11-30-2018

Nephromyces Encodes a Urate Metabolism Pathway and Predicted Peroxisomes, Demonstrating These Are Not Ancient Losses of Apicomplexans

Christopher Paight
University of Rhode Island

Claudio H. Slamovits

Mary Beth Saffo
University of Rhode Island

Christopher E. Lane
University of Rhode Island, clane@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/bio_facpubs

Nephromyces Encodes a Urate Metabolism Pathway and Predicted Peroxisomes, Demonstrating These Are Not Ancient Losses of Apicomplexans

Creative Commons License

Nephromyces encodes a urate metabolism pathway and predicted peroxisomes, demonstrating these are not ancient losses of apicomplexans

Christopher Paight[1], Claudio H. Slamovits[2], Mary Beth Saffo[1,3] & Christopher E Lane[1]*

[1] Department of Biological Sciences, University of Rhode Island, Kingston RI, 02881, USA.

[2] Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada

[3] Smithsonian National Museum of Natural History, Washington, DC 20560, USA.

Abstract

The Phylum Apicomplexa is a quintessentially parasitic lineage, whose members infect a broad range of animals. One exception to this may be the apicomplexan genus *Nephromyces*, which has been described as having a mutualistic relationship with its host. Here we analyze transcriptome data from *Nephromyces* and its parasitic sister taxon, *Cardiosporidium*, revealing an ancestral purine degradation pathway thought to have been lost early in apicomplexan evolution. The predicted localization of many of the purine degradation enzymes to peroxisomes, and the *in silico* identification of a full set of peroxisome proteins, indicates that loss of both features in other apicomplexans occurred multiple times. The degradation of purines is thought to play a key role in the unusual relationship between *Nephromyces* and its host. Transcriptome data confirm previous biochemical results of a functional pathway for the utilization of uric acid as a primary nitrogen source for this unusual apicomplexan.

Key words: Apicomplexan, tunicates, Peroxisomes, Purine degradation, *Nephromyces*,

Cardiosporidium

Apicomplexans are most well known for being parasites of humans and livestock.

Species in the genus *Plasmodium*, for instance, are the etiological agents of malaria.

Apicomplexan species show tremendous variation in transmission methods, life cycles,

host range, host manipulation strategies, cell-types infected, metabolic capabilities,

immune evasion strategies, and virulence (Roos 2005; Reid et al. 2012; Kemp et al. 2013;

Cardoso et al. 2016). Because of this variability, there are few apicomplexan characteristics

shared throughout the phylum. Among the few universal apicomplexan features are a

parasitic life-history and an inability to degrade purines (Janouskovec & Keeling 2016).

Nephromyces, a derived apicomplexan genus of uncertain phylogenetic placement, appears

to be an exception to both of these traits.

Nephromyces was misclassified as a fungus for more than a 100 years, based on long

hyphal-like cell structures, flagellated spores interpreted by some as chytrid zoospores and

cell walls made of a chitin (Giard 1888). It was not until the application of molecular

methods that *Nephromyces* was confirmed as a member of the derived apicomplexans

(Saffo et al. 2010). Although some analyses have tentatively placed it sister to adeleid

coccidia, or piroplasms, the precise phylogenetic position of *Nephromyces* remains

unresolved (Saffo et al. 2010; Janouškovec et al. 2015). *Nephromyces* species are

monoxenous (infecting a single host) and are found exclusively in the Molgulidae family of

tunicates (Saffo & Davis 1982). In a phylum composed of obligate parasites, the feature that

distinguishes *Nephromyces* is its apparent mutualistic relationship with its tunicate hosts.

2

The mutualistic relationship has been inferred based primarily on the nearly 100% infection rate and lack of clearance from the host (Saffo 1978, 1988, 1990, Saffo et al. 2010). We use this label with caution, given how complex host-symbiont dynamics can be, how the costs and benefits of both "harmful" and "beneficial" symbioses can be difficult to determine, and how they can vary with genomic changes in hosts and symbionts (Leung & Poulin 2008; Saffo 2104; Mushegian & Ebert 2016).

A shift in lifestyle from obligate parasite to mutualistic symbiont is quite rare, and completely unknown from deep within a eukaryotic lineage with such a long evolutionary history of parasitism. One common consequence of a parasitic lifestyle is a loss of genes essential to free living organisms (Greganova et al. 2013; Janouškovec et al. 2015; Zarowiecki & Berriman 2015; Petersen et al. 2015). In an intracellular environment, if precursor molecules can be scavenged, there is less selective pressure to maintain biosynthesis pathways, and many are consequently lost (Keeling 2004; Sakharkar et al. 2004; Morrison et al. 2007). In phyla such as Apicomplexa, these losses can be extreme and over half of the genes found in their photosynthetic sister group, chromerids, have been lost in apicomplexans (Woo et al. 2015).

With so many basic metabolic functions lost, and with such dependence on the host, it is difficult to see how the relationship between host and parasite could change to a mutualistic interaction. However, one way for an organism to rapidly change its metabolic capabilities is to take on a bacterial symbiont. *Nephromyces* has done just that, leading to the hypothesis that bacterial endosymbionts inside *Nephromyces* perform some of the metabolic functions lost in Apicomplexa, and potentially contribute something beneficial to the tunicate host (Saffo 1990; Saffo et al. 2010). Bacterial endosymbionts are common

3

across the tree of life (although rare in apicomplexans) and perform a wide variety of functions for their hosts (Nowack & Melkonian 2010). These include amino acid metabolism and vitamin metabolism (Moran et al. 2005), nitrogen metabolism (Lopez-Sanchez et al. 2009), defense (De Souza et al. 2009), chemotrophic energy production (Urakawa et al. 2005), and photosynthesis (Marin et al. 2005), to name a few.

A tempting hypothesis for the functional role of *Nephromyces* bacterial endosymbionts is the break down of purines to urea in the purine degradation pathway (Saffo 1990). In support of this hypothesis *Nephromyces* infected tunicates have quite high levels of the enzyme urate oxidase, which catalyzes conversion of uric acid to 5-hydroxyisourate, but the enzyme is undetectable in uninfected tunicates (Mahler et al. 1955; Saffo 1988). Coupled with the fact that all known apicomplexans and tunicates have lost the purine degradation pathway, these data were suggestive of a bacterial contribution to purine degradation.

In a yet unexplained quirk of tunicate biology, many tunicate species have localized deposits of uric acid (Lambert et al. 1998; Saffo & Lowenstam 1978; Goodbody 1965). Storage as a form of excretion, nitrogen storage for future release, and structural support, are among the proposed functions of tunicate urate deposits (Goodbody 1965; Saffo 1988; Lambert et al. 1998). Tunicates in the Molgulidae family have the largest uric acid deposits, which are localized to a specialized, ductless structure, called a renal sac (Saffo & Lowenstam 1978). These uric acid deposits occur regardless of infection status, indicating a tunicate origin of these purine deposits. Despite the name, the renal sac has many features (most notably, the absence of any ducts or macroscopic openings) atypical for an excretory organ, and its biological function has yet to be determined. *Nephromyces* infects

4

feeding molgulid tunicates after the post-metamorphic onset of host feeding and completes its entire lifecycle within the renal sac. Four factors led to the conclusion that the bacterial endosymbionts within *Nephromyces* are the source of urate oxidase activity in this system: 1) the colonization of *Nephromyces* within a structure with high concentrations of urate, 2) the absence of urate oxidase activity in the molgulid hosts (Saffo, 1988, 1991), 3) the high urate oxidase activity found in *Nephromyces* (including its bacterial symbionts: Saffo, 1988, 1991), coupled with 4) the lack of obvious ultrastructural evidence of peroxisomes in *Nephromyces* (Saffo, 1990).

It is logical to think that the addition of bacterial endosymbionts to *Nephromyces* might have been key to colonizing this novel purine–rich niche, and is how *Nephromyces* escaped the "evolutionary dead end" of a parasitic lifestyle. In order to test this directly, and examine the metabolic relationships between the tunicate host, *Nephromyces*, and its bacterial endosymbionts, we sequenced the community transcriptome. To identify possible evolutionary or physiological changes involved in coevolution of *Nephromyces* with its molgulid hosts, we also sequenced the transcriptiome of a sister taxon of *Nephromyces*, *Cardiosporidium cionae* (Ciancio et al., 2008; Saffo et al., 2010), an apicomplexan parasite found in the blood in a broad range of non-molgulid ascidian hosts, including *Ciona intestinalis, Styela clava, Halocynthia roretzi,* and *Ascidiella aspersa* (Ciancio et al. 2008; Dong et al. 2006)*.* Interestingly, *Cardiosporidium cionae* also harbors bacterial endosymbionts, which allows for a more direct comparison between *Nephromyces* and *Cardiosporidium*.

Here we confirm the exceptionally high levels of urate oxidase activity in tunicates with *Nephromyces,* and extend this result to include high expression levels of all the genes

in the purine degradation pathway (xanthine dehydrogenase, urate oxidase, 5-hydroxyisourate hydrolase, 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline decarboxylase, and allantoinase). The breakdown of purines starts by conversion to xanthine. Xanthine then enters the ureide pathway and the enzyme xanthine dehydrogenase catalyzes the reaction of xanthine to urate (Xi et al. 2000; Nishino et al. 2008). Urate oxidase catalyzes the oxidation of uric acid to 5-hydroxyisourate. Following conversion the enzyme 5-hydroxyisourate hydrolase catalyzes 5-hydroxyisourate to 5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate (Kahn & Tipton 1998). This is further processed into (s)-allantoin by the enzyme 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline decarboxylase (Jung et al. 2006). Allantoinase catalyzes (s)-allantoin into allantoate. From this point there are a few different pathways with different endpoints that organisms are able to shuttle allantoate to (Cusa et al. 1999). A common end point is to process allantoate into urea and ureidoglycolate, to be further converted into carbon dioxide and ammonia. Alternatively, ureidoglycolate can be converted to glyoxylate, or the urea may be excreted as waste (Schultz et al. 2001; Werner et al. 2009).

