



Metabolic Contributions of an Alphaproteobacterial Endosymbiont in the Apicomplexan *Cardiosporidium cionae*

Elizabeth Sage Hunter¹, Christopher Paight² and Christopher E. Lane^{1*}

¹ Department of Biological Sciences, University of Rhode Island, Kingston, RI, United States, ² Department of Ecology, Evolution & Marine Biology, University of California, Santa Barbara, Santa Barbara, CA, United States

OPEN ACCESS

Edited by:

Alexei Yu. Kostygov,
University of Ostrava, Czechia

Reviewed by:

Anzhelika Butenko,
Academy of Sciences of the Czech
Republic (ASCR), Czechia
Javier del Campo,
University of Miami, United States

*Correspondence:

Christopher E. Lane
clane@uri.edu

Specialty section:

This article was submitted to
Microbial Symbioses,
a section of the journal
Frontiers in Microbiology

Received: 06 July 2020

Accepted: 05 November 2020

Published: 01 December 2020

Citation:

Hunter ES, Paight C and Lane CE
(2020) Metabolic Contributions of an
Alphaproteobacterial Endosymbiont
in the Apicomplexan *Cardiosporidium*
cionae. *Front. Microbiol.* 11:580719.
doi: 10.3389/fmicb.2020.580719

Apicomplexa is a diverse protistan phylum composed almost exclusively of metazoan-infecting parasites, including the causative agents of malaria, cryptosporidiosis, and toxoplasmosis. A single apicomplexan genus, *Nephromyces*, was described in 2010 as a mutualist partner to its tunicate host. Here we present genomic and transcriptomic data from the parasitic sister species to this mutualist, *Cardiosporidium cionae*, and its associated bacterial endosymbiont. *Cardiosporidium cionae* and *Nephromyces* both infect tunicate hosts, localize to similar organs within these hosts, and maintain bacterial endosymbionts. Though many other protists are known to harbor bacterial endosymbionts, these associations are completely unknown in Apicomplexa outside of the Nephromycidae clade. Our data indicate that a vertically transmitted α -proteobacteria has been retained in each lineage since *Nephromyces* and *Cardiosporidium* diverged. This α -proteobacterial endosymbiont has highly reduced metabolic capabilities, but contributes the essential amino acid lysine, and essential cofactor lipoic acid to *C. cionae*. This partnership likely reduces resource competition with the tunicate host. However, our data indicate that the contribution of the single α -proteobacterial endosymbiont in *C. cionae* is minimal compared to the three taxa of endosymbionts present in the *Nephromyces* system, and is a potential explanation for the virulence disparity between these lineages.

Keywords: apicomplexa, parasite – host interactions, mutualism, parasitism, bacterial endosymbiont, alphaproteobacteria host-associated bacteria

INTRODUCTION

Apicomplexa includes a multitude of highly virulent pathogenic organisms, such as *Plasmodium falciparum*, *Cryptosporidium parvum*, and *Toxoplasma gondii*, the causative agents of malaria, cryptosporidiosis, and toxoplasmosis, respectively. Malaria claims about half a million human lives annually (Center for Disease Control, 2019), *T. gondii* is estimated to infect up to 60% of the human population in much of Europe (Pappas et al., 2009), and cryptosporidiosis causes 3–5 million cases of gastrointestinal disease annually in children in Africa and India alone (Sow et al., 2016).

These organisms represent major human health concerns, but as a result, our understanding of this phylum is largely based on a small subset of clinically relevant apicomplexans. Every metazoan likely plays host to at least one apicomplexan (Morrison, 2009), and this is probably an underestimation, as many species can host multiple apicomplexan species. Apicomplexans have been described in a vast array of vertebrates from avians to marine mammals (Jeurissen et al., 1996; Conrad et al., 2005), and also in cnidarians (Kwong et al., 2019), molluscs (Dyson et al., 1993; Suja et al., 2016), arthropods (Alarcón et al., 2017; Criado-Fornelio et al., 2017), and urochordates (Ciancio et al., 2008; Saffo et al., 2010). Their host range is enormous, and their diversity and adaptation to the parasitic lifestyle is unparalleled.

The long history of evolution and adaptation to life within a host has given rise to a series of characteristic genomic losses and the evolution of specialized cellular machinery in apicomplexans (McFadden and Waller, 1997; Soldati et al., 2001; Roos, 2005; Janouskovec and Keeling, 2016; Fréchal et al., 2017). Specific structural adaptations of these organisms include those for functions related to host infection and persistence; namely a remnant plastid (apicoplast) and apical complex (McFadden and Waller, 1997; Soldati et al., 2001). Genomic reductions associated with parasitism in apicomplexans include losses in gene families for the biosynthesis of purines, amino acids, sterols, various cofactors, the glyoxylate cycle, endomembrane components, and genes related to motility (Woo et al., 2015; Janouskovec and Keeling, 2016). Additionally, apicomplexans also show expansions in gene families related to infection and persistence within host cells (Janouskovec and Keeling, 2016). However, the assumption that these genomic signatures are associated with parasitism is based on limited information, since a direct comparison to closely related free-living sister taxa is not possible, and there are no known free-living apicomplexans (Janouskovec and Keeling, 2016). However, genomic data is available from the photosynthetic Chromerids (Woo et al., 2015), which likely diverged from apicomplexans 600–800 million years ago (Votýpka et al., 2016).

Despite the high pathogenicity and parasitic adaptations of many members, questions have emerged over whether Apicomplexa is an entirely parasitic group. Though this sentiment has long been mentioned in publications (Roos, 2005; Morrison, 2009; Gubbels and Duraisingh, 2012; Woo et al., 2015; Votýpka et al., 2016; McFadden and Yeh, 2017; Mathur et al., 2018), the current evidence suggests that the interactions between apicomplexans and their hosts are far more varied than previously recognized. In fact, it is likely that apicomplexans span the full spectrum from parasitism to commensalism, and even mutualism (Saffo et al., 2010; Kwong et al., 2019; Rueckert et al., 2019). However, what defines the boundaries along this continuum of symbiotic association is still a topic of much debate (Ewald, 1987; Leung and Poulin, 2008; Johnson and Oelmüller, 2009). Phylogenetic analysis indicates *Nephromyces* is sister to the hematozoan clade, and closely related to highly virulent genera such as *Plasmodium*, *Theilera*, and *Babesia* (Muñoz-Gómez et al., 2019a). Thus far, apicomplexan species with variable life strategies have been found in early branching groups, such as the Gregarina and Coralicods. However, the existence of this

reportedly mutualistic taxon deep within Apicomplexa, sister to a group of highly virulent blood parasites, suggests the unique biology of Nephromycidae might be responsible for such a shift to a commensal or mutualistic life strategy.

Cardiosporidium cionae was originally described in 1907 by Van Gaver and Stephan, who correctly identified it as a novel sporozoan parasite of the invasive tunicate *Ciona intestinalis*. This species wasn't mentioned again until it was observed by Scippa, Ciancio, and de Vincentiis in 2000, and then formally redescribed by Ciancio et al. (2008), a full century after its initial discovery. Similar to other hemosporidians such as *Plasmodium*, *C. cionae* is found in the blood of its host. It localizes to the heart and pericardial body, a collection of sloughed off cells that accumulates over the life of the tunicate inside the pericardium (Evans Anderson and Christiaen, 2016). *Ciona intestinalis* is highly invasive; this prolific species has spread globally traveling in the hulls and bilgewater of ships and is now found on every continent except Antarctica. While *C. cionae* infection has only been formally confirmed in The Gulf of Naples, Italy (Ciancio et al., 2008), and Narragansett Bay, Rhode Island, United States, it likely has a broad range as well. Additionally, TEM data from the redescription of *C. cionae* revealed a bacterial endosymbiont (Ciancio et al., 2008).

The closest relative of *C. cionae*, *Nephromyces*, was first described around the same time in 1888, though its unusual filamentous morphology caused it to be misclassified as a chytrid fungus until 2010 (Saffo et al., 2010). *Nephromyces* is found in the Molgulidae family of tunicates, in a ductless structure of unknown function adjacent to the heart, known as the renal sac. It is thought to be mutualistic due to a near 100% infection prevalence (Saffo et al., 2010) and is capable of utilizing the waste products that the host tunicate sequesters in the renal sac as a source of glycine, pyruvate, and malate (Paight et al., 2019). *Nephromyces* also houses three separate lineages of bacterial endosymbionts (Paight et al., 2020). Though endosymbiotic associations are commonly found in other protists such as ciliates, diatoms, and amoebas, bacterial endosymbiosis in Apicomplexa is unique to this lineage (Nowack and Melkonian, 2010), which only includes *Cardiosporidium* and *Nephromyces* (Muñoz-Gómez et al., 2019a).

Endosymbiotic bacteria allow eukaryotes to exploit an enormous range of environments they would otherwise be unable to inhabit. Endosymbionts span a wide variety of taxa, from the *Buchnera* endosymbionts of aphids, which provide essential vitamins and amino acids, to the chemotrophic bacteria at the base of the deep-sea hydrothermal vent food chain. The diversity of prokaryotic metabolic pathways (McCutcheon et al., 2019) drives the propensity of bacteria to colonize and exploit unusual habitats, including such extreme environments as radioactive waste (Fredrickson et al., 2004), highly acidic hot springs (Marciano-Cabral, 1988), or even the inside of a host. In multicellular hosts, bacterial endosymbionts are frequently sequestered to specific structures or tissues, but in protists they must reside directly in the cytoplasm, making these associations far more intimate (Nowack and Melkonian, 2010).

