Divergent Plastid Genomes in the Deepest-Branching Apicomplexan Parasites

Gordon Lax [1,2,*], Waldan K. Kwong [1,3], Ina Na [1], Victoria K.L. Jacko-Reynolds [1], Morelia Trznadel [1], Chloe S.E. House [1,4], Varsha Mathur [5], Kevin C. Wakeman [6], Patrick J. Keeling [1]

[1]Department of Botany, University of British Columbia, Vancouver, Canada

[2]Department of Biology, Dalhousie University, Halifax, Canada

[3]Instituto Gulbenkian de Ciência, Oeiras, Portugal

[4]Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, Canada

[5]Faculty of Life Sciences, University of Vienna, Vienna, Austria

[6]Faculty of Science, Institute for the Advancement of Higher Education, Hokkaido University, Japan

*Corresponding author: E-mail: gordon.lax@gmail.com.

Accepted: June 02, 2025

Abstract

Apicomplexans are widespread and diverse obligate symbionts of animals, and many—like the malaria-causing *Plasmodium*—are important human parasites. Some of the closest free-living relatives of apicomplexans are photosynthetic, and most apicomplexans retain a relict, nonphotosynthetic plastid called the apicoplast. The origin and evolution of this plastid has been studied extensively, but most data come from biomedically relevant taxa. There has been increasing interest in the diversity of plastids of other apicomplexans, which has revealed a complex picture of recurring, independent losses of plastids or their genomes in the deepest-branching apicomplexan lineages, *Cryptosporidium* and the gregarines. Indeed, the plastid is absent from many of these deep-branching apicomplexans, and those that retain it have almost all lost its genome. The single exception appears to be the archigregarines, the deepest-branching subgroup of gregarines, which are suspected to have retained a genome based on analyses of nuclear genes, but for which no plastid genome data are known. As the best potential representatives of plastid genomes in deep-branching apicomplexans, this lineage is an obvious missing piece of the evolutionary puzzle of the apicoplast. Here, we used single-cell sequencing to characterize the plastid genome from uncultivated representatives of all 4 known archigregarine lineages. The plastid genomes are all more divergent with lower GC-content than other apicoplast genomes, and encode a slightly different set of genes. There is no evidence of photosynthesis-related genes. These genomes fill a key gap in the diversity of apicomplexan plastid genomes, furthering our understanding of the complex evolution of the apicoplast.

Key words: plastome, gregarine, archigregarines, *Selenidium*.

Significance Statement

Apicomplexans are a widespread group of symbionts of animals, including humans (e.g. *Plasmodium*, the malaria parasite). Their ancestors were photosynthetic, and as such many apicomplexans retain a remnant and reduced plastid that is not capable of photosynthesis. This plastid has mainly been investigated in medically relevant apicomplexans, yet most diversity can be found in the gregarines. This early branching apicomplexan group has mostly been overlooked, and we know little about their plastid evolution apart from some members having completely lost the plastid, while others likely retained it. We used single-cell sequencing to generate plastid genomes from 4 diverse lineages of archigregarines isolated from marine invertebrates. The plastid genomes are highly reduced in length and gene content, are AT-rich, and vary substantially between the 4 archigregarine lineages. Understanding plastid evolution of archigregarines helps us understand the evolution and loss of the plastid in apicomplexans, and plastid evolution in general.

Introduction

Apicomplexa are a diverse group of obligate animal symbionts, best-known as biomedically important parasites like *Plasmodium* (malaria), *Toxoplasma* (toxoplasmosis), and *Cryptosporidium* (cryptosporidiosis). They evolved from a free-living photosynthetic ancestor, and most apicomplexans still harbor a remnant plastid (the apicoplast), which arose from an endosymbiosis with a red alga (Williamson et al. 1994; Janouškovec et al. 2010). Much of what we know of apicomplexan biology—and the apicoplast more specifically—comes from biomedically relevant species, but these represent only a narrow subset of apicomplexan diversity. Emerging data from other lineages shows an incredible diversity in many characteristics, including genome form, content, and function (Sato 2011; Janouškovec et al. 2019; Salomaki et al. 2021; Yazaki et al. 2021; Mathur et al. 2023; Na et al. 2024). One of the most diverse subgroups of apicomplexans are the gregarines, which alongside *Cryptosporidium*, have been repeatedly demonstrated to be the deepest-branching lineages of apicomplexans (Leander 2008; Rueckert and Leander 2008; Paskerova et al. 2018; Rueckert et al. 2019; Lax et al. 2024; Na et al. 2024).

All apicomplexans are nonphotosynthetic obligate symbionts, and consequently have a reduced plastid metabolic repertoire (Leander 2008; Janouškovec et al. 2019; Muñoz-Gómez et al. 2019; Mathur, Wakeman, et al. 2021; Salomaki et al. 2021; Mathur et al. 2023). No apicoplast retains any functions associated with photosynthetic activity, and the plastid is almost always retained to carry out a variety of essential biosynthetic pathways that are mediated by nucleus-encoded genes for plastid-targeted proteins (Janouškovec et al. 2019; Kwong et al. 2019; Muñoz-Gómez et al. 2019; Mathur, Kwong, et al. 2021; Salomaki et al. 2021; Mathur et al. 2023; Na et al. 2024). Most apicoplasts have a genome, which is thought to be retained due to the presence of vital iron–sulfur cluster synthesis genes (*sufB* and *clpC*), and indeed in cases where these genes are transferred to the nucleus, the plastid may lose its genome (Janouškovec et al. 2015).