We confirm that all the genes necessary for purine degradation are encoded by *Nephromyces* and *Cardiosporidium,* and not their endosymbiotic bacteria. Whereas the expression of urate oxidase by *Nephromyces* and *Cardiosporidium* is unexpected, a parallel issue is where the urate oxidase is physically located in the cell, given that apicomplexans reportedly lack peroxisomes (Schlüter et al. 2006). Urate oxidase activity is restricted to peroxisomes in eukaryotes, due to the numerous toxic byproducts that are produced in the break down of uric acid. Research into peroxisomes in Apicomplexa has a complex and contradictory history, with studies reporting both the presence (Kaasch & Joiner 2000;

6

Gabaldon et al. 2016) and absence (Ding et al. 2000; Schlüter et al. 2006; Gabaldon 2010) of peroxisomes in Apicomplexa. Recent work by Moog *et. al* (2017) and Ludewig-Klingner et al. (2018) demonstrates compelling support for peroxisomes in coccidians. Both studies present comprehensive bioinformatic (and also proteomic, in part) evidence for the presence of peroxisomal biogenesis factors (peroxins) and typical peroxisomal metabolic enzymes (including predicted relevant targeting signals) in coccidians (Moog et al. 2017; Ludewig-Klingner et al. 2018). However, neither paper provides explicit experimental evidence (for example, microscopic) for the formation of peroxisomes in these organisms. Although direct evidence is still absent, both studies point to (Lige et al. 2009)and their identification of peroxisome-like vesicles in *T. gondii*, for possible microscopic support.

Our data demonstrate that *Nephromyces* encodes a complete purine degradation pathway and a number of proteins predicted to be targeted to, or involved in, peroxisome biogenesis, maintenance and protein import, providing novel support of peroxisomes in Apicomplexa. Additionally, we propose the functional significance of purine degradation in *Nephromyces*, and reject the hypothesis that bacterial endosymbionts facilitated an escape from parasitism by providing genes in the purine degradation pathway.

Methods

<u>*Molgula manhattensis* collection and laboratory culture</u>

Molgula manhattensis tunicates were collected from a dock in Greenwich Bay, Rhode Island (41°39'22.7"N 71°26'53.9"W) on July 2014. For transcriptomic analysis, a single renal sac was separated from one tunicate, and all extraneous tissue removed. The intact renal sac was placed in liquid nitrogen for 5 min and then stored at -80°C for later RNA

7

extraction. Gonads were dissected from five, sexually mature, *M. manhattensis*, collected from the same population in Greenwich Bay, Rhode Island August 2014. Eggs and sperm were mixed with sterile seawater and divided evenly between two petri dishes. Plates were incubated at room temperature for two days with daily 100% water changes. Tunicate larvae attached to the bottom and sides of the petri dishes by day three. By day four, larvae had metamorphosed into adults and were actively feeding. Plates were moved to an incubator at 18° C with a 24 hr dark cycle to limit growth of contaminants. Tunicates were fed by 100% water exchange with cultures of *Isochrysis galbana* and *Chaetoceros gracilis* three days a week. After several weeks tunicates were moved to aerated beakers to meet their increased nutrient and gas exchange requirements. Feeding regimen remained the same except that food volume was increased with tunicate growth. Tunicates were grown for six months until they were ~10mm across. Each renal sac was placed into a 1.5ml Eppendorf tube and flash frozen in liquid nitrogen. PCR screens confirmed *Nephromyces* was absent from lab-raised individuals. Lab grown tunicates were split into two groups. Renal sacs were harvested from three tunicates to use as transcriptome controls. A second group was infected with *Nephromyces* oocysts to limit co-infections from multiple species, and raised for genomic analysis.

Cardiosporidium cionae collection, isolation and concentration

Ciona intestinalis were collected from docks in Snug Harbor RI (41.3890° N, 71.5201° W), in August 2017. Tunics were removed and the body wall was opened to allow access to the heart. A sterile syringe was used to remove cardiac blood as cleanly as possible. Blood was kept at 4° C until *Cardiosporidium* infection was verified using Giemsa stain to visualize

8

Cardiosporidium. Heavily infected samples were pooled together and centrifuged at 500 x g for 5 minutes. The resulting supernatant was removed and the samples were frozen in liquid nitrogen and stored at -80° C. Samples with low rates of infection were enriched for *Cardiosporidium* using sucrose gradients (Ogedengbe et al. 2015; Arrowood and Sterling 2016). Gradients of 20, 25, 30, 35, 40% sucrose solutions in phosphate buffer were layered together. Approximately 5 ml of tunicate blood was added to the column and centrifuged at 500 x g for 30 mins at 4° C. The 25% and 30% layers were collected (based on visual screens showing high *Cardiosporidium* cell density and low tunicate cell density), washed in PBS, pelleted and then frozen in liquid nitrogen and stored at -80° C.

RNA Extraction

RNA extraction buffer (Zymo Reasearh LLC. Irvine, CA) was added to samples and ground with a pestle. Following grinding, the Zymo Quick-RNA kit (Zymo Research LLC. Irvine, CA) was used and the manufacturer's protocol was followed. RNA was converted to cDNA and sequenced at the School of Medicine Genome Resource Center, University of Maryland. Five separate paired-end RNA libraries (two from infected renal sac, and three uninfected renal sac) were multiplexed on one lane of the Illumina HiSeq platform. Resulting in 326,299,923; 327,957,761 and 316,754,780 reads for the three renal sacs without *Nephromyces*, and 40,606,230 from the wild *M. manhattensis* renal sac. For *Cardiosporidium,* three samples of *C. intestinalis* blood were used: one with unseparated blood, one enriched with cells collected at the 25% sucrose gradient, and one enriched with cells from the 30% sucrose gradient were multiplexed on one lane of the Illumina HiSeq platform, resulting in 92,250,706; 109,023,104 and 110,243,954 respectively. Transcriptome data was

9

assembled and proteins were predicted with the Trinity/Trinotate pipeline version 2.4.0

run on the server at Brown University Center for Computation and Visualization (Haas et al.

2014). Reads assembled into 115,457; 388,535 109,446 contigs from infected, uninfected

samples, and *C. intestinalis* respectively. Protein sequences were predicted using

Transdecoder (Haas et al. 2014). Transcriptome completeness was assesed with Busco v3

against the Eukaryotic reference data sets (Simão et al. 2015).

DNA Extraction

The renal sacs from 8 lab grown *M. manhattensis* individuals were dissected and

their renal fluid was pooled in a 1.5ml Eppendorf tube. Contents were centrifuged at 8000 g

for 5 min. to pellet *Nephromyces* cells, and following centrifugation the renal fluid was

discarded. Five hundred microliters of CTAB buffer with 5ul of proteinase K and ceramic

beads were added to the pelleted *Nephromyces* cells. The sample was placed in a bead

beater for 3 min. and then on a rotator for 1.5hrs at room temp. Five hundred microliters of

chloroform was added, mixed gently and centrifuged for 5 min. The top layer was removed

and 2x the sample volume of ice cold 100% EtOH and 10% sample volume of 3M sodium

acetate were added to the sample and incubated a -20°C overnight. The sample was

centrifuged at 16000Xg for 30min. and the liquid was removed. Ice cold 70% EtOH was

added and centrifuged at 16000xg for 15min. Liquid was removed and sample air dried for

2 min. DNA was re-eluted in 50ul of deionized water.

Illumina Sequencing

A nanodrop (2000c, ThermoScientific) was used to assess DNA purity and DNA

concentration, and an agarose gel was run to assess genomic DNA fragmentation. Following

quality control, an Illumina library was constructed. Library prep and sequencing were

done at the URI Genomics and Sequencing Center (URIGSC). The completed library was sequenced on the Illumina Miseq platform at the URIGSC and the HiSeq platform at the University of Baltimore sequencing center on three lanes.

Pacific Biosciences Sequencing

Using the contents of 150 (done in batches of 10 then pooled) *M. manhattensis* renal sacs, the same DNA extraction protocol was performed as for Illumina sequencing. DNA was sequenced using three SMRT cells on the Pacific Biosciences platform at the University of Baltimore sequencing center.

Illumina sequence data assembly

One MiSeq lane and three lanes of HiSeq, all from the same library, were trimmed using Trimmomatic (Bolger et al. 2014) and then assembled using Spades assembler (Bankevich et al. 2012) on the URI server BlueWaves.

Pacific Biosciences sequence data assembly

Pacific Biosciences reads were error corrected using pbsuite/15.8.24 (English et al. 2012) on the Brown University server, Oscar. Reads were then assembled using Canu (Koren et al. 2014). Contigs generated by Canu were combined with Illumina MiSeq/HiSeq short reads with Abyss v2.02 (Jackman et al. 2017).

Sequence annotation

Genes in the urate pathway were identified initially using KEGG GhostKOALA and KASS and subsequently by BLASTP searches against NCBI's nr protein database (Kanehisa et al. 2016). All candidate genes were screened using InterProScan to predict function (Finn et al. 2017). A curated database of phylogenetically representative species with good

11

quality annotations for the three purine degradation genes and malate synthase were downloaded from NCBI. These genes were then used to construct gene trees.

Sequences were aligned with MAFFT (Katoh & Standley 2013) using FFT-NS-i. Maximum likelihood phylogenetic trees were constructed performed with RAxML (Stamatakis 2014) using the GAMMA model with 1000 seed trees and 1000 bootstrap replicates. Trees were viewed and modified using Figtree (v1.4.0, http://tree.bio.ed.ac.uk/software/figtree/).