Though these interactions appear beneficial, endosymbiosis is rooted in conflict (Keeling and McCutcheon, 2017;

McCutcheon et al., 2019). Many of the common endosymbiotic taxa, such as those within the order *Rickettsiales*, are closely related to pathogens. *Rickettsiales* is likely the sister taxon to the modern eukaryotic mitochondria (Fitzpatrick et al., 2006), and also contains *Wolbachia*, a genus of arthropod and nematode endosymbionts known to infect 25–70% of insects (Kozek and Rao, 2007). Endosymbiosis and pathogenesis are closely related due to host cell invasion and persistence mechanisms (Keeling and McCutcheon, 2017). However, the invading bacteria rarely see long term benefits from these interactions. Endosymbiont genomes are frequently found to be highly reduced due to the impact of Muller's ratchet, in which population bottlenecks in vertically transmitted endosymbionts cause an accumulation of deleterious mutations over time (Moran, 1996; Nowack and Melkonian, 2010; McCutcheon and Moran, 2012). With no gene flow between populations, endosymbionts are unable to recover from mutations and replication errors, which are more likely to occur in G/C rich regions, resulting in a characteristic A/T bias (McCutcheon et al., 2019). The net impact of these forces is the creation of highly reduced, A/T rich genomes, which have convergently evolved in the majority of vertically transmitted endosymbiont lineages (Moran, 1996; Nowack and Melkonian, 2010; McCutcheon and Moran, 2012; Keeling and McCutcheon, 2017; McCutcheon et al., 2019). Though the endosymbiont is fed and housed, it is also effectively incapacitated and permanently tied to its host.

Housing an endosymbiont is also costly for the host, and maintaining a foreign cell, rather than digesting or expelling it, indicates the endosymbiont confers a significant advantage. As part of a larger investigation of the Nephromycidae, here we focus on characterizing the role of the bacterial endosymbiont reported in *Ciona intestinalis* (Ciancio et al., 2008). Since *Cardiosporidium* and *Nephromyces* have maintained α -proteobacteria endosymbionts

since before they diverged, we hypothesize this lineage of endosymbiont must provide metabolic functions of high value to its host apicomplexans. The maintenance of bacterial endosymbionts could be reducing host dependency and resource competition by providing novel biosynthetic pathways, thereby reducing virulence in this unique lineage.

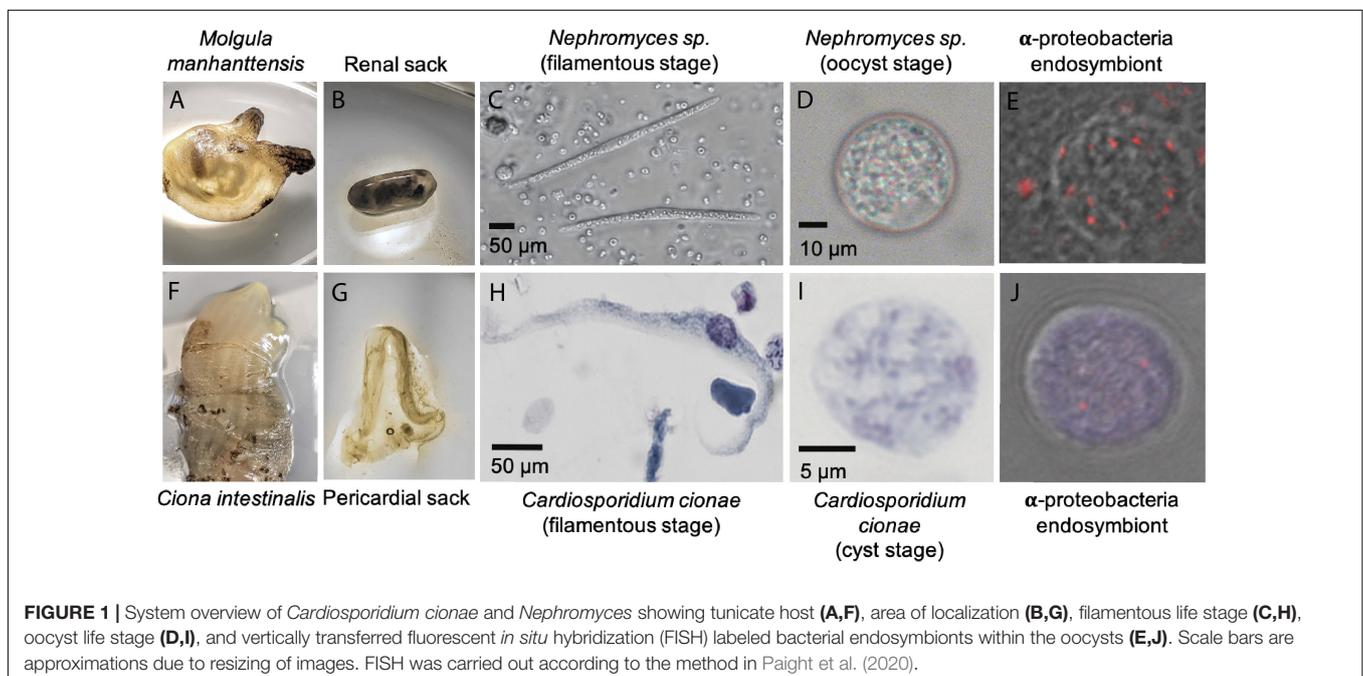
MATERIALS AND METHODS

Microscopy

Visual screens of *Ciona intestinalis* hemolymph were conducted using a 5% Giemsa/phosphate buffer stain with a thin smear slide preparation, as is commonly used to identify malarial infections (Moll et al., 2008). The filamentous life stage was identified during these screens due to its morphological similarity to *Nephromyces*. To confirm identity, three samples comprising 10–15 of the cell types of interest were manually picked and washed using stretched Pasteur pipettes and phosphate buffered saline. These samples were extracted, PCR amplified with *C. cionae* specific primers, and the resulting PCR product sequenced on the Sanger platform at the University of Rhode Island Genome Sequencing Center. Fluorescence *in situ* hybridization (FISH) with 16S rRNA class specific probes was used to localize the bacterial endosymbionts as shown in **Figures 1E,J**. The hybridization was conducted as described in Paight et al. (2020).

Extraction, Sequencing, Assembly, and Binning

The material for the *Cardiosporidium cionae* transcriptome was collected and isolated from wild *Ciona intestinalis* tunicates, as described in detail in Paight et al. (2019). A sucrose density



gradient was used to isolate *C. cionae* from tunicate hemolymph, and to enrich highly infected samples of hemolymph identified with microscopy. The gradient was composed of 20, 25, 30, 35, and 40% sucrose in phosphate buffer, loaded with approximately 5 mL of hemolymph, and centrifuged in a swinging bucket rotor on $500 \times g$ for 30 min, at 4°C (Paight et al., 2019). The 25 and 30% layers were then collected, pelleted, washed with phosphate buffered saline, and stored at -80°C . RNA was extracted from the pellets and the highly infected samples used the Zymo Quick-RNA kit (Zymo Research LLC, Irvine, CA). Three samples with unfiltered hemolymph, hemolymph enriched with the 25% layer, and hemolymph enriched with the 30% layer were shipped on dry ice to the University of Maryland, Baltimore Institute for Genome Sciences, and multiplexed on a single lane of an Illumina HiSeq. These samples produced 92,250,706, 109,023,104, and 110,243,954 reads (Paight et al., 2019). They were assembled with Trinity/Trinotate v2.4.0 (Haas et al., 2013) and binned iteratively with OrthoFinder v2.3.3 (Emms and Kelly, 2019) using a custom database of tunicates, Alveolates, and bacterial endosymbiont data to remove contamination from the host and environment.

For genomic sequencing, *C. intestinalis* were collected from Snug Harbor in South Kingstown, Rhode Island ($41^{\circ}23'13.4''\text{N}$, $71^{\circ}31'01.5''\text{W}$) in August and September 2018, following the same protocol for dissection and needle extraction of the tunicate hemolymph from the pericardial sac. The sucrose density gradient described above was also used to isolate *C. cionae* infected cells for genomic DNA, except that, in 3 of the 4 samples used, a 27% sucrose layer was substituted for the 25% layer to better capture *C. cionae* infected cells. In the fourth sample, a 30% layer was used. The layer of interest was centrifuged, collected, pelleted, and washed as described above, and in Paight et al. (2019). Filtered samples were used alone, rather than being incorporated into unfiltered hemolymph samples. Genomic DNA was immediately extracted using a 1% SDS lysis buffer, Proteinase K and phenol-chloroform extraction, followed by an overnight ethanol precipitation at -20°C . Samples were assessed for quality and concentration with gel electrophoresis, NanoDrop, and Qubit (broad range), and then stored at -20°C .

Samples from four separate gradient columns were individually prepared at the University of Rhode Island Genome Sequencing Center, and the resulting libraries run on a single lane of an Illumina HiSeq4000 at the University of Maryland, Baltimore Institute for Genome Sciences. These libraries were independently trimmed and assessed for quality using Trimmomatic v0.36 and FastQC v0.11.8 before being pooled and assembled with SPAdes v3.13.0 on the Brown University OSCAR server (Bolger et al., 2014; “FastQC A Quality Control Tool for High Throughput Sequence Data” n.d.; Bankevich et al., 2012).

