Phylogenomics reveals Adeleorina are an ancient and distinct subgroup of Apicomplexa

Ina Na a,*, Claudia Campos b, Gordon Lax a, Waldan K. Kwong a,b, Patrick J. Keeling a,*

a Department of Botany, University of British Columbia, Vancouver, BC, Canada
b Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal

ARTICLE INFO

Keywords:
Adeleorina
Apicomplexa
Phylogenomics
Adeleidae
Legerellidae
Transcriptome

ABSTRACT

Apicomplexans are a diverse phylum of unicellular eukaryotes that share obligate relationships with terrestrial and aquatic animal hosts. Many well-studied apicomplexans are responsible for several deadly zoonotic and human diseases, most notably malaria caused by Plasmodium. Interest in the evolutionary origin of apicomplexans has also spurred recent work on other more deeply-branching lineages, especially gregarines and sister groups like squirmids and chrompodellids. But a full picture of apicomplexan evolution is still lacking several lineages, and one major, diverse lineage that is notably absent is the adeleorinids. Adeleorina apicomplexans comprises hundreds of described species that infect invertebrate and vertebrate hosts across the globe. Although historically considered coccidians, phylogenetic trees based on limited data have shown conflicting branch positions for this subgroup, leaving this question unresolved. Phylogenomic trees and large-scale analyses comparing cellular functions and metabolism between major subgroups of apicomplexans have not incorporated Adeleorina because only a handful of molecular markers and a couple organellar genomes are available, ultimately excluding this group from contributing to our understanding of apicomplexan evolution and biology. To address this gap, we have generated complete genomes from mitochondria and plastids, as well as multiple deep-coverage single-cell transcriptomes of nuclear genes from two Adeleorina species, Klossia helicina and Legerella nova, and inferred a 206-protein phylogenomic tree of Apicomplexa. We observed distinct structures reported in species descriptions as remnant host structures surrounding adeleorinid oocysts. Klossia helicina and L. nova branched, as expected, with monoxenous adeleorinids within the Adeleorina and their mitochondrial and plastid genomes exhibited similarity to published organellar adeleorinid genomes. We show with a phylogenomic tree and subsequent phylogenetic analyses that Adeleorina are not closely related to any of the currently sampled apicomplexan subgroups, and instead fall as a sister to a large clade encompassing Coccidia, Protozoa, and Nephromycida, collectively. This resolves Adeleorina as a key independently-branching group, separate from coccidians, on the tree of Apicomplexa, which now has all known major lineages sampled.

1. Introduction

Apicomplexans are widespread obligate parasites and symbionts infecting diverse animal hosts around the world. Over 6,000 species have been described among the millions estimated to infect aquatic and terrestrial vertebrates and invertebrates (Votýpka et al., 2017; del Campo et al., 2019). As obligate parasites and symbionts, all apicomplexans must pass through at least one host to complete their life cycle, however their propensity to influence host development and health exists on a spectrum, from the highly virulent and deadly, such as the malaria-causing Plasmodium spp. (Sato, 2021), to less well-studied commensal and potentially mutualistic members such as Nephromyces (Saffo et al., 2010; Paight et al., 2019). Similar forms of obligate symbiosis with animal hosts have arisen at least three times in the apicomplexans and the closely-related squirmids and chrompodellids, all of which arose from a free-living photosynthetic ancestor. Within these groups, convergent evolution has been identified as a recurring theme, affecting their red-algae derived plastids, mitochondria, and apical structure used for feeding and infection (Mathur et al., 2019).

Major subgroups within the Apicomplexa are the Coccidia, Protozoa, Hematozoa, Nephromycida, Marosporida, Gregarininasina, and Cryptosporidium. Among these subgroups, the coccidians and...
hematozoans have been best-studied because they include model systems and human pathogens like *Toxoplasma* and *Plasmodium* (Kim and Weiss, 2004). More recently, there has been progress on the genomics of early-branching gregarines because they have provided insights into the origin and early evolution of apicomplexans (Mathur et al., 2019; Boisnard and Florent, 2020). This latter work was powered by significant advances in culture-free genomics and transcriptomics, which have enabled previously less well-studied major apicomplexan subgroups to be assessed with phylogenomic methods that generate more robust phylogeny by using hundreds of proteins instead of just one gene, like the SSU rDNA.

Culture-free genomics and phylogenomics have also recently been used to pin-point the phylogenetic position of other apicomplexans whose relationships have been historically enigmatic. In particular, the recent demonstration that several disparate groups that were not thought to be related actually collectively form a new class, Marosporida (Mathur et al., 2021). These advances have led to a very robust tree of genomes have been reported from a few adeleorinids (L. infect two or more hosts (Dactylosomatidae, Haemogregarinidae, Hep.)

