Research

Author for correspondence:
Gordon Lax
e-mail: gordon.lax@gmail.com

THE ROYAL SOCIETY PUBLISHING

Phylogenomic diversity of archigregarine apicomplexans

Gordon Lax[1], Eunji Park[1], Ina Na[1], Victoria Jacko-Reynolds[1], Waldan K. Kwong[4], Chloe S. E. House[1,2], Morelia Trznadel[1], Kevin Wakeman[5], Brian S. Leander[1,3] and Patrick Keeling[1]

[1]Department of Botany, [2]Department of Biochemistry and Molecular Biology, and [3]Department of Zoology, University of British Columbia, Vancouver, Canada
[4]Instituto Gulbenkian de Ciência, Oeiras, Portugal
[5]Institute for the Advancement of Higher Education, Hokkaido University, Sapporo, Hokkaido, Japan

GL, 0000-0003-3986-3068; PK, 0000-0002-7644-0745

Gregarines are a large and diverse subgroup of Apicomplexa, a lineage of obligate animal symbionts including pathogens such as *Plasmodium*, the malaria parasite. Unlike *Plasmodium*, however, gregarines are poorly studied, despite the fact that as early-branching apicomplexans they are crucial to our understanding of the origin and evolution of all apicomplexans and their parasitic lifestyle. Exemplifying this, the earliest branch of gregarines, the archigregarines, are particularly poorly studied: around 80 species have been described from marine invertebrates, but almost all of them were assigned to a single genus, *Selenidium*. Most are known only from light micrographs and largely unresolved rDNA phylogenies, where they exhibit a great deal of sequence variation, and fall into four subclades. To resolve the relationships within archigregarines, we sequenced 12 single-cell transcriptomes from species representing all four known subclades, as well as one blastogregarine (which frequently branch with *Selenidium*). A 190-gene phylogenomic tree confirmed four maximally supported individual clades of archigregarines and blastogregarines. These clades are discrete and distantly related, and also correlate with host identity. We propose the establishment of three novel genera of archigregarines to reflect their phylogenetic diversity and host range, and nine novel species isolated from a range of marine invertebrates.

1. Background

Apicomplexans are a diverse group of obligate symbionts of animals. Apart from biomedically relevant taxa like *Plasmodium* and *Toxoplasma*, the causative agents of malaria and toxoplasmosis, the vast majority of apicomplexans infect many different invertebrate animal phyla in marine and terrestrial environments, like corals [1], tunicates [2–4], crustaceans [5], annelids [6–8] and insects [9]. Many of these invertebrate parasites are gregarines, an early branching clade of apicomplexans [10,11] that probably occupy a range of symbioses from parasitism to commensalism [12]. Gregarines are very diverse, and the current state of their convoluted taxonomy and phylogeny reflects this [11,13]. A widely accepted view of gregarines classifies them into three distinct groups: eugregarines, blastogregarines and archigregarines [14]. Eugregarines probably represent the majority of gregarine diversity and include species with a very wide range of animal hosts from marine, freshwater and terrestrial environments [9,13,15]. There is evidence that the 'neogregarines', another historically common group, should be part of eugregarines based on molecular data [16]. In stark contrast to eugregarines, blastogregarines currently only contain three described species in the

genus *Siedleckia* and one in *Chattonaria* [14], which have been found to only infect Orbiniidae, a group of marine polychaetes. The known diversity of archigregarines is somewhat intermediate, with around 80 species, but they are currently virtually monogeneric—the vast majority of species are in the genus *Selenidium*, which contains upwards of 60 described species from a variety of different marine invertebrate hosts [7,8,17–21]. Most of the rest of the described species are nominally in another genus, *Selenidoides*, based on whether merogony was observed or not. The field has largely ignored this because of the dubious nature of this character [7,18], which is borne out by the fact that in small subunit ribosomal DNA (SSU rDNA) phylogenies, several species transferred to *Selenidoides* (*S. axiferens*, *S. hollandei* and *S. mesnili*) do not form a monophyletic group but are instead mixed with species in the genus *Selenidium* as has been proposed by Paskerova *et al.* [7] (see below).

Consequently, *Selenidium* represents a considerable amount of diversity: the various species are morphologically diverse, but this is most obvious at the molecular level, where different species of *Selenidium* can be more distantly related to each other than are entirely different classes of apicomplexans [6,21,22]. Despite recent advances, our understanding of the evolutionary relationships of archigregarines and blastogregarines remains murky at best, largely due to unresolved SSU rDNA phylogenies [7,23]. Four clades of archigregarines (all *Selenidium*) are routinely recovered in SSU phylogenies, but their phylogenetic relationships to each other, and to other gregarines, remain unclear [7,17,21,22]. These four correspond to the hosts they can be found in lineage Ag1 in several polychaete Sedentaria clades, Ag2 in sipunculids (peanut worms), Ag3 in cirratulids (a group of Sedentaria) and Ag4 in terebellids [7](another group of Sedentaria). Crucially, lineage Ag1 contains the type species *Selenidium pendula* Giard 1884, for which SSU rDNA data are available, making Ag1 the 'true' *Selenidium*.

