Metabolic contributions of an alphaproteobacterial endosymbiont in the apicomplexan *Cardiosporidium cionae*

Elizabeth Sage Hunter[1], Christopher J Paight[2], Christopher E Lane[1*]

[1]*Department of Biological Sciences, University of Rhode Island, Kingston, RI, 02881, USA.*
[2]*Department of Ecology, Evolution & Marine Biology, University of California, Santa Barbara, 93106 CA, USA.*

*** Correspondence:**
Dr. Christopher Lane
clane@uri.edu

Keywords: Apicomplexa, Parasite – Host interactions, Mutualism, Parasitism, Bacterial Endosymbiont, Alphaproteobacteria host-associated bacteria

Abstract

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.

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, Roussos, and Falagas 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 (Suja et al. 2016; Dyson, Grahame, and Evennett 1993), arthropods (Criado-Fornelio et al. 2017; Alarcón et al. 2017), and urochordates (Ciancio et al. 2008; Mary Beth 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 (Roos 2005; Janouskovec and Keeling 2016; McFadden and Waller 1997; Soldati, Dubremetz, and Lebrun 2001; Frénal 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, Dubremetz, and Lebrun 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 (Janouskovec and Keeling 2016; Woo et al. 2015). 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 (Morrison 2009; Roos 2005; Mathur et al. 2018; Gubbels and Duraisingh 2012; McFadden and Yeh 2017; Woo et al. 2015, Votýpka et al. 2016), 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 (Rueckert, Betts, and Tsaousis 2019; Kwong et al. 2019; Saffo et al. 2010). However, what defines the boundaries along this continuum of symbiotic association is still a topic of much debate (Leung and Poulin 2008; Johnson and Oelmüller 2009; Ewald 1987). Phylogenetic analysis indicates *Nephromyces* is sister to the haematozoan clade, and closely related to highly virulent genera such as *Plasmodium, Theiliera,* and *Babesia* (Muñoz-Gómez et al. 2019). Thus far, apicomplexan species with variable life strategies have been found in early branching groups, such as the Gregarina and Corallicods. 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. in 2008, a full century after its initial discovery. Similar to other haemosporidians 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. in 2008), and Narragansett Bay, Rhode Island, USA, 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 separate three 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. 2019).

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, Boyd, and Dale 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, Boyd, and Dale 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, Creevey, and McInerney 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; McCutcheon and Moran 2012; Nowack and Melkonian 2010). 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, Boyd, and Dale 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; McCutcheon and Moran 2012; Keeling and McCutcheon 2017; McCutcheon, Boyd, and Dale 2019; Nowack and Melkonian 2010). 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 endosymbionts reported from *C. cionae* (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 Fig.1 (panels e and i). 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 5mL of hemolymph, and centrifuged in a swinging bucket rotor on 500 x g for 30 minutes, at 4°C (Paight et al. 2019). The 25% and 30% layers were then collected, pelleted, and washed with phosphate buffered saline, and stored at -80°C. RNA was

163 extracted from the pellets and the highly infected samples used the Zymo Quick-RNA kit
164 (Zymo Research LLC, Irvine, CA). Three samples with unfiltered hemolymph, hemolymph
165 enriched with the 25% layer, and hemolymph enriched with the 30% layer were shipped
166 on dry ice to the University of Maryland, Baltimore Institute for Genome Sciences, and
167 multiplexed on a single lane of an Illumina HiSeq. These samples produced 92,250,706,
168 109,023,104, and 110,243,954 reads (Paight et al. 2019). They were assembled with
169 Trinity/Trinotate v2.4.0 (Haas et al. 2014) and binned iteratively with OrthoFinder v2.3.3
170 (Emms and Kelly 2019) using a custom database of tunicates, Alveolates, and bacterial
171 endosymbiont data to remove contamination from the host and environment.
172 For genomic sequencing, *C. intestinalis* were collected from Snug Harbor in South
173 Kingstown, Rhode Island (41°23'13.4"N, 71°31'01.5"W) in August and September 2018,
174 following the same protocol for dissection and needle extraction of the tunicate
175 hemolymph from the pericardial sac. The sucrose density gradient described above was
176 also used isolate *C. cionae* infected cells for genomic DNA, except that, in 3 of the 4 samples
177 used, a 27% sucrose layer was substituted for the 25% layer to better capture *C. cionae*
178 infected cells. In the fourth sample, a 30% layer was used. The layer of interest was
179 centrifuged, collected, pelleted, and washed as described above, and in Paight et al. (2019).
180 Filtered samples were used alone, rather than being incorporated into unfiltered
181 hemolymph samples. Genomic DNA was immediately extracted using a 1% SDS lysis buffer,
182 Proteinase K and phenol-chloroform extraction, followed by an overnight ethanol
183 precipitation at -20 °C. Samples were assessed for quality and concentration with gel
184 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, Lohse, and Usadel 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 | Bioinformatics Software for Sequence Data Analysis" 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 genomecov function (Langmead and Salzberg 2012; Quinlan and Hall 2010). 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 1kb, 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 estimations (Rodriguez and Konstantinidis 2016).

Organellar assemblies for both the apicoplast and the mitochondrion were generated with NOVOPlasty v3.7.2 (Dierckxsens, Mardulyn, and Smits 2016). The seed sequences for these assemblies were located using the apicoplast genomes of *Nephromyces* (Muñoz-Gómez et al. 2019), 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) (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. 2019). 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, Hubley, and Green 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+ (Kanehisa et al. 2017; Madden 2003). 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 orthologues 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 average nucleotide identity and average amino acid identity calculations using the web-based ANI and AAI calculator (Rodriguez 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.; Hyatt et al. 2010; Price, Dehal, and Arkin 2010; Edgar 2004; "TaxonKit - NCBI Taxonomy Toolkit" n.d.; Capella-Gutiérrez, Silla-Martínez, and Gabaldón 2009).

The apicoplast encoded genes of *C. cionae* were added to the dataset used in (Muñoz-Gómez et al. 2019) 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 (1000) and aLRT (1000) (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 1kb. Binning with CAT resulted in 3,641 contigs assigned to the superphylum Alveolata. Contigs assigned to Dinophyceae and Ciliophora were removed, leaving 2754 contigs, and 1790 of these contigs were larger than 1kb. The RNA-seq assisted coverage-based binning added an additional 935 contigs, 423 of which were unique and larger than 1kb. 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 *C. cionae* genome assembly is 57Mb in total, with an N50 of 54.04kb, and a G/C content of 34.4% (Table 1). This is smaller than some apicomplexan genomes such as coccidian *Toxoplasma gondi* (80Mb), but considerably larger than haemosporidian *Plasmodium falciparum* (22.9Mb) and the highly reduced *Cryptosporidium parvum* (9Mb) (Sibley and Boothroyd 1992; Abrahamsen 2004). Gene