The only other molecular data are from the organelles: an apicoplast protein phylogeny showed Hematozoa and Marosporidia forming a clade branching sister to Adeleorina (Pales Espinosa et al., 2019a). The group includes monoxenous lineages that infect one host (Adeleidae, Legerellidae, and Klossiellidae), known as the adelids, and a monophyletic heteroxenous lineage, known as the haemogregarines, where members infect two or more hosts (Dactylosomatidae, Haemogregarinidae, Hepatozoa, and Karyolysidae) (Barta, 2000; Votýpka et al., 2017; Barta et al., 2012). The families Adeleidae and Klossiellidae diverge near the root of this group in phylogenetic analyses but the position of Legerellidae is unresolved as this is the only known group within Adeleorina with no genetic data reported (Barta et al., 2012; Lévellé et al., 2019a). Indeed, molecular data for the group are scarce in general, in contrast to other long-established major apicomplexan clades. The majority of reported molecular data is restricted to SSU rDNA markers representing a small portion of all described species. Mitochondrial and apicoplast genomes have been reported from a few adelorinids (Lévellé et al., 2014; Lévellé et al., 2020; Zeldenrust and Barta, 2021; Lévellé et al., 2019a,b; Hrazdilová et al., 2021). No genome or transcriptome data has been reported from any member of Adeleorina.

Adeleorina apicomplexans have traditionally been considered coccidians (Votýpka et al., 2017), but their position in the tree of Apicomplexa is essentially unresolved (Barta et al., 2012; Maia et al., 2016). Phylogenies using the SSU rDNA from a wide range of adelorinid and other apicomplexan taxa have generated trees using three different phylogenetic analysis methods, but these could only conclude that Adeleorina formed a monophyletic clade in various possible positions related to Coccidia and Hematozoa (Barta et al., 2012). Subsequent SSU rDNA analyses have shown adelorinids in different positions: within the Hematozoa sister to Piromplasida (Lévellé et al., 2019b), sister to Coccidia (del Campo et al., 2019; Kopečná et al., 2006; Chagas et al., 2021; Zeldenrust and Barta, 2021), sister to Hematozoa and Coccidia collectively (Cavalier-Smith, 2014; Vázquez et al., 2022; Pales Espinosa et al., 2023), or, when combined with LSU rRNA, branching with Nephromycida and collectively sister to Coccidia (Kwong et al., 2020). The only other molecular data are from the organelles: an apicoplast protein phylogeny showed Hematozoa and Marosporida forming a clade branching sister to Adeleorina (Pales Espinosa et al., 2023), whereas a mitochondrial protein phylogeny showed Adeleorina branching sister to Coccidia (Kwong et al., 2020).

To provide a more comprehensive molecular dataset to compare adelorinids with other major apicomplexan groups, we have generated genomic and transcriptomic data from two species, *Klossia helicina* and *Legerella nova*. These are the first nuclear transcriptomes from any adelorinid, and *L. nova* also represents the only remaining family within Adeleorina without any molecular data to date. We also sequenced complete plastid and mitochondrial genomes from both species. Using phylogenomics we resolve the placement of Adeleorina in the phylogeny of Apicomplexa, revealing a strongly supported clade that is not closely related to any other major subgroup. Instead, the adelorinids branch after Marosporidia and sister to Coccidia, Protococcidia, Hematozoa, and Nephromycida collectively.

2. Materials and methods

2.1. Sample collection

The definitive host of *Klossia helicina*, the grove snail *Cepaea nemoralis*, were obtained from residential areas from the municipalities of Richmond, Delta, and Vancouver in British Columbia, Canada. *Cepaea nemoralis* were euthanized prior to dissection using a 2-step method involving anesthesia followed by euthanasia (Gilbertson and Wyatt, 2016). To ensure complete anesthesia of *C. nemoralis*, which are of a different species and size than the snails from the original method (Gilbertson and Wyatt, 2016), the ethanol concentration was increased to 10% and 80% for the anesthetization and euthanasia steps, respectively, and an incubation period of 1 h was used for each step. Euthanized snails were rinsed in water to remove ethanol. After removing the shell, the snail body was submerged in water to rinse off residual ethanol prior to dissection. Kidney tissue was transferred to a new dish with 1 × phosphate-buffered saline (PBS), teased apart using fine forceps, then observed under a Leica DM IL LED inverted microscope.

Cells of *K. helicina* were collected using hand-drawn glass micropetites and rinsed twice in 1 × PBS to wash off host cell debris. Images were taken at 63 × magnification with a Sony alpha 7RIII digital camera attached to the microscope. Cells were sorted by life cycle stage based on previous descriptions (Woodcock, 1911; Barta, 2000) and 4–20 cells were pooled into individual 20 µl thin-walled PCR tubes containing 2 µl of Smart-Seq2 cell lysis buffer (2 U/µl RNaseOUT Recombinant Ribonuclease Inhibitor [Invitrogen] in 0.2 % [vol/vol] Triton X-100) for single-cell transcriptome preparation, and 2 µl of ultrapure water for single cell whole genome amplification (WGA) (Barta, 2000; Picelli et al., 2014). Previous studies used liquid nitrogen and 37 °C water bath freeze–thaw cycles to break open oocysts of *K. helicina* for the extraction of genetic material (Barta et al., 2012). Our attempt at breaking open mixed cyst stages including oocysts using this method with the freezing temperature at ~80 °C was unsuccessful for RNA extraction. Instead, a 1000 µl pipette tip was melted at the tip over a Bunsen burner to form a closed ball approximately the size of the end of the 20 µl collection tube, at the same time sterilizing it. Once the pipette tip has cooled down, it was used as a pestle to manually break open cells post freeze–thaw to release cell material for downstream steps.