Protein sequences were used to search against PeroxisomeDB (Schlüter et al. 2009) and BLAST hits lower than e^{-20} were retained and used in a BLASTP query against NCBI's Refseq protein database (Schlüter et al. 2009). Additional peroxisomal genes were identified with KAAS (Moriya et al. 2007). As many of these peroxisome genes are encoded by *M. manhattensis*, all copies that had a closest hit to opisthokonts or bacteria were removed. Transcripts from uninfected *M. manhattensis* were used to screen additional tunicate genes using cd-hit at a 90% identity level (Li & Godzik 2006). Remaining genes were tested for signal motifs and subcellular location predictions with Wolf PSORTII, Ppero, TargetP, topcons, and Predotar (Supplementary Table 1)(Nakaia & Horton 1999; Wang et al. 2017; Emanuelsson et al. 2007; Bernsel et al. 2009; Small et al. 2004).

Nephromyces specific RNAseq reads were mapped to our genomic assembly using bowtie2 (Langmead & Salzberg 2012) with the `-very sensitive` flag set. Following mapping, Bedtools (Quinlan & Hall 2010) was used to quantify coverage across contigs, which were separated based on coverage levels. Contigs identified as *Nephromyces* were annotated using Maker2 with *ab initio* gene predictions from Augustus (Holt & Yandell 2011; Stanke et al. 2004).

12

Results

The contents of a single renal sac from an individual *Molgula manhattensis* resulted in 195,694 transcripts from *M. manhattensis*, *Nephromyces*, and the bacterial endosymbionts. After binning by species, 60,223 transcripts were attributed to *Nephromyces*. The cardiac fluid from 40 infected *Ciona intestinalis* individuals resulted in 109,446 transcripts, including 15,541 *Cardiosporidium* transcripts. The BUSCO algorithm was used to assess the completeness of the transcriptomes and reported 81.8% complete transcripts and 6.3% partial for the *Nephromyces* data and 69.7% complete and 11.9% partial for *Cardiosporidium*.

The *Nephromyces* genome assembly consists of 1176 contigs greater than 5kb with a maximum length of 287,191 bp and an average length of 36 kb (Paight *et al*, in prep). This dataset was used to search for purine degradation genes to determine their genomic context. All of the purine degradation genes, as well as malate synthase, were predicted and annotated in the genome by Maker2. All genes but URAD contained introns, and neighboring genes on the identified contigs had top BLAST hits to apicomplexans in all cases (Table 1), indicating that they are encoded in the *Nephromyces* genome, not the endosymbiotic bacteria or host *Molgula manhattensis*. Phylogenetic trees for xanthine dehydrogenase, uric oxidase, malate synthase and allantoicase consistently resolve the monophyly of *Nephromyces, Cardiosporidium,* and Chromerids (Figure 1). Chromerids are the photosynthetic and the closest free-living relatives of Apicomplexa (Moore et al. 2008), indicating a vertical inheritance of this pathway from the common ancestor of apicomplexans.

13

The presence of urate oxidase also provides further support for peroxisomes in some lineages of Apicomplexa (Moog et al. 2017; Ludewig-Klingner et al. 2018), because urate oxidase activity is confined to peroxisomes in eukaryotes (Usuda et al. 1994). In addition to urate oxidase, *Nephromyces* and *Cardiosporidium* encode more peroxisome-associated proteins than *Plasmodium,* and nearly the same complement of genes encoded by *Toxoplasma* (Table 2)*.* There are a few notable differences between *Toxoplasma* and *Nephromyces/Cardiosporidium*, including the absence of PEX3, PEX16, VLACS, and SCPX in *Nephromyces/Cardiosporidium* and the absence of PMP22, GSTK1, DHRS4, XDH, and UO in *Toxoplasma.* Additionally, *Nephromyces* encodes a copy of Malate synthase (MLS) absent in both *Cardiosporidium* and *Toxoplasma*. Malate synthase is a key gene in the glyoxylate cycle, a pathway maintained in the photosynthetic *Chromera velia* and *Vitrella brassicaformis*, but lost in all other apicomplexans (Ludewig-Klingner et al. 2018). *Nephromyces/Cardiosporidium* also encode the enzyme serine-pyruvate transaminase (AGXT), which also uses glyoxylate as a substrate. AGXT converts glyoxylate into glycine and pyruvate and is often localized to peroxisomes, however the localization of AGXT in *Nephromyces/Cardiosporidium* is unclear (Supplementary Table 1).

Discussion

The recent scrutiny by Moog et al (2017), and Ludewig-Klingner et al. (2018) has built a case for the presence of peroxisomes in some apicomplexan lineages. While some apicomplexans may have lost peroxisomes, it seems likely that it is not a universally shared trait in the phylum. Despite the extensive search for peroxisome-associated functions in apicomplexans, no genes involved in purine degradation were found in other sequenced

14

apicomplexan genomes, with the lone exception of allantoicase in *Plasmodium* (Gardner et al. 2002). Our *in silico* predictions indicate a complete purine degradation pathway in *Nephromyces* and *Cardiosporidium.* In addition to highly expressed transcripts for the genes involved, all of the identified purine degradation genes and MLS have been located on genomic contigs from *Nephromyces.* Based on neighboring genes and the presence of introns in the *Nephromyces* genes matching the expressed transcripts, these contigs almost certainly originate from the *Nephromyces* genome (Table 1)*.* Additionally, none of the purine degradation transcripts attributed to *Nephromyces* were detected in uninfected tunicates (Table 3). Phylogenetic trees of purine degradation genes are poorly supported at an inter-phylum level, indicating a rapid evolutionary rate. Whereas most genes are phylogenetically uninformative across the spectrum of eukaryotes, these gene trees have strong support for monophyly of purine degradation genes from *Nephromyces* and *Cardiosporidium* with Chromerids (Figure 1). The combination of gene trees, expression only when *Nephromyces* is present, and preliminary genomic assemblies strongly suggest that these genes were present since the divergence of Apicomplexa and Chromerida and have been vertically transmitted. Thus, these genes have been subsequently lost across apicomplexans, possibly multiple times. Although the exact placement of *Nephromyces* and *Cardiosporidium* is not certain (Saffo et al. 2010), multi-gene phylogenies place them in the subclass Hematozoa (Muñoz et al. in prep), suggesting that purine degradation was independently lost multiple times in Apicomplexa as well as maintained long after apicomplexans had become obligate parasites.

The presence of predicted purine degradation genes in *Nephromyces* and *Cardiosporidium,* adds a function not previously demonstrated in apicomplexan

15

peroxisomes (Table 2; Moog et al 2017; Ludewig-Klingner et al. 2018). While *Toxoplasma* and *Cardiosporidium/Nephromyces* share many of the same peroxisomal marker genes, no copy of PEX3 has been found in *Cardiosporidium/Nephromyces.* PEX3 (along with PEX10, PEX12, and PEX19) is one of the 4 genes reportedly required for peroxisome function (Schluter et al. 2006). However, the fundamentals of peroxisome biology have been described from a limited set of eukaryotes, and organisms such as ciliates have peroxisomes but also lack PEX3 (Ludewig-Klingner et al. 2017). Therefore, PEX3 may not be critical to peroxisome function for alveolates, and possibly other under-studied eukaryotic lineages. Extreme sequence conservation of PEX3 and PEX19 is only found in opisthokonts and sequence divergence in other lineages may indicate alternative functions (Hua et al. 2015).

Two other genes (Sterol carrier protein 2, SCPX and Solute carrier family 27, member 2, VLACS) missing from *Cardiosporidium/Nephromyces,* but found in *Toxoplasma*, are involved in β-fatty acid oxidation. Both *Cardiosporidium/Nephromyces* encode the seven other β-fatty acid oxidation genes encoded in *Toxoplasma*, suggesting β-fatty acid oxidation forms part of the functional capabilities of the *Cardiosporidium/Nephromyces* peroxisome. Fatty acid oxidation is often a central component of peroxisome function and has been hypothesized to be the impetus for the evolution of peroxisomes (Speijer 2011).

Based on transcript abundance, purine degradation in *Nephromyces* peroxisomes appears to be heavily utilized. Only 0.13% of genes had a higher transcription rate than urate oxidase in our data from wild collected *Nephromyces,* and the other genes in the purine degradation pathway are among the most highly expressed transcripts in both wild and lab grown *Nephromyces* samples (Table 3). This result aligns with the previously

16

reported high levels of urate oxidase protein in the renal sac of infected *Molgula* (Saffo 1988), indicating that the expression levels reported here do translate to protein. Much of this pathway is expressed over the 99th percentile of all transcripts in *Nephromyces,* which corresponds to the top 100 genes. Expression of purine degradation genes in *Cardiosporidium* is far lower, and in the 70-90 percentile range (Table 3). Such high expression in *Nephromyces* represents an enormous metabolic investment and it is unlikely that these transcripts go largely untranslated.

Both *Nephromyces* and *Molgula manhattensis* encode xanthine dehydrogenase, and are able to convert xanthine into uric acid. Since we have identified the tunicate host as the source of purines, this raises the question of why *Nephromyces* is expressing xanthine dehydrogenase in the 97.87th percentile, compared with similarly high tunicate expression (93.64th percentile). Although the percentile ranking between these two organisms cannot be directly compared, such high xanthine dehydrogenase expression in *Nephromyces* is surprising. It seems unlikely that so much xanthine dehydrogenase production is needed to convert only endogenous purines of *Nephromyces*. However, xanthine is only detected in the renal sac in small quantities, not nearly as abundant as uric acid, and xanthine dehydrogenase activity is restricted to the renal wall, not the renal lumen (Nolfi 1970). One possible explanation is that *Nephromyces* exports its xanthine dehydrogenase into the renal wall in order to drive the production of xanthine from hypoxanthine before the purine salvage enzymes adenine phosphoribosyltransferase and hypoxanthine-guanine phosphoribosyltransferase can salvage hypoxanthine into adenine and guanine.