While the divergence of the different *Selenidium* clades has been known for several years now, and efforts to further understand these relationships have been made by combining SSU and large subunit (LSU) rDNA phylogenies, this has fallen short of resolving the phylogenetic placement of archigregarines and blastogregarines [7,17]. This is probably due to the fact that archigregarine and blastogregarine species likely diverged a very long time ago, and the fact that the SSU and LSU rDNAs of some archigregarines are quite divergent, which is known to make accurate phylogenetic estimation difficult [11,17,24,25]. Multigene phylogenetics or phylogenomics have become a common way to resolve these difficult relationships. Recent studies using phylogenomics on apicomplexans have included a small number of species from three out of the four archigregarine groups [9,23], but were missing lineage Ag3, and other groups were mostly represented by only a single species. These phylogenies did not clearly resolve any relationships and presented only moderate support at best for any given topology among archigregarines and blastogregarines, and could not address whether the subgroups are each monophyletic.

To improve our understanding of archigregarine and blastogregarine phylogenetic relationships, we isolated single cells of 12 species of archigregarines representing all known four subclades, and an additional blastogregarine, and generated their single-cell transcriptomes. With these transcriptomes, we generated an updated SSU rDNA gene phylogeny to assess their relationship to species with available molecular data, as well as an updated 190-gene phylogenomic tree to provide a more comprehensive sampling of this group for the first time. We propose to transfer three of the four clades of *Selenidium* to three newly established genera *Lunidium, Devanium* and *Metzidium*, to reflect the diversity, phylogenetic relationships and host specificity of this group. We also propose eight new archigregarine species and one blastogregarine species.

2. Methods

2.1. Sampling, isolation and microscopy

Several host species of polychaetes (Sedentaria) and sipunculids were collected from marine benthic sites in British Columbia (Canada), Hokkaido (Japan) and Curaçao (electronic supplementary material, table S1). Host animals (table 1) were dissected and had their intestines removed, which were opened in sterile 0.2 µm filtered seawater and agitated by rapid pipetting. This slurry was examined for gregarines using a Leica DMIL inverted microscope. Once identified, gregarine cells were then isolated with a glass micropipette, washed 3–6 times in 0.2 µm filtered seawater, and imaged at 200× to 400× magnification with a Sony A7RIII camera. A single cell was then deposited in SmartSeq2 lysis buffer for single-cell transcriptomics [26].

2.2. Single-cell transcriptomics with SmartSeq2 and sequencing

Cells were subjected to two or three freeze/thaw cycles, after which cDNA was generated and amplified following the SmartSeq2 protocol [26], using 22–24 PCR cycles. The resulting cDNA was made into libraries with Illumina Nextera XT or Illumina DNA Prep kits, and sequenced on several Illumina NextSeq (2 × 150 bp paired end) or MiSeq runs (2 × 250 bp or 2 × 300 bp paired end; see electronic supplementary material, table S1). We isolated multiple cells (2–3) for half of the reported morphospecies and co-assembled them (see below).

2.3. Transcriptome assembly

Raw reads were subjected to read error correction with rcorrector v. 1.0.4 [27] using default settings. They were then adapter- and quality-trimmed with Trimmomatic v. 0.39 [28] using the TSO, oligo-dT, ISPCR, and transposase sequences and the following settings: ILLUMINACLIP:adapters.fasta:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:5:16 MINLEN:60.

Table 1. Isolated archigregarines and blastogregarines and their corresponding hosts, and whether COI host sequences were recovered. An asterisk (*) indicates COI sequence derived from PCR, all others were extracted from transcriptome assemblies using blastx. *Lunidium laculatum*, *L. shako* and *L. proboscidis* were all isolated from the same host individual.

species	host	host COI
Selenidium validusae	*Acrocirrus validus*	yes
Selenidium pherusae Ph226	Flabelligeridae sp.	yes*
Selenidium natalis nov. sp. SEL2980	*Spirobranchus giganteus*	–
Selenidium capillus nov. sp. Ph213	Cirratulidae sp.	yes*
Lunidium melongena	*Thelepus japonicus*	yes
Lunidium laculatum nov. sp. SNEK	*Eupolymnia* sp. (individual 1)	yes
Lunidium shako nov. sp. KNOB	*Eupolymnia* sp. (individual 1)	yes
Lunidium proboscidis nov. sp. SNEKD	*Eupolymnia* sp. (individual 1)	yes
Devanium robustum nov. sp.	*Cirratulus robustus*	yes
Devanium cincinnus nov. sp. Ph216	Cirratulidae sp.	yes*
Metzidium perlucensae nov. sp. SQU2901	*Phascolosoma perlucens*	yes
Siedleckia leitoscoloplosis nov. sp. BL3	*Leitoscoloplos pugettensis*	yes

Corrected and trimmed reads were assembled with rnaSPAdes v. 3.15.1 [29] using corrected and trimmed forward, reverse and unpaired Illumina reads. Where possible, we generated co-assemblies from identical cells isolated from the same host using rnaSPAdes (electronic supplementary material, table S1). Protein-coding sequences were predicted from the assemblies using Transdecoder v. 5.5.0 [30]. Mitochondrial cytochrome c oxidase (COI) sequences of the host animals were extracted from the protein-coding sequences with blastx, using publicly available polychaete and sipunculid COI sequences as queries. We conducted PCR using genomic DNA isolated from animal tissue to generate host COI sequences for some samples using primers LCO1490and HCO2198 (table 1), with conditions detailed in a previous study [31].