The host of *Legerella nova*, the pill millipede *Glomeris marginata*, were collected in Bonn, Germany in April 2022. Specimens were kept at 8 °C until dissection. *Glomeris marginata* were euthanized in a kill jar with acetone vapor, after which their Malpighian tubules were removed into 1 × PBS and inspected for the presence of oocysts under 63 × magnification with a Leica DM IL LED inverted microscope and a Fujifilm X-S10 camera. In total, 23 pill millipedes were dissected (3 males, 20 females, 2 undetermined sex), of which 4 (1 male, 2 females, 1 undetermined) had *L. nova* oocysts. Cells picked with glass micropipettes, washed with 1 × PBS, and combined into pools of up to 6 cells in 2 µl of cell lysis buffer (2 U/µl RNaseOUT Recombinant Ribonuclease Inhibitor [Invitrogen] in 0.2 % [vol/vol] Triton X-100) and frozen until Smart-Seq2 library preparation. A piece of Malpighian tubule with a high density of *L. nova* was collected in 40 µl of 1 × PBS for genomic DNA extraction.

2.2. Single cell genome and transcriptome extraction, sequencing, and assembly

Transcriptomes were generated from *K. helicina* and *L. nova* cells.
using the Smart-Seq2 (Illumina Inc.) protocol, with PCR amplification adjusted to 23 cycles due to low quantity of starting cell material (Picelli et al., 2014). Genomic DNA was extracted from K. helicina single cells using the 4BB TruePrime Single Cell WGA Kit (4basebio SLU). To obtain L. nova genomic DNA from Malpighian tubule tissue, we used the DNeasy PowerSoil Pro Kit (Qiagen) according to the manufacturer’s instructions and eluted the DNA in 50 µl of 1 mM Tris solution. DNA was quantified with the Qubit dsDNA HS Assay Kit on a Qubit 2.0 Fluorometer. For K. helicina and L. nova CDNA and K. helicina genomic DNA, samples above 2 ng/µl concentration were sequenced. For L. nova genomic DNA, a single sample of 1.5 ng/µl concentration was sequenced.

Sequencing libraries for genomic DNA and cDNA products were generated using the Nextera XT DNA Library Preparation Kit (Illumina Inc.). Sequencing was performed on the Illumina NextSeq platform using 150 bp paired-end reads at the University of British Columbia Sequencing Centre (cDNA libraries and K. helicina genomic DNA) and at the Instituto Gulbenkian de Ciência Genomics Facility (L. nova genomic DNA).

Raw reads from multiple samples and sequencing runs were merged to generate final transcriptome datasets. Recovered reads for K. helicina samples ranged from 8,342,353 (1,259,695,303 bp) to 16,617,794 (2,509,286,894 bp). Sequencing reads were repeated for two samples recovering 61,700,273 (9,316,741,233 bp) and 86,078,183 (12,997,805,633 bp) from each repeat. The original runs, repeated runs, and 6 additional K. helicina samples were assembled to merge the final K. helicina transcriptome dataset. The final L. nova transcriptome dataset was assembled by merging 13 samples with reads ranging from 7,254,148 (1,080,868,052 bp) to 14,663,520 (2,184,846,480 bp). Raw read quality was assessed using FastQC. TrimGalore v0.6.6 was used to remove adaptor sequences and low-quality bases. Transcriptome reads were assembled using SPAdes v3.15.1 with the -rna option. Klossia helicina genomic reads were assembled with SPAdes v3.15.1 and the -sc option for genome assembly. Legerella nova genomic reads were assembled with Megahit v1.2.9 (Li et al., 2015) with default options. To confirm that the sequenced samples indeed contained K. helicina and L. nova, SSU rRNA sequences from the assemblies were extracted and queried by BLASTn against the NCBI GenBank nr database. Transcriptome samples that passed this identification were reassembled together with SPAdes v3.15.1 to maximize completeness. Animal host contamination was filtered out from genomes and transcriptomes by processing assembled reads through megaBLAST searches against the NCBI nr nucleotide database followed by diamond BLASTX against the UniProt reference proteome database (Altschul et al., 1990; UniProt Consortium, 2021). Visualization of contig contaminants was achieved using BlobTools and contaminants belonging to bacterial, archaeal, and metazoan groups were removed (Laetch and Blaxter, 2017). Trans-Decoder v5.1.0 was used to predict open reading frames (ORFs) and coding genes and longest ORFs were searched against the Uniprot database using BLASTP to generate annotations based on similarity (Haas, 2015; UniProt Consortium, 2021). Estimations of assembly completeness were made using BUSCO v5.4.3, using the alveolate database (Simão et al., 2015).