The SPAdes metagenomic assembly was binned by assigning taxonomy to contigs with CAT (von Meijenfeldt et al., 2019). *Rickettsiales* sequences were confirmed using MetaBAT (Kang et al., 2015), and the resulting contigs inspected for contamination and reassembled with Geneious v9.1.8 (Geneious, 2017). Additional apicomplexan sequences were identified by mapping trimmed and binned transcriptomic reads to the full metagenomic assembly using Bowtie2 v2.3.5.1, and contig

coverage calculated with the bedtools v2.26.0 genomcov function (Quinlan and Hall, 2010; Langmead and Salzberg, 2012). The resulting file was sorted with R to extract contigs with greater than 50% coverage of *C. cionae* transcripts. Both the *C. cionae* and α -proteobacterial endosymbiont genomic assemblies were trimmed to a minimum length of 1 kb, as contigs smaller than this were unlikely to be reliably binned. Genome assembly graphs were also visualized with Bandage v0.8.1 (Wick et al., 2015) and clusters of interest were identified with BLAST. The α -proteobacteria cluster was identified with BLAST, exported, and compared to the CAT binned bacterial assembly with average nucleotide identity (ANI) estimations (Rodriguez-R and Konstantinidis, 2016).

Organellar assemblies for both the apicoplast and the mitochondrion were generated with NOVOPlasty v3.7.2 (Dierckxsens et al., 2016). The seed sequences for these assemblies were located using the apicoplast genomes of *Nephromyces* (Muñoz-Gómez et al., 2019a), and Sanger sequences of the *C. cionae* cytochrome C oxidase subunit one (COX-1) gene generated with PCR with local BLASTN databases (Madden, 2003).

Gene Prediction and Annotation

Annotation of the α -proteobacteria endosymbiont genome and *C. cionae* mitochondria was carried out using Prokka v1.14.5 (Seemann, 2014). Closely related *Rickettsia* genomes (as indicated by the preliminary 16S phylogeny) were retrieved from NCBI and used to generate a custom database for α -proteobacteria genome annotation (Seemann, 2014; **Supplementary Table S1**). Annotation of the apicoplast was carried out in Geneious v9.1.8 using a custom database of *Nephromyces* apicoplast annotations (Muñoz-Gómez et al., 2019a). Inverted repeat regions were identified with Repeat Finder plugin.

Cardiosporidium cionae genomic contigs were annotated with the MAKER v2.31.10 pipeline (Holt and Yandell, 2011). Repeats were soft masked using RepeatMasker v4.0.9 (Smit et al., 2013). *Ab initio* predictions and species training parameters were generated with both WebAugustus (Hoff and Stanke, 2013) and SNAP-generated hidden markov models (Korf, 2004). This process was repeated iteratively, and the AED values indicating the fit of gene prediction to the model were analyzed to ensure high quality predictions. Predicted proteins from both organisms were functionally classified with the Kyoto Encyclopedia of Genes and Genomes (KEGG) and NCBI BLASTP v2.7.0+ (Madden, 2003; Kanehisa et al., 2017). Coding sequences were searched for homologous domains with InterProScan (Mitchell et al., 2019). Individual genes of interest were screened using BLAST databases.

Analysis

Completeness of the *C. cionae* genome and transcriptome were assessed with BUSCO using the eukaryotic database (Simão et al., 2015). Homologs identified as multicopy by BUSCO were manually screened with NCBI-BLAST to confirm they did not represent contamination in the finished assembly. Proteins were annotated using orthologs from EuPathDB, PFAM, Kegg, and Interpro. Transcripts with multiple predicted isoforms in the

transcriptome were filtered and selected based on completeness, Interpro score, and length. Completeness and contamination of the *C. cionae* α -proteobacteria genome was assessed with the Microbial Genome Atlas (MiGA) and CheckM (Parks et al., 2015; Rodriguez-R et al., 2018). Candidate pseudogenes were located with Pseudofinder with default parameters, using a custom BLAST database of proteins from the 785 complete alphaproteobacterial genomes available on NCBI (Syberg-Olsen et al., 2018). Visual representations of the metabolic pathways were constructed for both *C. cionae* and the α -proteobacteria using functional annotations from KEGG (Kanehisa et al., 2017). Visual representations of the α -proteobacteria were generated using Circos (Krzywinski et al., 2009) with annotation data from Prokka and functional annotations from KEGG. In addition to functional comparisons using KEGG annotations, the α -proteobacterial endosymbiont genome was also compared to the *Nephromyces* endosymbiont with similarity estimations and orthologous gene content. Similarity was compared with ANI and average amino acid identity (AAI) calculations using the web-based ANI and AAI calculator (Rodriguez-R and Konstantinidis, 2016). Orthologous gene content comparisons between the *C. cionae* α -proteobacteria and all of the endosymbionts in the *Nephromyces* system was carried out with OrthoFinder v2.3.3 (Emms and Kelly, 2019). The resulting overlaps were calculated using the R package limma (Ritchie et al., 2015), and the final figure generated with Venn Diagram (Chen and Boutros, 2011), also in R. Functional gene overlap was based on KEGG annotations and generated using the same R packages.

Phylogenetics

Bacterial phylogenies were constructed using the predicted taxonomy from MiGA, which assigned the endosymbiont to the class alphaproteobacteria with a *p*-value of 0.25 (Parks et al., 2015; Rodriguez-R et al., 2018). To confirm this result, all complete bacterial proteome accessions belonging to this class were retrieved from the NCBI database (712 in total). These data were searched using the α -proteobacteria HMM single copy gene set comprised of 117 proteins, aligned, and the tree constructed using the GToTree workflow (Lee, 2019; “Accelerated Profile HMM Searches” n.d.; Edgar, 2004; Hyatt et al., 2010; Price et al., 2010; “TaxonKit - NCBI Taxonomy Toolkit, n.d.; Capella-Gutiérrez et al., 2009).

The apicoplast encoded genes of *C. cionae* were added to the dataset used in Muñoz-Gómez et al. (2019a) to confirm monophyly with *Nephromyces*, previously indicated with COI and 18S gene trees. Protein homologs were identified using local BLAST-P searches and concatenated with the existing dataset. These sequences were aligned with MAFFT v7, trimmed in Geneious, and concatenated (Madden, 2003; Katoh and Standley, 2013). Species phylogeny was inferred with Maximum Likelihood using IQ-TREE (v1.6) and the LG + G model. Statistical support at branches was estimated using ultrafast bootstrap (1,000) and aLRT (1,000) (Nguyen et al., 2015).

Parameters

The specific scripts and settings used for bioinformatic analysis of *C. cionae* and its endosymbiont have been

deposited in a publicly accessible GitHub repository (github.com/liz-hunter/cardio_project).

RESULTS

Cardiosporidium cionae

Genomic sequencing of the pooled *C. cionae* libraries yielded a total of 320,000,000 paired reads. After trimming and assembly, this resulted in 656,251 contigs, 176,701 of which were larger than 1 kb. Binning with CAT resulted in 3,641 contigs assigned to the superphylum Alveolata. Contigs assigned to Dinophyceae and Ciliophora were removed, leaving 2,754 contigs, and 1,790 of these contigs were larger than 1 kb. The RNA-seq assisted coverage-based binning added an additional 935 contigs, 423 of which were unique and larger than 1 kb. Further manual curation using OrthoFinder eliminated 7 additional contigs. This resulted in a total of 2,206 contigs assigned to *C. cionae*. Of the remaining 174,496 contigs, 221 were assigned to the order Rickettsiales, and 147,793 contigs assigned to the class Ascidiacea (tunicate). The remaining contigs that were classified by CAT were either unclassified or identified as contamination from a variety of marine organisms, particularly bacteria and other protists. Though we cannot rule out seawater as a source of contamination, this likely arises from the physiology of tunicates such as *Ciona intestinalis*, which are filter feeders with an open circulatory system. The *C. cionae* genome assembly is 57 Mb in total, with an N50 of 54.04 kb, and a G/C content of 34.4% (Table 1). This is smaller than some apicomplexan genomes such as coccidian *Toxoplasma gondii* (80 Mb), but considerably larger than hemosporidian *Plasmodium falciparum* (22.9 Mb) and the highly reduced *Cryptosporidium parvum* (9 Mb) (Sibley and Boothroyd, 1992; Abrahamsen et al., 2004). Gene prediction resulted in 4,674 proteins (Table 1). The binned transcriptome assembly yielded a total of 15,077 proteins assigned to *C. cionae*, including all isoforms. When filtered to remove redundancy, this dataset was reduced to 6,733 unique proteins.

The final binned *C. cionae* genome assembly is estimated to be 63.7% complete by BUSCO, with 13.9% of the BUSCOs duplicated. The transcriptome is slightly more complete, with a BUSCO estimate of 68.3% complete orthologs, and 12.5% partial (Paight et al., 2019). When the isoforms were filtered for annotation, this completeness value dropped slightly to 60.0% with 4.6% duplication (Table 1). Despite this, the *C. cionae* assembly contains genes from all of the expected core biosynthetic pathways for a hematozoan (Supplementary Figure S3). *Cardiosporidium* has a suite of basic metabolic pathways including complete or nearly complete functional predictions for glycolysis, gluconeogenesis, pyruvate oxidation, the pentose phosphate cycle, and the citric acid cycle (Figure 2 and Supplementary Table S2). It also encodes a handful of unexpected pathways, including the entire *de novo* IMP biosynthetic pathway. *Cardiosporidium* contains the genes for fatty acid biosynthesis and elongation in the endoplasmic reticulum, as well degradation to produce acetyl-CoA (Figure 2 and Supplementary Table S2).

TABLE 1 | Statistics for the genomic and transcriptomic datasets presented.

