/
bowtie2 v2.4.2	Langmead et al.[75]	https://github.com/BenLangmead/bowtie2
minimap2 v2.18	Li et al.[76]	https://github.com/lh3/minimap2/releases
Blobtools v1.0.1	Laetsch and Blaxter[74]	https://zenodo.org/record/845347
PROKKA v1.12	Victorian Bioinformatics Consortium[80]	http://www.vicbioinformatics.com/software.prokka.shtml
RAST webserver	Aziz et al.[103]	rast.nmpdr.org
Mfannot	Université de Montréal	http://megasun.bch.umontreal.ca/apps/mfannot/
GeSeq tool	Tillich et al.[85]	http://chlorobox.mpimp-golm.mpg.de/geseq
AlphaFold2	DeepMind[86]	https://www.deepmind.com/open-source/alphafold
Foldseek Search		https://search.foldseek.com/search
BLAST	NCBI[104]	http://blast.ncbi.nlm.nih.gov/Blast.cgi/

(Continued on next page)

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CheckM v1.0.18	Parks et al.[78]	https://kbase.us/applist/apps/kb_Msuite/run_checkM_lineage_wf/release
HHpred	Max Planck Institute for Biology[81,82]	https://toolkit.tuebingen.mpg.de/tools/hhpred
HMMER	Potter et al.[83]	http://hmmer.org/
KEGG Automated Annotation Server		https://www.genome.jp/kegg/kaas/
Pathway Tools	BioCyc[89]	http://bioinformatics.ai.sri.com/ptools/
TXSScan	Abby et al.[90]	http://galaxy.pasteur.fr/
Phobius	Stockholm Bioinformatics Center[91]	http://phobius.sbc.su.se/
Pseudofinder	Syberg-Olsen et al.[92]	github.com/filip-husnik/pseudo-finder
OrthoFinder	Emms and Kelly[94]	https://github.com/davidemms/OrthoFinder
WebMGA	Wu et al.[93]	weizhong-lab.ucsd.edu/webMGA/server/
NLSdb	Nair et al.[95]	https://rostlab.org/services/nlsdb/
NLStradamus	Nguyen Ba et al.[97]	http://www.moseslab.csb.utoronto.ca/NLStradamus/
AliView	Uppsala University[100]	http://www.ormbunkar.se/aliview/
ANI tool	Environmental Microbial Genomics Laboratory	
IQ-TREE v1.5.4	Nguyen et al.[101]	http://www.iqtree.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Emma George, (3mma6eorg3@gmail.com)

Materials availability

This study did not generate any new unique reagents.

Data and code availability

- Sequence data and genomes have been deposited at NCBI and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cryptomonas sp. SAG 25.80 was obtained from the Culture Collection of Algae at the University of Göttingen, Germany (SAG) and cultured in Desmidiacean medium (Micrasterias+Erddekokt+VitaminB_{12}) SAG v11.2008 at 20°C in 12 hour light-dark cycles.

METHOD DETAILS

DNA isolation and sequencing

DNA was prepared using both QIAGEN Power Biofilm and QIAamp DNA mini kits, and the quality and quantity of each sample was recorded by NanoDrop and Qubit (Thermo Fisher Scientific) readings. DNA library preparations were performed with the Illumina DNA prep kit and sequenced using Illumina NexSeq with 2x150 bp paired-end reads (160,052,767 in total). In addition, Oxford Nanopore MinION 1D ligation library was also sequenced, generating 788,590 reads. Illumina DNA library preparation and sequencing services were performed at the University of British Columbia Sequencing Consortium (Vancouver, Canada).

DNA was also isolated by Invisorb® Spin Plant Mini Kit (STRATEC Molecular GmbH, Berlin, Germany), and subjected to library construction with TruSeq Nano DNA Kit (insert size 350 bp) and paired-end Illumina NovaSeq 6000 platform sequencing. Finally, genomic DNA isolated by the modified plant DNA minipreparation protocol[72] was subjected to SMRTbell library construction kit with 10 kb insert size and sequenced on the SMRTcells on PacBio RSII platform. DNA library preparation and sequencing services were performed at Macrogen, Inc. (Seoul, South Korea).

