Highly Reduced Genomes of Protist Endosymbionts Show Evolutionary Convergence

Highlights
- Unrelated bacterial symbionts from marine diplonemids show convergent evolution
- The symbionts have reduced genomes with similar content and metabolic potential
- The symbionts contain secretion systems including the type VI secretion system
- Diverse symbionts from a large range of eukaryotic hosts have similar modified cellular machinery

Authors
Emma E. George, Filip Husnik, Daria Tashyreva, ..., Waldan K. Kwong, Julius Lukeš, Patrick J. Keeling

Correspondence
EmmaEGeorge3@gmail.com

In Brief
Bacterial endosymbionts have evolved multiple times independently across the tree of life. George et al. provide an example of convergent evolution in the endosymbionts of marine protists and reveal the invisible interactions between these bacteria and their hosts.
Highly Reduced Genomes of Protoplast Endosymbionts Show Evolutionary Convergence

Emma E. George,1,4,6* Filip Husnik,1,4 Daria Tashyrev,2,4 Galina Prokopchuk,2,4 Aleš Horák,2,3 Waldan K. Kwong,1 Julius Lukes,2,3,8 and Patrick J. Keeling1,9

1University of British Columbia, Department of Botany, Vancouver, BC V6T 1Z4, Canada
2Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 370 05 České Budějovice, Czech Republic
3University of South Bohemia, Faculty of Science, 370 05 České Budějovice, Czech Republic
4These authors contributed equally
5Senior author
6Lead Contact
*Correspondence: 3mma6eorg3@gmail.com


SUMMARY

Genome evolution in bacterial endosymbionts is notoriously extreme: the combined effects of strong genetic drift and unique selective pressures result in highly reduced genomes with distinctive adaptations to hosts [1–4]. These processes are mostly known from animal endosymbionts, where nutritional endosymbioses represent the best-studied systems. However, eukaryotic microbes, or protists, also harbor diverse bacterial endosymbionts, but their genome reduction and functional relationships with their hosts are largely unexplored [5–7]. We sequenced the genomes of four bacterial endosymbionts from three species of diplonemids, poorly studied but abundant and diverse heterotrophic protists [8–12]. The endosymbionts come from two bacterial families, Rickettsiaceae and Holosporaceae, that have invaded two families of diplonemids, and their genomes have converged on an extremely small size (605–632 kilobase pairs [kbp]), similar gene content (e.g., metabolite transporters and secretion systems), and reduced metabolic potential (e.g., loss of energy metabolism). These characteristics are generally found in both families, but the diplonemid endosymbionts have evolved greater extremes in parallel. They possess modified type VI secretion systems that could function in manipulating host metabolism or other intracellular interactions. Finally, modified cellular machinery like the ATP synthase without oxidative phosphorylation, and the reduced flagellar apparatus present in some diplonemid endosymbionts and nutritional animal endosymbionts, indicates that intracellular mechanisms have converged in bacterial endosymbionts with various functions and from different eukaryotic hosts across the tree of life.

RESULTS

Diplonemid Endosymbionts Represent Phylogenetically Divergent Species

Three diplonemid species from previously studied cultures that were known to contain endosymbionts [11, 12] were screened to confirm the presence and identity of their full complement of endosymbionts by fluorescent in situ hybridization (FISH) and sequencing of the bacterial 16S rRNA gene (see STAR Methods). Diplonema aggregatum YPF1605 and YPF1606 and Diplonema japonicum YPF1603 and YPF1604 were confirmed to contain the Holosporaceae species, Cytomitobacter indipagum, and Cytomitobacter primus, respectively (bacterial taxa will be referred to without the Candidatus prefix) [11]. Interestingly, in D. japonicum a second and more abundant but previously undetected endosymbiont was found that was closely related to Cytomitobacter (Figure S1) but distinct enough to warrant a new genus (~87% 16S rRNA sequence identity to Cytomitobacter, well below the 94.5% gene sequence identity threshold for genera [13]) (Figure S2). We propose the new species, Nesclobacter abundans gen. nov., sp. nov. for this novel Holosporaceae endosymbiont (see full description below). Finally, the more distantly related diplonemid, Namystinia karyoxenos YPF1621, was confirmed to contain the Rickettsiaceae species Sneabacter namystus [12] (Figure S1).