2.3. SSU rRNA phylogenetic analyses

Nuclear SSU rRNA genes of K. helicina and L. nova were extracted from assembled transcriptomes. Published sequences were retrieved from NCBI GenBank. To maximize phylogenetic resolution, sequences < 1000 bp in length were excluded, except for Haemogregarina bigemina which was previously thought to be an adeleorinid but instead branches with other unidentified marine fish-infecting apicomplexans sister to coralliloids on phylogenetic trees and is the only described member of this clade (Hayes and Smit, 2019). Sequences were aligned with Mafft v7.481 (Katoh et al., 2002), then unaligned 5’ and 3’ ends were manually trimmed before trimming with trimAl v1.4 using the automated1 method (Capella-Gutiérrez et al., 2009). A maximum likelihood tree was built using IQ-TREE v1.6.12 with 1000 non-parametric bootstrap replicates and a GTR + F + R5 substitution model chosen by ModelFinder (Nguyen et al., 2014; Kalyaanamoorthy et al., 2017).

2.4. Apicoplast and mitochondrial genome assembly and annotation

Klossia helicina and L. nova genome assemblies were used to search for apicoplast contigs using BLAST (Altschul et al., 1990) with published Adeleorina apicoplast sequences as queries. Contigs identified as belonging to the apicoplast were then used as templates which all trimmed raw reads were mapped against them with Bowtie2 v2.4.2 (Langmead and Salzberg, 2012). Reads that successfully mapped were extracted with SAMtools (Danecek et al., 2021), and manually reassembled and visualized using Tablet v1.12.02.08 (Müllner et al., 2010). This procedure was repeated to assemble and close (circularize) the genome. Apicoplast genome annotation was completed manually on Geneious Prime 2022.0.1 (Biomatters Ltd.) using tRNA predictions from tRNAscan-SE 2.0 (Chan and Lowe, 2019) and rRNA, ribosomal protein genes, and other protein encoding gene predictions from a combination of RNAmmer v1.2 (Lagesen et al., 2007), MFannot (https://megasun.bch.umontreal.ca/apps/mfannot/), and Geneious Prime’s find open reading frame and annotation prediction function with apicomplexan apicoplast reference genomes.

Similarly, mitochondrial contigs were recovered from the K. helicina and L. nova genome assemblies by selecting high-confidence low GC content contigs and verifying against related apicomplexan mitochondrial sequences using BLAST. Multiple K. helicina mitochondrial contigs were identified but could not be circularized using the technique described to close the apicoplast genome. The single L. nova mitochondrial contig we identified was circularizable, suggesting it represents a complete mitochondrial genome. The L. nova mitochondrial genome was annotated manually on Geneious Prime 2022.0.1 (Biomatters Ltd.) using predictions of protein coding genes from MFannot with translation table 4 (Mold, Protozoan, and Coelenterate Mitochondrial) (https://megasun.bch.umontreal.ca/apps/mfannot/) and predictions of small and large rDNA fragments made using Geneious Prime’s annotation prediction function with K. helicina (MT084563), K. razorbacki (NC_058856), and Klossia helicina (MH203050) mitochondrial reference genomes. The putative fragments were checked for sequence similarity using BLAST and predictions recovering 97.5–100 percent identity to apicomplexan mitochondrial genomes retained.

2.5. Phylogenomic tree construction and analyses

Newly generated transcriptomic data from K. helicina and L. nova and additional taxa with publicly available genomic and transcriptomic data were added to an existing set of 263 genes used previously for apicomplexan phylogenomic analyses containing representatives from most eukaryotic supergroups (Mathur et al., 2023). BLASTP was used to search coding sequences predicted by TransDecoder v5.1.0 during assembly for 263 genes used in the set with an e-value threshold of 1e-20 and query coverage threshold of 50% (Haas, 2015). To identify poorly aligned sequences to extract from newly identified genes, BLAST outputs were searched against the UniProt database using BLASTP (UniProt Consortium, 2021). After poorly aligned sequences were removed the final 263 gene set was aligned using MAFFT L-INS-i v7.481 (Katoh et al., 2002). The alignments were used to construct single gene trees with IQ-TREE v1.6.12 and viewed with FigTree v1.4.4 to manually screen for contaminants and paralogs (Nguyen et al., 2015; Rambaut, 2014). Cleaned tree sets that had a maximum of 40 % missing genes were trimmed with the automated1 method in trimAl. The initial dataset included all taxa, including long-branching taxa. To account for the effects of extremely divergent members within Apicomplexa like long branch attraction or other artefacts, long-branching taxa were removed from the main dataset used to generate the main figures. The final
A dataset was concatenated in SCaFoS v4.55 resulting in a concatenated alignment consisting of 67 taxa, 206 proteins, and 51,460 amino acid sites (Roure et al., 2007). A phylogenomic maximum likelihood tree was built using IQ-TREE v1.6.12 under the LG + C60 + F + G4 mixture model with 1,000 ultrafast bootstrap replicates (UFB) (Nguyen et al., 2015; Minh et al., 2013).