303 prediction resulted in 4,674 proteins (Table 1). The binned transcriptome assembly
304 yielded a total of 15,077 proteins assigned to *C. cionae,* including all isoforms. When filtered
305 to remove redundancy, this dataset was reduced to 6,733 unique proteins.
306 The final binned *C. cionae* genome assembly is estimated to be 63.7% complete by
307 BUSCO, with 13.9% duplication. The transcriptome is slightly more complete, with a BUSCO
308 estimate of 68.3% complete orthologs, and 12.5% partial (Paight et al. 2019). When the
309 isoforms were filtered for annotation, this completeness value dropped slightly to 60.0%
310 with 4.6% duplication (Table 1). Despite this, the *C. cionae* assembly contains genes from
311 all of the expected core biosynthetic pathways for a haematozoan. *Cardiosporidium* has a
312 suite of basic metabolic pathways including complete or nearly complete functional
313 predictions for glycolysis, gluconeogenesis, pyruvate oxidation, the pentose phosphate
314 cycle, and the citric acid cycle (Fig. 2, Table S2). It also encodes a handful of unexpected
315 pathways, including the entire *de novo* IMP biosynthetic pathway. *Cardiosporidium* contains
316 the genes for fatty acid biosynthesis and elongation in the endoplasmic reticulum, as well
317 degradation to produce acetyl-CoA (Fig. 2, Table S2).
318 *Cardiosporidium* also encodes the majority of the pathway of triacylglycerol
319 biosynthesis, and partial pathways for cholesterol and ketone body synthesis. It completely
320 lacks any evidence of biosynthetic genes for eight of the twenty-one amino acids but does
321 encode amino acid conversion pathways that other Hematozoa lack. These include the
322 conversion of phenylalanine to tyrosine, and homocysteine to methionine (Fig. 2, Table S2).
323 Additionally, *C. cionae* is able to generate serine from multiple sources (glycerate-3P and
324 glycine), as well as degrade it to pyruvate. The genomic data we recovered only encodes
325 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* (Fig. 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* (Fig. 3). This analysis differs with the placement results for *Nephromyces* published by Muñoz-Gómez et al. 2019, 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 (Fig. S1).

<u>**α**- proteobacteria</u>

Of the contigs assigned to Rickettsiales, 45 were larger than 1kb. Reassembly yielded 31 contigs, and manual curation resulted in a final 29 contigs. The final **α**-endosymbiont assembly is 1.05Mb in total, with an N50 of 250.39kb, 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 average nucleotide identity (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 (Fig. 4). This organism is predicted to encode just 906 genes by Prokka (Table 1), and 37 of which were identified as candidate pseudogenes by Pseudofinder. These genes were largely hypothetical proteins, but 15 genes were also identified as likely being nonfunctional. These included a permease, transposase, thioesterase, phosphodiesterase, and multiple transferases. Pseudofinder also joined 13 ORFs, leaving only 865 predicted functional genes. With so few functional genes, it is not surprising that this **α**-proteobacteria has a sparse number of complete metabolic pathways (Fig. 4, Fig. 2). The genome is slightly smaller than closely related alphaproteobacterial endosymbionts, such as *Candidatus* Phycorickettsia trachydisci sp. nov. (1.4MB), *Orientia tsutsugamushi* (2MB), and other protist associated Rickettsiales lineages (1.4-1.7MB) (Yurchenko et al. 2018; Nakayamak et al. 2010; Muñoz-Gómez et al. 2019). 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 (Fig. 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 average nucleotide identity (ANI), and they were instead compared with average amino acid identity (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* (Fig. 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 (Fig. S2).

Data Availability

All data associated with this project is deposited in GenBank under the BioProject PRJNA664590. The *Cardiosporidium cionae* whole genome shotgun projected 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 GIVE00000000. The versions described in this paper are versions JADAQX010000000, JADAQY010000000, and GIVE01000000.

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 (Fig. 2, 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 *Crypthecodinium 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 (Fig. 2, 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 generate 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 genomic data suggest their morphology and metabolism are also remarkably similar.

Cardiosporidium cionae and *Nephromyces* (Nephromycidae) branch within Haematozoa, a group of obligate, parasitic, intracellular apicomplexans (Muñoz-Gómez et al. 2019; Mathur et al. 2019). All previously described members of Haemotozoa, 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 (Fig. 1). *Nephromyces* is completely extracellular (Saffo and Nelson 1982), 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 has 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 (Fig. 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 (Fig. S2). Primarily, these unique genes are related to energy metabolism (Fig. 4, Fig. 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. 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* (Tymoshenko et al. 2015; Schuster 2002). 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 is likely absent within Nephromcyidae. 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* (Fig. 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 (Fig. 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 haematozoan 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).

BIBLIOGRAPHY

Abrahamsen, Mitchell S, Thomas J Templeton, Shinichiro Enomoto, Juan E Abrahante, Guan Zhu, Cheryl A Lancto, Mingqi Deng, et al. 2004. "Complete Genome Sequence of the Apicomplexan Cryptosporidium Parvum" 304: 6.

"Accelerated Profile HMM Searches." n.d. Accessed February 7, 2020. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002195.

Alarcón, M. E., A. Jara-F, R. C. Briones, A. K. Dubey, and C. H. Slamovits. 2017. "Gregarine Infection Accelerates Larval Development of the Cat Flea Ctenocephalides Felis (Bouché)." *Parasitology* 144 (4): 419–25. https://doi.org/10.1017/S0031182016002122.

"Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data." n.d. Accessed November 30, 2019. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19 (5): 455–77. https://doi.org/10.1089/cmb.2012.0021.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. "TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25 (15): 1972–73. https://doi.org/10.1093/bioinformatics/btp348.

Center for Disease Control. 2019. "Impact of Malaria." January 29, 2019. https://www.cdc.gov/malaria/malaria_worldwide/impact.html.

Chen, Hanbo, and Paul C. Boutros. 2011. "VennDiagram: A Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R." *BMC Bioinformatics* 12 (1): 35. https://doi.org/10.1186/1471-2105-12-35.

Ciancio, A., S. Scippa, M. Finetti-Sialer, A. De Candia, B. Avallone, and M. De Vincentiis. 2008. "Redescription of Cardiosporidium Cionae (Van Gaver and Stephan, 1907) (Apicomplexa: Piroplasmida), a Plasmodial Parasite of Ascidian Haemocytes." *European Journal of Protistology* 44 (3): 181–96. https://doi.org/10.1016/j.ejop.2007.11.005.

Conrad, P.A., M.A. Miller, C. Kreuder, E.R. James, J. Mazet, H. Dabritz, D.A. Jessup, Frances Gulland, and M.E. Grigg. 2005. "Transmission of Toxoplasma: Clues from the Study of Sea Otters as Sentinels of Toxoplasma Gondii Flow into the Marine Environment." *International Journal for Parasitology* 35 (11–12): 1155–68. https://doi.org/10.1016/j.ijpara.2005.07.002.