The relative abundance and distribution of C. primus and N. abundans were compared in D. japonicum by using FISH probes (Figure S3) that distinguished the endosymbionts. Both endosymbionts were present in all examined host cells, where they were sporadically distributed beneath the host cell’s surface (Figure 1). The population size of both endosymbiont species and their ratio changed depending on the host life stage. The total abundance of endosymbionts was significantly higher and more variable in the larger trophic (feeding) hosts with 24–108 bacteria per host (mean ± SD [standard deviation of mean], 53.1 ± 17.8; n = 127), compared to swimming (starved) hosts, with 16–48 bacteria per host (32.7 ± 6.9; n = 120). The higher abundance of endosymbionts in the trophic cells was also previously reported [11]. The host cell size varied between life stages: host cell size was smaller during the swimming stage (12.7 ± 1.0 [n = 10] μm long and 4.6 ± 0.4 μm wide, mean ± SD [standard deviation of mean]) than the trophic stage...
Diplonemid Endosymbionts Have Small Genomes with Many Uncharacterized Proteins

The metagenome of each diplonemid species was sequenced, from which we retrieved complete endosymbiont genomes (Figure S3). These genomes were among the smallest recorded for protist endosymbionts. The *Holosporaceae* genomes ranged from 615,988 to 625,897 base pairs (bp) with 29.7%–30.0% G+C content and 505–550 protein-coding genes. The *S. namystus* genome was composed of two elements, a 605,311-bp chromosome and a 27,632-bp plasmid with 34.9% G+C content and 613 predicted protein-coding genes (Figure 2). All endosymbiont genomes were gene dense, with very few pseudogenes or mobile elements (Table S1). The genomes shared 223 orthologous genes, and the *Cytomitobacter* spp. and *N. abundans* genomes shared an additional 58 orthologs (Figure 2). Classifying proteins into cluster of orthologous groups (COGs), we found that the endosymbionts had highly similar relative abundances of most functional categories, with a few exceptions such as motility (Figure 2). Each genome contained numerous hypothetical proteins or proteins with unknown functions, altogether making up 22%–50% of the predicted protein-coding genes (Table S1) and included many putative secreted proteins, which could function in bacterial-host or inter-bacterial interactions.

Symbiont-Mediated Nutritional Provisioning Is Unlikely

Most highly reduced endosymbionts characterized to date function as nutritional mutualists (Table S2). To determine whether the diplonemid endosymbionts were providing their hosts with specific metabolites, the genomes were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. All diplonemid endosymbionts had severely reduced energy metabolism with no genes for glycolysis, tricarboxylic acid (TCA) cycle, or oxidative phosphorylation complexes I–IV (only a partial complex V [ATP synthase] was present) (Figure 3). The *Holosporaceae* diplonemid endosymbionts encoded a partial non-oxidative branch of the pentose phosphate pathway. The retained enzymes of these energy-generating pathways likely serve only for producing biosynthetic intermediates; *N. abundans* and *S. namystus* contained a pyruvate dehydrogenase complex that converts pyruvate to acetyl-CoA, and genes for other acetyl-CoA/pyruvate interconversion enzymes were present in all endosymbiont genomes (Figure 1).

The absence of glycolysis is well known in *Rickettsiaceae*, where the endosymbionts import metabolites from their host [14], but the loss of glycolysis was only recently discovered in certain *Holosporaceae* species [15]. Although most *Holosporaceae* genomes encoded reduced glycolytic pathways, a clade containing *Holospora*, *Hepatobacter*, and the four diplonemid endosymbionts had completely lost glycolysis along with the respiratory chain complexes III and IV (Figure 4). Oxidative phosphorylation was even further reduced in both *Holosporaceae* and *Rickettsiaceae* diplonemid endosymbionts as well.
as in Holospora, where the loss of the TCA cycle coincided with the absence of NADH dehydrogenase (complex I) and succinate dehydrogenase (complex II). Despite the absence of the respiratory chain, ATP synthase was retained in all diplonemid-endosymbionts (Figure 4), and this pattern has been observed in other highly reduced endosymbionts of various hosts [1, 16], where it could be used to hydrolyze ATP to generate a proton gradient [17, 18].

The most complete metabolic pathways in the diplonemid endosymbionts were for the biosynthesis of peptidoglycans, fatty acids, lipids, and iron-sulfur clusters (Figure 3). Other partial pathways in all endosymbionts included myo-inositol biosynthesis and thioredoxin recycling pathways as well as vitamin degradation and salvage pathways. However, there were no complete synthesis pathways for essential metabolites, such as amino acids or vitamins, which the endosymbionts could provide to their diplonemid hosts. Because of their extreme metabolic diminution, the endosymbionts likely depend on the import of many metabolites from their host.