Several subsequent analyses were performed to explore the support and robustness of the position where Adeleorina emerged. The former dataset that still contained long-branching taxa was analyzed in parallel to the main dataset using the same parameters in all additional analyses including the generation of a phylogenomic maximum likelihood tree with 74 taxa, 203 proteins, and 48,966 amino acid sites. The initial maximum likelihood tree was used as the input guide tree to run the Posterior Mean-Site Frequency (PMSF) model under the LG + C60 + F + G4 model using 200 non-parametric bootstrap replicates (BS) in IQ-TREE v1.6.12 for rapid approximation of the profile mixture model to evaluate node support (Wang et al., 2017). Bayesian analyses was performed with PhyloBayes-MPI 20,201,026 using the CAT + GTR model with 4 parallel chains allowed to run simultaneously until a cutoff point of minimum 10,000 cycles each was reached (Lartillot et al., 2009). The first 10 % of trees generated in each chain were discarded as burn-in before calculating differences in bipartition frequencies with -bpcomp.

Fig. 1. Light micrographs of *Klossia helicina* and *Legerella nova* cells, pictures of hosts, and phylogeny of the Adeleorina. A. From left to right the micrographs show a *Klossia helicina* sporulated oocyst, unsporulated oocyst or macrogamete associated with a microgamete, unsporulated oocyst or macrogamete, and the snail host *Cepaea nemoralis*. B. Light micrographs of *Legerella nova* from left to right show oocysts with naked sporozoites, oocysts in Malpighian tubules of the pill millipede host, and the pill millipede host *Glomeris marginata*. Scale bar = 50 µm. Black arrow = host nucleus; White arrow = host nucleus or microgamete; White triangle = inner membrane originating from parasite; Black triangle = outer membrane originating from host. C. Maximum-likelihood SSU rDNA tree of 37 apicomplexan taxa, with a focus on Adeleorina subgroups, generated under the GTR + F + R5 model with 1000 non-parametric bootstrap replicates. Heteroxenous lineages of Adeleorina (haemogregarines) are shaded in purple. Monoxenous Adeleorina (adelines) are shaded in green. All other major apicomplexan subgroups are shaded in grey. Newly added adeleorinids *Klossia helicina* and *Legerella nova* are indicated in bold. The branch supporting *Haemoproteus balmorali* has been reduced to 1/6 of its original length. The tree is rooted on *Cryptosporidium parvum*. Bootstrap supports under 70 % are omitted, nodes with full bootstrap support are denoted by a black circle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
to assess convergence of the four chains and manually inspecting consensus trees from each chain (maxdiff = 1). We removed fast-evolving sites in 5% incremental steps until a maximum 95% was removed (Irwin, 2021). The site-specific substitution rates were generated using the -wrr option in IQ-TREE v1.6.12. A maximum likelihood phylogenetic tree was made for each increment with IQ-TREE v1.6.12 using the LG + G60 + F + G4 model and 1,000 ultrafast bootstrap replicates. To assess topology robustness the approximately unbiased (AU) test (Shimodaira, 2002) was performed on 12 trees representing a range of possible topologies possible, some based on the original maximum likelihood tree, maximum likelihood tree including long-branching apicomplexans, positions in published phylogenetic trees (Léveillé et al., 2014; Léveillé et al., 2020; Zeldenrust and Barta, 2021; Léveillé et al., 2019a; Léveillé et al., 2019b; Hrazdílová et al., 2021).

3. Results and discussion

3.1. Identification and isolation of Klossia helicina and Legerella nova

Klossia helicina and L. nova were manually isolated from dissected kidney tissue of grove snails and the Malpighian tubules of pill millipedes (Fig. 1A, B), respectively, and the material was used to generate transcriptomes and organellar genomes. This study is the first report of K. helicina in grove snails in Canada. Previous attempts at finding the adeleorinid in Eastern Canada were unsuccessful and until now reports of K. helicina were limited to grove snails originating from Europe (Barta et al., 2012). Sporulated oocysts and cells that may be unsporulated oocysts or macrogametes of K. helicina were spherical and ranged from 70 to 80 µm in length (Fig. 1A). Sporozoites within sporulated oocysts measured 12.5 µm in diameter (Fig. 1A). A distinctly elongated cell that was either an unsporulated oocyst or macrogamete measured 90 µm in length and 51 µm in width (Fig. 1A). Legerella nova oocysts with naked sporozoites measured up to 30 µm (Fig. 1B). This was consistent with previous reports of comparatively large adeleorinid oocysts and macrogametes, relative to invertebrate host cells (Barta, 2000; Volkmann, 1967; Vincent, 1927).

Both adeleorinids had distinct inner and outer membranes with a structure situated in between (Fig. 1A, B). Previous descriptions identify the inner membrane originating from the adeleorinid while the outer membrane is what remains of the host membrane highly distended well beyond its original size (Vincent, 1927; Volkmann, 1967). The structure situated between the membranes in K. helicina and L nova sporulated oocysts is likely to be the remaining hypertrophied host nucleus (Fig. 1A, B) (Volkmann, 1967; Vincent, 1927). However, distinguishing this structure as a hypertrophied host nucleus beside an unsporulated oocyst or, alternatively, a microgamete attached to a macrogamete in syzygy in K. helicina is difficult under light microscopy as both a microgamete and a hypertrophied host nucleus would appear as small structures directly adjacent to a larger pigmented cell (Fig. 1A) (Volkmann, 1967; Schulte, 1971, Scholtynsek et al., 1971). The mechanism behind the manipulation, maintenance, and function of these host structures and the adeleorinid parasite requires further investigation to be understood.