High expression of purine degradation genes in *Nephromyces* is clear, but the purpose is uncertain. It does indicate purine degradation is an important pathway for

17

Nephromyces, however*,* the functional significance is not immediately obvious. Pathway

analysis predicts that *Nephromyces* is able to convert xanthine into urea and

ureidoglycolate, however neither compound is biologically useful without further

conversion. We propose that the products of purine degradation in *Nephromyces* are

converted to glyoxylate.

One possible route is the conversion of ureidoglycolate into glyoxylate. There are two

known enzymes able to catalyze this conversion: ureidoglycolate lyase, found in fungi and

bacteria, which catalyzes (s)-ureidoglycolate to glyoxylate and urea, and ureidoglycolate

amidohydrolase, found in plants and bacteria, which catalyzes (s)-ureidoglycolate to

glyoxylate, carbon dioxide, and ammonia (Muñoz et al. 2006; Percudani et al. 2013; Wells &

Lees 1991; Werner et al. 2010; Shin et al. 2012; Serventi et al. 2010). Both ureidoglycolate

lyase and ureidoglycolate amidohydrolase are amidohydrolases - hydrolases that use

amide bonds as substrates. No orthologs to either ureidoglycolate lyase and

ureidoglycolate amidohydrolase have been found in the *Nephromyces* transcriptome.

However, an amidohydrolase is present, which is predicted to be structurally similar to the

ureidoglycolate amidohydrolase found in *Arabidopsis*, including similar location and

number of zinc binding domains*.* This amidohydrolase also has a similarly high expression

level as the other purine degradation enzymes (Table 3). In order to determine if the

amidohydrolase found in *Nephromyces* is capable of catalyzing (s)-ureidoglycolate,

functional assays will need to be performed.

While the functionality of this particular amidohydrolase has yet to be determined,

its ability to act on an (s)-ureidoglycolate is an attractive hypothesis for a few reasons. One,

there are two known enzymes capable of breaking the amide bond in (s)-ureidoglycolate

18

that have independently evolved: ureidoglycolate lyase and ureidoglycolate amidohydrolase. This pathway has not been widely explored across eukaryotes and the modification to a class of molecules able to break amide bonds to accommodate the structure of (s)-ureidoglycolate may not be a complex evolutionary step. Two, (s)-ureidoglycolate is unstable and will spontaneously convert to glyoxylate, albeit without the stereospecific conversion present when catalyzed by ureidoglycolate amidohydrolase (Werner et al. 2010). Spontaneous conversion of glyoxylate results in a 50% loss of efficiency *versus* enzymatic conversion, presumably creating strong evolutionary pressure to enzymatically degrade (s)-ureidoglycolate to maintain stereochemistry.

Glyoxylate is a common substrate for a number of enzymes including glyoxylate oxidase, which catalyzes glyoxylate with water and oxygen to form oxalate and hydrogen peroxide (Kasai et al. 1963). Notably, no copy of glyoxylate oxidase has been identified in *Nephromyces*, which is surprising given that another common component of the renal sac is calcium oxalate (Saffo & Lowenstam 1978). We have not identified any genes suggesting that *Nephromyces* or its bacterial endosymbionts can produce or process oxalate, oxalate is also found in uninfected hosts indicating that the source of the calcium oxalate is likely the tunicate. Another enzyme that uses glyoxylate as a substrate, which is present in *Nephromyces/Cardiosporidium*, is serine-pyruvate transaminase (AGXT), which can be localized to peroxisomes or mitochondria, and catalyzes glyoxylate to glycine and pyruvate (Takada & Noguchi 1985). An alternative enzyme for processing glyoxylate is malate synthase (MLS), which is also targeted to the peroxisome and missing from apicomplexans, including *Cardiosporidium*, but is found in *Nephromyces* (Figure 1).

19

Malate synthase is one of two genes integral to the glyoxylate cycle, an alternative pathway for part of the citrate cycle. In the glyoxylate cycle, isocitrate is converted into glyoxylate and succinate by isocitrate lyase (McFadden & Howes 1965). Glyoxylate is combined with acetyl-CoA to create malate (Molina et al. 1994). This cycle allows for the creation of glucose from fatty acids directly (Kornberg & Krebs 1957). The presence of malate synthase indicates at least a piece of the glyoxylate cycle is present in *Nephromyces*. No copy of isocitrate lyase is predicted from the *Nephromyces* transcriptome, and only a small fragment of a possible isocitrate synthase has been identified in *Cardiosporidium*. However, under the model proposed here, the generation of glyoxylate is from uric acid, and isocitrate synthase would not be required.

Both AGXT and MLS (in *Nephromyces*) show similarly high expression as the purine degradation genes (Table 3), which is consistent with our proposed uric acid to glyoxylate pathway. In particular, AGXT is among the most highly expressed *Nephromyces* transcripts, with consistently higher expression than MLS, possibly indicating it is the primary route of glyoxylate conversion. The products of AGXT, glycine and pyruvate, are versatile substrates and used by a number of pathways. Glycine is the simplest amino acid and an essential component of many important biological compounds, as a nitrogen source in a readily useable form. Pyruvate is extremely versatile and involved in several critical biological pathways. A non-inclusive list includes amino acid biosynthesis, acetal-CoA biosynthesis, fatty acid biosynthesis, and the citric acid cycle. These pathways represent both carbon and energy acquisition (Figure 2). Additionally, *Nephromyces* has the ability to use MLS to convert glyoxylate and acetal-CoA into malate, a compound central to the citric acid cycle, allowing for another mechanism of carbon and energy acquisition (Figure 2).

The hypothesized conversion of uric acid to glyoxylate in *Nephromyces* creates several possibilities. First, it allows for the metabolic waste product, uric acid, to be converted into glycine, pyruvate, and malate (Figure 2). Second, it provides an explanation for the exceptionally high expression of the purine degradation pathway. Third, it gives *Nephromyces* access to a primary carbon, nitrogen, and an energy source at no cost to its host. And finally, this change in primary carbon, nitrogen, and energy could conceivably reduce the impact of *Nephromyces* on its host, allowing *Nephromyces* densities to increase while decreasing virulence. Reduction in virulence would have been a necessary first step toward mutualism.

Uric acid as a primary carbon and energy source is not completely unknown. Bacterial species have been found in chicken hutches that were able to grow solely on uric acid (Rouf & Lomprey 1968; Thong-On et al. 2012), and some species of fungi are able to grow on media solely containing uric acid (Middelhoven et al. 1989). However, this is a novel substrate for an apicomplexan to grow on, and while it is unlikely that *Nephromyces* could survive on uric acid alone, it is a promising base for both carbon and nitrogen acquisition. It is possible that the *Nephromyces* bacterial endosymbionts (Sabree et al. 2009; Potrikus & Breznak 1980) are contributing to the proposed purine to glucose pathway, but that is not currently supported by our data.

As the adaptive significance of uric acid deposits in tunicates, and particularly in *Molgula*, are unknown, it is difficult to speculate on the effects of *Nephromyces* uric acid degradation to the host. If these renal sac deposits are a form of excretion by storage, as has been hypothesized (Goodbody 1965), then having a symbiont that is capable of digesting uric acid may be beneficial simply by digesting an indigestible metabolite and converting

21

uric acid into urea. Alternatively, once the uric acid has been broken down, the tunicate may benefit from metabolites derived from uric acid previously unavailable to the tunicate. If *Nephromyces* is overexpressing xanthine dehydrogenase in order to outcompete adenine phosphoribosyltransferase and hypoxanthine-guanine phosphoribosyltransferase, diverting hypoxanthine from purine salvage to purine degradation, there could be a potential cost to the host under purine-limited conditions.

Our data demonstrate that both the proposed mutualistic *Nephromyces* and parasitic *Cardiosporidium* encode the genes for purine degradation, which have been lost in other apicomplexans, sequenced to date. Additionally, these genes share a common ancestry with chromerid genes, indicating they are not the product of a recent horizontal gene transfer from bacteria. These data also add support to the growing body of evidence that indicate the presence of peroxisomes in apicomplexans. *Nephromyces* and *Cardiosporidium* are predicted to have peroxisomes and, unlike any other apicomplexan, are capable of preforming both purine degradation and part of the glyoxylate cycle. The presence of purine degradation, AGXT, and MLS allow for the intriguing possibility of conversion of uric acid into a primary nitrogen, carbon and energy source. This predicted metabolic activity would be a completely novel substrate for an apicomplexan and may have been an important factor in the reduction of virulence in *Nephromyces*.

Acknowledgements

Arrowood MJ, Sterling CR. 2016. Isolation of *Cryptosporidium* Oocysts and Sporozoites Using Discontinuous Sucrose and Isopycnic Percoll Gradients Author (s): Michael J . Arrowood and Charles R . Sterling Published by : Allen Press on behalf of The American Society of Parasitologists St. J. Parasitol. 73:314–319.

Bankevich A 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.

Bernsel A, Viklund H, Hennerdal A, Elofsson A. 2009. TOPCONS: Consensus prediction of membrane protein topology. Nucleic Acids Res. 37:465–468. doi: 10.1093/nar/gkp363.

Bolger A, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114.

Cardoso R, Soares H, Hemphill A, Leitão A. 2016. Apicomplexans pulling the strings: manipulation of the host cell cytoskeleton dynamics. Parasitology. 1–14. doi: 10.1017/S0031182016000524.

Ciancio A et al. 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.