Each species encoded between two to four ADP/ATP translocases (tlc), which enable the direct import of ATP from the host cytosol. Several other transporters were also present including amino acids and metabolite/drug transporters (Table S3). Considering their lack of pathways for energy production, and their localization near the host mitochondria [11], these endosymbionts might participate in “energy parasitism,” as reported for other members of the Holosporaceae and Rickettsiaceae [21].
Numerous Secretion Systems Might Mediate Intercellular Interactions

In contrast to their dramatic metabolic reduction, the diplonemid endosymbiont genomes encoded a large number of genes dedicated to protein secretion. The number of genes with predicted signal peptides (i.e., proteins targeted to membranes or the periplasm) ranged from 44 in N. abundans to 106 in C. indipagum (Table S1). A reduced general secretion system was present in all diplonemid endosymbionts (Table S3), and a small number of type II secretion system (T2SS) and type IV pilus (T4P) genes were found in Cytomitobacter and Nesciobacter genomes (Figure 3), suggesting a possible T2SS/T4P hybrid secretion system [22]. Sneabacter namystus retained a T1SS and additionally contained a reduced flagellar type III secretion system (T3SS) (Figures 3 and 4). The flagellar basal body was present, but the hook, filament, hook–filament junction, and cap were missing along with several T3SS proteins (Table S3). This reduction has been observed in the Buchnera endosymbionts of aphids [23–25], where flagellar basal bodies cover the surface [24] and have likely been repurposed for secretion of unknown effectors instead of flagellin [26, 27]. This is also the most plausible function in S. namystus, thus demonstrating evolutionary convergence of cellular machinery between unrelated endosymbionts of protists and animals.

Furthermore, all diplonemid endosymbionts possessed a modified type VI secretion system (T6SS). T6SSs are known for membrane puncturing and toxin delivery in bacterial competition and phagosome evasion [28, 29] and hence could play a role in interactions between intracellular endosymbionts. This system appeared to be ancestral to Cytomitobacter and Nesciobacter, but, in Sneabacter, a highly divergent T6SS was completely encoded on the plasmid. T6SS genes were absent in all other Rickettsiaceae genomes (Figure 4), although some T6SS genes were found in a metagenome-assembled Rickettsiales genome (Figure S4). Therefore, the T6SS was likely acquired horizontally in S. namystus, and the acquisition might be relatively ancient because of the similar GC content of the plasmid (33.8%) and chromosome (34.9%), along with the high sequence divergence between this T6SS and all other known T6SSs (Figure S4). Interestingly, all the diplonemid endosymbionts retained the majority of canonical T6SS components (Figure 3; Table S3), but the inner tube (TssD) and outer membrane complex (TssJ) were missing in all four species, as well as in related Holosporaceae species (Figure 3), suggesting these bacteria use a specialized, modified...
T6SS. A TssI (YwqK family) immunity protein was also encoded next to the VgrG tip protein in all diplonemid endosymbionts and the *S. namystus* plasmid contained two VapC toxins encoded next to two antitoxins.

Symbiont Genomes Are Relatively Stable and Retain DNA Repair Pathways

Genomic erosion can be accelerated by the loss of DNA repair and recombination mechanisms, which can lead to a runaway accumulation of deleterious mutations, i.e., Muller’s ratchet [30, 31]. In the diplonemid endosymbionts, the RecA-dependent RecFOR pathway was present for double-strand break repair, but the RecBCD recombination pathway was absent. Other DNA repair mechanisms were also intact: UvrABC and RuvABC were found in *S. namystus*, and MutSL and DNA polymerase I were identified in all endosymbionts, although several domains were missing in the DNA polymerase I genes. Additionally, the endosymbionts encoded the majority of cell-cycle and division genes (Figure 2).

Description of “*Candidatus Nesciobacter*”

“*Candidatus Nesciobacter,*** gen. nov., belonging to the family *Holosporaceae*.

Type Species

“*Candidatus Nesciobacter abundans***”
Diagnosis
Obligate endosymbiont of Diplonema japonicum YPF1604; reside freely in host cytoplasm; short rods (0.9 to 1.2 μm long and 0.5–0.7 μm wide); Gram-negative cell wall organization; flagella absent; granular homogeneous electron-dense cytoplasm; no visible inclusions or internal membrane structures.

Etymology
The genus name is derived from the Latin word nescio meaning “unknown” or “I do not know” and bacter referring to bacteria.

Description of “Candidatus Neschiobacter abundans”
“Candidatus Neschiobacter abundans,” sp. nov.