3.2. Phylogenetic placement based on SSU rDNA

Small subunit rRNA sequences were extracted from K. helicina and L. nova transcriptomes and used to infer a Maximum-likelihood (ML) phylogenetic tree with 37 taxa representing all 7 known sub-groups of Adeleorina (Fig. 1C, Table S1) (Barta, 2000). Legerella nova is the only taxon representing Legerellidae since it is the first member of the group with available molecular data. The genera Cyrilida and Dieriera are considered to be members of the Haemogregarinidae, but could not be included in the adeleorinid tree as presently no molecular data have been obtained from any members of these genera. Adeleorinid taxa from a variety of definitive and intermediate hosts as well as host geographical location were selected to represent adeleorinid prevalence and distribution, as they have been found on every continent except Antarctica so far (Fig. 1C, Table S1).

Klossia helicina from grove snails from Western Canada branched with K. helicina obtained from grove snails in Europe with strong support (Fig. 1C, 98 % BS), forming a Klossia clade with K. razorbacki from the North American magnolia three-toothed snail Triodopsis hopotomensis (Fig. 1C, 65 % BS) (Zeldenrust and Barta, 2021). Legerella nova branched within Adeleorina, but without significant support and within the Adeleidae clade sister to the Klossia clade (Fig. 1C, 65 % BS). Availability of published molecular data from monoxenous adeleorinids is scarce compared to heteroxenous adeleorinids although many species have been described from the former group, and from Legerellidae more specifically (Barta et al., 2000; Vincent et al., 1927; Tuzet and Manieri, 1957). Obtaining molecular data from additional monoxenous adeleorinids, especially members of the Legerellidae, in combination with phylogenetic analysis using more data than just the SSU rRNA (e.g., phylogenomics or at least the mitochondrial COI protein) may improve the resolution of the position of Legerellidae within Adeleorina.

Interestingly, a greater sampling of taxa from Adeleorina not only has the potential of illuminating phylogenetic relationships within Adeleorina, but also across Apicomplexa as a whole. Phylogenetic analysis of the marine fish haemogregarine, Haemogregarina bigeminia, revealed it is not really a haemogregarine at all: instead, it was found to branch outside of Adeleorina as sister to the marine, coral-infecting coralicolid (Fig. 1C) (Hayes and Smit, 2019; Kwong et al., 2021). Only a limited number of arthropods and gastropods have been inspected for presence of adeleorinids, and marine sampling of invertebrate-infecting apicomplexans is very poor (Rueckert et al., 2019), altogether suggesting that these abundant and species-rich animal groups probably harbour a greater diversity of new lineages.

3.3. Apicoplast and mitochondrial genomes

Apicoplast genomes extracted from K. helicina and L. nova are extremely similar to each other, and to the only other known apicoplast genome from Adeleorina. The apicoplast genomes of K. helicina and L. nova were circular-mapping like the apicoplast genome of the adeleorinid Hepatozoon canis (Léveillé et al., 2019b). The K. helicina and L. nova apicoplast genomes were 33,281 bp (21.9 % GC content) and 32,955 bp (22.6 % GC content), respectively. Both genomes retained inverted repeat regions that included the LSU (23S) and SSU (16S) rRNA (Fig. 2A), like the apicoplast genome of H. canis (Léveillé et al., 2019b). No genes were found that are involved in photosynthesis or chlorophyll biosynthesis (chlA, chlB, and acsF) (Kwong et al., 2019).

A mitochondrial genome sequence was obtained from L. nova, and a fragmented genome was also found in K. helicina. The L. nova mitochondrial genome was 6,750 bp long (36.3 % GC content), circular-mapping, and contained large and small ribosomal RNA (rns and rml) fragments and protein coding genes: cytochrome c oxidase subunit I (COI), cytochrome c oxidase subunit III (COIII), and cytochrome b (COB) (Fig. 2B). Four contigs that had high percent identity to a previously published K. helicina mitochondrial genome were also identified (Fig. 2B). These contigs included the COI, COB, and COIII genes, rnl and rns fragments, and shared 87.93–94.63 % identity to the previous K. helicina mitochondrial genome (MT084563). Unlike the previous K. helicina and K. razorbacki mitochondrial genomes, we could not fully circularize our K. helicina mitochondrial genome (Zeldenrust and Barta, 2021).