2.4. SSU rDNA phylogenetics

SSU rDNA sequences of archigregarines were extracted from transcriptome assemblies with barrnap v. 0.9 (https://github.com/tseemann/barrnap). To identify apicomplexan SSU rDNA sequences from contaminants, these extracted sequences were then checked against the NCBI nt database via megablast. We generated a comprehensive SSU rDNA dataset of publicly available apicomplexans, biased towards archigregarines and blastogregarines. Extracted sequences identified as apicomplexan were aligned with this dataset using MUSCLE [32] and trimmed using Gblocks [33], as implemented in SeaView v. 5.0.4 [34]. The final trimmed alignment consisted of 115 sequences with 1129 sites (accessions and lengths are listed in electronic supplementary material, table S2). A phylogenetic tree was estimated with RAxML-NG v. 1.1.0 under the GTR+G model and 1000 non-parametric bootstrap replicates [35].

2.5. Phylogenomics

To generate a multigene phylogeny, we used a previously published phylogenomic pipeline and 263-gene dataset of eukaryotes, with a comprehensive sampling of apicomplexans [9,23,36]. As part of this pipeline, single-gene trees for each of the 263 genes were generated and checked by hand for obvious contaminant, paralogous or otherwise aberrant sequences, which were then removed from the dataset. Only genes in which at least 40% of all taxa were present were used in the final concatenation, done with SCaFoS v. 1.25 [37].

A final dataset of 190 genes (38 577 sites) and 63 taxa (see electronic supplementary material, table S3 for details) was used to estimate a phylogenomic tree under the LG+C60+F+G model with 1000 Ultrafast bootstraps [38] (UFB) and 200 non-parametric bootstraps under LG+C60+F+G+PMSF [39], using IQTree2 [40]. To test the impact of missing data on our phylogenetic inference, we generated two additional datasets with the same 63 taxa but differing numbers of genes. The 129-gene dataset included all genes that had at least 70% taxon coverage, and the 22-gene dataset with at least 90% coverage across all taxa. Both of these datasets were run in IQTree2 under model LG+C60+F+G and 1000 UFBs.

We also conducted fast-site removal (FSR) and heterotacheous-site removal (HSR) analyses using scripts from PhyloFisher [41]. These scripts each removed 3000 sites at each step from our main dataset until exhaustion, and we estimated a LG+C20+F+G tree from each removal step with 1000 UFBs. The support values for relevant nodes from each step were then mapped against the number of removed sites. To reduce potential coding bias, we recoded our main 190-gene dataset to SR4, using the aa_recoder.py script in PhyloFisher, and ran the resulting matrix in IQTree2 under model GTR+R6+F and 1000 UFB. gGene concordance factor (gCF) and site concordance factor (sCF) analyses were run on our 190-gene dataset under default parameters in IQTree2 [42,43] (http://www.iqtree.org/doc/Concordance-Factor).

3. Results

3.1. Morphology

3.1.1. *Selenidium capillus* nov. sp.

The trophozoite is vermiform, 216 μm long and 14 μm wide, with a pointed anterior end (figure 1*a*). The cell is translucent, and the nucleus is situated around the midpoint of the cell. The cell moved by bending and twisting. Transcriptome data were derived from the cell imaged. Isolated from the intestinal lumen of polychaete worm Cirratulidae sp. (Annelida, Polychaeta, Sedentaria, Terebellida, Cirratulidae), which was collected from subtidal surface sediment (8–11 m below sea level) at Hyacinthe Bay, BC, Canada.

3.1.2. *Selenidium pherusae*

The trophozoite is vermiform, 212 μm long and 18 μm wide (at its mucron), with a club-shaped mucron (figure 1*b*). A nucleus was not clearly visible, and the cell moved by bending and twisting. Transcriptome data were derived from the cell imaged. Isolated from the intestinal lumen of a bristle-cage worm Flabelligeridae sp. (Annelida, Polychaeta, Sedentaria, Terebellida, Flabelligeridae), which was collected from subtidal surface sediment (8–11 m below sea level) at Hyacinthe Bay, BC, Canada.

3.1.3. *Selenidium natalis* nov. sp.