Type Strain
1604HC

Diagnosis
With characteristics of the genus. Genome GC content 29.78%. NCBI GenBank accession number: CP043314.1

Etymology
The species name is derived from the Latin word for abundant and describes the higher abundance of “Ca. Neschiobacter abundans” as compared to the other endosymbiont in the same host cell.

DISCUSSION

Many eukaryotes harbor prokaryotic endosymbionts, but the study of these—and particularly those with the most severely reduced genomes—has been historically biased toward those found in animal hosts, especially nutritional endosymbionts of insects [4, 32]. For example, of the approximately 210 endosymbionts with sequenced genomes under 1 Mb, 87% are engaged in nutritional mutualisms (Table S2). These systems provide important evolutionary contexts and a great deal of the theoretical basis for our understanding of the impacts of endosymbiosis. However, the scope of microbial symbioses at the lower limits of cellular and genomic complexity are likely much more diverse and more ambiguous than has been found thus far.

In protists, the Kinetoplastid endosymbionts (742–833 kb) of trypanosomatids aid in the synthesis of heme co-factors, amino acids, and B vitamins [33], retaining many metabolic genes. In contrast, the recently discovered endosymbiont of a Euplotes ciliate, Pinguicoccus supinus (at 163 kb, the smallest protist endosymbiont genome to date) lacks most metabolic pathways but interacts closely with host lipid droplets for unknown reasons [34]. Fokinia solitaria (837 kb) is another mysterious protist endosymbiont [35], which bears many similarities to the diplomend endosymbionts; these include the lack of many central metabolic pathways, the presence of ADP/ATP translocases, and a diverse arsenal of protein secretion systems [35]. These commonalities indicate that particular intracellular lifestyles can be converged upon from different evolutionary starting points, and we show that the diplomend endosymbionts have converged on very similar genome sizes and contents starting from two distinctly related bacterial lineages.

Interestingly, a large portion (13%–25%) of the reduced diplomend endosymbiont genomes were dedicated to protein secretion, including the presence of an extensive arsenal of effectors and putative toxins. However, the total number of secretion systems were diminished in comparison to relatives with larger genomes and tended to have smaller or modified gene complements (Figures 3 and 4). For example, two major T6SS components, the inner tube and outer membrane complex, were absent in the diplomend endosymbionts and other Holosporaceae endosymbionts, but also in S. namystus (Figures 3 and 4). These components are required for the extension and membrane-puncturing capability of the T6SS in bacterial competition and phagosome evasion [28], but those T6SS functions might not be necessary for intracellular endosymbionts that live outside host vesicles. Proteins with leucine-rich repeat (LRR) domains, known to be involved in interactions with eukaryotic proteins [36, 37], were present upstream and downstream of T6SS effectors in Cytomitobacter and Neschiobacter, and other LRRs were found throughout all diplomend endosymbiont genomes (Table S1); this suggests that secreted proteins could be used for host interaction.

The functions of secreted effectors and putative antimicrobial toxins are largely unknown in the Holosporaceae and the Rickettsiaceae. Their presence hints at a critical role of cellular interactions initiated by the endosymbionts, but the target of these manipulations remains unclear. Speculatively, secreted effectors might be used against the host to establish stable colonization or to mediate intracellular spatial positioning, but some could also target the mitochondria [11], or even other intracellular bacteria. In the latter case, this could be beneficial to the diplomend host, because it could protect it against infection by bacterial parasites or pathogens, as has been found for amoebae endosymbionts [5, 38, 39]. Other possible beneficial functions include osmotic stress resistance [40], heat tolerance [41], or even nutritional supplementation that is not yet obvious, given the large number of uncharacterized proteins encoded in their genomes (Figure 2). Yet, the endosymbionts could also be parasites that are no longer able to spread through horizontal transmission, leading to further reduction of their genomes, or could even be bacterial “free-loaders” that have little to no effect on their host. All these hypotheses will require further experimental tests to verify.

The co-occurrence of two different Holosporaceae endosymbionts in Diplonema japonicum (Figure 1) raises further questions about diplomend endosymbiont evolution. Many Diplonema species found thus far lack endosymbionts [11, 12], but the hosts might have lost symbionts due to initial antibiotic treatments used to establish the diplomend cultures. However, even with the limited sampling of symbionts now available, C. primus and N. abundans are not found to be sister species, altogether arguing against the conclusion that they speciated within D. japonicum. Multiple endosymbiont losses in the relatives of D. japonicum could have also occurred, and additional sampling could find that endosymbionts are more common in diponemids than previously shown. The presence of co-occurring diplomend endosymbionts also questions the function and interaction of the two Holosporaceae endosymbionts. In animal systems, such co-occurrence has been
observed to lead to partitioning of essential functions [1, 3], but this is most obviously applicable to nutritional supplementation, and there is no evidence for the partitioning of any pathway in C. primitus and N. abundans.