3.4. Phylogenomic analyses

Phylogenomics was used to place Adeleorina on a multigene tree of Apicomplexa. Transcriptomes of K. helicina and L. nova generated in this study, and genomes and transcriptomes from Besnoitia besnoiti, Cystoisospora suis, Eimeria necatrix, Eimeria brunetti, Cyclospora caytanensis, Goussia janae, and Cardiopodirum cionae extracted from online
databases, were added to an existing dataset (Mathur et al., 2023) (Table S2). In a 67-taxa 206-protein ML phylogenomic tree, the two Adeleorina species formed a monophyletic clade that was strongly supported as sister to a large group consisting of Coccidia, Protococcidia, Hematozoa, and Nephromycida (Fig. 3, 98 % UFB; 100 % PMSF-BS; Fig. S2A, 1 pp). Marosporida, another less well-studied group, branched sister to this clade and Adeleorina (Fig. 3, 100 % UFB; 100 % PMSF-BS; Fig. S2A, 1 pp). This position of adelorinds was also observed in some previous analyses of SSU rDNA (Cavalier-Smith, 2014; Vazquez et al., 2022; Espinosa et al., 2023), but with poor support and with members of the Marosporida in different positions. The position of Adeleorina remained the same in our 74-taxa 203-protein ML tree that included long-branching apicomplexans in its dataset (Fig. S1, 92 % UFB; 100 % PMSF-BS; Fig. S2B 1 pp).

In Bayesian analysis none of the four chains for each dataset converged, but none of the differences between chains were related to the position of Adeleorina which was fully supported in all cases (Fig. S2, 1 pp). In all four chains generated using the main dataset (Fig. S2A) and main dataset plus long branching taxa (Fig. S2B) Adeleorina remained sister to Coccidia, Protococcidia, Hematozoa, and Nephromycida collectively and Marosporida remained sister to Adeleorina, Coccidia, Protococcidia, Hematozoa, and Nephromycida collectively. Bootstrap support values for Adeleorina normalized to 100 % UFB after the first 5 % incremental removal step of fast-evolving sites in trees made from both datasets and remained at 100 % UFB until 40 % of fast-evolving sites were removed (Fig. S3). The support value for Adeleorina in the original dataset lowered to 99 % UFB when 40–50 % fast-evolving sites were removed, then rose to 100 % UFB again when 55–65 % fast-evolving sites were removed. The dataset that included long-branching taxa lowered to 99 % at 65 % fast-evolving site removal, and, along with the original dataset, lowered to 98 % once 70 % fast-evolving sites were removed. Tree topology significantly diverged from published literature when 75 % or greater fast-evolving sites were removed and were thus not included as part of the analysis. The position of Adeleorina was also strongly supported in the AU tests where 11 out of 12 tested topologies were rejected: only the topology matching the maximum-likelihood tree where Adeleorina branches sister to Coccidia, Protococcidia, Hematozoa, and Nephromycida was not rejected at a 95 % confidence interval (Fig. S4, Table S3).

Phylogenomic trees made using multiple slow-evolving proteins can more robustly represent phylogenetic relationships and have been used to resolve conflicting branch patterns in Apicomplexa as well as inform evolutionary history (Mathur et al., 2021; Salomaki et al., 2021). The phylogenomic positions of Marosporida, Gregarines and now Adeleorina were resolved within Apicomplexa using multigene phylogenomic methods, just as the demonstration that squirmids were not apicomplexans was also made with multigene phylogenies, altogether changing our interpretation of apicomplexan evolution by demonstrating how parasitism, plastid loss, and loss of photosynthesis occurred multiple times (Mathur et al., 2021; Salomaki et al., 2021). The importance of
parallel evolutionary changes is also underscored by our analyses, as the phylogenetic tree shows the transition from a monoxenous parasitic lifestyle to a heteroxenous lifestyle, the emergence of blood parasitism, and the transmission by invertebrate blood vectors have also emerged multiple times independently within Apicomplexa. Both monoxenous and heteroxenous taxa are represented in Coccidia, Adeleorina, Marosporida, and Gregarines, but a pattern of strictly monoxenous taxa close to the root within these subgroups to heteroxenous taxa at the tips of these subgroups appear in the Coccidia, Adeleorina, and Gregarines. Among apicomplexans, intracellular invasion of blood cells appears in two major subgroups that branch in different positions of the phylogenetic tree: the Hematozoa and the Adeleorina (and also seemingly in the lineage of marine fish-infecting apicomplexans previously mis-identified as Haemogregarines). Blood-infecting apicomplexans in these two major subgroups are also transmitted by blood vectors such as mosquitoes and ticks (and gnathiid isopods and leeches in the case of the marine fish-infecting group (Hayes and Smit, 2019)).

Phylogenomics can help us understand deeper relationships among groups within the Apicomplexa, while mitochondrial and apicoplastidal data appear to be more effective in investigating more closely related taxa. Other drawbacks to the organelle genomes are that mitochondrial and/or apicoplast genomes have been lost in cryptosporidians and multiple clades of gregarines (Putignani et al., 2004; Zhu et al., 2000; Toso and Omoto, 2007; Mathur et al., 2019; Janoušková et al., 2019). Further, using apicoplast data alone has been shown to generate less robust phylogenies of Apicomplexa (Muñoz-Gómez et al., 2019; Mathur et al., 2021). Sampling for more transcriptome and genome data from members of the Adeleorina can contribute to a better understanding of their phylogenetic relationships and biological diversity as well as investigate species of interest that infect endangered, domesticated, or economically important animals. Compared to heteroxenous members of the Adeleorina, monoxenous adeleorinids are more easily isolated and