Crawford, Michael J, Nadine Thomsen-Zieger, Manisha Ray, Joachim Schachtner, David S Roos, and Frank Seeber. 2006. "Toxoplasma Gondii Scavenges Host-Derived Lipoic Acid despite Its de Novo Synthesis in the Apicoplast." *The EMBO Journal* 25 (13): 3214–22. https://doi.org/10.1038/sj.emboj.7601189.

Criado-Fornelio, A., C. Verdú-Expósito, T. Martin-Pérez, I. Heredero-Bermejo, J. Pérez-Serrano, L. Guàrdia-Valle, and M. Panisello-Panisello. 2017. "A Survey for Gregarines (Protozoa: Apicomplexa) in Arthropods in Spain." *Parasitology Research* 116 (1): 99–110. https://doi.org/10.1007/s00436-016-5266-0.

Dierckxsens, Nicolas, Patrick Mardulyn, and Guillaume Smits. 2016. "NOVOPlasty: De Novo Assembly of Organelle Genomes from Whole Genome Data." *Nucleic Acids Research*, October. https://doi.org/10.1093/nar/gkw955.

Dyson, J., J. Grahame, and P.J. Evennett. 1993. "The Mucron of the Gregarine Digyalum Oweni (Protozoa: Apicomplexa), Parasitic in Littorina Species (Mollusca: Gastropoda)." *Journal of Natural History* 27 (3): 557–64. https://doi.org/10.1080/00222939300770311.

Edgar, Robert C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Emms, David M., and Steven Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20 (1): 238. https://doi.org/10.1186/s13059-019-1832-y.

Evans Anderson, Heather, and Lionel Christiaen. 2016. "Ciona as a Simple Chordate Model for Heart Development and Regeneration." Journal of Cardiovascular Development and Disease 3 (3): 25. https://doi.org/10.3390/jcdd3030025.

Fitzpatrick, David A., Christopher J. Creevey, and James O. McInerney. 2006. "Genome Phylogenies Indicate a Meaningful α-Proteobacterial Phylogeny and Support a Grouping of the Mitochondria with the Rickettsiales." *Molecular Biology and Evolution* 23 (1): 74–85. https://doi.org/10.1093/molbev/msj009.

Fredrickson, James K., John M. Zachara, David L. Balkwill, David Kennedy, Shu-mei W. Li, Heather M. Kostandarithes, Michael J. Daly, Margaret F. Romine, and Fred J. Brockman. 2004. "Geomicrobiology of High-Level Nuclear Waste-Contaminated Vadose Sediments at the Hanford Site, Washington State." *Applied and Environmental Microbiology* 70 (7): 4230–41. https://doi.org/10.1128/AEM.70.7.4230-4241.2004.

Frénal, Karine, Jean-François Dubremetz, Maryse Lebrun, and Dominique Soldati-Favre. 2017. "Gliding Motility Powers Invasion and Egress in Apicomplexa." *Nature Reviews Microbiology* 15 (11): 645–60. https://doi.org/10.1038/nrmicro.2017.86.

"Geneious | Bioinformatics Software for Sequence Data Analysis." 2017. Geneious. 2017. https://www.geneious.com/.

Gubbels, Marc-Jan, and Manoj T. Duraisingh. 2012. "Evolution of Apicomplexan Secretory Organelles." *International Journal for Parasitology* 42 (12): 1071–81. https://doi.org/10.1016/j.ijpara.2012.09.009.

Günther, Svenja, Lynsey Wallace, Eva-Maria Patzewitz, Paul J. McMillan, Janet Storm, Carsten Wrenger, Ryan Bissett, Terry K. Smith, and Sylke Müller. 2007. "Apicoplast Lipoic Acid Protein Ligase B Is Not Essential for Plasmodium Falciparum." *PLOS Pathogens* 3 (12): e189. https://doi.org/10.1371/journal.ppat.0030189.

Haas, Brian J, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. "De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis." *Nature Protocols* 8 (8): 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hoff, Katharina J., and Mario Stanke. 2013. "WebAUGUSTUS—a Web Service for Training AUGUSTUS and Predicting Genes in Eukaryotes." *Nucleic Acids Research* 41 (Web Server issue): W123–28. https://doi.org/10.1093/nar/gkt418.

Holt, Carson, and Mark Yandell. 2011. "MAKER2: An Annotation Pipeline and Genome-Database Management Tool for Second-Generation Genome Projects." *BMC Bioinformatics* 12 (1): 491. https://doi.org/10.1186/1471-2105-12-491.

Hyatt, Doug, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, and Loren J Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11 (1). https://doi.org/10.1186/1471-2105-11-119.

Janouskovec, Jan, and Patrick J. Keeling. 2016. "Evolution: Causality and the Origin of Parasitism." *Current Biology* 26 (4): R174–77. https://doi.org/10.1016/j.cub.2015.12.057.

Jeurissen, S. H. M., E. M. Janse, A. N. Vermeulen, and L. Vervelde. 1996. "Eimeria Tenella Infections in Chickens: Aspects of Host-Parasite: Interaction." *Veterinary Immunology and Immunopathology*, Proceedings of the International Veterinary Immunology Symposium, 54 (1): 231–38. https://doi.org/10.1016/S0165-2427(96)05689-9.

Kanehisa, Minoru, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. 2017. "KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs." *Nucleic Acids Research* 45 (D1): D353–61. https://doi.org/10.1093/nar/gkw1092.

Kang, Dongwan D., Jeff Froula, Rob Egan, and Zhong Wang. 2015. "MetaBAT, an Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities." *PeerJ* 3 (August): e1165. https://doi.org/10.7717/peerj.1165.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30 (4): 772–80. https://doi.org/10.1093/molbev/mst010.

Keeling, Patrick J., and John P. McCutcheon. 2017. "Endosymbiosis: The Feeling Is Not Mutual." *Journal of Theoretical Biology* 434 (December): 75–79. https://doi.org/10.1016/j.jtbi.2017.06.008.

Korf, Ian. 2004. "Gene Finding in Novel Genomes." *BMC Bioinformatics*, 9.

Kozek, Wieslaw J., and Ramakrishna U. Rao. 2007. "The Discovery of Wolbachia in Arthropods and Nematodes – A Historical Perspective." Wolbachia: A Bug's Life in Another Bug 5: 1–14. https://doi.org/10.1159/000104228.

Kwong, Waldan K., Javier del Campo, Varsha Mathur, Mark J. A. Vermeij, and Patrick J. Keeling. 2019. "A Widespread Coral-Infecting Apicomplexan with Chlorophyll Biosynthesis Genes." *Nature* 568 (7750): 103–7. https://doi.org/10.1038/s41586-019-1072-z.

Lambert, Charles C, Gretchen Lambert, Guy Crundwell, and Katherine Kantardjieff. 1998. "Uric Acid Accumulation in the Solitary Ascidian," 10.

Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Lee, Michael D. 2019. "GToTree: A User-Friendly Workflow for Phylogenomics." Edited by Yann Ponty. *Bioinformatics* 35 (20): 4162–64. https://doi.org/10.1093/bioinformatics/btz188.

Madden, Tom. 2003. *The BLAST Sequence Analysis Tool*. National Center for Biotechnology Information (US). https://www.ncbi.nlm.nih.gov/books/NBK21097/.

Marciano-Cabral, F. 1988. "Biology of Naegleria Spp." *Microbiological Reviews* 52 (1): 114–33.

Mathur, Varsha, Javier del Campo, Martin Kolisko, and Patrick J. Keeling. 2018. "Global Diversity and Distribution of Close Relatives of Apicomplexan Parasites." *Environmental Microbiology* 20 (8): 2824–33. https://doi.org/10.1111/1462-2920.14134.

Mathur, Varsha, Martin Kolísko, Elisabeth Hehenberger, Nicholas A. T. Irwin, Brian S. Leander, Árni Kristmundsson, Mark A. Freeman, and Patrick J. Keeling. 2019. "Multiple Independent Origins of Apicomplexan-Like Parasites." Current Biology 29 (17): 2936-2941.e5. https://doi.org/10.1016/j.cub.2019.07.019.

McCutcheon, John P., Bret M. Boyd, and Colin Dale. 2019. "The Life of an Insect Endosymbiont from the Cradle to the Grave." *Current Biology* 29 (11): R485–95. https://doi.org/10.1016/j.cub.2019.03.032.

McCutcheon, John P., and Nancy A. Moran. 2012. "Extreme Genome Reduction in Symbiotic Bacteria." *Nature Reviews Microbiology* 10 (1): 13–26. https://doi.org/10.1038/nrmicro2670.

McFadden, Geoffrey I., and Ross F. Waller. 1997. "Plastids in Parasites of Humans." *BioEssays* 19 (11): 1033–40. https://doi.org/10.1002/bies.950191114.

McFadden, Geoffrey Ian, and Ellen Yeh. 2017. "The Apicoplast: Now You See It, Now You Don't." *International Journal for Parasitology* 47 (2–3): 137–44. https://doi.org/10.1016/j.ijpara.2016.08.005.

Meijenfeldt, F. A. Bastiaan von, Ksenia Arkhipova, Diego D. Cambuy, Felipe H. Coutinho, and Bas E. Dutilh. 2019. "Robust Taxonomic Classification of Uncharted Microbial Sequences and Bins with CAT and BAT: Supplementary Table 1." *BioRxiv*, January. https://doi.org/10.1101/530188.

Mitchell, Alex L., Teresa K. Attwood, Patricia C. Babbitt, Matthias Blum, Peer Bork, Alan Bridge, Shoshana D. Brown, et al. 2019. "InterPro in 2019: Improving Coverage, Classification and Access to Protein Sequence Annotations." *Nucleic Acids Research* 47 (D1): D351–60. https://doi.org/10.1093/nar/gky1100.

Moll, Kirsten, Inger Ljungström, Hedvig Perlmann, Artur Scherf, and Mats Wahlgren. 2008. "Methods in Malarial Research" 5: 351.

Moran, N. A. 1996. "Accelerated Evolution and Muller's Rachet in Endosymbiotic Bacteria." *Proceedings of the National Academy of Sciences* 93 (7): 2873–78. https://doi.org/10.1073/pnas.93.7.2873.

Morrison, David A. 2009. "Evolution of the Apicomplexa: Where Are We Now?" *Trends in Parasitology* 25 (8): 375–82. https://doi.org/10.1016/j.pt.2009.05.010.

Muñoz-Gómez, Sergio A., Keira Durnin, Laura Eme, Christopher Paight, Christopher E. Lane, Mary B. Saffo, and Claudio H. Slamovits. 2019. "Nephromyces Represents a Diverse and Novel Lineage of the Apicomplexa That Has Retained Apicoplasts." *Genome Biology and Evolution* 11 (10): 2727–40. https://doi.org/10.1093/gbe/evz155.

Muñoz-Gómez, Sergio A, Sebastian Hess, Gertraud Burger, B Franz Lang, Edward Susko, Claudio H Slamovits, and Andrew J Roger. 2019. "An Updated Phylogeny of the Alphaproteobacteria Reveals That the Parasitic *Rickettsiales* and *Holosporales* Have Independent Origins." ELife 8 (February): e42535. https://doi.org/10.7554/eLife.42535.

Nakayama, K., K. Kurokawa, M. Fukuhara, H. Urakami, S. Yamamoto, K. Yamazaki, Y. Ogura, T. Ooka, and T. Hayashi. 2010. "Genome Comparison and Phylogenetic Analysis of *Orientia Tsutsugamushi* Strains." DNA Research 17 (5): 281–91. https://doi.org/10.1093/dnares/dsq018.

Nguyen L.-T., Schmidt, H.A., Haeseler, A. von, Minh, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.*, 32:268-274. https://doi.org/10.1093/molbev/msu300.

Nolfi, James R. 1970. "Biosynthesis of Uric Acid in the Tunicate, *Molgula Manhattensis*, with a General Scheme for the Function of Stored Purines in Animals." Comparative Biochemistry and Physiology 35 (4): 827–42. https://doi.org/10.1016/0010-406X(70)90078-2.

Nowack, Eva C. M., and Michael Melkonian. 2010. "Endosymbiotic Associations within Protists." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1541): 699–712. https://doi.org/10.1098/rstb.2009.0188.

Paight, Christopher, Claudio H. Slamovits, Mary Beth Saffo, and Christopher E. Lane. 2019. "Nephromyces Encodes a Urate Metabolism Pathway and Predicted Peroxisomes, Demonstrating That These Are Not Ancient Losses of Apicomplexans." *Genome Biology and Evolution* 11 (1): 41–53. https://doi.org/10.1093/gbe/evy251.

Paight, Christopher, Elizabeth Sage Hunter, and Christopher E. Lane. 2020. Codependence in the Nephromyces species swarm depends on heterospecific bacterial endosymbionts. BioRxiv [Preprint]. Available at: https://www.biorxiv.org/content/10.1101/2020.10.18.344572v1 (Accessed October 18, 2020).

Pappas, Georgios, Nikos Roussos, and Matthew E. Falagas. 2009. "Toxoplasmosis Snapshots: Global Status of Toxoplasma Gondii Seroprevalence and Implications for Pregnancy and Congenital Toxoplasmosis." *International Journal for Parasitology* 39 (12): 1385–94. https://doi.org/10.1016/j.ijpara.2009.04.003.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. "CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes." *Genome Research* 25 (7): 1043–55. https://doi.org/10.1101/gr.186072.114.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. "FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments." *PLOS ONE* 5 (3): e9490. https://doi.org/10.1371/journal.pone.0009490.

Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7): e47–e47. https://doi.org/10.1093/nar/gkv007.