The extreme genome reduction in the diplonemid endosymbionts also complements our understanding of evolutionary mechanisms known from animal symbioses. Strong genetic drift, along with increased mutation rates and released selection pressure on non-essential genes in the intracellular environment, has led to the smallest bacterial genome found thus far [1, 32], and the reduced diplonemid endosymbionts are also a result of these evolutionary mechanisms. The diplonemid endosymbionts, along with other reduced symbionts, are susceptible to the evolutionary consequences of genome reduction and likely experience population bottlenecks because of factors like host cell division and starvation. The decreased endosymbiont load in starving diplonemid hosts (Figure 1) suggests that small endosymbiont populations occur in nature and that a sharp reduction in population size can lead to the inevitable fixation of deleterious mutations [30, 31]. The retained recombination and repair mechanisms in the Rickettsiaceae and Holosporaceae endosymbionts might correct deleterious mutations, thus slowing Muller’s ratchet [30]; however, pseudogenization and gene loss can still occur if mutations accumulate in asexual populations.

The question of why the diplonemid endosymbionts have converged on similar genome content also necessitates comparisons to well-studied animal symbioses. Host selection on endosymbionts of animals often leads to the retention of specific metabolic pathways [1, 3, 42], especially in nutritional mutualisms, but the extent of host selection on the diplonemid endosymbionts can only be speculated because of the unknown function of the endosymbionts. If the endosymbionts are mutualists, then the diplonemid host pressure could have selected for similar genome content in the Holosporaceae and Rickettsiaceae endosymbionts. On the other hand, the endosymbionts could be parasites, in which case intracellular pressures within the hosts could have also led to similar genome contents. Convergence likely due to intracellular selection can be observed in both insect and diplonemid endosymbionts, where similar reduced cellular systems such as the flagellar basal bodies [24] and ATP synthase without the respiratory chain [1] have been retained. Thus, genome content and metabolic potential of endosymbionts stem from host and intracellular selection that, in turn, depend on the type of symbiosis (e.g., mutualism, commensalism, or parasitism).

Overall, the genome reduction of Rickettsiaceae and Holosporaceae diplonemid endosymbionts show that extremely streamlined endosymbionts can be common associates of many single-celled eukaryotes. Although they possess many cellular characteristics common to other Rickettsiaceae and Holosporaceae endosymbionts, the diplonemid endosymbionts have all evolved to greater extremes in parallel. The commonalities between the diplonemid endosymbionts and unrelated endosymbionts of animals indicate that particular cellular machinery, such as the retention of modified secretion systems, is advantageous in a large range of hosts. Therefore, this study provides a model of convergent evolution in endosymbionts and reveals the invisible interactions between bacteria and one of the most abundant marine protists.

**STAR+METHODS**

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

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
  - DNA extraction, genome sequencing, and assembly
  - FISH and DNA staining
  - Detection of the second D. japonicum endosymbiont
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Endosymbiont abundance and morphological analyses
- **DATA AND CODE AVAILABILITY**

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2019.12.070.

**ACKNOWLEDGMENTS**

We thank Vittorio Boscaro for his advice and knowledge on endosymbionts of protists. We also thank Sergio Muñoz-Gómez and Bruce Curtis for sharing their multi-protein alignment and Z-trimming script. This work was supported by a grant from the Natural Sciences and Engineering Council of Canada (RGPIN-2014-03994). J.L. was supported from European Research Council CZ (LL1601) and the European Report on Development Funds, project OPVVV 16_019/0000769; E.E.G. was supported by the University of British Columbia International Fellowship, Canada; F.H. was supported by the European Molecular Biology Organization Long-Term Fellowship; W.K.K. was funded by the Killam Postdoctoral Research Fellowship, Canada; and A.H. and G.P. were supported by the Czech Grant Agency, project 18-23787S.