Assembly	Contigs	Size (Mb)	N50 (kb)	N90 (kb)	G/C (%)	Proteins	Completeness (%)	Duplication (%)
<i>C. cionae</i> (DNA)	2,213	57	54.12	15.21	34.4	4,692	63.7	13.9
α -proteobacteria	29	1.05	250.39	36.72	29.1	906	91.9	1.9
<i>C. cionae</i> (RNA)	–	–	–	–	–	6,733	60.0	4.6

MIGA was used to assess the bacterial genome completeness, and BUSCO was used to assess the *Cardiosporidium cionae* genome and transcriptome.

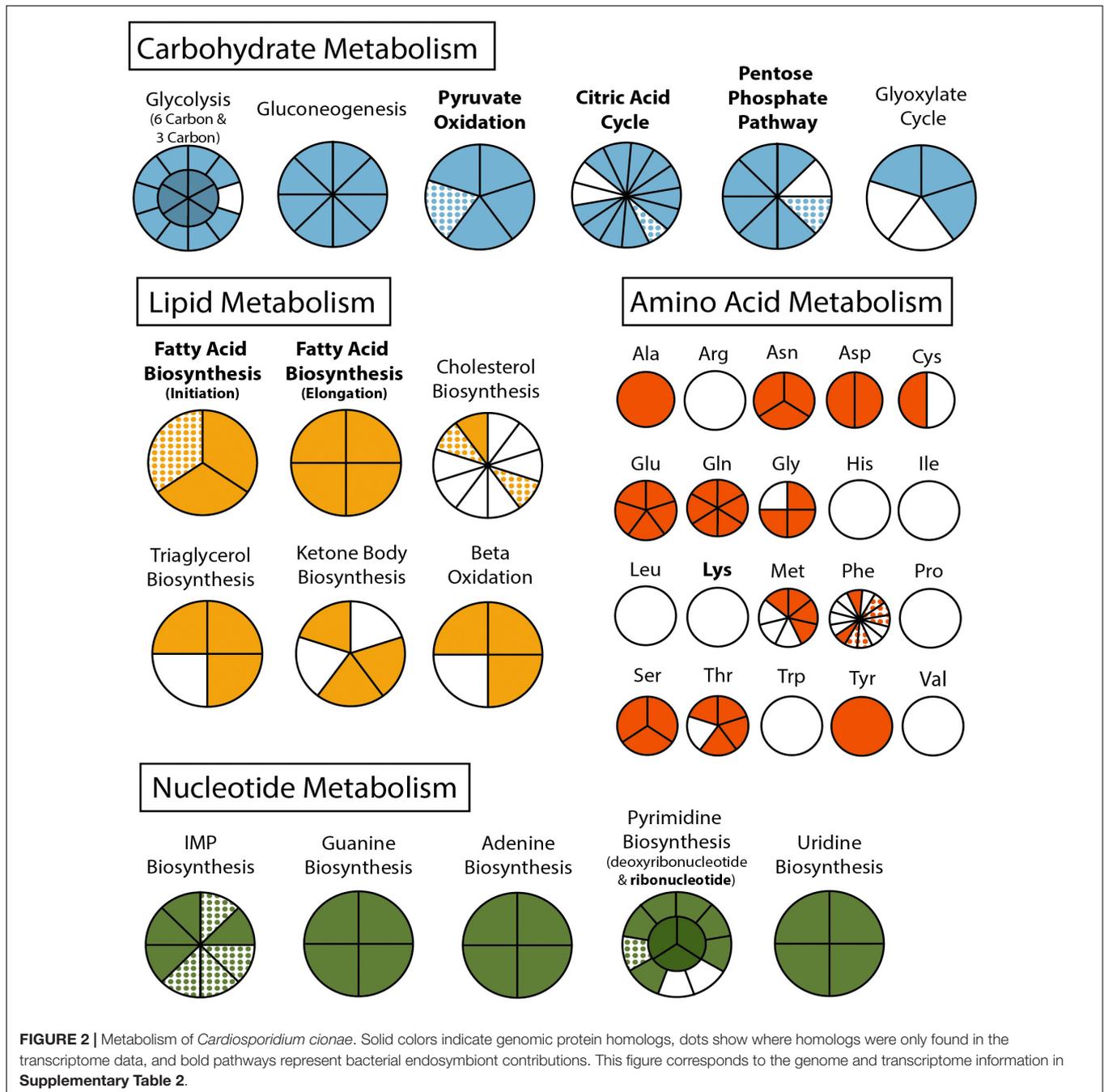
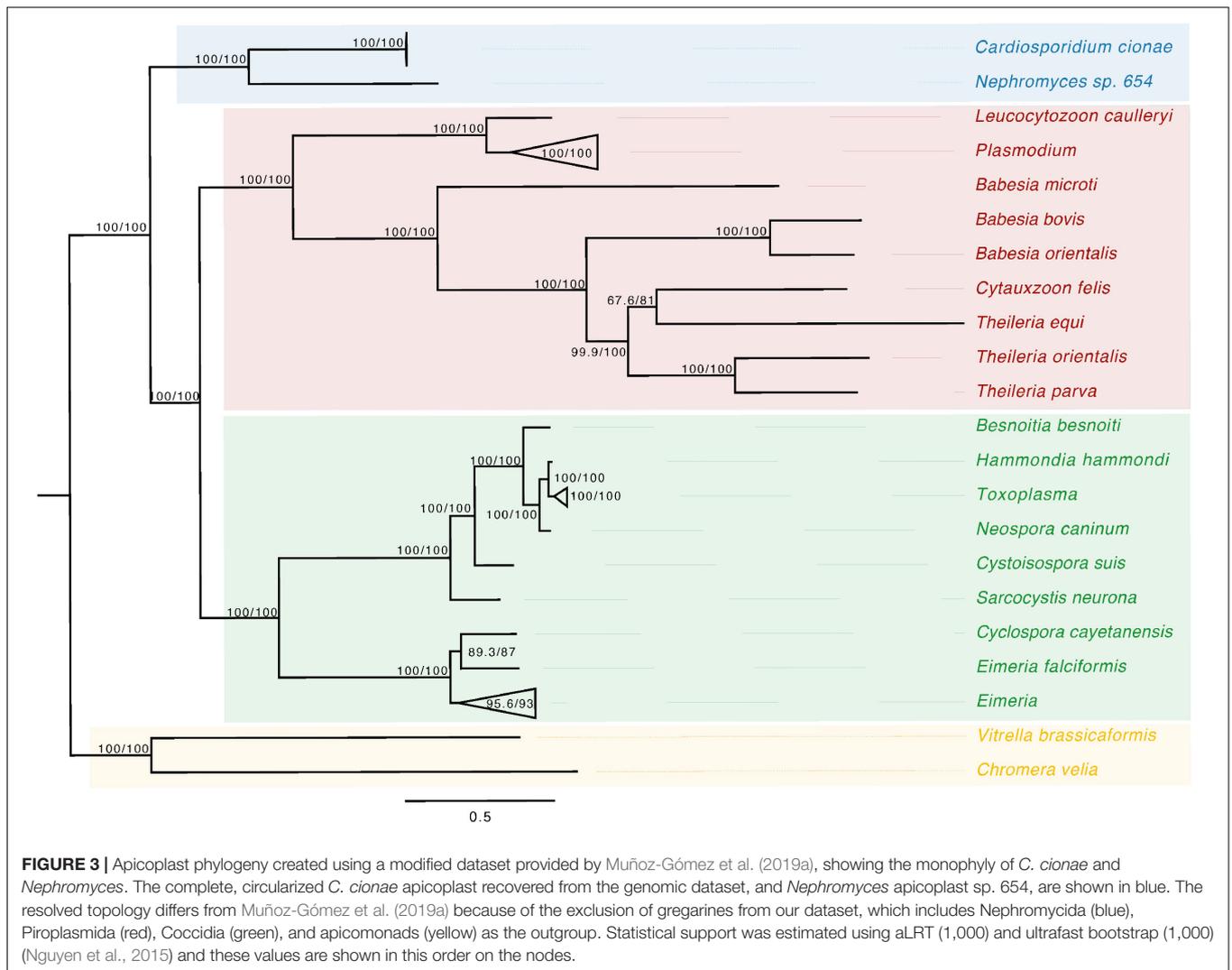


FIGURE 2 | Metabolism of *Cardiosporidium cionae*. Solid colors indicate genomic protein homologs, dots show where homologs were only found in the transcriptome data, and bold pathways represent bacterial endosymbiont contributions. This figure corresponds to the genome and transcriptome information in **Supplementary Table 2**.

Cardiosporidium also encodes the majority of the pathway of triacylglycerol biosynthesis, and partial pathways for cholesterol and ketone body synthesis. It completely lacks any evidence of biosynthetic genes for eight of the twenty-one amino acids,

but it does encode amino acid conversion pathways that other Hematozoa lack. These include the conversion of phenylalanine to tyrosine, and homocysteine to methionine (**Figure 2** and **Supplementary Table S2**). Additionally, *C. cionae* is able to



generate serine from multiple sources (glycerate-3P and glycine), as well as degrade it to pyruvate. The genomic data we recovered only encodes partial pathways for riboflavin, and heme synthesis, and also lacks genes for biotin, thiamine, ubiquinone, and cobalamin synthesis. However, we identified both C5 and C10-20 isoprenoid biosynthesis pathways. This genome also supports the presence of the purine degradation pathway previously identified in the transcriptome of *Nephromyces* (Paight et al., 2019).

Visual screens with thin-smear Giemsa staining indicate that *C. cionae* maintains a very low density inside its host. These microscopy screens further revealed the presence of a large, extracellular filamentous life stage analogous to the filamentous life-stage in *Nephromyces* (Figure 1). Single cell isolation, extraction, and PCR confirmed these cell types were indeed a life-stage of *C. cionae*.