Rodriguez R., Luis M, Santosh Gunturu, William T Harvey, Ramon Rosselló-Mora, James M Tiedje, James R Cole, and Konstantinos T Konstantinidis. 2018. "The Microbial Genomes Atlas (MiGA) Webserver: Taxonomic and Gene Diversity Analysis of Archaea and Bacteria at the Whole Genome Level." *Nucleic Acids Research* 46 (Web Server issue): W282–88. https://doi.org/10.1093/nar/gky467.

Rodriguez-R, Luis M., and Konstantinos T. Konstantinidis. 2016. "The Enveomics Collection: A Toolbox for Specialized Analyses of Microbial Genomes and Metagenomes." e1900v1. PeerJ Inc. https://doi.org/10.7287/peerj.preprints.1900v1.

Roos, David S. 2005. "Themes and Variations in Apicomplexan Parasite Biology." *Science, New Series* 309 (5731,): 72–73.

Rueckert, Sonja, Emma L. Betts, and Anastasios D. Tsaousis. 2019. "The Symbiotic Spectrum: Where Do the Gregarines Fit?" *Trends in Parasitology*, July, S1471492219301643. https://doi.org/10.1016/j.pt.2019.06.013.

Rueckert, Sonja, Kevin C. Wakeman, Holger Jenke-Kodama, and Brian S. Leander. 2015. "Molecular Systematics of Marine Gregarine Apicomplexans from Pacific Tunicates, with Descriptions of Five Novel Species of Lankesteria." *International Journal of Systematic and Evolutionary Microbiology* 65 (8): 2598–2614. https://doi.org/10.1099/ijs.0.000300.

Saffo, M. B. 1990. "Symbiosis within a Symbiosis: Intracellular Bacteria within the Endosymbiotic Protist Nephromyces." *Marine Biology* 107 (2): 291–96. https://doi.org/10.1007/BF01319828.

Saffo, Mary Beth, Adam M. McCoy, Christopher Rieken, and Claudio H. Slamovits. 2010. "*Nephromyces*, a Beneficial Apicomplexan Symbiont in Marine Animals." *Proceedings of the National Academy of Sciences of the United States of America* 107 (37): 16190–95. https://doi.org/10.1073/pnas.1002335107.

Saffo, Mary Beth, and Rebecca Nelson. 1983. "The cells of Nephromyces: developmental stages of a single life cycle." *Canadian Journal of Botany* 61 (12): 3230–39. https://doi.org/10.1139/b83-360.

Seemann, Torsten. 2014. "Prokka: Rapid Prokaryotic Genome Annotation." *Bioinformatics* 30 (14): 2068–69. https://doi.org/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.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12. https://doi.org/10.1093/bioinformatics/btv351.

Soldati, Dominique, Jean Francois Dubremetz, and Maryse Lebrun. 2001. "Microneme Proteins: Structural and Functional Requirements to Promote Adhesion and Invasion by the Apicomplexan Parasite Toxoplasma Gondii." *International Journal for Parasitology* 31 (12): 1293–1302. https://doi.org/10.1016/S0020-7519(01)00257-0.

Sow, Samba O., Khitam Muhsen, Dilruba Nasrin, William C. Blackwelder, Yukun Wu, Tamer H. Farag, Sandra Panchalingam, 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 Neglected Tropical Diseases* 10 (5). https://doi.org/10.1371/journal.pntd.0004729.

Spalding, Maroya D., and Sean T. Prigge. 2010. "Lipoic Acid Metabolism in Microbial Pathogens." *Microbiology and Molecular Biology Reviews : MMBR* 74 (2): 200–228. https://doi.org/10.1128/MMBR.00008-10.

Suja, Girija, V. Kripa, Kolliyil Mohamed, Shamal Pothodi, and N K Sanil. 2016. "Nematopsis Sp. (Apicomplexa: Porosporidae) Infection in Crassostrea Madrasensis and Its Associated Histopathology." *Journal of the Marine Biological Association of India* 58 (January): 29–33. https://doi.org/10.6024/jmbai.2016.58.1.1890-04.

Syberg-Olsen, M., A. Garber, J. McCutcheon, & F. Husnik 2018: Pseudofinder, GitHub repository: https://github.com/filip-husnik/pseudo-finder/.

"TaxonKit - NCBI Taxonomy Toolkit." n.d. Accessed February 7, 2020. https://bioinf.shenwei.me/taxonkit/.

Tymoshenko, Stepan, Rebecca D. Oppenheim, Rasmus Agren, Jens Nielsen, Dominique Soldati-Favre, and Vassily Hatzimanikatis. 2015. "Metabolic Needs and Capabilities of Toxoplasma Gondii through Combined Computational and Experimental Analysis." Edited by Costas D. Maranas. *PLOS Computational Biology* 11 (5): e1004261. https://doi.org/10.1371/journal.pcbi.1004261.

Votýpka, Jan, David Modrý, Miroslav Oborník, Jan Šlapeta, and Julius Lukeš. 2016. "Apicomplexa." In *Handbook of the Protists*, edited by John M. Archibald, Alastair G.B. Simpson, Claudio H. Slamovits, Lynn Margulis, Michael Melkonian, David J. Chapman, and John O. Corliss, 1–58. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-32669-6_20-1.

Wick, Ryan R., Mark B. Schultz, Justin Zobel, and Kathryn E. Holt. 2015. "Bandage: Interactive Visualization of *de Novo* Genome Assemblies: Fig. 1." *Bioinformatics* 31 (20): 3350–52. https://doi.org/10.1093/bioinformatics/btv383.

Woo, Yong H, Hifzur Ansari, Thomas D Otto, Christen M Klinger, Martin Kolisko, Jan Michálek, Alka Saxena, et al. 2015. "Chromerid Genomes Reveal the Evolutionary Path from Photosynthetic Algae to Obligate Intracellular Parasites." Edited by Magnus Nordborg. *ELife* 4 (July): e06974. https://doi.org/10.7554/eLife.06974.

Yurchenko, Tatiana, Tereza Ševčíková, Pavel Přibyl, Khalid El Karkouri, Vladimír Klimeš, Raquel Amaral, Veronika Zbránková, et al. 2018. "A Gene Transfer Event Suggests a Long-Term Partnership between Eustigmatophyte Algae and a Novel Lineage of Endosymbiotic Bacteria." The ISME Journal 12 (9): 2163–75. https://doi.org/10.1038/s41396-018-0177-y.

Table 1: Statistics for the genomic and transcriptomic datasets presented. MiGA was used to assess the bacterial genome completeness, and BUSCO was used to assess the *Cardiosporidium cionae* genome and transcriptome.