**AUTHOR CONTRIBUTIONS**


**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: July 24, 2019
Revised: September 24, 2019
Accepted: December 23, 2019
Published: January 23, 2020

**REFERENCES**


10. Luke


12. Prokopchuk, G., Tashyreva, D., Yabuki, A., Hora´ k, A., Masa


16. Luke

17. Tashyreva, D., Prokopchuk, G., Voty´ pka, J., Yabuki, A., Hora´ k, A., and Masa


**STAR METHODS**

**KEY RESOURCES TABLE**

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biological Samples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diplonema japonicum</td>
<td>This study</td>
<td>Strains YPF1603 and YPF1604</td>
</tr>
<tr>
<td>Diplonema aggregatum</td>
<td>This study</td>
<td>Strains YPF1605 and YPF1606</td>
</tr>
<tr>
<td>Namystynia karyoxenos</td>
<td>This study</td>
<td>Strain YPF1621</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProLong Gold antifade reagent with DAPI</td>
<td>Life Technologies</td>
<td>Cat# P36931</td>
</tr>
<tr>
<td>FISH Eub338 probe: GCTGCCTCCGTAAGGAGT</td>
<td>[49]</td>
<td>N/A</td>
</tr>
<tr>
<td>FISH HHC117 probe: CCCTCCATATGCGAATT CCC</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>FISH HLC36 probe: CAGCGTTGTTCTGAGCCGC</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>HLC_FISH_16-35_helper: CAGCGTTGTTCTGAGCCGC</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>HLC_FISH_55-79_helper: GAAAACATAACTCCGTTCGACTTG</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>HHC_FISH_93-116_helper: ATGTATTTACTACCCGTTGACCTG</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>HHC_FISH_138-160_helper: CCTGCGTTTTCAAAACTATCC</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNAeasy PowerBiofilm kit</td>
<td>QIAGEN</td>
<td>Cat# 24000-50</td>
</tr>
<tr>
<td>Nextera XT DNA Library Preparation Kit</td>
<td>Illumina</td>
<td>Cat# FC-131-1096</td>
</tr>
<tr>
<td>TruSeq RNA Library Prep Kit</td>
<td>Illumina</td>
<td>Cat# RS-122-2001</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytomitobacter primus genome</td>
<td>This paper</td>
<td>BioSample: SAMN12491242; GenBank: CP043316</td>
</tr>
<tr>
<td>Cytomitobacter indipagum genome</td>
<td>This paper</td>
<td>BioSample: SAMN12491243; GenBank: CP043315</td>
</tr>
<tr>
<td>Nesciobacter abundans genome</td>
<td>This paper</td>
<td>BioSample: SAMN12491244; GenBank: CP043314</td>
</tr>
<tr>
<td>Sneabacter namystus genome</td>
<td>This paper</td>
<td>BioSample: SAMN12491245; GenBank: CP043312; GenBank: CP043313</td>
</tr>
<tr>
<td><strong>Software and Algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BlobTools v1.0.1</td>
<td>[45]</td>
<td><a href="https://zenodo.org/record/845347">https://zenodo.org/record/845347</a></td>
</tr>
<tr>
<td>Unicycler v0.4.7</td>
<td>[46]</td>
<td><a href="https://github.com/rwick/Unicycler/">https://github.com/rwick/Unicycler/</a></td>
</tr>
<tr>
<td>RStudio v3.4.2</td>
<td>R</td>
<td><a href="https://www.rstudio.com/">https://www.rstudio.com/</a></td>
</tr>
<tr>
<td>IQ-TREE v1.5.4</td>
<td>[49]</td>
<td><a href="http://www.iqtree.org/">http://www.iqtree.org/</a></td>
</tr>
<tr>
<td>TXSScan</td>
<td>[52]</td>
<td><a href="http://galaxy.pasteur.fr/">http://galaxy.pasteur.fr/</a></td>
</tr>
<tr>
<td>SecReT6</td>
<td>[53]</td>
<td><a href="http://db-mml.sjtu.edu.cn/SecReT6/">http://db-mml.sjtu.edu.cn/SecReT6/</a></td>
</tr>
<tr>
<td>Pseudofinder</td>
<td>[55]</td>
<td>github.com/filip-husnik/pseudo-finder</td>
</tr>
</tbody>
</table>

**LEAD CONTACT AND MATERIALS AVAILABILITY**

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). This study did not generate new unique reagents.
EXPERIMENTAL MODEL AND SUBJECT DETAILS

Five different species/strains of diplonemids containing endosymbionts were analyzed: *Diplonema japonicum* YPF1603 and YPF1604, *D. aggregatum* YPF1605 and YPF1606, and *Namystynia karyoxenos* YPF1621. *Diplonema* spp. were grown axenically in seawater-based Hemi medium at 15°C [11]. *Namystynia karyoxenos* was grown axenically at 21°C in 12 hour light and dark cycles. In a rich nutrient medium supplemented with horse serum, *Diplonema* spp. existed in the form of a trophic stage with short flagella, which upon the depletion of nutrients (in old batch cultures or after transferring into a serum-free medium), transformed into a highly motile smaller swimming stage. All strains were from previous studies [11, 12] and the cultures were established in 2016.