Plastid genome data are now available for most major apicomplexan groups; however, we lack data from the deepest-branching lineages. This is partially because *Cryptosporidium* and most gregarine apicomplexans investigated to date have either lost their plastid genomes, or the entire organelle (Zhu et al. 2000; Salomaki et al. 2021; Mathur et al. 2023). The only exception is the archigregarines, the deepest-branching lineage of gregarines, which are thought to have retained a plastid genome (Cavalier-Smith 2014; Janouškovec et al. 2019; Mathur et al. 2023). Because of their deep phylogenetic position, archigregarines are crucial in understanding the evolution of apicomplexans as a whole, but their retention of apicoplast genomes makes them particularly important to understand how this strange genome evolved and why it was retained. Using single-cell genomics, here we characterized the plastid genomes of representatives from all known archigregarine clades (*Selenidium*, *Lunidium*, *Metzidium*, and *Devanium*), and show them to have evolved unique and derived features.

Results and Discussion

Archigregarine Apicoplast Genome Structure and Content

Twenty-seven individual cells from 8 distinct archigregarine species representing all 4 currently recognized genera were manually isolated from their hosts (supplementary table S1, Supplementary Material online), and total genomic DNA from each cell was subjected to whole genome amplification and sequencing. For each cell, the identity of the parasite was first confirmed by phylogenetic analysis of the nuclear SSU rDNA. Both previous phylogenomic and SSU rRNA phylogenies have shown that archigregarines branch into 4 distinct groups: *Selenidium*, *Metzidium*, *Devanium*, and *Lunidium* (Paskerova et al. 2018; Wakeman 2020; Lax et al. 2024). Each of our SSU rDNA sequences was found to fall in the expected position, with the genus corresponding to their identification based on morphology and host, and with the expected species in cases where SSU