Fig. 3. Maximum likelihood phylogeny of 67 taxa inferred from 206 genes (51,460 amino acid sites) under the LG + C60 + F + G4 mixture model with 1000 UFBs. Major apicomplexan subgroups are separated by colour. The tree is rooted on Dinoflagellates. Adeleorinids Klossia helicina and Legerella nova from this work are in bold. Black dots indicate nodes with 100 % ultrafast bootstrap support, and 100 % posterior mean site frequency (PMSF) bootstrap support from 200 non-parametric bootstrap replicates. Values otherwise are indicated in this order from left to right. The bar graph on the right shows the percentage of genes represented in the phylogeny for each taxon with black bars representing the newly added adeleorinids from this work.
can be cultured (Molmann, 1980). This enables the use of techniques requiring manipulation of live cells to be completed with these taxa, making adeleorinids—particularly *K. helicina*—a potential candidate model organism for future research. Strong phylogenomic support for Adeleorina forming an independent subgroup sister to Coccidia, Proto- coccidia, Hematozoa, and Nephromycida puts the adeleorinids in an interesting position, diverging around the time when apicomplexans appear to have been transitioning from mostly-extracellular and benign symbionts to more intracellular and more likely to be pathogenic, making them interesting cases in the study of the origin of parasitism that are largely overlooked.

### 4. Conclusions

Adeleorina, a major apicomplexan subclade infecting a large variety of vertebrate and invertebrate animals across the world, has previously been excluded from phylogenomic analyses due to unavailable molecular data. Here, we obtained transcriptomes from two members of the Adeleorina to use in a phylogenomic analysis, resolving its position as a major group branching in a position sister to a clade formed by Coccidia, Protococcidia, Hematozoa, and Nephromycida. This demonstrates that the Adeleorina are not Coccidia, as in classical taxonomy, but instead, like Marosporida, diverged from other apicomplexans early in the transition from a mostly-extracellular to a mostly-intracellular infection strategy was emerging. From this updated tree of Apicomplexa multiple independent instances of a transition from monoxenous to heteroxenous parasitic lifestyle, blood parasitism, and transmission through an animal blood-feeding vector appear to have arisen within multiple subgroups. *Legerella nova* is the only member of the *Legerellidae* in the Adeleorina thus far to have genetic material sequenced, demonstrating the need for further sampling from monoxenous members of Adeleorina to better resolve this area of the Adeleorina tree. *Klossia helicina* was previously believed to be endemic to *C. nemoralis* and *C. hortensis* in Europe, but were revealed to be present in hosts that have been transferred to North America and therefore dispersed more widely than was previously known. *Klossia helicina* and *L. nova* retain apicoplasts closely resembling other Adeleorina with circular-mapping apicoplast genomes. A circular-mapping mitochondrial genome could be extracted from *L. nova*, but the mitochondrial contigs extracted from *K. helicina* could not be circularized. The tree of Apicomplexa continues to be refined by using phylogenomics and adding previously unsampled taxa, largely through culture-free genomics and transcriptomics methods. Resolving the position of the historically established, widespread, and extensively described Adeleorina on the tree of Apicomplexa as an independent subgroup changes how we must interpret characteristics attributed to adeleorinids, and point to the need to further develop these methods and datasets to really understand apicomplexan evolution and diversity.

### Data availability

Transcriptome and genome raw reads are available on NCBI SRA under BioProject PRJNA1032149. SSU rDNA sequences are available at the NCBI GenBank Nucleotide Database under accession numbers OR680781 (*Klossia helicina*) and OR689432 (*Legerella nova*). The plastid and mitochondrial data generated in this study are available at the NCBI GenBank Nucleotide Database under the following accession numbers: OR693319, OR693320, OR693321, PP329795, PP329796, PP329797, PP329798. The alignments used to generate our phylogenomic trees are available on Mendeley Data under the DOI: 10.17632/nmmjgckb4.1.

### Acknowledgements

We are grateful to Thomas Wesener and Thorsten Klug from the Alexander Koenig Zoological Research Museum for providing *G. marginata* samples, Varsha Mathur and Anna Cho for advice on data processing, and Elizabeth Cooney for advice on phylogenomic analysis and visualization. This work was supported by a grant from the Gordon and Betty Moore Foundation (https://doi.org/10.37807/GMBP9201). IN was supported by the National Sciences and Engineering Research Council (NSERC) of Canada Postgraduate Scholarship-Doctoral (PGSD) and the University of British Columbia (UBC) Botany Four-Year Fellowship. CC and WK was supported by the European Molecular Biology Organization Installation Grant 5045-2022, Fundação para a Ciência e a Tecnologia grant CEECIND/01358/2021, and by Fundação Calouste Gulbenkian. GL was supported by Genome BC grant R02MSE.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2024.108060.

### References


I. Na et al.  Molecular Phylogenetics and Evolution 195 (2024) 108060
I. Na et al.

Molecular Phylogenetics and Evolution 195 (2024) 108060