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

Genome assemblies and annotation

The sequencing reads were assembled with SPAdes v3.11.1,[73] after which contigs belonging to the host, organelles, bacterial endosymbionts, and phage were identified in BlobTools v1.0.1[74] using G+C content and coverage thresholds. Illumina and Nanopore reads were then mapped to the endosymbiont and phage contigs with bowtie2 v2.4.2[75] and minimap2 v2.18,[76] and all mapped reads were reassembled with Unicycler v0.4.7.[77] Bacterial genome completeness was determined by CheckM v1.0.18[78] on the KBase web server.[79] PROKKA v1.12[80] and the RAST web server (rast.nmpdr.org) were used for gene prediction and preliminary functional annotation of the bacterial and phage genomes, and protein-encoding genes were also annotated with HHpred[81,82] and HHMER.[83] Protein domains were investigated using the Pfam v31 database.[84] The genomes of *Grellia numerosa* (*Midichloriaceae*), *Megaira polyxenophila* (*Rickettsiaceae*) and MAnkyphage (phage of *Megaira polyxenophila*; *Caudoviricetes*) were sequenced from a culture of *Cryptomonas* sp. SAG 25.80. The mitochondrial, plastid and nucleomorph genomes of *Cryptomonas* sp. SAG 25.80 were also assembled and annotated with the online GeSeq tool (chlorobox.mpimp-golm.mpg.de/geseq)[85] and Mfannot (megasun.bch.umontreal.ca/apps/mfannot/). The initial Unicycler assembly of the nucleomorph genome was divided into four contigs, none of them reaching the telomere, but two of them could be joined by filling in a gap between them by recruiting a scaffold from the SPAdes assembly, resulting in three chromosomes as expected for a cryptomonad nucleomorph. The terminal regions of the choromosomes were assembled manually by considering overlaps between a series of shorter SPAdes scaffolds (including one with a putative telomeric repeat region uniquely consisting of the GAAAAAAA units intermingled with the TAAAAAAA units), the sequence coverage of the repeated subtelomeric regions relative to the coverage of the unique regions of the nucleomorph genome, and the known conserved organization of subtelomeric regions in cryptomonad nucleomorph chromosomes.

The identity of the MAnkyphage proteins was further tested by predicting their tertiary structure using AlphaFold2[86] as implemented at ColabFold server[87]; the default settings were used (i.e. prediction without any template information). The highest-ranking model as provided for each protein by AlphaFold was compared to a combined database of experimentally determined and bioinformatically predicted protein structures using the Foldseek Search server (https://search.foldseek.com/search).[88] Bacterial metabolic pathways were constructed in Pathway Tools[89] and the KEGG Automated Annotation Server (genome.jp/kegg/kaas). Secretion systems were annotated using TXSScan,[90] and signal peptides were identified with Phobius (http://phobius.sbc.su.se/).[91] Pseudogenes and mobile elements were identified with Pseudofinder,[92] and cluster of orthologous groups (COGs) and Pfam categories were classified with WebMGA[93] using the COG database from the National Center for Biotechnology Information (NCBI) [ncbi.nlm.nih.gov/research/cog]. Finally, orthologous gene comparison was conducted with OrthoFinder,[94] and nuclear localization signals were identified using the NLSdb (rostlab.org/services/nlsdb/)[95,96] and NLStradamus (moseslab.csb.utoronto.ca/NLStradamus/)[97] web servers.

Fluorescence in situ hybridization (*FISH*)