Cusa E et al. 1999. Genetic Analysis of a Chromosomal Region Containing Genes Required for Assimilation of Allantoin Nitrogen and Linked Glyoxylate Metabolism in *Escherichia coli* Genetic Analysis of a Chromosomal Region Containing Genes Required for Assimilation of Allantoin. 181:7479–7484.

Ding M, Clayton C, Soldati D. 2000. *Toxoplasma gondii* catalase: are there peroxisomes in toxoplasma? J. Cell Sci. 113 (Pt 1:2409–2419.

Dong LC et al. 2006. First report on histology and ultrastructure of an intrahemocytic paramyxean parasite (IPP) from tunicate *Halocynthia roretzi* in Korea. Dis. Aquat. Organ. 72:65–69. doi: 10.3354/dao072065.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2:953–971. doi: 10.1038/nprot.2007.131.

English AC et al. 2012. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. PLoS One. 7:1–12. doi: 10.1371/journal.pone.0047768.

Finn RD et al. 2017. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 45:D190–D199. doi: 10.1093/nar/gkw1107.

Gabaldon T. 2010. Peroxisome diversity and evolution. Philos. Trans. R. Soc. B Biol. Sci. 365:765–773. doi: 10.1098/rstb.2009.0240.

Gabaldon T, Ginger ML, Michels PAM. 2016. Peroxisomes in parasitic protists. Mol. Biochem. Parasitol. 209:33–45. doi: 10.1016/j.molbiopara.2016.02.005.

Gardner MJ et al. 2002. Europe PMC Funders Group Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 419:3–9. doi: 10.1038/nature01097.Genome.

Giard MA. 1888. On *Nephromyces*, a new genus of Fungi parasitic in the kidney of the Molgulidæ. Ann. Mag. Nat. Hist. Ser. 6. 1:386–388. doi: 10.1080/00222938809460753.

Goodbody BYI. 1965. Nitrogen Excretion in Ascidiacea. Enzyme. 34:299–305.

Greganova E, Steinmann M, Mäser P, Fankhauser N. 2013. In silico ionomics segregates parasitic from free-living eukaryotes. Genome Biol. Evol. 5:1902–1909. doi:

10.1093/gbe/evt134.

Haas BJ et al. 2014. *reference generation and analysis with Trinity*. doi: 10.1038/nprot.2013.084.De.

Holt C, 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.

Hua R, Gidda SK, Aranovich A, Mullen RT, Kim PK. 2015. Multiple Domains in PEX16 Mediate Its Trafficking and Recruitment of Peroxisomal Proteins to the ER. Traffic. 16:832–852. doi: 10.1111/tra.12292.

Jackman SD et al. 2017. ABySS 2 . 0 : resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 27:768–777. doi: 10.1101/gr.214346.116.Freely.

Janouškovec J et al. 2015. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. Proc. Natl. Acad. Sci. U. S. A. 112:10200–7. doi: 10.1073/pnas.1423790112.

Janouskovec J, Keeling PJ. 2016. Evolution: Causality and the origin of parasitism. Curr. Biol. 26:R174–R177. doi: 10.1016/j.cub.2015.12.057.

Jung D-K et al. 2006. Structural and functional analysis of PucM, a hydrolase in the ureide pathway and a member of the transthyretin-related protein family. Proc. Natl. Acad. Sci. 103:9790–9795. doi: 10.1073/pnas.0600523103.

Kaasch AJ, Joiner KA. 2000. Targeting and Subcellular Localization of. J. Biol. Chem. 275:1112–1118.

Kahn K, Tipton PA. 1998. Spectroscopic characterization of intermediates in the urate

oxidase reaction. Biochemistry. 37:11651–11659. doi: 10.1021/bi980446g.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference

resource for gene and protein annotation. Nucleic Acids Res. 44:D457–D462. doi:

10.1093/nar/gkv1070.

Kasai T, Suzuki I, Asai T. 1963. Glyoxylate Oxidation in Acetobacter with Refenence to the

formation of Oxalic Acid. Applyed Microbiol. 9:49–58.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1152716&tool=pmcentrez&r

endertype=abstract.

Katoh K, Standley DM. 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 PJ. 2004. Reduction and compaction in the genome of the apicomplexan parasite

Cryptosporidium parvum. Dev. Cell. 6:614–616. doi: 10.1016/S1534-5807(04)00135-2.

Kemp LE, Yamamoto M, Soldati-Favre D. 2013. Subversion of host cellular functions by the

apicomplexan parasites. FEMS Microbiol. Rev. 37:607–631. doi: 10.1111/1574-6976.12013.

Koren S et al. 2014. Canu: scalable and accurate long-read assembly via adaptive k-mer

weighting and repeat separation. Genome Res. 1–11. doi: 10.1101/gr.215087.116.Freely.

Kornberg HL, Krebs HA. 1957. Synthesis of cell constituents from C2-units by a modified

tricarboxylic acid cycle. Nature. 179:988–991. doi: 10.1038/179988a0.

Lambert CC, Lambert G, Crundwell G, Kantardjieff 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, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357. http://dx.doi.org/10.1038/nmeth.1923.

Leung TLF, Poulin R. 2008. Parasitism, commensalism, and mutualism: Exploring the many shades of symbioses. Vie Milieu. 58:107–115.

Li W, Godzik A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22:1658–1659. doi: 10.1093/bioinformatics/btl158.

Lige B, Jayabalasingham B, Zhang H, Pypaert M, Coppens I. 2009. 'Role of an Ancestral D-Bifunctional Protein Containing' 'Two Sterol-Carrier Protein-2 Domains in Lipid Uptake and Trafficking in *Toxoplasma*'. Mol. Biol. Cell. 20:658–672.

Lopez-Sanchez MJ et al. 2009. Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. PLoS Genet. 5. doi: 10.1371/journal.pgen.1000721.

Ludewig-Klingner A-K, Michael V, Jarek M, Brinkmann H, Petersen J. 2018. Distribution and Evolution of Peroxisomes in Alveolates (Apicomplexa, Dinoflagellates, Ciliates). Genome Biol. Evol. 10:1–13. doi: 10.1093/gbe/evx250.

Mahler BYHR, The W, Assistamze T, Germille OF. 1955. Studies On Uricase. J. Biol. Chem.

Marin B, Nowack ECM, Melkonian M. 2005. A plastid in the making: Evidence for a second primary endosymbiosis. Protist. 156:425–432. doi: 10.1016/j.protis.2005.09.001.

27

McFadden B, Howes W. 1965. Crystallization and Some Properties of Isocitrate Lyase from *Pseudomonas indigofera**. J. Biochem. 58:116–122.

Middelhoven WJ, De Hoog GS, Notermans S. 1989. Carbon assimilation and extracellular antigens of some yeast-like fungi. Antonie Van Leeuwenhoek. 55:165–175. doi: 10.1007/BF00404756.

Molina I, Pellicer M, Badia J, Aguilar J, Baldoma L. 1994. Molecular Characterization of Escherichia coli Malate Synthase G: Differentiation with the Malate Synthase A Isoenzyme. Eur. J. Biochem. 224:541–548. doi: 10.1111/j.1432-1033.1994.00541.x.

Moog D, Przyborski JM, Maier UG. 2017. Genomic and proteomic evidence for the presence of a peroxisome in the apicomplexan parasite *Toxoplasma gondii* and other Coccidia. Genome Biol. Evol. 1–39. doi: 10.1093/gbe/evx231/4596564.

Moore RB et al. 2008. A photosynthetic alveolate closely related to apicomplexan parasites. Nature. 452:900–900. doi: 10.1038/nature06871.

Moran NA, Dunbar HE, Wilcox JL. 2005. Regulation of transcription in a reduced bacterial genome: Nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. J. Bacteriol. 187:4229–4237. doi: 10.1128/JB.187.12.4229-4237.2005.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:182–185. doi: 10.1093/nar/gkm321.

Morrison HG et al. 2007. Genomic Minimalism in the Early Diverging Intestinal Parasite *Giardia lamblia*. Science (80-.). 317:1921–1926. doi: 10.1126/science.1143837.

Muñoz A, Raso MJ, Pineda M, Piedras P. 2006. Degradation of ureidoglycolate in French bean (Phaseolus vulgaris) is catalysed by a ubiquitous ureidoglycolate urea-lyase. Planta. 224:175–184. doi: 10.1007/s00425-005-0186-8.

Mushegian AA, Ebert D. 2016. Rethinking 'mutualism' in diverse host-symbiont communities. BioEssays. 38:100–108. doi: 10.1002/bies.201500074.

Nakaia K, Horton P. 1999. PSORT : a program for detecting sorting signals in proteins and predicting their subcellular localization. Comput. CORNER. 0004:1336–1337.

Nishino T, Okamoto K, Eger BT, Pai EF, Nishino T. 2008. Mammalian xanthine oxidoreductase - Mechanism of transition from xanthine dehydrogenase to xanthine oxidase. FEBS J. 275:3278–3289. doi: 10.1111/j.1742-4658.2008.06489.x.

Nolfi JR. 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 ECM, Melkonian M. 2010. Endosymbiotic associations within protists. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 365:699–712. doi: 10.1098/rstb.2009.0188.

Ogedengbe ME, Qvarnstrom Y, da Silva AJ, Arrowood MJ, Barta JR. 2015. A linear mitochondrial genome of *Cyclospora cayetanensis* (Eimeriidae, Eucoccidiorida, Coccidiasina, Apicomplexa) suggests the ancestral start position within mitochondrial genomes of eimeriid coccidia. Int. J. Parasitol. 45:361–365. doi: 10.1016/j.ijpara.2015.02.006.

Percudani R, Carnevali D, Puggioni V. 2013. Ureidoglycolate hydrolase, amidohydrolase, lyase: How errors in biological databases are incorporated in scientific papers and vice versa. Database. 2013:1–9. doi: 10.1093/database/bat071.