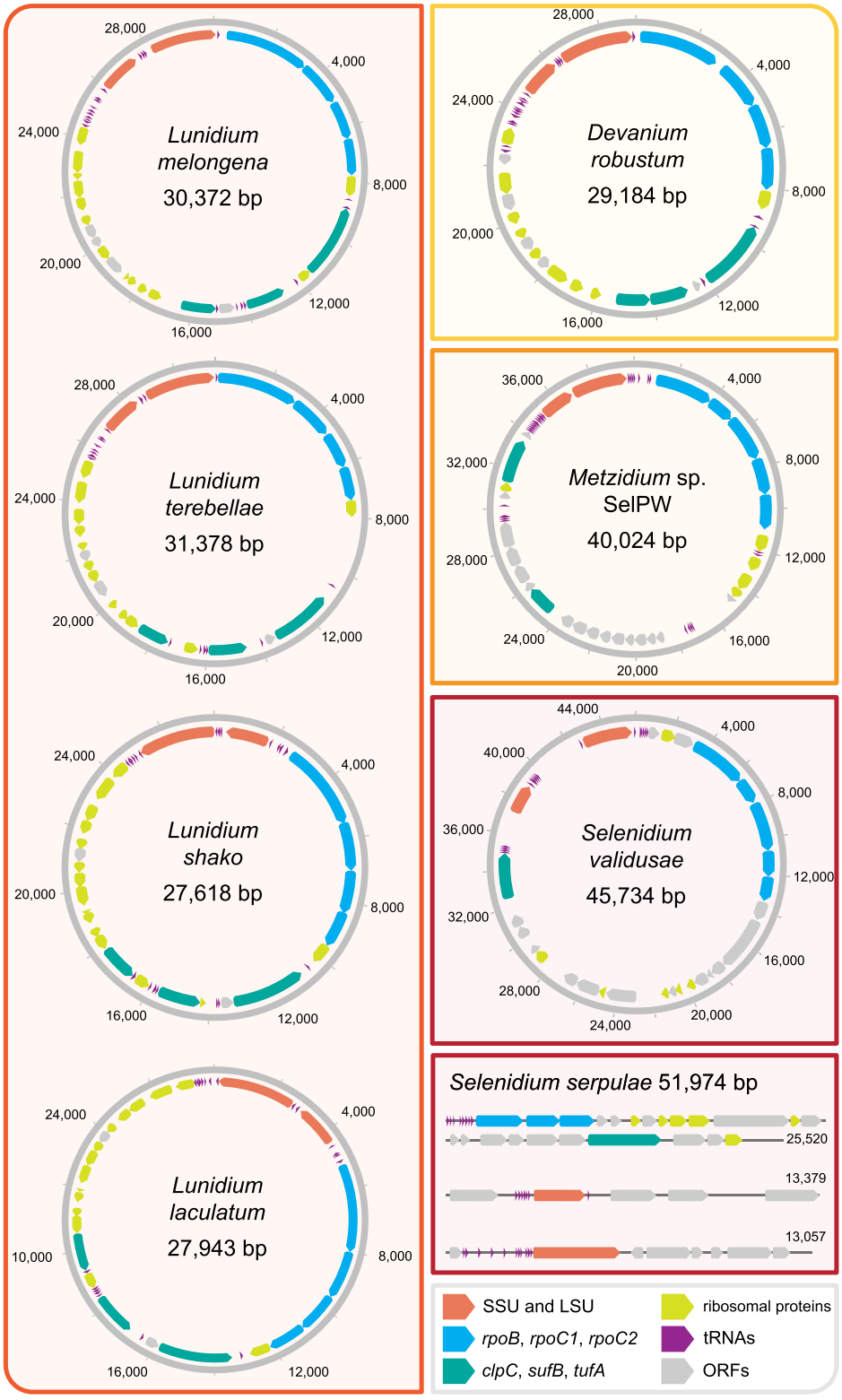

Fig. 1. Apicoplast genomes of archigregarines. Apicoplast genomes from each of the 4 identified clades of archigregarines.

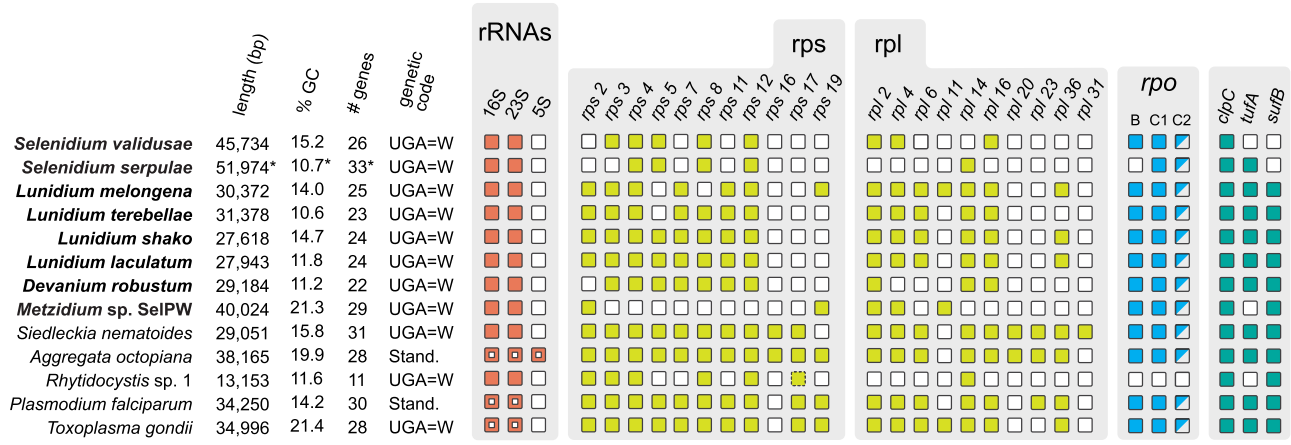

Fig. 2. Apicoplast statistics and gene content of archigregarine apicoplast genomes. Apicoplast genomes generated in this study are in bold. Solid boxes denote presence of a gene, boxes with a dotted line denote a pseudogene, and split boxes denote a split gene for *rpoC2*. A white dot inside a box denotes an inverted repeat (IR) for plastidial rRNAs. The asterisk (*) next to *S. serpulae* denotes a fragmented assembly.

rRNA was already available (supplementary fig. S1, Supplementary Material online).

Assembling plastid genomes for 7 of the species yielded deep-coverage, circularized, and complete genomes (Fig. 1; supplementary fig. S2 and table S1, Supplementary Material online). The *Selenidium serpulae* assembly was fragmented (supplementary fig. S2, Supplementary Material online). The genomes ranged in size from 27,618 bp (*Lunidium shako*) to 45,734 bp (*Selenidium validusae*; Figs. 1 and 2), and the genomes from each genus shared a high level of synteny. We found no evidence of inverted rRNA repeats in any assembly, and in *L. laculatum* and *L. shako* the SSU and LSU rRNA genes were inverted in order when compared with the canonical rRNA operon.

The gene content (including predicted ORFs and rRNAs) ranged from 22 in *Devanium robustum* to 29 in *Metzidium* (Fig. 2). Many of the unidentified ORFs were short and fall in the same position as ribosomal proteins in other apicoplasts, and therefore possibly correspond to those genes, but because the sequences in all 8 genomes were extremely divergent, they shared no detectable sequence similarity with their putative orthologs. The overall GC-content was also extreme, ranging from a high of only 21% to a low of 9% in *S. serpulae*. Interestingly, we found alternative codon usage in all 8 genomes, where UGA encodes for tryptophan instead of a stop-codon (Fig. 2). Alternative codon usage is widespread in eukaryotic organellar genomes (Smith and Keeling 2015), and has previously been documented in apicomplexan plastid genomes as well (Mathur, Kwong, et al. 2021). There is no indication that alternative codon usage has any functional reason, and it is assumed this process is neutral (Smith and Keeling 2015).