The trophozoite is >138 μm (cell ruptured during isolation) and 15 μm wide (figure 1*c*). The posterior forms into a pointed end, and the ovoid nucleus is situated around the midpoint of the cell. The cell is light brown in colour. Transcriptome data were derived from the cell imaged. Isolated from intestinal lumen of the Christmas tree worm *Spirobranchus giganteus* (Annelida, Polychaeta, Sedentaria, Sabellida and Serpulidae), which was found in the reef in front of CARMABI research station, Curaçao.

3.1.4. *Metzidium perlucensae* nov. gen. nov. sp.

The trophozoite is vermiform, 117 μm long and 23 μm wide at its widest point, with a pointed posterior end and rounded anterior end (figure 1*d*). Longitudinal folds cover the whole cell, and the elongated nucleus is situated around the midpoint of the cell, closer to the anterior end. The cell appears light brown, and transcriptome data were derived from the cell imaged. Isolated from the intestinal lumen of the peanut worm *Phascolosoma perlucens* Baird, 1868 (Annelida, Sipuncula, Phascolosomatidae), found in the reef in Piscadera Bay, in front of the CARMABI research station, Curaçao.

3.1.5. *Lunidium laculatum* nov. gen. nov. sp.

The vermiform trophozoites are 115–123 μm long and 17 μm wide (*n* = 2), with the ovoid nucleus situated around midpoint, closer to the anterior of the cell (figure 1*e–f*). Both anterior and posterior are narrowing to a blunt point. The cell is translucent and has clearly defined helical epicytic folds running across the surface, making the cell seem to have a cross-hatched pattern. Movement by bending and twisting. Two cells have been sequenced and their assemblies combined, but only one cell is pictured in figure 1. Isolated from the intestinal lumen of a terebellid polychaete *Eupolymnia* sp. (Annelida, Polychaeta, Sedentaria, Terebellida, Terebellidae), found in the reef in Piscadera Bay, in front of the CARMABI research station, Curaçao.

3.1.6. *Lunidium proboscidis* nov. gen. nov. sp.

The trophozoites are vermiform, 110–170 μm long, 18–19 μm wide (*n* = 2), either with a blunt, almost square mucron resembling an elephant's trunk (figure 1*g,h*) and a posterior ending in a blunt point, or both ends ending in a blunt point. The round nucleus is in the centre of the cell, its widest point. Distinct epicytic folds run longitudinally along the whole length of the cells. The cells move via bending and twisting. Two cells have been sequenced and their assemblies combined, but only one cell is pictured in figure 1. Isolated from the intestinal lumen of a terebellid polychaete *Eupolymnia* sp. (Annelida, Polychaeta, Sedentaria, Terebellida, Terebellidae), found in the reef in Piscadera Bay, in front of the CARMABI research station, Curaçao.

3.1.7. *Lunidium shako* nov. gen. nov. sp.