Assembly	Contigs	Size (Mb)	N50 (kb)	N90 (kb)	G/C (%)	Proteins	Completeness (%)	Duplication (%)
C. cionae (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
C. cionae (RNA)	-	-	-	-	-	6,733	60.0	4.6

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

***Rickettsia* reference genomes**	**RefSeq Accession**
Rickettsia conorii str. Malish 7	GCF_000007025.1
Rickettsia felis str. Pedreira	GCF_000964665.1
Rickettsia japonica YH	GCF_000283595.1
Rickettsia massiliae MTU5	GCF_000016625.1
Rickettsia peacockii str. Rustic	GCF_000021525.1
Rickettsia prowazekii str. Breinl	GCF_000367405.1
Rickettsia typhi str. Wilmington	GCF_000008045.1
Rickettsia bellii RML369-C	GCF_000012385.1

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.

Pathway/Step	Description	KNUM	DNA	RNA
Glycolysis, glucose => pyruvate				
1	hexokinase	K00844	x	x
2	glucose-6-phosphate isomerase	K01810	x	x
3	6-phosphofructokinase 1	K00805		
4	fructose-1,6-bisphosphate aldolase	K01623	x	x
5	triosephosphate isomerase (TIM)	K01803	x	x
6	glyceraldehyde 3-phosphate dehydrogenase	K00134	x	x
7	phosphoglycerate kinase	K00927	x	x
8	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	K01834	x	x
9	enolase	K01689	x	x
10	pyruvate kinase	K00873	x	x
Gluconeogenesis, oxaloacetate => fructose-6P				
1	phosphoenolpyruvate carboxykinase (ATP)	K01610	x	x
2	enolase	K01689	x	x
3	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	K01834	x	x
4	phosphoglycerate kinase	K00927	x	x
5	glyceraldehyde 3-phosphate dehydrogenase	K00134	x	x
6	triosephosphate isomerase (TIM)	K01803	x	x
7	fructose-bisphosphate aldolase, class I	K01623	x	x
8	fructose-1,6-bisphosphatase I + II	K03841	x	x
Pyruvate Oxidation, pyruvate => acetyl-CoA				
1	pyruvate dehydrogenase E1 component alpha subunit	K00161	x	x
2	pyruvate dehydrogenase E1 component	K00162	x	

	beta subunit			
3	pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase)	K00627	x	x
4	dihydrolipoamide dehydrogenase	K00382		x
5	dihydrolipoamide dehydrogenase-binding protein of pyruvate dehydrogenase complex	K13997	x	
Citric Acid Cycle				
1	citrate synthase	K01647	x	x
2	aconitate hydratase	K01681	x	x
3	isocitrate dehydrogenase	K00031	x	x
4	2-oxoglutarate dehydrogenase E1 component	K00164	x	x
5	2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase)	K00658	x	x
6	pyruvate dehydrogenase complex subunit PDH-E3I	K00382		x
7	succinyl-CoA synthetase alpha subunit	K01899	x	x
8	succinyl-CoA synthetase beta subunit	K01900	x	x
9	succinate dehydrogenase (ubiquinone) flavoprotein subunit	K00234	x	x
10	succinate dehydrogenase (ubiquinone) iron-sulfur subunit	K00235	x	x
11	succinate dehydrogenase (ubiquinone) cytochrome b560 subunit	K00236		
12	succinate dehydrogenase (ubiquinone) membrane anchor subunit	K00237		
13	fumarate hydratase, class II	K01679	x	
14	malate dehydrogenase	K00024	x	x
Pentose Phosphate Pathway				
1	glucose-6-phosphate 1-dehydrogenase	K00036	x	x
2	6-phosphogluconolactonase	K01057		
3	6-phosphogluconate dehydrogenase	K00033		x
4	ribulose-phosphate 3-epimerase	K01783	x	
5	ribose 5-phosphate isomerase A	K01807	x	x
6	transketolase	K00615	x	x
7	transaldolase	K00616	x	x

8	glucose-6-phosphate isomerase	K01810	x	x
Glyoxylate Cycle				
1	citrate synthase	K01647	x	x
2	aconitate hydratase	K01681	x	x
3	isocitrate lyase	K01637		
4	malate synthase	K01638		
5	malate dehydrogenase	K00024	x	x
Fatty Acid Biosynthesis (Initiation)				
1	acetyl-CoA carboxylase / biotin carboxylase 1	K11262	x	x
2	[acyl-carrier-protein] S-malonyltransferase	K00645	x	x
3	3-oxoacyl-[acyl-carrier-protein] synthase III	K00648		x
Fatty Acid Biosynthesis (Elongation)				
1	3-oxoacyl-[acyl-carrier-protein] synthase II	K09458	x	x
2	3-oxoacyl-[acyl-carrier protein] reductase	K00059	x	
3	beta-hydroxyacyl-acyl carrier protein dehydratase (FABZ)	K02372	x	x
4	enoyl-[acyl-carrier protein] reductase I	K00208	x	x
Cholesterol Biosynthesis, squalene 2,3-epoxide => cholesterol				
1	lanosterol synthase	K01852		
2	sterol 14alpha-demethylase	K05917		
3	Delta14-sterol reductase	K00222		
4	methylsterol monooxygenase	K07750		x
5	sterol-4alpha-carboxylate 3-dehydrogenase (decarboxylating)	K07748		
6	17beta-estradiol 17-dehydrogenase / 3beta-hydroxysteroid 3-dehydrogenase	K13373		
7	Delta24-sterol reductase	K09828		
8	cholestenol Delta-isomerase	K01824		
9	C-5 sterol desaturase, putative	K00227	x	

10	7-dehydrocholesterol reductase, putative	K00213	x	x
Triaglycerol Biosynthesis				
1	glycerol-3-phosphate O-acyltransferase / dihydroxyacetone phosphate acyltransferase	K13507	x	x
2	glycerol-3-phosphate O-acyltransferase	K00655		x
2	1-acyl-sn-glycerol-3-phosphate acyltransferase	K22831	x	
3	1-acylglycerol-3-phosphate O-acyltransferase			
4	diacylglycerol O-acyltransferase 1	K11155	x	x
Ketone Body Biosynthesis, acetyl-CoA => acetoacetate/3-hydroxybutyrate/acetone				
1	acetoacetate decarboxylase	K01574		
2	3-hydroxybutyrate dehydrogenase	K00019	x	
3	acetyl-CoA C-acetyltransferase	K00626	x	x
4	hydroxymethylglutaryl-CoA synthase	K01641		
5	hydroxymethylglutaryl-CoA lyase	K01640	x	
Beta Oxidation				
1	acyl-CoA oxidase	K00232	x	x
2	enoyl-CoA hydratase	K07511	x	x
3				
4	acetyl-CoA acyltransferase	K07513	x	x
IMP Biosynthesis				
1	amidophosphoribosyltransferase	K00764		x
2	phosphoribosylamine---glycine ligase	K01945	x	
3	Trifunctional GART (fragment)	K11787		x
4	phosphoribosylformylglycinamidine synthase	K01952		x
5	Trifunctional GART (fragment)	K11787		x
6	phosphoribosylaminoimidazole-succinocarboxamide synthase	K01923	x	x
7	adenylosuccinate lyase	K01756	x	