The surprising discovery that obligate intracellular parasites like *Plasmodium* and *Toxoplasma* had plastid organelles set off a long hunt to better understand the origin, evolution, and function of their apicoplasts. We now know they arose from a red alga, and share a common ancestry with the plastids of dinoflagellates and a handful of apicomplexan sister lineages, squirmids, and chrompodellids (Moore et al. 2008; Janouškovec et al. 2010; Obornik et al. 2012; Mathur et al. 2019). Comparative analyses to reconstruct this evolution depend on sampling a wide diversity of apicomplexans and their relatives, but 1 stubborn blind spot has remained: the deepest-branching apicomplexan lineages, *Cryptosporidium* and the gregarines. Most of these species retain plastids, and the plastid function has been examined using transcriptome sequencing of nucleus-encoded genes from a few representatives (Janouškovec et al. 2019; Mathur et al. 2023). But even among those with a plastid, only the archigregarine lineage is predicted to have retained its plastid genome, and to date no genomic sequencing has been available from archigregarines. One plastid genome has been assembled from transcriptome data from the blastogregarine *Siedleckia nematoides* (Janouškovec et al. 2019; Mathur, Kwong, et al. 2021), a lineage possibly related to archigregarines, but whose position in the tree has remained uncertain (Lax et al. 2024).

Here, we characterized plastid genomes from 8 species representing all currently identified genera (Lax et al. 2024), and show that archigregarines possess a reduced plastid genome with an extremely low GC-content (10-21%), and encode a reduced set of genes (23-29; Fig. 2). The gene order in archigregarine apicoplast genomes varies considerably between lineages, particularly when compared to other apicomplexan groups. In Coccidia and Hematozoa for example, the number of genes and their synteny is largely conserved (Cai et al. 2003; Sato 2011; Arisue et al. 2012). The phylogenetic distance between the 4 archigregarine lineages is actually larger than the distance between Hematozoa and Coccidia (Lax et al. 2024), so it is perhaps not surprising to find a more

divergent apicoplast structure within archigregarines, despite there only being 4 genera. While it is possible other mechanisms are at work (e.g. structural instability due to extremely low GC-content), we consider the considerable phylogenetic distances a key reason why archigregarine plastid genomes vary so much in terms of gene content and structure. Within genera, on the other hand, there is more evidence of synteny, particularly among *Lunidium* where we have the broadest sampling and the gene content and synteny are highly conserved between closely related taxa (*L. terebellae, L. melongena, L. shako,* and *L. laculatum*; Fig. 2). Beyond synteny, even these close relatives show a high degree of variation: genome sizes in *Lunidium* range from 27 to 31 kbp, and the GC-content is similarly variable between 10.6% and 14.7%. All archigregarine apicoplasts lack inverted repeats, but in *L. shako* and *L. laculatum*, the plastid SSU and LSU rRNA gene sequences also appear to be inverted (Fig. 2). Inverted repeats can be found in many apicomplexan lineages, but not all (Mathur, Kwong, et al. 2021). Curiously, presence or absence does not seem to follow any obvious pattern, phylogenetic or otherwise. The archigregarines appear to be a lineage fully missing inverted repeats on their plastid genomes.

There is also considerable variation in the number of predicted apicoplast genes between archigregarine genera (Fig. 2). This may be due to gene loss or transfer to the nucleus, but many unannotated ORFs may also represent highly divergent variants of genes identified in other archigregarine genomes.

All the apicoplast assemblies reported here share a complete lack of any photosynthesis-related genes, which is consistent with previous analyses of nucleus-encoded genes including members of *Lunidium*, *Metzidium*, *Selenidium*, and *Siedleckia* based on the hyperLOPIT plastid proteome of *Toxoplasma* (Mathur et al. 2023). The apicoplast genomes of some corallicolid apicomplexans have been shown to contain chlorophyll biosynthesis genes but none encoding photosystem proteins (Kwong et al. 2019; Bonacolta et al. 2024; Jacko-Reynolds et al. 2025), the distribution of which makes it possible they might also have been retained in archigregarines, but we found no evidence for any of these genes either.

Interestingly, *tufA* was not identified in either *Metzidium* sp. or *S. validusae*. This is not unprecedented, as *Rhytidocystis* sp. was also found to have lost *tufA* (a translation elongation factor), and there seems to be no evidence of its transfer to the nucleus (Janouškovec et al. 2019). It has been proposed that the transfer of *clpC* (a subunit of the ATP-dependent Clp protease) and *sufB* (a subunit of the iron–sulfur cluster assembly) genes to the nucleus is the last step to rendering the apicoplast genome dispensable (Janouškovec et al. 2015), and the *S. validusae* apicoplast is lacking *sufB*. We were not able to find any evidence of nuclear *sufB* sequences in the transcriptome (Lax et al. 2024) or nuclear genome sequences of *S. validusae*, which could be due to insufficient sequencing depth. If *clpC* were to be lost or moved to the nucleus, it suggests *S. validusae* would likely also lose its plastid genome. In all apicomplexan plastid genomes sequenced so far, *rpoC2* is split into *rpoC2A* and *rpoC2B* (Janouškovec et al. 2019), and we found this to also be the case in all our new assemblies, suggesting this is an ancestral feature of apicoplast genomes (Figs. 1 and 2).