The vermiform trophozoites measure 108–143 μm in length, 13–19.5 μm in width (*n* = 2), with the anterior ending in a 'knob' or hat-like mucron (figure 1*i,j*) and the posterior ending in a blunt point. The oval nucleus is situated roughly at the midpoint of the cell, and distinct epicytic folds run across the whole length of the cell in a helical pattern. The cells move via bending and twisting. Both cells pictured in figure 1 have been sequenced and their assemblies combined. Isolated from the intestinal lumen of a terebellid polychaete *Eupolymnia* sp. (Annelida, Polychaeta, Sedentaria, Terebellida, Terebellidae), found in the reef in Piscadera Bay, in front of the CARMABI research station.
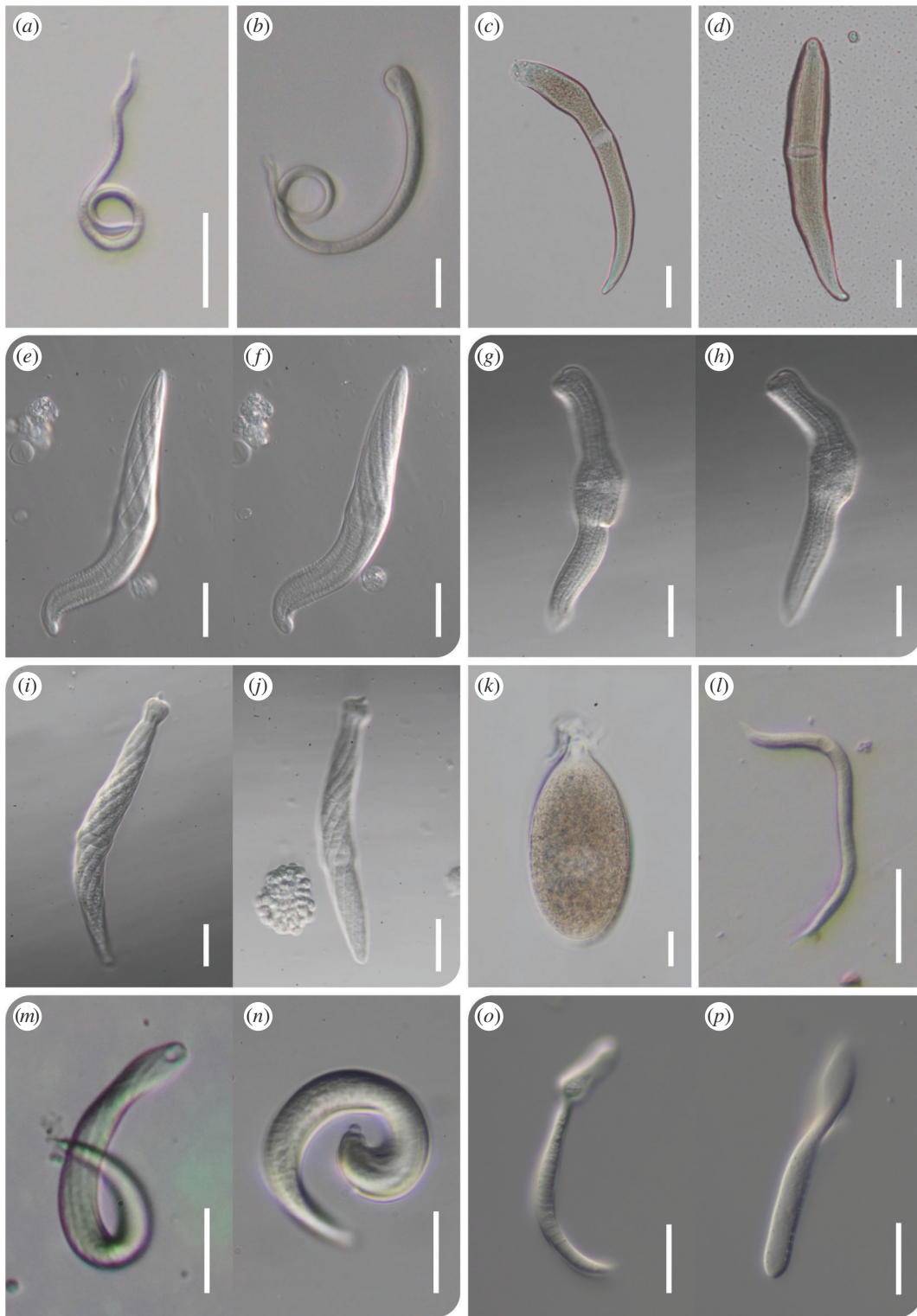
Figure 1. Micrographs of trophozoites of isolated archigregarines and blastogregarines. (*a*) *Selenidium capillus*, isolated cell. (*b*) *Selenidium pherusae* (Ph226), isolated cell. (*c*) *Selenidium natalis*, isolated cell. (*d*) *Metzidium perlucensae*, isolated cell. (*e,f*) *Lunidium laculatum*, same cell, one of three isolated and sequenced cells. (*g,h*) *Lunidium proboscidis*, same cell, one of two isolated and sequenced cells. (*i,j*) *Lunidium shako*, same cell, one of three isolated and sequenced cells. (*k*) *Lunidium melongena*. (*l*) *Devanium cincinnus*, isolated cell. (*m,n*) *Devanium robustum*, different cells from same host. (*o,p*) *Siedleckia leitoscoloplosis*, different cells from same host. Scale bars are 20 µm for all, except 50 µm for A, L and M.

3.1.8. *Lunidium melongena*

Trophozoites are oval, measuring 116–136 µm in length and 45–64 µm in width (*n* = 3), with a round nucleus in the centre of the cell (figure 1*k*). The mucron is 'neck-shaped' and often has host material still attached to it. The cells are commonly found in the coelom of the host, attaching to the intestine from the outside. Three cells were sequenced and coassembled, but only one is pictured in figure 1. The host *Thelepus japonicus* (Annelida, Polychaeta, Sedentaria, Terebellida, Terebellidae) was found in sediment between sea grass at Clover Point, Victoria, BC, Canada.

3.1.9. *Devanium cincinnus* nov. gen. nov. sp.

Trophozoite is vermiform and measures 183 × 9.5 μm, with both anterior and posterior ending in sharp points. Faint longitudinal epicytic folds run along the whole cell (figure 1*l*). The round nucleus is located in the anterior quarter of the cell, and the gregarine moves by bending and twisting. Transcriptome data were derived from the cell imaged. Isolated from a cirratulid bristle worm Cirratulidae sp. (Annelida, Polychaeta, Sedentaria, Terebellida, Cirratulidae) from subtidal surface sediment (8–11 m below sea level) at Hyacinthe Bay, BC, Canada.

3.1.10. *Devanium robustum* nov. gen. nov. sp.

Several trophozoites were found in the intestinal lumen of the same individual of the cirratulid bristle worm *Cirratulus robustus* (Annelida, Polychaeta, Sedentaria, Terebellida, Cirratulidae), with three cells isolated for single-cell transcriptomics. The cells are vermiform with a pointed posterior end, an anterior ending in a 'knob-like' mucron (figure 1*m,n*), and range from 104 to 215 μm in length (s.d. = 42.2 μm; *n* = 6) and 10.5–20 μm in width (s.d. = 3.2 μm; *n* = 6). Faint longitudinal epicytic folds are running down the whole cell. The cells move by twisting and bending. The host was found in sediment between sea grass at Clover Point, Victoria, BC, Canada.