8	phosphoribosylaminoimidazolecarboxamide formyltransferase	K00602	x	
Guanine Biosynthesis				
1	IMP dehydrogenase	K00088	x	x
2	GMP synthase (glutamine-hydrolysing)	K01951	x	x
3	guanylate kinase	K00942	x	x
4	pyruvate kinase	K00873	x	x
4	nucleoside-diphosphate kinase	K00940	x	
Adenine Biosynthesis				
1	adenylosuccinate lyase	K00939	x	
2	adenylosuccinate synthase	K01756	x	x
3	adenylate kinase	K01939	x	x
4	nucleoside-diphosphate kinase	K00940	x	
4	pyruvate kinase	K00873	x	x
Pyrimidine Biosynthesis (ribonucleotide)				
1	UMP-CMP kinase	K13800	x	x
2	nucleoside-diphosphate kinase	K00940	x	
3	CTP synthase	K01937	x	x
Pyrimidine Biosynthesis (deoxyribonucleotide)				
1	ribonucleoside-diphosphate reductase subunit M1	K10807	x	x
2	ribonucleoside-diphosphate reductase subunit M2	K10808	x	x
3	nucleoside-diphosphate kinase	K00940	x	
4				
5	dCTP deaminase	K01494		
6	dUTP pyrophosphatase	K01520	x	x
7	dihydrofolate reductase / thymidylate synthase	K13998		x
8	thymidylate kinase	K00943	x	x
9	nucleoside-diphosphate kinase	K00940	x	
Uridine Biosynthesis				

1	aspartate carbamoyltransferase		x	
1	carbamoyl-phosphate synthase	K11541		x
1	dihydroorotase	K01465	x	x
2	dihydroorotate dehydrogenase	K00254	x	x
3	orotate phosphoribosyltransferase	K00762	x	
4	orotidine-5'-phosphate decarboxylase	K01591	x	x
Alanine, pyruvate => alanine (L)				
1	alanine dehydrogenase	K00259	x	
Arginine	N/A			
Asparagine, pyruvate => asparagine				
1	pyruvate carboxylase	K01958	x	
2	aspartate aminotransferase, cytoplasmic	K14454	x	x
2	aspartate aminotransferase, mitochondrial	K14455	x	x
3	asparagine synthase	K01953	x	x
Aspartic Acid, pyruvate => aspartate				
1	pyruvate carboxylase	K01958	x	
2	aspartate aminotransferase, cytoplasmic	K14454	x	x
2	aspartate aminotransferase, mitochondrial	K14455	x	x
Cysteine biosynthesis, homocysteine + serine => cysteine				
1	cystathionine beta-synthase	K01697	x	
2	cystathionine gamma-lyase	K01758		
Glutamic Acid				
1	pyruvate carboxylase	K01958	x	
2	citrate synthase	K01647	x	x
3	aconitate hydratase	K01681	x	x
4	isocitrate dehydrogenase	K00031	x	x
5	aspartate aminotransferase, cytoplasmic	K14454	x	x
Glutamine, pyruvate => glutamine				

1	pyruvate carboxylase	K01958	x	x
2	citrate synthase	K01647	x	x
3	aconitate hydratase	K01681	x	x
4	isocitrate dehydrogenase	K00031	x	x
5	aspartate aminotransferase, cytoplasmic	K14454	x	x
6	glutamine synthetase	K01915	x	x
Glycine, glyoxylate => glycine				
1	alanine-glyoxylate transaminase / serine-glyoxylate transaminase / serine-pyruvate transaminase	K00830		x
Glycine, 3P-D-Glycerate => glycine				
1	D-3-phosphoglycerate dehydrogenase / 2-oxoglutarate reductase	K00058	x	x
2	phosphoserine aminotransferase	K00831	x	
3	phosphoserine phosphatase	K01079	x	x
4	glycine hydroxymethyltransferase	K00600		
Histidine	N/A			
Isoleucine	N/A			
Leucine	N/A			
Lysine	N/A			
Methionine biosynthesis, apartate => homoserine => methionine				
1	bifunctional aspartokinase / homoserine dehydrogenase 1	K12524	x	x
2	aspartate-semialdehyde dehydrogenase	K00133	x	x
3	bifunctional aspartokinase / homoserine dehydrogenase 1	K12524	x	x
4				
5	cystathionine gamma-synthase	K01739		
6				
7	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	K00549	x	x
Phenylalanine, phosphoenol-pyruvate =>				

phenylalanine				
1	3-deoxy-7-phosphoheptulonate synthase	K03856		
2	3-dehydroquinate synthase	K01735		x
3	pentafunctional AROM polypeptide	K13830		x
4	pentafunctional AROM polypeptide	K13830		x
5				
6				
7	pentafunctional AROM polypeptide	K13830		x
8	pentafunctional AROM polypeptide	K13830		x
9	chorismate synthase	K01736	x	
10				
11				
12				
13	aspartate aminotransferase, cytoplasmic	K14454	x	x
14				
Proline	N/A			
Serine biosynthesis, glycerate-3P => serine				
1	D-3-phosphoglycerate dehydrogenase / 2-oxoglutarate reductase	K00058	x	x
2	phosphoserine aminotransferase	K00831	x	
3	phosphoserine phosphatase	K01079	x	x
Threonine biosynthesis, aspartate => homoserine => threonine				
1	bifunctional aspartokinase / homoserine dehydrogenase 1	K12524	x	x
2	aspartate-semialdehyde dehydrogenase	K00133	x	x
3	bifunctional aspartokinase / homoserine dehydrogenase 1	K12524	x	x
4				
5	threonine synthase	K01733	x	x
Tryptophan	N/A			
Tyrosine, phenylalanine =>				

tyrosine				
1	phenylalanine-4-hydroxylase, putative	K00500	x	x
Valine	no genes present			

Figure 1: System overview of *Cardiosporidium cionae* and *Nephromyces* showing tunicate host (a and e), area of localization (b and f), filamentous life stage (c and g), oocyst life stage (d and h), and vertically transferred fluorescent *in situ* hybridization (FISH) labeled bacterial endosymbionts within the oocysts (e and i). Scale bars are approximations due to resizing of images. FISH was carried out according to the method in Paight et al. 2020.

Figure 2: Overview of the 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.