Our *S. serpulae* data did not assemble into a circularized apicoplast (Fig. 1; supplementary fig. S2, Supplementary Material online). It is possible this genome is largely present in linear fragments, as they are in some proportion in *Toxoplasma* and *Plasmodium* (Williamson et al. 2001; Sato 2011), but it is also possible our assemblies did not circularize due to the extremely low GC-content (around 9%), repeats, and insufficient sequencing depth.

Implications for Understanding Apicomplexan Evolution

Why archigregarines (and presumably the related blastogregarines) have retained their plastid genomes while all other gregarines have lost it appears to be consistent with the hypothesis that the retention of *clpC* and *sufB* are sufficient and necessary reasons to maintain the genome. Whether some consistent factor other than simple chance led to the loss or transfer of these genes in other gregarines remains unclear; possibly the animal host environment is different, or availability of metabolites inside the host, or the physical relationship with the host cell (e.g. being predominantly inside it or outside) (Janouškovec et al. 2015). Such a factor might even explain other features, like the relatively normal state of archigregarine mitochondria, when compared with those of other gregarines (Mathur, Wakeman, et al. 2021; Keeling et al. 2024). However, there is no obvious candidate for such a factor to distinguish between eugregarine and archigregarine environments or ecology, as some eugregarines can even be found in the same annelid clades and same body part of the same host as archigregarines (Leander 2007; Park and Leander 2024).

The presence of plastid genomes in early branching archigregarines is nevertheless further direct evidence that the pattern of plastid loss and plastid genomes across Apicomplexa is very complex (Janouškovec et al. 2015, 2019; Mathur et al. 2019, 2023; Mathur, Kwong, et al. 2021), and this complexity extends to other functional features (Fig. 3). The convergent loss of photosynthesis genes is intriguing in particular: photosynthesis has been lost several times in related lineages, and in corallicolid apicomplexans the recent characterization of chlorophyll biosynthesis genes shows the pattern of loss is quite complex (Jacko-Reynolds et al. 2025).

It is currently unclear why some groups retain chlorophyll biosynthesis genes yet they have been disposed of in closely related groups. The gregarine ancestor possessed an apicoplast genome (likely without chlorophyll biosynthesis

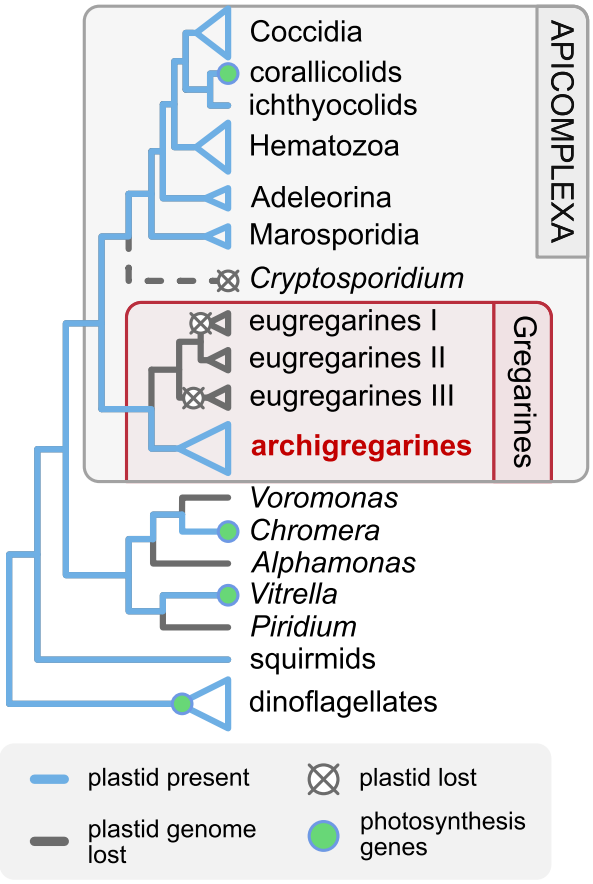

Fig. 3. Plastid loss among apicomplexan groups. Plastid presence and loss, genome loss, and photosynthesis genes are indicated on a tree of Apicomplexa. Archigregarines are in bold, and the uncertain, but possible placement of *Cryptosporidium* is marked with a dotted line (placement in concordance with Lax et al. 2024).

pathways), which was then subsequently fully lost in the eugregarines, but retained in archigregarines. Still, inferring plastid loss in all eugregarines is likely premature, since it is one of the largest and most diverse groups of apicomplexans, highlighting the need for exploration of plastid genomes across gregarines, and other apicomplexan lineages.

Materials and Methods

Sampling, Isolation, and Microscopy

Polychaetes (Sedentaria) and peanut worms (Sipuncula) were collected from marine benthic sites in British Columbia (Canada), Hokkaido (Japan), and Curaçao (supplementary table S1, Supplementary Material online). The intestines of the animals were removed and dissected in sterile 0.2 µm-filtered seawater, and the contents agitated by rapid pipetting. The resulting slurry was examined for archigregarines using a Leica DMIL inverted microscope. Identified cells were then isolated with a glass micropipette, washed 3-6 times in 0.2 µm-filtered seawater. A single cell was then deposited in ultrapure H_2O for single-cell genomics. These cells were isolated at the same time as cells destined for single-cell transcriptomics, as reported previously (Lax et al. 2024).