3.1.11. *Siedleckia leitoscoloplosis* nov. sp.

The trophozoite is vermiform with a club-like mucron and measures 70–84 μm × 7.5–8 μm (*n* = 2; figure 1*o,p*). Some cell bodies appear 'layered' (figure 1*o*), and the mucron has a large vesicle with a granular interior. The cell is translucent and moves by bending and twisting. Two different cells are pictured, but only the cell in figure 1*o* was sequenced. Found in the polychaete *Leitoscoloplos pugettensis* (Annelida, Polychaeta, Sedentaria and Orbiniidae) which was collected from subtidal surface sediment (8–11 m below sea level) at Hyacinthe Bay, BC, Canada.

3.2. SSU rDNA phylogenetics

In our SSU rDNA tree (figure 2), archigregarines do not form a single clade but branch into four distinct lineages: *Selenidium* (lineage Ag1), *Metzidium* nov. gen. (Ag2), *Devanium* nov. gen. (Ag3) and *Lunidium* nov. gen. (Ag4). Each clade is receiving maximum bootstrap support, and blastogregarines are highly supported with 97% (*Siedleckia* and *Chattonaria*). The exact relationships between these five clades remain unclear since the backbone of the topology is completely unsupported (nodes without any reported bootstrap support are ≤50%).

3.3. Multigene phylogenetics

Our 190-gene phylogenomic tree shows both Apicomplexa and gregarines with full support (figure 3). Eugregarines and the archigregarine+blastogregarine group each form fully supported branches, but the archigregarine clade alone is poorly supported in our main analysis (44% UFB and 83% PMSF) and not at all in many of our additional analyses (22-gene = 18%, FSR = 0–93%, HSR = 0–31%,; electronic supplementary material, figures S2, S5 and S6). As in the SSU rDNA phylogeny, archigregarines separate into four distinct and maximally supported clades representing lineages Ag1-4, but the relationships between them are equally unclear since the archigregarine backbone is unsupported (UFB and PMSF values generally ≤70%). *Siedleckia* forms a fully supported branch that branches sister to the archigregarine group in our main analysis, although this relationship was not strongly supported. In most of our additional analyses *Siedleckia* branches inside archigregarines, sister to *Lunidium* with low to moderate support with UFB (129-gene = 76%, 22-gene = 59%, SR4 = 54%, HSR = 59–98%; electronic supplementary material, figures S1–S3 and S5). A gCF and site-concordance factor analysis (sCF) yielded low support values for the relationships between archigregarine genera and blastogregarines, mirroring the results from bootstrapping (electronic supplementary material, figure S4). gCF values stayed below 10% for any relationships between archi- and blastogregarine clades but ranged from 74% to 94% for individual clades. sCF values followed the same trend, with individual clades ranging from 59% to 93%, but inter-clade relationships hovering around the generally lowest possible value for sCF analyses [43] (32–36%).