In order to confirm the presence of the two bioinformatically identified endosymbionts within *Cryptomonas* sp. SAG 25.80 cells, FISH experiments were conducted following the procedure described by Hugenholtz et al..[98] The oligonucleotide probe Megenus_487 (5′-GCCGGGGCTTTTTCTGTTGGT-3′)[67] labelled with a 5’-FAMfluorescent dye was used for the detection of *Megaira polyxenophila*. The oligonucleotide probe BanNum_173 (5’-CCTCTCGGCAATATACAGTA-3’)[70] labelled with a 5’-Cy3 fluorescent dye was applied for the detection of *Grellia numerosa*. Bothfluorescent oligonucleotides were synthesized by Eurofins Genomics (Ebersberg, Germany). Cell pellets were fixed for 30 min with 4% paraformaldehyde in phosphate-buffered saline (PBS), washed and air-dried on glass slides. To reduce the fluorescence of phycobiliproteins interfering with the fluorescence of the Cy3-labelled probe, slides were exposed to UV-radiation (Transilluminator 4000, Stratagene) for 2 min. Following dehydration in an increasing ethanol series (50%, 80% and 96% v/v), cells were hybridized in formamide-free buffer (900 mM NaCl, 20 mM Tris/HCl pH 7.2, 0.01% SDS) for 2 hours at 46°C. The probes were removed by incubating slides with the washing buffer (900 mM NaCl, 20 mM Tris/HCl, 0.01% SDS) for 30 min at 46°C. The slides were air-dried and mounted in ProLong® Gold Antifade Reagent with 4′,6-diamidino-2-phenylindole (or DAPI, Life Technologies), and observed with an AxioPlan 2 fluorescence microscope (Carl Zeiss Microscopy GmbH, Jena, Germany) under F36-670 (DAPI), Chroma F31-01 (5’-FAM) and F31-002 (Cy3) filter sets. Cells were photographed at several focal planes, and the resulting images were merged in GIMP v. 2.10.14 [gimp.org].

Transmission electron microscopy

Cryptomonas cells were examined with transmission electron microscopy (TEM) using two methods of preservation: chemical fixation and high-pressure freezing (HPF). For chemical fixation, 20 mL of *Cryptomonas* cell cultures were fixed in 2.5% (v/v) glutaraldehyde for 1 hour at room temperature, rinsed three times in fresh culture medium, post-fixed in 1% (w/v) OsO_4 at room temperature, and embedded in resin. Cells were dehydrated in an increasing ethanol series (30%, 50%, 70%, 90%, 95%, and three times in 100% v/v) for 10 min at each step. Infiltration with Spurr’s resin followed a graded ascending series in acetone (33%, 50%, 66%, and three time in 100% v/v) for a minimum of 3 hours at each step. Samples were polymerized at 60°C for 48 hours. For HPF, 5 mL of *Cryptomonas* cell culture was centrifuged at 2,000 rpm for two min, and 4.5 mL of supernatant was discarded. The remaining 500 μL pellet was gently resuspended by pipette and used in the HPF workflow. Cell culture (1.4 μL) was transferred to brass planchets and immediately frozen using a Leica HPM100, and six total planchets were frozen across two different rounds of HPF. The frozen planchets were transferred to cryovials containing 2% (w/v) OsO_4 and 0.1% (w/v) uranyl acetate (UA) in anhydrous acetone and processed in an automatic freeze-substitution system (Leica AFS2) according to the following schedule: 96 hours at -90°C, heated to -50°C over 12 hours, stayed at -50°C for 8 hours, heated to -20°C over 12 hours, stayed at -20°C for 8 hours, then heated to 22°C over 12 hours. Samples were rinsed 3 times in anhydrous acetone, and then infiltrated in an ascending graded series of Spurr’s resin in acetone

Please cite this article in press as: George et al., A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.010

(20%, 40%, 60%, 80%, then three infiltrations of 100% v/v) for a minimum of 3 hours at each step. Samples were polymerized at 60°C for 48 hours.

Sections (65 nm) were cut from Spurr's embedded blocks using a Leica UC7 Ultramicrotome and mounted on copper grids coated with 0.3% (w/v) formvar. Sections were post-stained with 2% (w/v) aqueous UA and lead citrate for 12 and 6 min, respectively. Imaging of sections was performed using a Tecnai Spirit TEMoperating at 80 kV with a DVC1500M side-mounted camera. Images were collected from three blocks, one from each round of HPF, and one from the chemically fixed samples. Both HPF-preserved (19 cells) and chemically fixed (40 cells) *Cryptomonas* were imaged.

QUANTIFICATION AND STATISTICAL ANALYSIS

Maximum likelihood (ML) trees of bacterial 16S rRNA genes and other bacterial and viral sequences (DNA and AA) aligned with MUSCLE[99] in AliView[100] were inferred using IQ-TREE v1.5.4,[101] and appropriate DNA or aminoacid substitution models were determined with model-testing in IQ-TREE. All sequences used in the alignments were selected from the top 100 BLAST hits to NCBI databases. Average nucleotide identity (ANI) was calculated with the Environmental Microbial Genomics Laboratory online ANI tool.[102]