METHOD DETAILS

DNA extraction, genome sequencing, and assembly

For DNA extraction, diplonemid cultures were grown to a maximum concentration of 1x10^5 cells/ml to 7x10^5 cells/ml. The genomes of four diplonemid endosymbionts sequenced included *Candidatus* Cytomitobacter primus, Ca. Cytomitobacter indigapag, Ca. Nesciobacter abundans (*Holosporaceae*), and Ca. Sneabacter namystus (*Rickettsiaceae*) that will be referred to without the *Candidatus* prefix. A QIAGEN Power Biofilm kit was used for DNA extractions 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 Nextera XT and TruSeq library kits (*N. abundans* and *C. primus*) and sequenced using Illumina MiSeq (*N. abundans* and *C. primus*) and Illumina HiSeq 2500 (all other species). The type of reads used for HiSeq was 2 x 125 bp and average coverage was 69 x (*S. namystus*), 164 x (*C. primus*), 310 x (*N. abundans*) and 3423 x (*C. indigapag*).

Each genome was assembled in SPAdes v3.11.1 [44] and host and containment contigs were removed from endosymbiont contigs in BlobTools v1.0.1 [45] using G+C content and coverage thresholds (Figure S3). In addition, Oxford Nanopore MinIon 1D ligation library was sequenced for *S. namystus* and *C. indigapag* and these genomes were closed using Unicycler v0.4.7 [46]. PROKKA v1.12 [47] and the RAST server [rast.nmpdr.org] were used for functional annotation. Orthologous gene comparison was conducted with EggNOG v4.5.1 [48] and protein family annotation was conducted with the Pfam v31 database using a Hidden Markov Model (HMM) search.

FISH and DNA staining

Fluorescence in situ hybridization (FISH) probes were used to determine endosymbiont abundance and spatial localization within *Diplonema japonicum*. Probes were designed on the basis of full-length 16S rRNA sequences and were designed to distinguish the endosymbionts from one another and from other sequences found in the metagenome: HHCl17 (5'-CCCTCCATATGGCAG ATTCCC-3') specific to *N. abundans* within the *D. japonicum* study system, and HLC36 (5'-CATGTGTTAAGGCCGCGGC-G3') specific to *C. primus* within the *D. japonicum* study system, were 5’-labeled with FITC and Cy3 fluorescent dyes, respectively. In addition, Eub338 probe targeting most groups of bacteria (5’-GCTGCCCTCCGGATGG-3’) 5’-labeled with Cy3, was used for total bacteria counts [43].

The specificity and efficiency of HHCl17 and HLC36 probes were confirmed in situ using mathfish.cee.wisc.edu tools and experimentally tested by hybridization in buffers containing 20%, 25%, 30% and 35% (v/v) formamide. Hybridization efficiency was improved by the addition of unlabeled helper oligonucleotides targeting the 16S rRNA regions adjacent to both 3’ and 5’ ends of HHCl17 and HLC36 probes. To determine if the probes matched other bacterial 16S rRNA in the *D. japonicum* culture system, 16S rRNA reads from the *D. japonicum* metagenome were analyzed using phylotrack v3.3 [56] and mapped against the endosymbionts’ 16S rRNA full length sequences using BBMap v37 [57]. No other full length 16S rRNA sequences were assembled in phylotrack and no other bacterial sequences matched the probe regions (Figure S3). The probes were also tested for other bacterial 16S rRNA matches using the Probe Match tool from The Ribosomal Database Project. The HHCl17 probe (*N. abundans*) showed zero matches while the HLC36 probe (*C. primus*) resulted in 9571 hits. Therefore, the probes were used solely to distinguish between the two endosymbionts within *D. japonicum* and future studies would need to design *N. abundans*-specific and *C. primus*-specific probes to identify the same endosymbionts in other study systems.