4. Discussion

4.1. Archigregarines and blastogregarines are monophyletic and sister to eugregarines

Archigregarines are pivotal to our understanding of the early evolution of apicomplexans because of their phylogenetic position as sister to all other gregarines, which are together with *Cryptosporidium* the most early diverging lineages of apicomplexans [11]. Only recently have multigene phylogenetics been applied to diverse apicomplexans, and gregarines specifically [9,10,23,44,45], and some gregarine groups remain undersampled, including the archigregarines and blastogregarines. Some recent studies have included multiple species of archigregarines, but a broad sampling reflecting more accurately the diversity of the group has been missing up until now [10,23,45]. In our 190-gene tree, archigregarines and blastogregarines form a maximally supported clade, which in turn branches with eugregarines with full support.
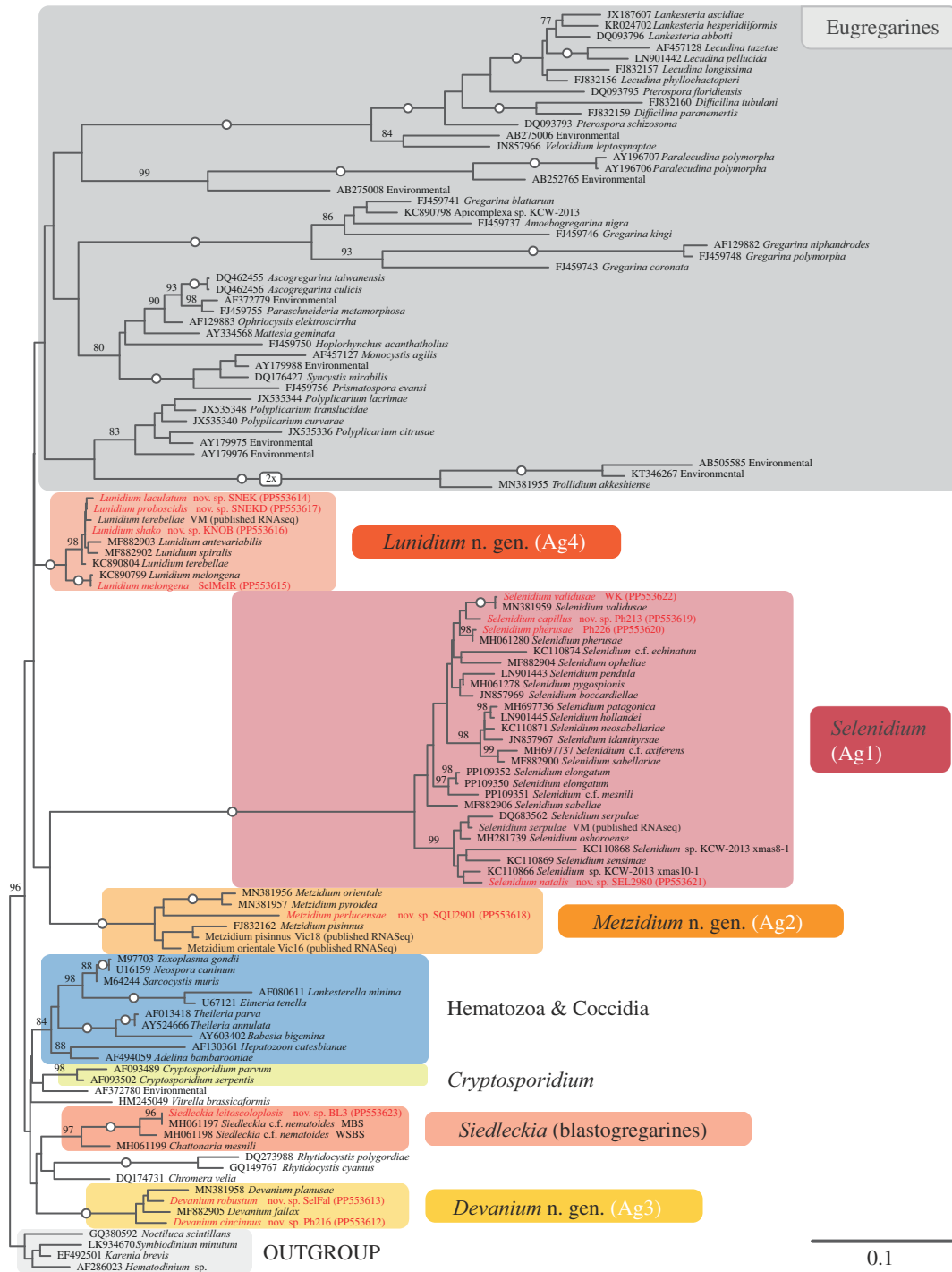
Figure 2. Maximum-likelihood SSU rDNA tree of apicomplexans, generated under the GTR+G model with 1000 non-parametric bootstrap replicates. Sequences from this study are in red and bold. Only support values >75% are shown, and maximally supported internodes (=100%) are marked with a circle.