29

Petersen G, Cuenca A, Møller IM, Seberg O. 2015. Massive gene loss in mistletoe (*Viscum*, *Viscaceae*) mitochondria. Sci. Rep. 5:1–7. doi: 10.1038/srep17588.

Potrikus CJ, Breznak JA. 1980. Uric acid-degrading bacteria in guts of termites (*Reticulitermes flavipes* (Kollar)). Appl. Environ. Microbiol. 40:117–124. doi: 16345587.

Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841–842. doi: 10.1093/bioinformatics/btq033.

Reid AJ et al. 2012. Comparative genomics of the apicomplexan parasites *Toxoplasma gondii* and *Neospora caninum*: Coccidia differing in host range and transmission strategy. PLoS Pathog. 8. doi: 10.1371/journal.ppat.1002567.

Roos DS. 2005. Themes and Variations in Apicomplexan Parasite Biology. Science (80-.). 309:72–73. doi: 10.1126/science.1115252.

Rouf MA, Lomprey RF. 1968. Degradation of uric acid by certain aerobic bacteria. J. Bacteriol. 96:617–622.

Sabree ZL, Kambhampati S, Moran N a. 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. Proc. Natl. Acad. Sci. U. S. A. 106:19521–19526. doi: 10.1073/pnas.0907504106.

Saffo M. 1990. Symbiosis within a symbiosis: Intracellular bacteria within the endosymbiotic protist *Nephromyces.* Mar. Biol. 107:291–296. doi: 10.1007/BF01319828.

Saffo M, McCoy A, Rieken C, Slamovits C. 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 MB. 1988. Nitrogen Waste or Nitrogen Source? Urate Degradation in the Renal Sac of Molgulid Tunicates. Biol. Bull. 175:403. doi: 10.2307/1541732.

Saffo MB, Davis WL. 1982. Modes of Infection of the Ascidian *Molgula manhattensis* by Its Endosymbiont *Nephromyces giard*. Biol. Bull. 162:105. doi: 10.2307/1540974.

Saffo MB, Lowenstam H a. 1978. Calcareous deposits in the renal sac of a molgulid tunicate. Science. 200:1166–1168. doi: 10.1126/science.200.4346.1166.

Sakharkar KR, Kumar Dhar P, Chow VVTK. 2004. Genome reduction in prokaryotic obligatory intracellular parasites of humans: A comparative analysis. Int. J. Syst. Evol. Microbiol. 54:1937–1941. doi: 10.1099/ijs.0.63090-0.

Schluter A et al. 2006. The evolutionary origin of peroxisomes: An ER-peroxisome connection. Mol. Biol. Evol. 23:838–845. doi: 10.1093/molbev/msj103.

Schlüter A, Real-Chicharro A, Gabaldón T, Sánchez-Jiménez F, Pujol A. 2009. PeroxisomeDB 2.0: An integrative view of the global peroxisomal metabolome. Nucleic Acids Res. 38:800–805. doi: 10.1093/nar/gkp935.

Schultz AC, Nygaard P, Saxild HH. 2001. Functional analysis of 14 genes that constitute the purine catabolic pathway in *Bacillus subtilis* and evidence for a novel regulon controlled by the PucR transcription activator. J. Bacteriol. 183:3293–3302. doi: 10.1128/JB.183.11.3293-3302.2001.

Serventi F et al. 2010. Chemical basis of nitrogen recovery through the ureide pathway: Formation and hydrolysis of S-ureidoglycine in plants and bacteria. ACS Chem. Biol. 5:203–214. doi: 10.1021/cb900248n.

31

Shin I, Percudani R, Rhee S. 2012. Structural and functional insights into (S)-ureidoglycine

aminohydrolase, keyenzyme of purine catabolism in *Arabidopsis thaliana*. J. Biol. Chem.

287:18796–18805. doi: 10.1074/jbc.M111.331819.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO:

Assessing genome assembly and annotation completeness with single-copy orthologs.

Bioinformatics. 31:3210–3212. doi: 10.1093/bioinformatics/btv351.

Small I, Peeters N, Legeai F, Lurin C. 2004. Predotar: A tool for rapidly screening proteomes

for N-terminal targeting sequences. Proteomics. 4:1581–1590. doi:

10.1002/pmic.200300776.

De Souza DJ, Bézier A, Depoix D, Drezen JM, Lenoir A. 2009. *Blochmannia* endosymbionts

improve colony growth and immune defence in the ant *Camponotus fellah*. BMC Microbiol.

9:1–8. doi: 10.1186/1471-2180-9-29.

Speijer D. 2011. Oxygen radicals shaping evolution: Why fatty acid catabolism leads to

peroxisomes while neurons do without it: FADH2/NADH flux ratios determining

mitochondrial radical formation were crucial for the eukaryotic invention of peroxisomes

and catabolic tissue. BioEssays. 33:88–94. doi: 10.1002/bies.201000097.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of

large phylogenies. Bioinformatics. 30:1312–1313. doi: 10.1093/bioinformatics/btu033.

Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: A web server for gene

finding in eukaryotes. Nucleic Acids Res. 32:309–312. doi: 10.1093/nar/gkh379.

Takada Y, Noguchi T. 1985. Characteristics of alanine: glyoxylate aminotransferase from

Saccharomyces cerevisiae, a regulatory enzyme in the glyoxylate pathway of glycine and

serine biosynthesis from tricarboxylic acid-cycle intermediates. Biochem. J. 231:157–63. doi: 10.1042/bj2310157.

Thong-On A et al. 2012. Isolation and Characterization of Anaerobic Bacteria for Symbiotic Recycling of Uric Acid Nitrogen in the Gut of Various Termites. Microbes Environ. 27:186–192. doi: 10.1264/jsme2.ME11325.

Urakawa H et al. 2005. Hydrothermal vent gastropods from the same family (Provannidae) harbour E- and y -proteobacterial endosymbionts. Environ. Microbiol. 7:750–754. doi: 10.1111/j.1462-2920.2005.00753.x.

Usuda N et al. 1994. Uric acid degrading enzymes, urate oxidase and allantoinase, are associated with different subcellular organelles in frog liver and kidney. J. Cell Sci. 107 (Pt 4:1073–81. http://www.ncbi.nlm.nih.gov/pubmed/8056832.

Wang J, Wang Y, Gao C, Jiang L, Guo D. 2017. PPero, a computational model for plant PTS1 type peroxisomal protein prediction. PLoS One. 12:1–12. doi: 10.1371/journal.pone.0168912.

Wells XE, Lees EM. 1991. Ureidoglycolate amidohydrolase from developing french bean fruits (*Phaseolus vulgaris* [L.].). Arch. Biochem. Biophys. 287:151–159. doi: 10.1016/0003-9861(91)90400-D.

Werner AK, Romeis T, Witte CP. 2010. Ureide catabolism in *Arabidopsis* thaliana and *Escherichia coli*. Nat. Chem. Biol. 6:19–21. doi: 10.1038/nchembio.265.

Woo YH et al. 2015. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. Elife. 4:1–41. doi: 10.7554/eLife.06974.

33

Xi H, Schneider BL, Reitzer L. 2000. Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. J. Bacteriol. 182:5332–5341. doi: 10.1128/JB.182.19.5332-5341.2000.

Zarowiecki M, Berriman M. 2015. What helminth genomes have taught us about parasite evolution. Parasitology. 142 Suppl:S85-97. doi: 10.1017/S0031182014001449.

Table 1. Genomic context of the annotated purine degradation genes and malate synthase, in the *Nephromyces* genomic assembly. The phylogenetic affiliation of neighboring genes on each contig was identified by top hit against the NCBI nr database using BLASTp. Every contig encoding a target gene included other apicomplexan genes, and genes that did not hit apicomplexans had no strong affinity for other organisms.

Gene	Introns in gene	Contig	Contig length (kb)	Predicted Genes on Contig	Genes with top apicomplexan BLAST hits
XDH	4	Neph_3686418	24.5	4	2
UO	7	Neph_3687015	82.5	12	7
uraH	2	Neph_3685393	94.3	6	4
URAD	0	Neph_3687674	30.9	6	5
ALLC	10	Neph_3687655	116.3	16	11
MS	6	Neph_3671841	7	2	1

Table 2. Peroxiomal genes identified in *Nephromyces* and *Cardiosporidium* and their functional catagory. (X) denotes presence of gene and (-) absence. *Vitrella brassicaformis* (Vb) *Chromera velia* (Cv) *Cardiosporidium (C) Nephromyces (N) Toxoplasma gondii (Tg) Plasmodium falciparum (Pf) Cryptosporidium parvum (Cp).* Table modified based on Ludwig-Klinger et al. (2017).