Whole Genome Amplification and Sequencing

Isolated cells were subjected to 2-4 freeze/thaw cycles (−80°C and room temperature). Genomic DNA of each cell was amplified with a TruePrime Single Cell WGA kit (4basebio, #351025) following the manufacturer's instructions, but increasing the amplification time to 12 h. Products were cleaned using Ampure XP beads and then quantified with a Qubit HS DNA assay. Libraries of the amplified DNA were prepared with Illumina DNA Prep and sequenced on several Illumina NextSeq 2 × 150 bp or MiSeq 2 × 250 bp runs (supplementary table S1, Supplementary Material online).

Genome Assembly

Illumina reads were trimmed with trimmomatic version 0.39 with: ILLUMINACLIP:NexteraPE-PE.fa:2:40:15 SLIDINGWINDOW:4:20 MINLEN:25 (Bolger et al. 2014) . These trimmed reads were assembled using SPAdes version 3.15.1 (Bankevich et al. 2012) using trimmed forward, reverse, and unpaired Illumina reads, generating co-assemblies from identical cells when possible (supplementary table S1, Supplementary Material online).

SSU rDNA Phylogenetics

SSU rDNA sequences of archigregarines were extracted from the genome assemblies with barrnap version 0.9 (https://github.com/tseemann/barrnap). To distinguish apicomplexan SSU rDNA sequences from contaminants, these extracted sequences were then checked against the NCBI nt database via megablast. Extracted sequences identified as apicomplexan were aligned with MUSCLE (Edgar 2004) against a previously published comprehensive archigregarine SSU rDNA dataset (Lax et al. 2024), and trimmed using Gblocks (Castresana 2000), as implemented in SeaView version 5.0.4 (Gouy et al. 2010). The final trimmed alignment consisted of 111 sequences with 1,128 sites. A phylogenetic tree was estimated with RAxML-NG version 1.1.0 under the GTR + G model and 1,000 nonparametric bootstrap replicates (Kozlov et al. 2019).

Plastid Genome Analysis

For initial identification of plastid contigs, we used BlobToolKit version 2.3.3 (Challis et al. 2020) on each assembly, running both blastx and diamondblast for taxonomic assignment. Putative plastid contigs were extracted based on their taxonomic assignment to Apicomplexa, their low GC content

(5% to 25%), and relatively high coverage. After extraction, we used MFannot (Lang et al. 2023) to check whether any plastid-encoded genes could be annotated on these contigs. The longest of these extracted and MFannot-confirmed contigs was then used as a seed for NOVOplasty version 4.3.1, a plastid-genome assembler (Dierckxsens et al. 2017). An exception was *S. validusae*, as we were able to recover a complete apicoplast genome sequence from the initial SPAdes assembly. For the NOVOplasty assemblies, we used all adapter-trimmed Illumina reads we had available from a single species, combining sequencing data from several runs of the same archigregarine.

We obtained several fully circularized plastid genomes, which were annotated with MFannot using genetic code 4. Additionally, ORFs of the apicoplast genome were predicted with transdecoder version 5.5.0 (Grabherr et al. 2011), and blasted against NCBI's nt database, and added to the final annotation. The annotations were manually curated in Geneious Prime version 2024.0.5. The *S. serpulae* assembly failed to produce a circularized genome, but we were able to recover several fragmented low-GC contigs from the initial SPAdes assembly, which were also annotated with MFannot (supplementary fig. S2, Supplementary Material online).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Mark Vermeij at CARMABI in Curaçao, and the Sequencing + Bioinformatics Consortium at the University of British Columbia. We also thank Oshoro Marine Station at Hokkaido University.

Funding

W.K.K. was supported by a Fundação para a Ciência e a Tecnologia grant (CEECIND/01358/2021). I.N. was supported by a Natural Sciences and Engineering Research Council of Canada Graduate Scholarship and the University of British Columbia's Four-Year Fellowship. V.K.L.J.-R. was supported by a University of British Columbia Indigenous Graduate Fellowship. K.C.W. was supported by the Japanese Society for the Promotion of Science (No. 18K14774). This work was supported by an Investigator Award to P.J.K. from the Gordon and Betty Moore Foundation (https://www.moore.org/grant-detail?grantId=GBMF9201).

Data Availability

Raw Illumina reads are deposited on NCBI SRA under BioProject PRJNA1127646. Nuclear SSU rDNA sequences are deposited on NCBI GenBank under accessions PQ586239 to PQ586245. Apicoplast genome assemblies are deposited on Borealis under accession.

Literature Cited

Arisue N, Hashimoto T, Mitsui H, Palacpac NMQ, Kaneko A, Kawai S, Hasegawa M, Tanabe K, Horii T. The plasmodium apicoplast genome: conserved structure and close relationship of *P. ovale* to rodent malaria parasites. Mol Biol Evol. 2012:29(9):2095–2099. https://doi.org/10.1093/molbev/mss082.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comp Biol. 2012:19(5):455–477. https://doi.org/10.1089/cmb.2012.0021.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014:30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bonacolta AM, Krause-Massaguer J, Smit NJ, Sikkel PC, del Campo J. A new and widespread group of fish apicomplexan parasites. Curr Biol. 2024:34(12):2748–2755.e3. https://doi.org/10.1016/j.cub.2024.04.084.