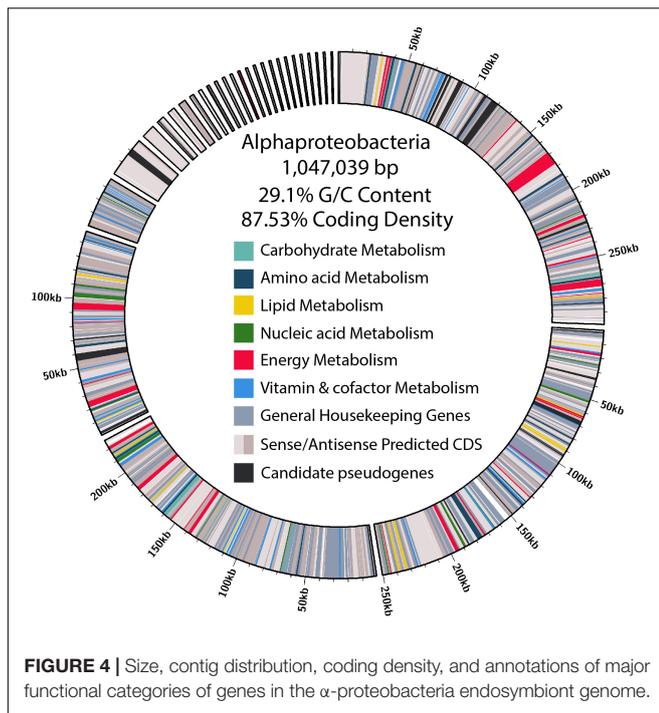
Phylogenetic analysis of the apicoplast encoded proteins supported the monophyly of *Nephromyces* and *C. cionae* (Figure 3). This analysis differs with the placement results for *Nephromyces* published by Muñoz-Gómez et al. (2019a), due to maximized data and the omission of early branching taxa. This

taxon sampling caused Nephromycidae to branch outside of the Hematozoa. The *C. cionae* apicoplast is structurally very similar to those of *Nephromyces* in terms of gene content, size, and organization (Supplementary Figure S1).

α -Proteobacteria

Of the contigs assigned to Rickettsiales, 45 were larger than 1 kb. Reassembly yielded 31 contigs, and manual curation resulted in a final 29 contigs. The final α -endosymbiont assembly is 1.05 Mb in total, with an N50 of 250.39 kb, and a G/C content of 29.1%. Gene prediction and annotation resulted in 906 proteins (Table 1). The Bandage cluster was shown to have an ANI (two-way ANI) of 99.95% (SD.81%) based on 4,878 fragments when compared with the CAT binned assembly. This provided independent validation for the bacterial genome assembly binning. The α -endosymbiont assembly is estimated to be 91.9% complete with 1.8% contamination by MiGA, and 95.5% complete with 2.1% contamination by CheckM.

Characteristic of bacterial endosymbionts, it has a low G/C content and high coding density (Figure 4). This



organism is predicted to encode just 906 genes by Prokka (Table 1), 37 of which were identified as candidate pseudogenes by Pseudofinder. Most of these pseudogenes are derived from genes of unknown function, while some are derived from genes encoding a permease, transposase, thioesterase, phosphodiesterase, and multiple transferases. Pseudofinder also joined 13 ORFs, leaving only 865 predicted genes. With so few genes, it is not surprising that this α -proteobacteria has a sparse number of complete metabolic pathways (Figures 2, 4 and Supplementary Figure S3). The genome is slightly smaller than closely related alphaproteobacterial endosymbionts, such as *Candidatus* Phycorickettsia trachydisci sp. nov. (1.4 MB), *Orientia tsutsugamushi* (2 MB), and other protist associated Rickettsiales lineages (1.4–1.7 MB) (Nakayama et al., 2010; Yurchenko et al., 2018; Muñoz-Gómez et al., 2019b). The α -proteobacteria genomes in both *C. cionae* and *Nephromyces* encode pathways for the biosynthesis of fatty acids, pyrimidines, lipoic acid, heme, glutamine, lysine, ubiquinone, and the citric acid cycle. Only the *C. cionae* α -proteobacteria maintains the genes for asparagine biosynthesis, glycolysis, and the pentose phosphate pathway (Supplementary Figure S2), while only the *Nephromyces* α -proteobacteria can complete glutamic acid biosynthesis.

When the α -proteobacteria in both *Cardiosporidium* and *Nephromyces* were compared for similarity, the results showed these taxa were too divergent to be compared with ANI, and they were instead compared with average AAI. A two-way AAI analysis of 656 proteins showed 47.61% (SD: 12.51%) similarity between these genomes, which is consistent with the phylogenetic analysis that indicates considerable evolutionary distance between these two taxa. This multigene phylogeny of

the α -endosymbionts is congruent with the preliminary 16S gene trees, which places these species in the order Rickettsiales. They belong to the family Rickettsiaceae and are sister to the genus *Rickettsia* (Figure 5).

Ortholog comparisons between the α -endosymbionts indicate these taxa share the majority of their core functions, but the *Cardiosporidium* system α -endosymbiont maintains more unique genes. This taxon also shares greater ortholog and functional overlap with the two additional endosymbionts present in the *Nephromyces* system: betaproteobacteria and Bacteroides (Supplementary Figure S2).

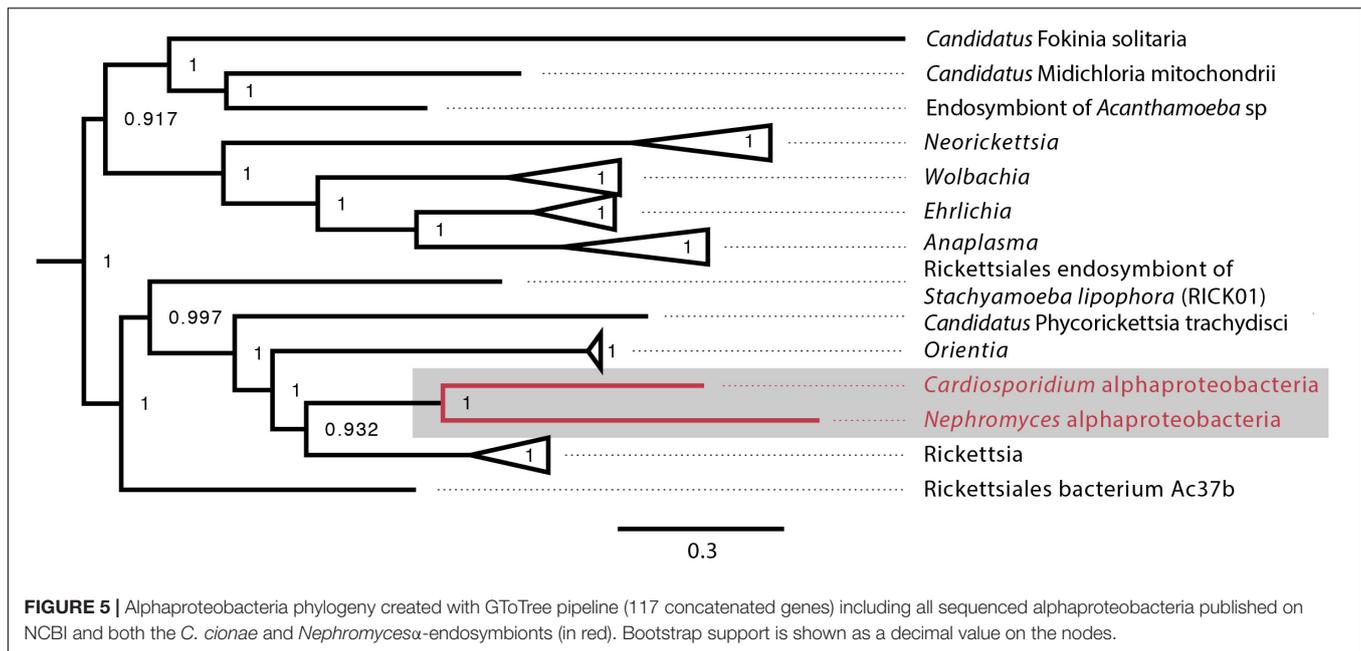
Data Availability

All data associated with this project is deposited in GenBank under the BioProject PRJNA664590. The *Cardiosporidium cionae* whole genome shotgun project has been deposited under the accession JADAQX000000000, and the alphaproteobacterial endosymbiont genome is deposited under the accession JADAQY000000000, and the transcriptome is deposited under the accession GIVE000000000. The versions described in this paper are versions JADAQX010000000, JADAQY010000000, and GIVE010000000.

DISCUSSION

Metabolically, *C. cionae* is similar to other sequenced hematozoans. However, it also encodes some unusual pathways. *Cardiosporidium cionae*, like *Nephromyces*, encodes the *de novo* purine biosynthesis pathway (Figure 2 and Supplementary Table S2), which has been lost in all other sequenced apicomplexans (Janouskovec and Keeling, 2016). These genes resolve with *Nephromyces*, *Vitrella brassicaformis* and dinoflagellates such as *Cryptothecodinium cohnii* in phylogenetic analysis, demonstrating this was not a recent horizontal gene transfer event (Paight et al., 2020). Instead, these data indicate that both genera within Nephromycidae have maintained the ancestral pathway found in free-living Chromerids, and the genes for purine biosynthesis have been lost independently in all other apicomplexan lineages. *De novo* biosynthesis of purines in *C. cionae* and *Nephromyces* reduces dependence on preformed purine metabolites from their respective hosts, potentially enabling the persistence of the extracellular life stages in both of these lineages. *Nephromyces* and *C. cionae* are also able to degrade purines (Paight et al., 2019), and we suspect this aspect of their metabolism is related to the physiology of tunicates, which are incapable of metabolizing uric acid, a purine waste product. Even though the tunicate hosts are unable to degrade uric acid, they inexplicably accumulate it (Nolfi, 1970; Lambert et al., 1998).