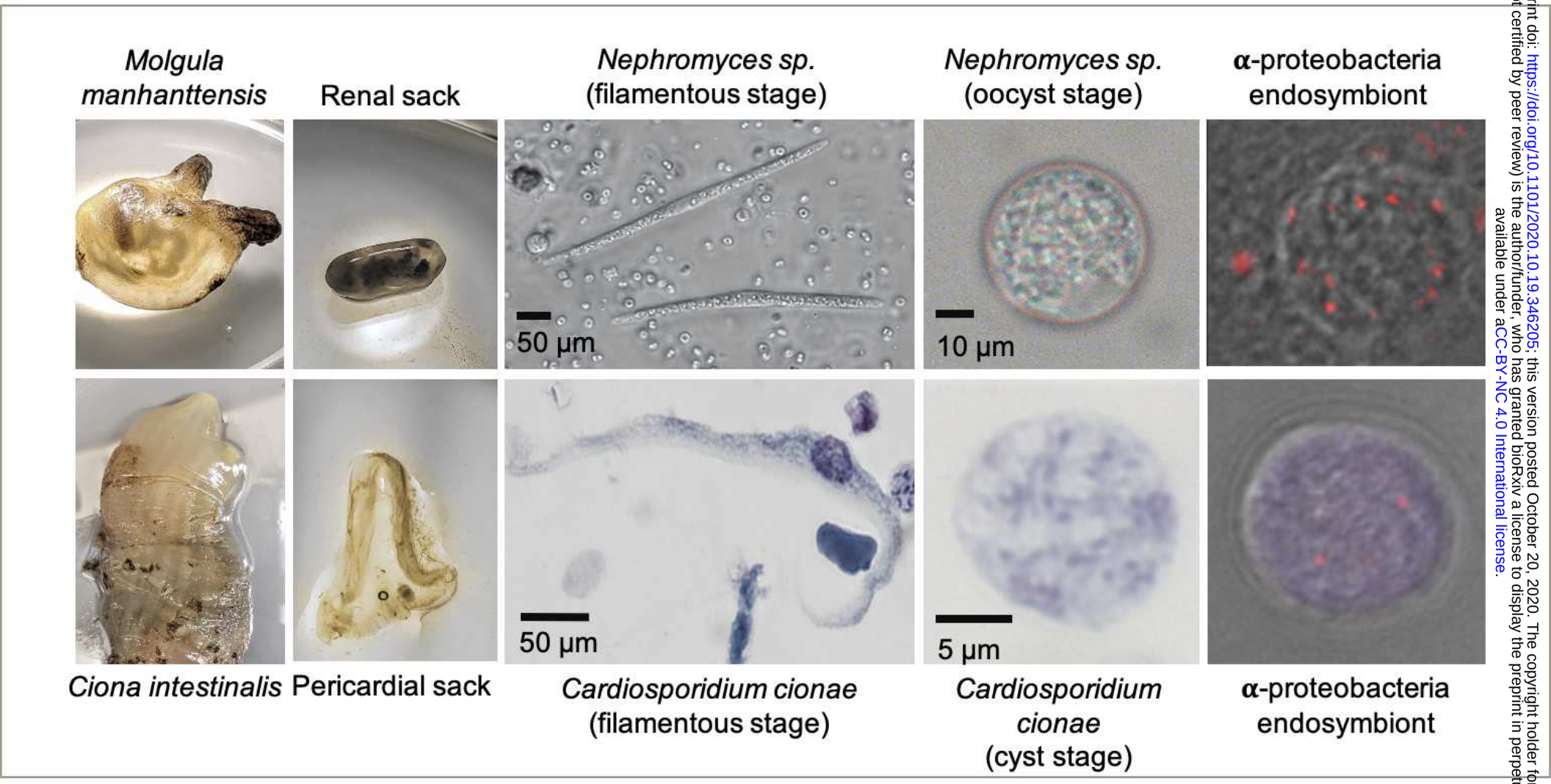
Figure 3: Apicoplast phylogeny created using a modified dataset provided by Muñoz-Gómez et al. 2019, showing the monophyly of *C. cionae* and *Nephromyces*. The complete, circularized *C. cionae* apicoplast recovered from the genomic dataset, and *Nephromyces* apicoplast sp. 654, are shown in blue. Statistical support was estimated using aLRT (1000) and ultrafast bootstrap (1000) (Nguyen et al. 2015) and these values are shown in this order on the nodes.

Figure 4: Overview of size, contig distribution, coding density, and annotations of major functional categories of genes in the **α**-proteobacteria endosymbiont genome.

Figure 5: Alphaproteobacteria phylogeny created with GToTree pipeline (117 concatenated genes) including all sequenced alphaproteobacteria published on NCBI and both the *C. cionae* and *Nephromyces* **α**-endosymbionts (in red). Bootstrap support is shown as a decimal value on the nodes.

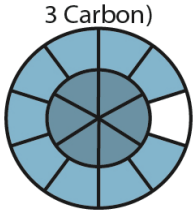
Supplementary Figure 1: The annotated, circularized *Cardiosporidium cionae* apicoplast. Two *C. cionae* apicoplasts were recovered that were 99.55% similar overall and contained identical gene organization, with the SNPs localized to the sufB gene. The *C. cionae* apicoplasts are similar in size, organization, 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.

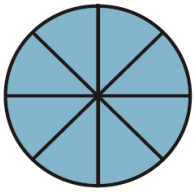


Carbohydrate Metabolism

Glycolysis (6 Carbon & 3 Carbon)

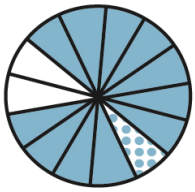


Gluconeogenesis

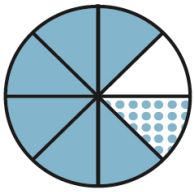


Pyruvate Oxidation

Citric Acid Cycle



Pentose Phosphate Pathway

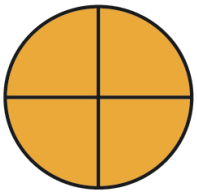


Glyoxylate Cycle

Lipid Metabolism

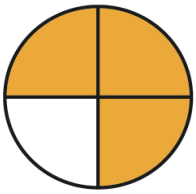
Fatty Acid Biosynthesis (Initiation)

Fatty Acid Biosynthesis (Elongation)



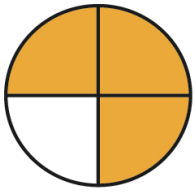
Cholesterol Biosynthesis

Triaglycerol Biosynthesis



Ketone Body Biosynthesis

Beta Oxidation



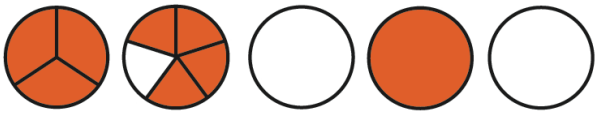
Amino Acid Metabolism

Ala Arg Asn Asp Cys

Glu Gln Gly His Ile

Leu **Lys** Met Phe Pro

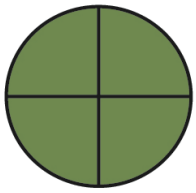
Ser Thr Trp Tyr Val



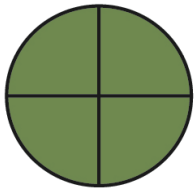
Nucleotide Metabolism

IMP Biosynthesis

Guanine Biosynthesis



Adenine Biosynthesis



Pyrimidine Biosynthesis (deoxyribonucleotide & **ribonucleotide**)

Uridine Biosynthesis