Cai X, Fuller AL, McDougald LR, Zhu G. Apicoplast genome of the coccidian *Eimeria tenella*. Gene. 2003:321:39–46. https://doi.org/10.1016/j.gene.2003.08.008.

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000:17(4):540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Cavalier-Smith T. Gregarine site-heterogeneous 18S rDNA trees, revision of gregarine higher classification, and the evolutionary diversification of Sporozoa. Eur J Protistol. 2014:50(5):472–495. https://doi.org/10.1016/j.ejop.2014.07.002.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. G3 (Bethesda). 2020:10(4):1361–1374. https://doi.org/10.1534/g3.119.400908.

Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017:45(4):e18–e18. https://doi.org/10.1093/nar/gkw955.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004:32(5):1792–1797. https://doi.org/10.1093/nar/gkh340.

Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010:27(2):221–224. https://doi.org/10.1093/molbev/msp259.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011:29(7):644–652. https://doi.org/10.1038/nbt.1883.

Jacko-Reynolds VKL, Kwong WK, Livingston SJ, Trznadel M, Lax G, Shivak J, Irwin NAT, Keeling PJ. Phylogenomics of coral-infecting corallicolids reveal multiple independent losses of chlorophyll biosynthesis in apicomplexan parasites. Curr Biol. 2025:35(5):1156–1163.e4. https://doi.org/10.1016/j.cub.2025.01.028.

Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc Natl Acad Sci U S A. 2010:107(24):10949–10954. https://doi.org/10.1073/pnas.1003335107.

Janouškovec J, Paskerova GG, Miroliubova TS, Mikhailov KV, Birley T, Aleoshin VV, Simdyanov TG. Apicomplexan-like parasites are