Whereas the complete pathways for pentose phosphate cycle, citric acid cycle, and gluconeogenesis mirror other hematozoans (Figure 2 and Supplementary Table S2), *C. cionae* also encodes a handful of genes that suggest it is able to produce glyoxylate. Paight et al. (2019) reported transcripts for a number of peroxisomal proteins in both *C. cionae* and *Nephromyces*, and predicted a novel metabolic pathway.



Despite their numerous metabolic similarities, *C. cionae* and *Nephromyces* appear to have distinct pathways for central carbon metabolism (specifically the citric acid cycle), and part of the closely linked glyoxylate cycle. Both *Nephromyces* and *C. cionae* possess a uniformly highly expressed purine degradation cycle that converts ureidoglycolate to glyoxylate using a novel amidohydrolase, and generates glycine and serine. However, only *Nephromyces* can feed glyoxylate back into the citric acid cycle using malate synthase (Paight et al., 2019). This pathway is a product of the unusual renal sac environment where *Nephromyces* makes its home, which contains an abundance of uric acid sequestered by the host tunicate. In *C. cionae*, malate synthase is conspicuously absent in both the genome and transcriptome, indicating the carbon cycling in these closely related organisms is likely distinct, and potentially one of the differences that accounts for the virulence disparity between *C. cionae* and *Nephromyces*. However, the list of differences, which also includes host species and organellar localization, is relatively short. Though it was known that these taxa have similar life history traits (Ciancio et al., 2008; Saffo et al., 2010), these data suggest their morphology and metabolism are also remarkably similar.

Cardiosporidium cionae and *Nephromyces* (Nephromycidae) branch within Hematozoa, a group of obligate, parasitic, intracellular apicomplexans (Muñoz-Gómez et al., 2019a; Mathur et al., 2019). All previously described members of Hematozoa and sister taxon Coccidia, are intracellular and obligately parasitic. Despite their phylogenetic position within an obligately intracellular clade, members of the Nephromycidae have large, filamentous, extracellular life stages (Figure 1). *Nephromyces* is completely extracellular (Saffo and Nelson, 1983), while *C. cionae* has both intracellular and extracellular life stages. Though morphologically similar to the more basal gregarine apicomplexans (Rueckert et al., 2015), these groups are phylogenetically distant. The Nephromycidae have evolved from

intracellular ancestors and transitioned to the extracellular environment. In *Nephromyces*, this transition is complete, whereas *C. cionae* has both intracellular and extracellular life stages. We believe that extracellularity in this group is related to another unusual characteristic: the maintenance of bacterial endosymbionts in both *C. cionae* and *Nephromyces*.

The maintenance of monophyletic α -endosymbionts in both the *C. cionae* and *Nephromyces* lineages indicates that this endosymbiont is providing something vital to the system. However, at first glance, these endosymbionts are contributing very little to their host apicomplexans. Like its counterpart in *Nephromyces*, the *C. cionae* α -endosymbiont contains only a handful of biosynthetic pathways (Figure 4). Overall, the α -endosymbiont in *C. cionae* does contain more unique orthologs and functional genes when compared with its counterpart in the *Nephromyces* system (Supplementary Figure S2). Primarily, these unique genes are related to energy metabolism (Figure 4 and Supplementary Figure S2), and their presence is likely a result of the heightened evolutionary pressure to maintain critical genes in a system with a single endosymbiont, compared to the three types of endosymbionts present in *Nephromyces* communities. The α -endosymbiont encoded pathways for energy and carbon cycling, while possibly advantageous to *C. cionae*, are likely not critical contributions because they can be completed by the apicomplexan, independent of the endosymbiont (Supplementary Figure S3). The maintenance of an endosymbiont is costly, and it is unlikely to be preserved for a redundant function (McCutcheon and Moran, 2012).

A handful of pathways have been maintained in both α -endosymbiont lineages and are also absent in the host apicomplexans. The only apparently critical functions that cannot be replaced by the apicomplexan metabolism are lysine biosynthesis, and lipoic acid biosynthesis. Lysine is an essential amino acid, and plays an important role in protein biosynthesis. Lysine is an essential media component for the growth of

P. falciparum and is predicted to be scavenged from the host by *T. gondii* (Schuster, 2002; Tymoshenko et al., 2015). Lysine biosynthesis is also absent in the *Nephromyces* genome and transcriptome (Paight et al., 2020). Like *Nephromyces*, *T. gondii*, and *P. falciparum*, our data indicate *C. cionae* cannot synthesize its own lysine and is dependent on host scavenging. Though we cannot exclude the possibility that *C. cionae* encodes lysine biosynthesis with an incomplete genome, based on the genomes of other apicomplexans, lysine biosynthesis is likely absent within Nephromycidae. Lysine is also essential for the host tunicate, *Ciona intestinalis* (Kanehisa et al., 2017), and both organisms requiring environmental sources of lysine puts them in constant competition for the resource. *Cardiosporidium cionae* appears to have circumvented this conflict by maintaining a bacterial endosymbiont that contains the pathway for *de novo* lysine biosynthesis. Rather than compete with the host tunicate for lysine, *C. cionae* cultivates an intracellular source for the essential amino acid, reducing host dependency and potentially virulence.

Lipoic acid is an aromatic sulfur compound that is an essential cofactor for a series of vital metabolic functions. These include the citric acid cycle and alpha keto dehydrogenase complexes, such as the pyruvate dehydrogenase complex and the glycine conversion system. In eukaryotes, lipoic acid is exclusively localized to the mitochondria and the plastid. Apicomplexans localize lipoic acid biosynthesis to the apicoplast, having lost the mitochondrial pathway after the acquisition of the plastid (Crawford et al., 2006). Instead, an alternative scavenging pathway is used to produce the lipoic acid required for the citric acid cycle and glycine conversion system in the mitochondria, and both the scavenging and biosynthetic pathways are considered essential (Günther et al., 2005). Functional studies have shown that when the lipoic acid biosynthetic pathways are knocked out, *P. falciparum* will compensate by scavenging more lipoic acid from the host and shuttling it to the apicoplast (Günther et al., 2007). Similarly, *T. gondii* growth is inhibited by lipoate-deficient media, suggesting scavenging is essential (Crawford et al., 2006). Metabolic modeling also indicates that even apicomplexans that maintain this pathway require supplemental lipoic acid from their host organisms (Blume and Seeber, 2018). Though lipoic acid is produced by the host tunicates, we speculate that there is limited availability for an extracellular organism because it is both produced and used in the mitochondria. This likely means *C. cionae* is dependent on this α -endosymbiont for the production of key compounds such as lipoic acid, for the persistence of a stable extracellular life stage. In this way, maintaining the α -endosymbiont as an internal cofactor source further reduces resource competition between *C. cionae* and its host.

The Nephromycidae have evolved from a clade of an obligately parasitic intracellular apicomplexans, and have transitioned to a mostly extracellular lifestyle. We hypothesize that, by obtaining bacterial endosymbionts, these apicomplexans have acquired metabolic capabilities that enabled this transition. Though *Nephromyces* shares an α -endosymbiont lineage with *C. cionae* (Figure 5), it also has betaproteobacteria and Bacteroides endosymbionts. With this bacterial taxonomic diversity comes metabolic diversity, and though the α -endosymbiont in *C. cionae*

has more unique functional proteins and orthologs than its counterpart, this is dwarfed by the number of unique proteins and orthologs contributed by the two additional taxa present in the *Nephromyces* system (Supplementary Figure S2). We believe the sole endosymbiont in *C. cionae* provides a dedicated source of the essential metabolites lysine and lipoic acid, which likely reduces competition with its host compared to its hematozoan relatives, and makes extracellular life stages possible. In this way, *Cardiosporidium cionae* represents a potential intermediate in the transition to mutualism that has been described in *Nephromyces* (Saffo et al., 2010).

DATA AVAILABILITY STATEMENT

The authors acknowledge that the data presented in this study must be deposited and made publicly available in an acceptable repository, prior to publication. Frontiers cannot accept a manuscript that does not adhere to our open data policies.

AUTHOR CONTRIBUTIONS

EH generated and analyzed the *C. cionae* and bacterial endosymbiont genomic dataset and carried out the comparative analyses. CP assembled the *C. cionae* transcriptome and the *Nephromyces* bacterial genomes. CL provided funding, experimental framework, and expertise. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Science Foundation (#1541510) and the National Institute of Health (R03AI124092).

ACKNOWLEDGMENTS

This material is based upon work conducted at a Rhode Island NSF EPSCoR research facility, the Genomics and Sequencing Center (GSC), supported in part by the National Science Foundation EPSCoR Cooperative Agreement #OIA-1655221. In particular, we would like to thank Janet Atoyan at the GSC for her support and expertise.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.580719/full#supplementary-material>

Supplementary Figure 1 | The annotated, circularized *Cardiosporidium cionae* apicoplast. Two *C. cionae* apicoplasts were recovered that were 99.55% similar

overall and contained identical gene organization, with the SNPs localized to the *subF* gene. The *C. cionae* apicoplasts are similar in size, organization, and gene content to the *Nephromyces* apicoplasts.