	Protein	Abbr.	Vb	Cv	C	N	Tg	Pf	Cp
Glyoxylate pathway									
	Isocitrate lyase	ICL	x	x	-	-	-	-	-
	Malate synthase	MLS	x	x	-	x	-	-	-
	Citrate synthase	CS	x	x	x	x	x	x	-
	Aconitase	ACO	x	x	x	x	x	x	-
	Malate dehydrogenase	MDH	x	x	x	x	x	x	x
Peroxisome									
	Biogenesis factor 1	Pex1	x	x	-	x	x	-	-
	Biogenesis factor 2	Pex2	x	x	x	x	x	-	-
	Biogenesis factor 3	Pex3	x	x	-	-	x	-	-
	Ubiquitin carrier protein	Pex4	x	x	x	x	x	x	x
	Biogenesis protein 5	Pex5	x	x	x	x	x	-	-
	Biogenesis protein 6	Pex6	x	x	x	x	x	-	-
	Biogenesis protein 7	Pex7	x	x	x	x	x	-	-
	Biogenesis protein 10	Pex10	x	x	x	x	x	-	-
	Biogenesis factor 11	Pex11	x	x	x	x	x	-	-
	Biogenesis protein 12	Pex12	x	x	-	x	x	-	-
	Biogenesis factor 13	Pex13	-	-	-	-	-	-	-
	Membrane protein 14	Pex14	x	x	x	x	x	-	-
	Membrane protein 15	Pex15	-	-	-	-	-	-	-
	Biogenesis factor 16	Pex16	x	x	-	-	x	-	-
	Membrane protein receptor	Pex19	x	-	-	-	-	-	-
	Biogenesis protein 22	Pex22	x	x	x	x	x	x	x
	Biogenesis factor 26	Pex26	-	-	-	-	-	-	-
	Membrane channel	PMP22	x	x	x	x	-	-	-
	Membrane protein 4	PMP27	-	-	-	-	-	-	-
	ATP/ADP-transporter	PMP34	x	x	x	x	x	-	-
	Fatty acid ABC-transporter	PMP70	x	x	x	x	x	-	-
	ROS metabolism	MPV17	x	x	x	x	x	x	-
	Protein	Abbr.	Vb	Cv	C	N	Tg	Pf	Cp
Fatty acid oxidation									
a-oxidation	2-Hydroxyacyl-CoA lyase	HPCL2	-	x	-	-	-	x	-
	Phytanoyl-CoA hydrolase	PHYH	x	x	-	-	-	-	-
B-oxidation	a-Methylacyl-CoA-racemase	AMACR	-	-	-	-	-	-	-
	Acyl-CoA-oxidase	ACOX	x	x	x	x	x	-	-
	Multifunctional protein	DBP	x	x	x	x	x	-	-
	Sterole carrier protein 2	SCPX	-	x	-	-	x	-	-

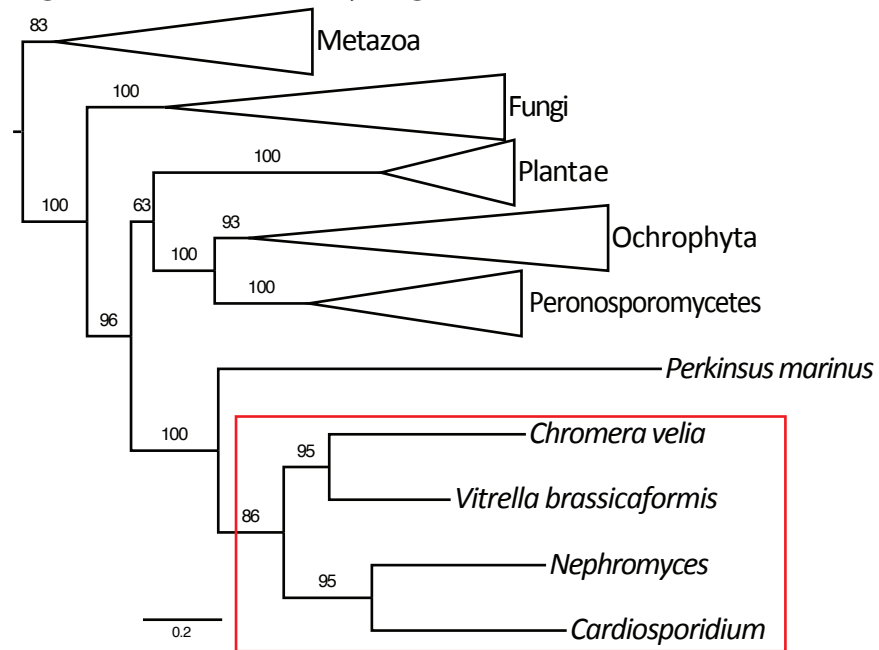|			Vb	Cv	C	N	Tg	Pf	Cp
	Multifunctional protein	PBE	-	-	-	-	-	-	-
	Acetyl-CoA acyltransferase 1	ACAA1	x	x	x	x	x	-	-
	2, 4-dienoyl-CoA reductase	PDCR	x	x	x	x	x	-	-
	d(3, 5)-d(2, 4)-dienoyl-CoA isomerase	ECH	x	x	x	x	x	-	-
	ATP-binding cassette, subfamily D	ABCD	x	x	x	x	x	-	-
	Long-chain acyl-CoA synthetase	ACSL	x	x	x	x	x	x	-
	Solute carrier family 27, member 2	VLACS	x	x	-	-	x	-	-
Other oxidation	Acyl-CoA thioesterase 8	PTE	-	-	-	-	-	-	-
	Nucleoside disphosphate-linked m.	NUDT19	-	x					
Amino acid metabolism							-	-
	Multifunctional protein	AGT	x	x	x	x	x	-	-
	D-Amino-acid oxidase	DAO	-	-	-	-	-	-	-
	Isocitrate dehydrogenase	IDH	x	x	x	x	x	x	-
	N1-acetylpolyamine oxidase	PAOX	x	x	-	-	-	-	-
	L-Pipecolate oxidase	PIPOX	x	x	-	-	-	-	-
	hydroxymethylgluatryl-CoA lyase	HMGCL	x	x	x	x	x	-	-
	(S)-2-hydroxy-acid oxidase	HAO	x	x	x	x	-	-	-
Antioxidant system							-	-
Hydrogen peroxide metabolism	Catalase	CAT	x	-	x	x	x	-	-
	Superoxide dismutase	SOD	-	-	x	x	x	x	x
	Nitric-oxide synthase, inducible	INOS	-	-	-	-	-	-	-
	Peroxiredoxin 1	PRDX1	x	x	-	x	x	x	x
	Peroxiredoxin 5	PRDX5	-	-	-	-	-	-	-
Glutathione metabolism								
	Glutathione S-transferase kappa 1	GSTK1	x	x	x	x	-	-	-
	Protein	Abbr.	Vb	Cv	C	N	Tg	Pf	Cp
Etherphospholipid biosynthesis								
	Dihydroxyacetone phosphate acyltr.	DHAPAT	x	x	x	x	x	-	-
	Alkyldihydroxyacetonephosphate syn	AGPS	x	x	x	-	-	-	-
	Fatty acyl-CoA reductase	FAR	x	x	-	x	x	-	-
Purine metabolism								
	Xanthine dehydrogenase	XDH	x	x	x	x	-	-	-
	Urate oxidase	UO	x	x	x	x	-	-	-
Retinol metabolism								
	Dehydrogenase/reductase SDR family	DHRS4	x	x	x	x	-	-	-
Sterol precursor biosynthesis								
	Mevalonate kinase	MVK	-	-	-	-	-	-	-
	Phosphomevalonate kinase	PMVK	-	-	-	-	-	-	-

Table 3. Expression percentile ranking of purine degredation genes, from total expressed transcripts in *Nephromyces (*Neph*),*
*Cardiosporidium (*Cardio*)* and *Molgula (*Mm*)*. The wild *Nephromyces* and *Molgula manhattensis* data originate from the same RNA
extraction and were bioinformatically separated. Data was also generated from labrotory grown tunicates, artificially infected with
Nephromyces (Lab grown Neph 1 & 2). *Cardiosporidium* fractions represent 1) unfiltered pericadial fluid, 2) the 25% and 3) 30% fractions
extracted from a sucrose gradient, and may contain different proportions of *Cardiosporidium* life stages. The three uninfected *Molgula*
manhattensis were raised from gametes in the lab and never exposed to *Nephromyces* infection. The (-) denotes the transcript was not
recovered in that dataset whereas (N/A) indicates the transcript was assembled, but the transcripts per million (TPM) was <1.

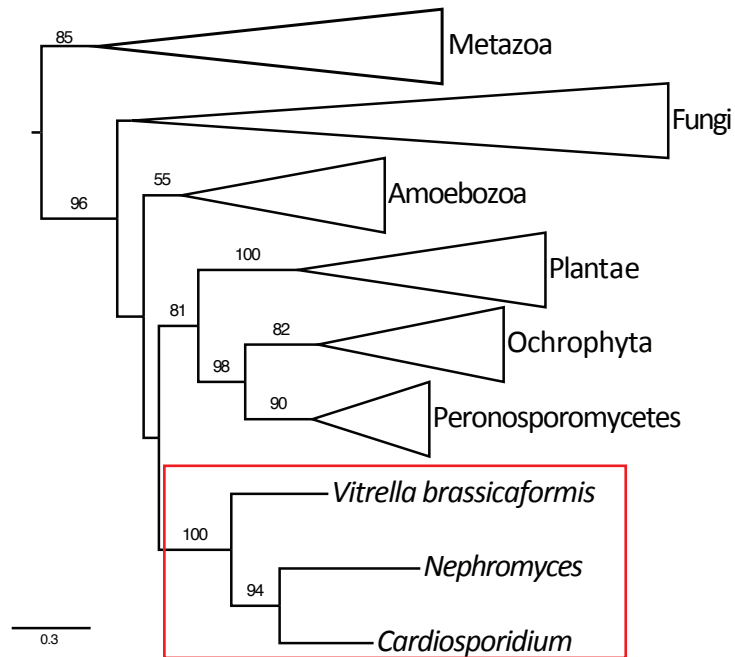| Gene	Wild Neph	Lab grown Neph 1	Lab grown Neph 2	Cardio Fraction 1	Cardio Fraction 2	Cardio Fraction 3	Mm	Uninfected Mm 1	Uninfected Mm 2	Uninfected Mm 3
xanthine dehydrogenase	97.87	93.17	94.83	none	76.88	69.5	93.64	N/A	N/A	N/A
urate oxidase	99.87	99.44	99.54	86.75	87.24	70.98	-	-	-	-
5-hydroxyisourate hydrolase	99.16	91.31	88.41	87.67	83.27	79.1	-	-	-	-
OHCU decarboxylase	93.38	-	-	-	-	-	-	-	-	-
allantoinase	99.09	98.38	98.23	73.61	90.32	71.89	-	-	-	-
aminodohydrolase	99.75	79.25	89.18	87.43	92.27	92.08	-	-	-	-
malate synthase	59.17	93.81	93.11	-	-	-	-	-	-	-
serine-pyruvate transaminase	99.85	99.57	99.79	84.64	80.81	77.79	85.65	91.17	71.85	75.05