polyphyletic and widely but selectively dependent on cryptic plastid organelles. Elife. 2019:8:e49662. https://doi.org/10.7554/eLife.49662.
Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Kolisko M, Mylnikov AP, Keeling PJ. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. Proc National Acad Sci U S A. 2015:112(33):10200–10207. https://doi.org/10.1073/pnas.1423790112.
Keeling PJ, Mtawali M, Trznadel M, Livingston SJ, Wakeman KC. Parallel functional reduction in the mitochondria of apicomplexan parasites. Eur J Protistol. 2024:94:126065. https://doi.org/10.1016/j.ejop.2024.126065.
Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019:35(21):4453–4455. https://doi.org/10.1093/bioinformatics/btz305.
Kwong WK, Campo J, Mathur V, Vermeij MJA, Keeling PJ. A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes. Nature. 2019:568(7750):103–107. https://doi.org/10.1038/s41586-019-1072-z.
Lang BF, Beck N, Prince S, Sarrasin M, Rioux P, Burger G. Mitochondrial genome annotation with MFannot: a critical analysis of gene identification and gene model prediction. Front Plant Sci. 2023:14:1222186. https://doi.org/10.3389/fpls.2023.1222186.
Lax G, Park E, Na I, Jacko-Reynolds V, Kwong WK, House CSE, Trznadel M, Wakeman K, Leander BS, Keeling P. Phylogenomic diversity of archigregarine apicomplexans. Open Biol. 2024:14(9):240141. https://doi.org/10.1098/rsob.240141.
Leander BS. Molecular phylogeny and ultrastructure of *Selenidium serpulae* (Apicomplexa, Archigregarinia) from the calcareous tubeworm Serpula vermicularis (Annelida, Polychaeta, Sabellida). Zool Scr. 2007:36(2):213–227. https://doi.org/10.1111/j.1463-6409.2007.00272.x.
Leander BS. Marine gregarines: evolutionary prelude to the apicomplexan radiation? Trends Parasitol. 2008:24(2):60–67. https://doi.org/10.1016/j.pt.2007.11.005.
Mathur V, Kolisko M, Hehenberger E, Irwin NAT, Leander BS, Kristmundsson A, Freeman MA, Keeling PJ. Multiple independent origins of apicomplexan-like parasites. Curr Biol. 2019:29(17):2936–2941.e5. https://doi.org/10.1016/j.cub.2019.07.019.
Mathur V, Kwong WK, Husnik F, Irwin NAT, Kristmundsson A, Gestal C, Freeman M, Keeling PJ. Phylogenomics identifies a new Major subgroup of apicomplexans, marosporida class nov., with extreme apicoplast genome reduction. Genome Biol Evol. 2021:13(2):evaa244. https://doi.org/10.1093/gbe/evaa244.
Mathur V, Salomaki ED, Wakeman KC, Kwong NI, Kolisko M, Keeling PJ. Reconstruction of plastid proteomes of apicomplexans and close relatives reveals the major evolutionary outcomes of cryptic plastids. Mol Biol Evol. 2023:40(1):msad002. https://doi.org/10.1093/molbev/msad002.
Mathur V, Wakeman KC, Keeling PJ. Parallel functional reduction in the mitochondria of apicomplexan parasites. Curr Biol. 2021:31(13):2920–2928.e4. https://doi.org/10.1016/j.cub.2021.04.028.
Moore RB, Oborník M, Janouškovec J, Chrudimský T, Vancová M, Green DH, Wright SW, Davies NW, Bolch CJS, Heimann K, et al. A photosynthetic alveolate closely related to apicomplexan parasites. Nature. 2008:451(7181):959–963. https://doi.org/10.1038/nature06635.
Muñoz-Gómez SA, Durnin K, Eme L, Paight C, Lane CE, Saffo MB, Slamovits CH. Nephromyces represents a diverse and novel lineage of the Apicomplexa that has retained apicoplasts. Genome Biol Evol. 2019:11(10):2727–2740. https://doi.org/10.1093/gbe/evz155.
Na I, Campos C, Lax G, Kwong WK, Keeling PJ. Phylogenomics reveals Adeleorina are an ancient and distinct subgroup of Apicomplexa. Mol Phylogenetics Evol. 2024:195:108060. https://doi.org/10.1016/j.ympev.2024.108060.
Obornik M, Modrý D, Lukeš M, Černotíková-Stříbrná E, Cihlář J, Tesařová M, Kotabová E, Vancová M, Prášil O, Lukeš J. Morphology, ultrastructure and life cycle of vitrella brassicaformis n. sp., n. gen., a novel chromerid from the great barrier reef. Protist. 2012:163:306–323. https://doi.org/10.1016/j.protis.2011.09.001.
Park E, Leander B. Coinfection of slime feather duster worms (Annelida, Myxicola) by different gregarine apicomplexans (Selenidium) and astome ciliates reflects spatial niche partitioning and host specificity. Parasitology. 2024:151(4):400–411. https://doi.org/10.1017/S0031182024000209.
Paskerova GG, Miroliubova TS, Diakin A, Kováčiková M, Valigurová A, Guillou L, Aleoshin VV, Simdyanov TG. Fine structure and molecular phylogenetic position of two marine gregarines, Selenidium pygospionis sp. n. and S. pherusae sp. n., with notes on the phylogeny of archigregarinida (apicomplexa). Protist. 2018:169(6):826–852. https://doi.org/10.1016/j.protis.2018.06.004.
Rueckert S, Betts EL, Tsaousis AD. The symbiotic Spectrum: where do the gregarines fit? Trends Parasitol. 2019:35(9):687–694. https://doi.org/10.1016/j.pt.2019.06.013.
Rueckert S, Leander BS. Morphology and phylogenetic position of two novel marine gregarines (apicomplexa, eugregarinorida) from the intestines of North-eastern Pacific ascidians. Zoöl Scr. 2008:37(6):637–645. https://doi.org/10.1111/j.1463-6409.2008.00346.x.
Salomaki ED, Terpis KX, Rueckert S, Kotyk M, Varadínová ZK, Čepička I, Lane CE, Kolisko M. Gregarine single-cell transcriptomics reveals differential mitochondrial remodeling and adaptation in apicomplexans. BMC Biol. 2021:19(1):77. https://doi.org/10.1186/s12915-021-01007-2.
Sato S. The apicomplexan plastid and its evolution. Cell Mol Life Sci. 2011:68(8):1285–1296. https://doi.org/10.1007/s00018-011-0646-1.
Smith DR, Keeling PJ. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc Natl Acad Sci U S A. 2015:112(33):10177–10184. https://doi.org/10.1073/pnas.1422049112.
Wakeman KC. Molecular phylogeny of marine gregarines (apicomplexa) from the sea of Japan and the northwest Pacific including the description of three novel Species of selenidium and trollidium akkeshiense n. gen. n. sp. Protist. 2020:171(1):125710. https://doi.org/10.1016/j.protis.2019.125710.
Williamson DH, Denny PW, Moore PW, Sato S, McCready S, Wilson RJM. The in vivo conformation of the plastid DNA of Toxoplasma gondii: implications for replication. J Mol Biol. 2001:306(2):159–168. https://doi.org/10.1006/jmbi.2000.4385.
Williamson DH, Gardner MJ, Preiser P, Moore DJ, Rangachari K, Wilson RJM. The evolutionary origin of the 35 kb circular DNA of Plasmodium falciparum: new evidence supports a possible rhodophyte ancestry. Mol Gen Genet. 1994:243(2):249–252. https://doi.org/10.1007/BF00280323.
Yazaki E, Miyata R, Chikami Y, Harada R, Kawakubo T, Tanifuji G, Nakayama T, Yahata K, Hashimoto T, Inagaki Y. Signs of the plastid: enzymes involved in plastid-localized metabolic pathways in a eugregarine species. Parasitol Int. 2021:83:102364. https://doi.org/10.1016/j.parint.2021.102364.
Zhu G, Marchewka MJ, Keithly JS. Cryptosporidium parvum appears to lack a plastid genome. Microbiology. 2000:146(2):315–321. https://doi.org/10.1099/00221287-146-2-315.

Associate editor: Gwenael Piganeau