Supplementary Figure 2 | Comparisons of the bacterial endosymbiont genomes in *Cardiosporidium cionae* (AlphaC) and *Nephromyces* (AlphaN, Beta, Bac). The left Venn diagram depicts orthologous groups predicted by OrthoFinder, while the right shows functional overlap predicted with KEGG.

Supplementary Figure 3 | Global metabolic map of the *Cardiosporidium cionae* genome and its α -proteobacterial endosymbiont, with pathways exclusive to *C. cionae* shown in blue, exclusive to the α -endosymbiont shown in red, and

the overlapping metabolic pathways shown in yellow. This figure was generated with iPath 3 (Darzi et al., 2018).

Supplementary Table 1 | Closely related reference genomes used for annotation of the *Cardiosporidium cionae*-endosymbiont, downloaded from GenBank, refseq.

Supplementary Table 2 | Key functional genes annotations from the genome and transcriptome of *Cardiosporidium cionae*, corresponding to those seen in **Figure 2**, beginning with the top middle slice and moving clockwise. Annotations are based on functional predictions from KEGG and referenced by the corresponding K-Number.

REFERENCES

- Abrahamsen, M. S., Templeton, T. J., Enomoto, S., Abrahante, J. E., Zhu, G., Lancto, C. A., et al. (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304, 441–445. doi: 10.1126/science.1094786
- Alarcón, M. E., Jara-F, A., Briones, R. C., Dubey, A. K., and Slamovits, C. H. (2017). Gregarine infection accelerates larval development of the cat flea *Ctenocephalides Felis* (Bouché). *Parasitology* 144, 419–425. doi: 10.1017/S0031182016002122
- Babraham Bioinformatics (n.d.). *FastQC A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed November 30, 2019).
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Blume, M., and Seeber, F. (2018). Metabolic interactions between *Toxoplasma Gondii* and its host. *F1000Res* 7:1719. doi: 10.12688/f1000research.16021.1
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Center for Disease Control (2019). *Impact of Malaria*. Available online at: https://www.cdc.gov/malaria/malaria_worldwide/impact.html (accessed January 29, 2019).
- Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12:35. doi: 10.1186/1471-2105-12-35
- Ciaccio, A., Scippa, S., Finetti-Sialer, M., De Candia, A., Avallone, B., and De Vincentiis, M. (2008). Redescription of *Cardiosporidium cionae* (Van Gaver and Stephan, 1907) (Apicomplexa: Piroplasmida), a plasmodial parasite of ascidian haemocytes. *Eur. J. Protistol.* 44, 181–196. doi: 10.1016/j.ejop.2007.11.005
- Conrad, P. A., Miller, M. A., Kreuder, C., James, E. R., Mazet, J., Dabritz, H., et al. (2005). Transmission of toxoplasma: clues from the study of sea otters as sentinels of *Toxoplasma gondii* flow into the marine environment. *Int. J. Parasitol.* 35, 1155–1168. doi: 10.1016/j.ijpara.2005.07.002
- Crawford, M. J., Thomsen-Zieger, N., Ray, M., Schachtner, J., Roos, D. S., and Seeber, F. (2006). *Toxoplasma gondii* scavenges host-derived lipoic acid despite its de novo synthesis in the apicoplast. *EMBO J.* 25, 3214–3222. doi: 10.1038/sj.emboj.7601189
- Criado-Fornelio, A., Verdú-Expósito, C., Martín-Pérez, T., Heredero-Bermejo, I., Pérez-Serrano, J., Guardia-Valle, L., et al. (2017). A survey for gregarines (Protozoa: Apicomplexa) in arthropods in Spain. *Parasitol. Res.* 116, 99–110. doi: 10.1007/s00436-016-5266-0
- Darzi, Y., Letunic, I., Bork, P., and Yamada, T. (2018). iPath3.0: interactive pathways explorer v3. *Nucleic Acids Res.* 46, W510–W513. doi: 10.1093/nar/gky299
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2016). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45:e18. doi: 10.1093/nar/gkw955
- Dyson, J., Grahame, J., and Evennett, P. J. (1993). The mucron of the gregarine *Digyalum oweni* (Protozoa: Apicomplexa), parasitic in *Littorina* species (Mollusca: Gastropoda). *J. Nat. Hist.* 27, 557–564. doi: 10.1080/00222939300770311
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y
- Evans Anderson, H., and Christiaen, L. (2016). *Ciona* as a simple chordate model for heart development and regeneration. *J. Cardiovasc. Dev. Dis.* 3:25. doi: 10.3390/jcdd3030025
- Ewald, P. W. (1987). Transmission modes and evolution of the parasitism-mutualism continuum. *Ann. N. Y. Acad. Sci.* 503, 295–306. doi: 10.1111/j.1749-6632.1987.tb40616.x
- Fitzpatrick, D. A., Creevey, C. J., and McInerney, J. O. (2006). Genome phylogenies indicate a meaningful α -proteobacterial phylogeny and support a grouping of the mitochondria with the rickettsiales. *Mol. Biol. Evol.* 23, 74–85. doi: 10.1093/molbev/msj009
- Fredrickson, J. K., Zachara, J. M., Balkwill, D. L., Kennedy, D., Li, S. M., Kostandarithes, H. M., et al. (2004). Geomicrobiology of high-level nuclear waste-contaminated vadose sediments at the Hanford site, Washington State. *Appl. Environ. Microbiol.* 70, 4230–4241. doi: 10.1128/AEM.70.7.4230-4241.2004
- Frénal, K., Dubremetz, J.-F., Lebrun, M., and Soldati-Favre, D. (2017). Gliding motility powers invasion and egress in Apicomplexa. *Nat. Rev. Microbiol.* 15, 645–660. doi: 10.1038/nrmicro.2017.86
- Geneious (2017). *Geneious | Bioinformatics Software for Sequence Data Analysis*. Available online at: <https://www.geneious.com/> (accessed August 2018).
- Gubbels, M.-J., and Duraisingh, M. T. (2012). Evolution of apicomplexan secretory organelles. *Int. J. Parasitol.* 42, 1071–1081. doi: 10.1016/j.ijpara.2012.09.009
- Günther, S., McMillan, P. J., Wallace, L. J. M., and Müller, S. (2005). Plasmodium falciparum possesses organelle-specific α -keto acid dehydrogenase complexes and lipoylation pathways. *Biochem. Soc. Trans.* 33, 977–980. doi: 10.1042/BST0330977
- Günther, S., Wallace, L., Patzewitz, E.-M., McMillan, P. J., Storm, J., Wrenger, C., et al. (2007). Apicoplast lipoic acid protein ligase B is not essential for *Plasmodium falciparum*. *PLoS Pathog.* 3:e189. doi: 10.1371/journal.ppat.0030189
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-Seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hoff, K. J., and Stanke, M. (2013). WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res.* 41, W123–W128. doi: 10.1093/nar/gkt418
- Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. doi: 10.1186/1471-2105-12-491
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Janouskovec, J., and Keeling, P. J. (2016). Evolution: causality and the origin of parasitism. *Curr. Biol.* 26, R174–R177. doi: 10.1016/j.cub.2015.12.057