Figure 1. Maximum likelihood protein trees of A) Xanthine Dehydrogenase, B) Urate oxidase, C) Allantoicase D) Malate synthase. Genes A-C are involved in purine degradation and their position supports an ancestral apicomplexan purine degradation pathway in *Nephromyces/Cardiosporidium.* Malate synthase, D), acts on glyoxylate and acetyl-CoA to produce malate to complete the pathway. Stramenopiles are paraphyletic in the malate synthase phylogeny, possibly indicating a deep gene duplication. Whereas the support for deeper nodes is variable among all four genes, there is consistent support for a monophyletic origin of *Nephromyces/Cardiosporidium* genes with Chromerids (red box). Major lineages have been collapsed for presentation. Support values are percentage bootstrap support above 50%*.*

Figure 2. Predicted purine degradation pathway in *Nephromyces*, within the peroxisome and cytosol. Dark blue arrows represent enzymes identified in the *Nephromyces* transcriptome. The light blue arrow represents the highly expressed amidohydrolase (red box) predicted to convert ureidoglycolate into glyoxylate. Enzymes on the left side are localized to peroxisomes, the right side to the cytosol, with the green vertical line representing the peroxisomal membrane. The predicted pathway is able to convert uric acid into glyoxylate, and subsequent conversion by serine-pyruvate transaminase (AGXT) or malate synthase, creates glycine and pyruvate or malate respectively. The * by AGXT indicates ambiguous predicted localization, to either peroxisomes or mitochondria.
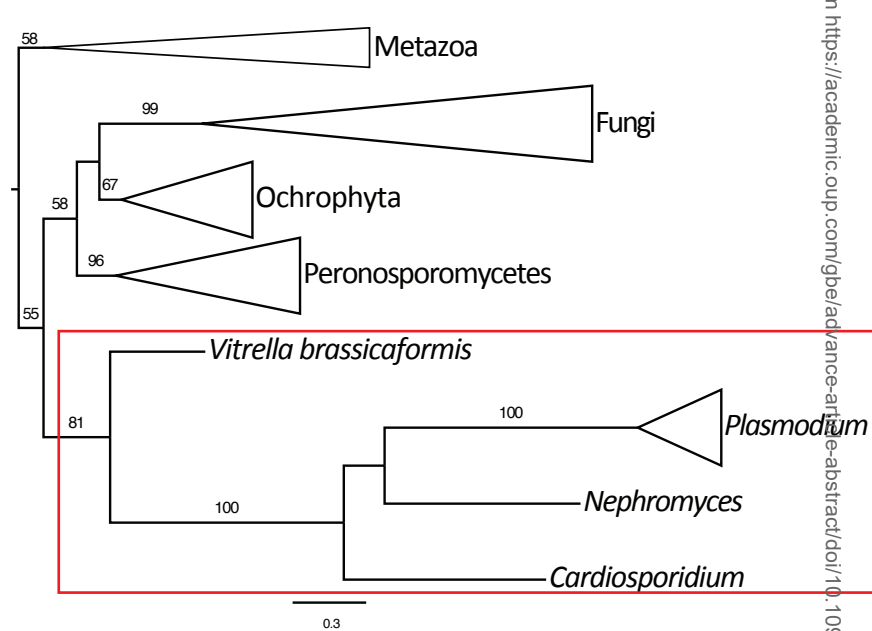
Figure 1 A) Xanthine Dehydrogenase XDH; B) Urate Oxidase UO; C) Allantoicase ALLC; D) Malate Synthase MLS
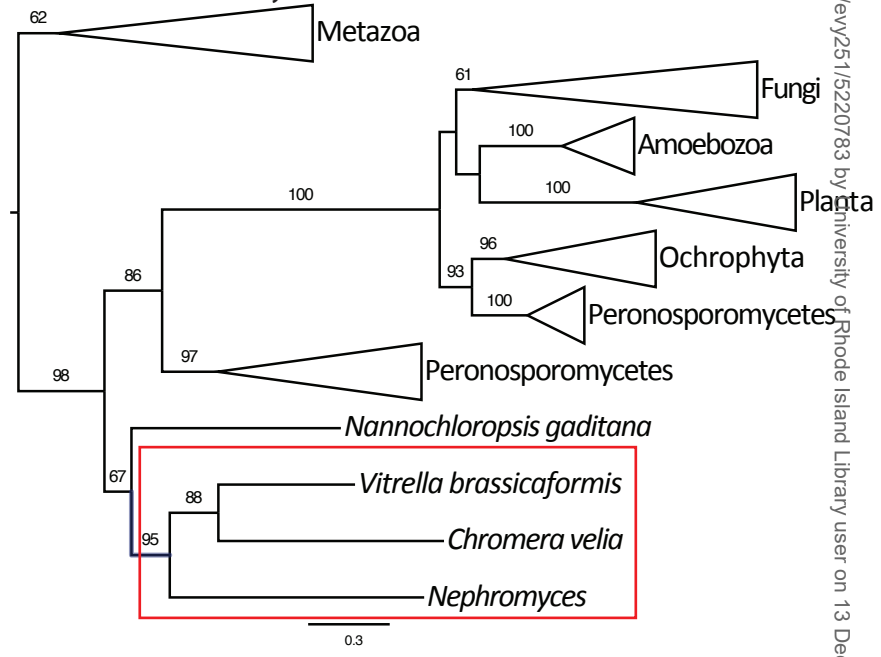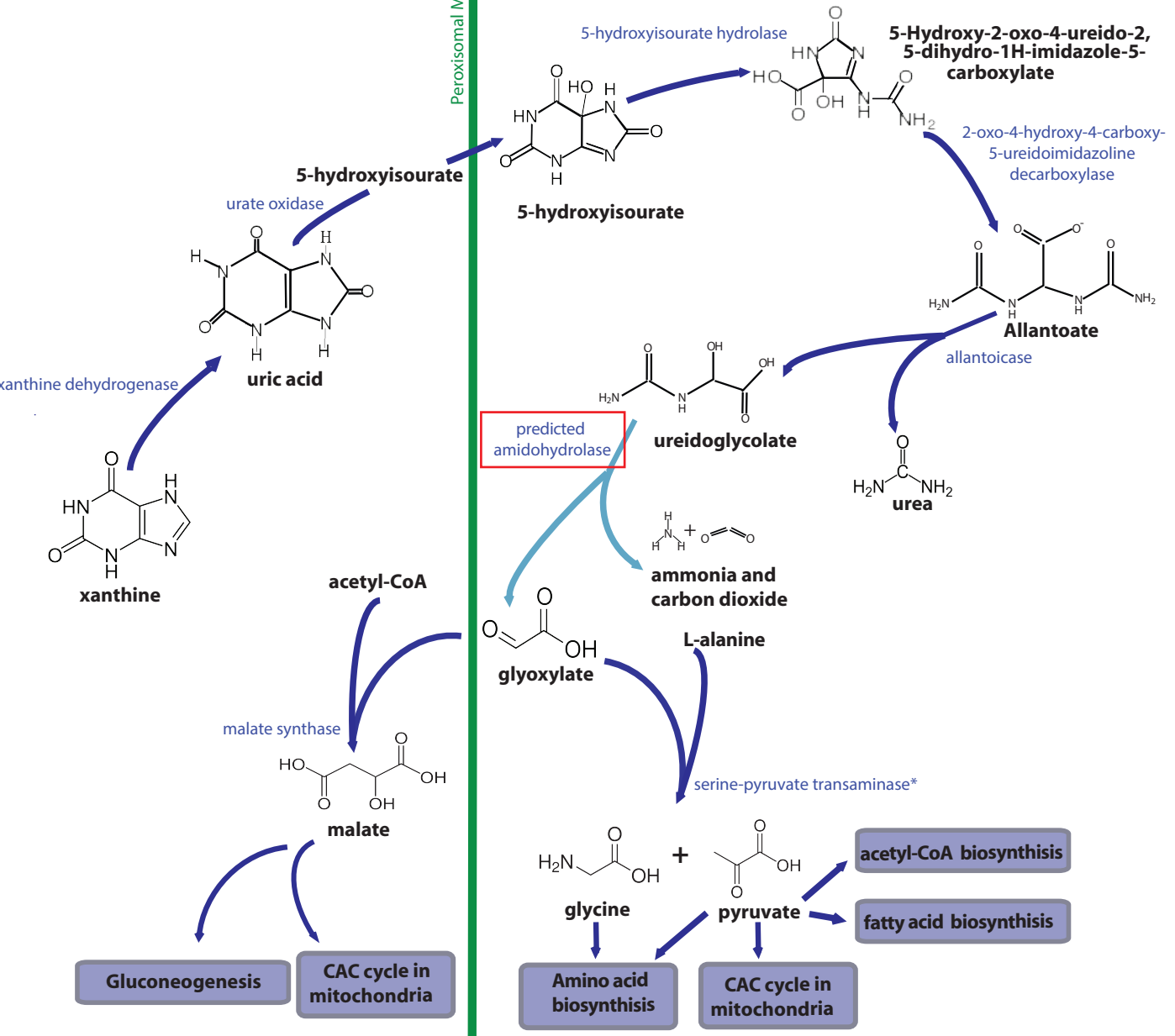
Figure 2