Both trophic and swimming cells were cultured in triplicates. Cell pellets were fixed with 4% paraformaldehyde in seawater for 30 min, rinsed with dH2O and air-dried on poly-L-lysine-coated glass slides. Adhered cells were dehydrated with 50%, 80% and 96% ethanol solutions for 3 min each. The slides were incubated simultaneously with 250 nM Eub338, HHCl17 and HLC36 probes in hybridization buffer (900 mM NaCl, 20 mM Tris/HCl, 0.01% SDS) containing 20% (v/v) formamide at 46°C for 30 min, rinsed with dH2O and air-dried on poly-L-lysine-coated glass slides. Adhered cells were dehydrated with 50%, 80% and 96% ethanol solutions for 3 min each. The slides were incubated simultaneously with 250 nM Eub338, HHCl17 and HLC36 probes in hybridization buffer (900 mM NaCl, 20 mM Tris/HCl, 0.01% SDS) containing 20% (v/v) formamide at 46°C for 30 min. The probes were removed by incubation in washing buffer (225 mM NaCl, 20 mM Tris/HCl, 0.01% SDS) on a shaker at 48°C for 30 min. FISH-labeled samples were air-dried and mounted in ProLong Gold antifade reagent (Life Technologies) containing 4’,6-diamidino-2-phenylindole (DAPI).

Detection of the second *D. japonicum* endosymbiont

In a previous study [11], 16S rRNA amplicon sequencing showed the presence of only one bacterial endosymbiont sequence in *Diplonema japonicum* YPF1603 and YPF1604. In the current study, the second endosymbiont *N. abundans* was detected by metagenomic sequencing and genome assembly. This study also found two mismatches of the 16S rRNA gene sequence in the
second endosymbiont, *N. abundans*, at the forward primer anneal site and within the reverse primer, likely the reason that the second endosymbiont was not detected by PCR in the previous study [11]. Additionally, the universal FISH probe (Eub338) along with the DAPI stain hybridized to both Cytomitobacter and Nesciobacter in the previous study due to no mismatching probes. Finally, TEM images failed to distinguish the two endosymbionts because all endosymbionts were nearly identical short rods with low variability in size and no notable differences in ultrastructure.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Maximum likelihood trees of bacterial 16S rRNA and other bacterial genes were made in IQ-TREE v1.5.4 [49] and metabolic pathways were constructed in Pathway Tools [50] and the KEGG Automated Annotation Server [genome.jp/kegg/kaas]. Secretion systems were identified with BLAST [51] and TXSScan [52], and type VI secretion systems were further analyzed using the SecReT6 database [53]. Signal peptides and transmembrane domains were found with the Phobius webserver [phobius.sbc.su.se] [54]. Pseudogenes were estimated using Pseudofinder [github.com/filip-husnik/pseudo-finder] [55]. Genome trait comparisons were conducted using BLAST and all Rickettsiales and Rhodospirillales genomes used in the comparisons were publicly available and downloaded from the National Center for Biotechnology Information (NCBI) database. Phage genes were identified by PROKKA and RAST annotations and phage domains were identified with the Pfam v31 database using an HMM search.

Both Cytomitobacter species had been previously described, and included *C. primus* found in *D. japonicum* YPF1604 and *C. indipagum* found in *D. aggregatum* YPF1606 [11]. Endosymbiont genomes from clonal *D. japonicum* YPF1603 and *D. aggregatum* YPF1605 strains were also assembled but the genomes had nearly identical sequences to *C. primus*, *N. abundans* and *C. indipagum* (Figure S2), and were not analyzed further.

**Endosymbiont abundance and morphological analyses**

FISH-labeled slides were viewed with the AxioPlan 2 fluorescence microscope (Zeiss, Germany) equipped with differential interference contrast (DIC) and Chroma F31-01 (FITC), F31-002 (Cy3) and F41-008 (Cy5) filters. Images were taken with DP72 digital camera at 1600 × 1200-pixel resolution using CellSens software v. 1.11 (Olympus) and processed with GIMP v. 2.8.14, IrfanView v. 4.41 and ImageJ v. 1.51 software. In each replicate, the number of *N. abundans* and *C. primus* endosymbionts were counted in 35 to 50 cells. The statistical analyses were conducted in R Studio v3.4.2 using simple t-test analyses in the R Stats Package. The fluorescence images were produced by overlaying 2 to 6 images focused on different planes of cells. After the FISH protocol with all 3 probes was completed, photos were taken at 100× magnification and the length and width of the endosymbiont cells were measured using ImageJ (50–60 host cells). TEM images from previously publish work [11] were also re-analyzed for ultrastructural characteristics (none were found).

**DATA AND CODE AVAILABILITY**

The endosymbiont genomes generated during this study are available in the NCBI GeneBank under the NCBI BioProject: PRJNA556273. The accession numbers for the sequences in this paper are GenBank: CP043316 (*Cytomitobacter primus*), GenBank: CP043315 (*Cytomitobacter indipagum*), GenBank: CP043314 (*Nesciobacter abundans*), GenBank: CP043312, and GenBank: CP043313 (*Sneabacter namystus*).