- Jeurissen, S. H., Janse, E. M., Vermeulen, A. N., and Vervelde, L. (1996). *Eimeria tenella* infections in chickens: aspects of host-parasite: interaction. *Vet. Immunol. Immunopathol.* <refvol> 54, 231–238. doi: 10.1016/S0165-2427(96)05689-9
- Johnson, J., and Oelmüller, R. (2009). Mutualism or parasitism: life in an unstable continuum. What can we learn from the mutualistic interaction between *Piriformospora indica* and *Arabidopsis thaliana*? – Review. *Endocytobiosis Cell Res.* 19, 81–111.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092
- Kang, D. D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. doi: 10.7717/peerj.1165
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Keeling, P. J., and McCutcheon, J. P. (2017). Endosymbiosis: the feeling is not mutual. *J. Theor. Biol.* 434, 75–79. doi: 10.1016/j.jtbi.2017.06.008
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5:59. doi: 10.1186/1471-2105-5-59
- Kozek, W. J., and Rao, R. U. (2007). “The discovery of wolbachia in arthropods and nematodes – a historical perspective,” in *Wolbachia: A Bug's Life in Another Bug*, Vol. 5, eds A. Hoerauf and R. U. Rao (Basel: Karger Publishers), 1–14. doi: 10.1159/000104228
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Kwong, W. K., del Campo, J., Mathur, V., Vermeij, M. J. A., and Keeling, P. J. (2019). A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes. *Nature* 568, 103–107. doi: 10.1038/s41586-019-1072-z
- Lambert, C. C., Lambert, G., Crundwell, G., and Kantardjiev, K. (1998). Uric acid accumulation in the solitary ascidian *Corella inflata*. *J. Exp. Zool.* 282, 323–331. doi: 10.1002/(sici)1097-010x(19981015)282:3<323::aid-jez5>3.0.co;2-o
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lee, M. D. (2019). GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35, 4162–4164. doi: 10.1093/bioinformatics/btz188
- Leung, T. L. F., and Poulin, R. (2008). *Parasitism, Commensalism, and Mutualism: Exploring the Many Shades of Symbioses*. Vol. 58. France: Vie et Milieu, 107–115.
- Madden, T. (2003). *The BLAST Sequence Analysis Tool*. Bethesda, MD: National Center for Biotechnology Information.
- Marciano-Cabral, F. (1988). Biology of *Naegleria* Spp. *Microbiol. Rev.* 52, 114–133. doi: 10.1128/mmr.52.1.114-133.1988
- Mathur, V., del Campo, J., Kolisko, M., and Keeling, P. J. (2018). Global diversity and distribution of close relatives of apicomplexan parasites. *Environ. Microbiol.* 20, 2824–2833. doi: 10.1111/1462-2920.14134
- Mathur, V., Kolisko, M., Hehenberger, E., Irwin, N. A. T., Leander, B. S., Kristmundsson, Á., et al. (2019). Multiple independent origins of apicomplexan-like parasites. *Curr. Biol.* 29, 2936–2941.e5. doi: 10.1016/j.cub.2019.07.019
- McCutcheon, J. P., Boyd, B. M., and Dale, C. (2019). The life of an insect endosymbiont from the cradle to the grave. *Curr. Biol.* 29, R485–R495. doi: 10.1016/j.cub.2019.03.032
- McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi: 10.1038/nrmicro2670
- McFadden, G. I., and Waller, R. F. (1997). Plastids in parasites of humans. *BioEssays* 19, 1033–1040. doi: 10.1002/bies.950191114
- McFadden, G. I., and Yeh, E. (2017). The apicoplast: now you see it, now you don't. *Int. J. Parasitol.* 47, 137–144. doi: 10.1016/j.ijpara.2016.08.005
- Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47, D351–D360. doi: 10.1093/nar/gky1100
- Moll, K., Ljungström, I., Perlmann, H., Scherf, A., and Wahlgren, M. (2008). *Methods in Malarial Research*. Vol. 5. Malaria Research and Reference Reagent Resource Center (MR4), 351.
- Moran, N. A. (1996). Accelerated evolution and muller's ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 93, 2873–2878. doi: 10.1073/pnas.93.7.2873
- Morrison, D. A. (2009). Evolution of the Apicomplexa: where are we now? *Trends Parasitol.* 25, 375–382. doi: 10.1016/j.pt.2009.05.010
- Muñoz-Gómez, S. A., Durnin, K., Eme, L., Paight, C., Lane, C. E., Saffo, M. B., et al. (2019a). *Nephromyces* represents a diverse and novel lineage of the Apicomplexa that has retained apicoplasts. *Genome Biol. Evol.* 11, 2727–2740. doi: 10.1093/gbe/evz155
- Muñoz-Gómez, S. A., Hess, S., Burger, G., Lang, B. F., Susko, E., Slamovits, C. H., et al. (2019b). An updated phylogeny of the Alphaproteobacteria reveals that the parasitic *Rickettsiales* and *Holosporales* have independent origins. *eLife* 8:e42535. doi: 10.7554/eLife.42535
- Nakayama, K., Kurokawa, K., Fukuhara, M., Urakami, H., Yamamoto, S., Yamazaki, K., et al. (2010). Genome comparison and phylogenetic analysis of *Orientia tsutsugamushi* strains. *DNA Res.* 17, 281–291. doi: 10.1093/dnares/dsq018
- 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. doi: 10.1093/molbev/msu300
- Nolfi, J. R. (1970). Biosynthesis of uric acid in the tunicate, *Molgula manhattensis*, with a general scheme for the function of stored purines in animals. *Comp. Biochem. Physiol.* 35, 827–842. doi: 10.1016/0010-406x(70)90078-2
- Nowack, E. C. M., and Melkonian, M. (2010). Endosymbiotic associations within protists. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 699–712. doi: 10.1098/rstb.2009.0188
- Paight, C., Hunter, E. S., and Lane, C. E. (2020). Codependence in the *Nephromyces* species swarm depends on heterospecific bacterial endosymbionts. *BioRxiv [Preprint]* doi: 10.1101/2020.10.18.344572v1
- Paight, C., Slamovits, C. H., Saffo, M. B., and Lane, C. E. (2019). *Nephromyces* encodes a urate metabolism pathway and predicted peroxisomes, demonstrating that these are not ancient losses of apicomplexans. *Genome Biol. Evol.* 11, 41–53. doi: 10.1093/gbe/evy251
- Pappas, G., Roussos, N., and Falagas, M. E. (2009). Toxoplasmosis snapshots: global status of *Toxoplasma gondii* seroprevalence and implications for pregnancy and congenital toxoplasmosis. *Int. J. Parasitol.* 39, 1385–1394. doi: 10.1016/j.ijpara.2009.04.003
- 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. doi: 10.1101/gr.186072.114
- PLoS Computational Biology (n.d.). *Accelerated Profile HMM Searches*. Available online at: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pone.0009490> (accessed February 7, 2020).
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). *Limma* powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Rodríguez-R, L. M., Gunturu, S., Harvey, W. T., Rosselló-Mora, R., Tiedje, J. M., Cole, J. R., et al. (2018). The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of *Archaea* and *Bacteria* at the whole genome level. *Nucleic Acids Res.* 46, W282–W288. doi: 10.1093/nar/gky467
- Rodríguez-R, L. M., and Konstantinidis, K. T. (2016). The envomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1
- Roos, D. S. (2005). Themes and variations in apicomplexan parasite biology. *Science* 309, 72–73. doi: 10.1126/science.1115252
- Rueckert, S., Betts, E. L., and Tsaousis, A. D. (2019). The symbiotic spectrum: where do the gregarines fit? *Trends Parasitol.* 35, 687–694. doi: 10.1016/j.pt.2019.06.013
- Rueckert, S., Wakeman, K. C., Jenke-Kodama, H., and Leander, B. S. (2015). Molecular systematics of marine gregarine apicomplexans from pacific tunicates, with descriptions of five novel species of *Lankesteria*. *Int. J. Syst. Evol. Microbiol.* 65, 2598–2614. doi: 10.1099/ijs.0.000300

- Saffo, M. B., McCoy, A. M., Rieken, C., and Slamovits, C. H. (2010). *Nephromyces*, a beneficial apicomplexan symbiont in marine animals. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16190–16195. doi: 10.1073/pnas.1002335107
- Saffo, M. B., and Nelson, R. (1983). The cells of *Nephromyces*: developmental stages of a single life cycle. *Can. J. Bot.* 61, 3230–3239. doi: 10.1139/b83-360
- Schuster, F. L. (2002). Cultivation of *Plasmodium* Spp. *Clin. Microbiol. Rev.* 15, 355–364. doi: 10.1128/CMR.15.3.355-364.2002
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Sibley, L. D., and Boothroyd, J. C. (1992). Construction of a molecular karyotype for *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* 51, 291–300. doi: 10.1016/0166-6851(92)90079-y
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Smit, A. F. A., Hubley, R. and Green, P. (2013). *RepeatMasker Open-4.0, 2013–2015*. Available online at: <http://www.repeatmasker.org/>
- Soldati, D., Dubremetz, J. F., and Lebrun, M. (2001). Microneme proteins: structural and functional requirements to promote adhesion and invasion by the apicomplexan parasite *Toxoplasma gondii*. *Int. J. Parasitol.* 31, 1293–1302. doi: 10.1016/s0020-7519(01)00257-0
- Sow, S. O., Muhsen, K., Nasrin, D., Blackwelder, W. C., Wu, Y., Farag, T. H., et al. (2016). The burden of *Cryptosporidium* diarrheal disease among children < 24 months of age in moderate/high mortality regions of sub-Saharan Africa and South Asia, utilizing data from the global enteric multicenter study (GEMS). *PLoS Negl. Trop. Dis.* 10:e0004729. doi: 10.1371/journal.pntd.0004729
- Suja, G., Kripa, V., Mohamed, K., Pothodi, S., and Sanil, N. K. (2016). *Nematopsis* Sp. (Apicomplexa: Porosporidae) infection in *Crassostrea madrasensis* and its associated histopathology. *J. Mar. Biol. Assoc. India* 58, 29–33. doi: 10.6024/jmbai.2016.58.1.1890-04
- Syberg-Olsen, M., Garber, A., McCutcheon, J., and Husnik, F. (2018). *Pseudofinder, GitHub Repository*. Available online at: <https://github.com/filip-husnik/pseudo-finder/> (accessed August 2018).
- TaxonKit - NCBI Taxonomy Toolkit (n.d.). *A Cross-Platform and Efficient NCBI Taxonomy Toolkit*. Available online at: <https://bioinf.shenwei.me/taxonkit/> (accessed February 7, 2020).
- Tymoshenko, S., Oppenheim, R. D., Agren, R., Nielsen, J., Soldati-Favre, D., and Hatzimanikatis, V. (2015). Metabolic needs and capabilities of *Toxoplasma gondii* through combined computational and experimental analysis. *PLoS Comput. Biol.* 11:e1004261. doi: 10.1371/journal.pcbi.1004261
- von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H., and Dutilh, B. E. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT: supplementary table 1. *BioRxiv [Preprint]* doi: 10.1101/530188
- Votýpka, J., Modrý, D., Oborník, M., Šlapeta, J., and Lukeš, J. (2016). “Apicomplexa,” in *Handbook of the Protists*, eds J. M. Archibald, A. G. B. Simpson, C. H. Slamovits, L. Margulis, M. Melkonian, D. J. Chapman, et al. (Cham: Springer International Publishing), 1–58. doi: 10.1007/978-3-319-32669-6_20-1
- Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of *de novo* genome assemblies: fig. 1. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383
- Woo, Y. H., Ansari, H., Otto, T. D., Klinger, C. M., Kolisko, M., Michálek, J., et al. (2015). Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *eLife* 4:e06974. doi: 10.7554/eLife.06974
- Yurchenko, T., Ševčíková, T., Příbyl, P., El Karkouri, K., Klimeš, V., Amaral, R., et al. (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. doi: 10.1038/s41396-018-0177-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Hunter, Paight and Lane. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.