The support for a monophyletic archigregarines is, by contrast, low, which is congruent with a recently published phylogeny that included a much smaller sampling and was missing *Devanium* [23]. It is noteworthy that all archi- and blastogregarine clades are relatively long-branching, but branches connecting the individual clades are short (figure 2; electronic supplementary material, figures S1–S3), which can complicate accurate phylogenetic estimation [42,43]. gCF/sCF yielded the same unresolved topology between clades in this part of the tree, as support values for gCF and sCF were low for the relationships between archigregarine clades but high for the individual clades (electronic supplementary material, figure S4). This probably means that no particular set of genes or sites is driving the observed topology, but that all underlying single-gene data are equally conflicting [42]. Supporting this is also the fact that the smaller and more restrictive 129-gene datasets do not increase support values among the archigregarine–blastogregarine clade significantly (electronic supplementary material, figures S1, 69–76% UFB). We found no evidence of amino-acid composition bias in any of the archigregarines and *Siedleckia* (electronic supplementary material, table S4, generated with PhyloFisher's aa_comp_calculator.py tool), which can influence and distort phylogenetic inference [46]. Overall, whether blastogregarines branch sister to archigregarines (as in our main phylogeny in figure 3) or alternatively fall within the archigregarines (as in most of our additional analyses in electronic supplementary material, figures S1, S5 and S6) is unclear, as was observed in a previous study [23]. If future analyses confirm blastogregarines branching within
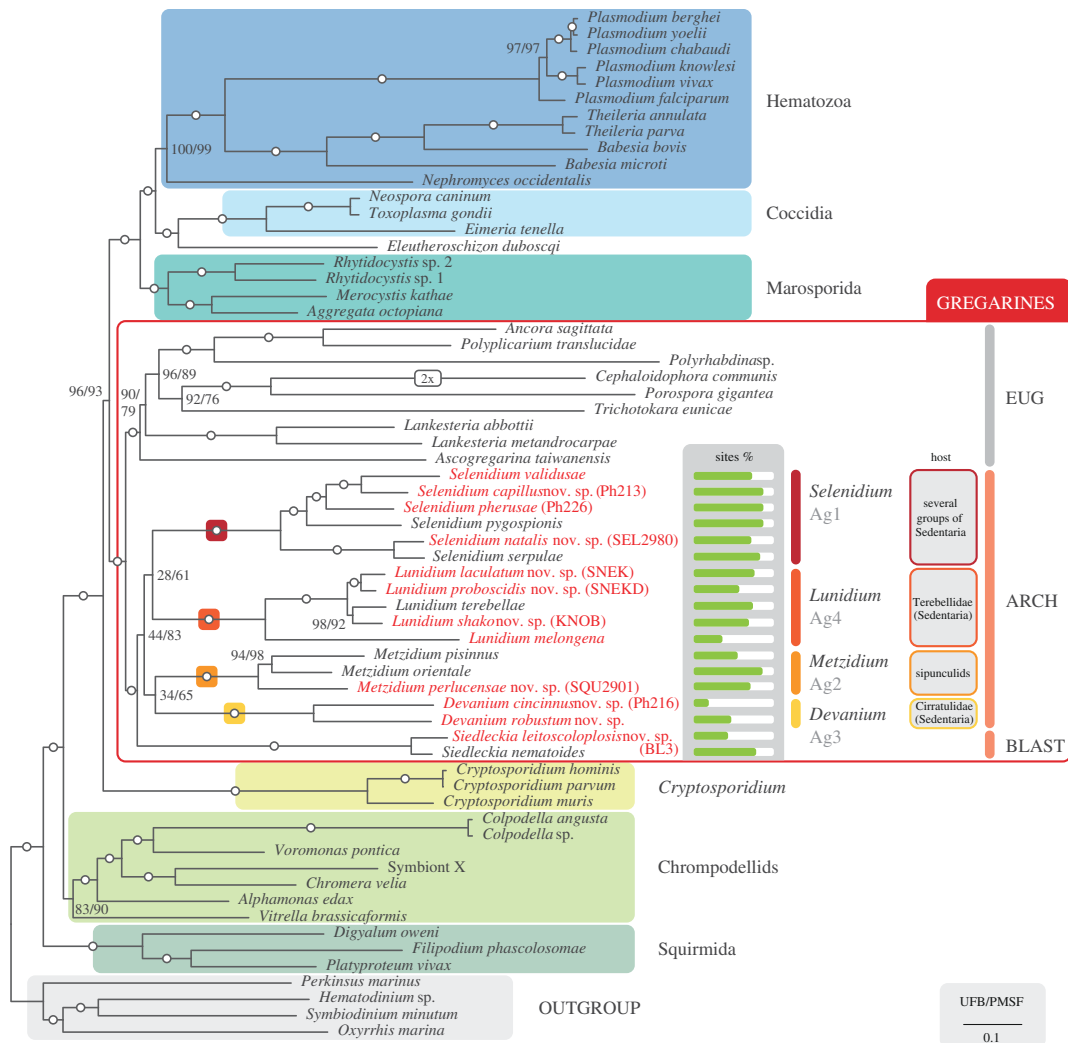
Figure 3. 190-gene phylogenomic tree of apicomplexans, generated under LG+C60+F+G with maximum likelihood. Newly added taxa are in red and bold. The first value on internodes is support from 1000 ultrafast bootstraps (UFB), the second from 200 non-parametric bootstraps under LG+C60+F+G+PMSF, and circles mark full bootstrap support with both. The percentage of sites used to estimate the tree is shown for archigregarines and blastogregarines (out of 38 577 amino acid sites).

archigregarines, the taxonomic descriptions of archigregarines (Archigregarinorida Grassé 1953) could be amended to include blastogregarines rather than keeping them as a separate taxonomic entity of similar level.

It is also possible that the phylogeny may be affected in the future by the discovery and characterization of archigregarine clades that have not yet been discovered and/or sequenced. Here the host-specificity may provide some clues: One species of archigregarine has been described from a hemichordate and two from ascidians [19], yet no molecular sequencing data exists for them, and the taxa in our tree have relatively strong host-specificity. Sequencing archigregarines from host taxa outside those investigated here is perhaps most likely to lead to new subgroups, which in turn may fill in gaps in the topology and help construct a more robust phylogeny, but for now the relationships remain unresolved.

4.2. New genera reflect the phylogenetic diversity of archigregarines

Each of the four clades of archigregarines identified in rDNA phylogenies is individually fully supported in phylogenomics, although the exact relationships between them remain unresolved [7,14,17,22]. When identifying these four clades, Paskerova *et al.* stopped short of any wider taxonomic changes, likely due to the lack of data, the fact that only two genes (large and small subunit rDNA) were available at the time, and that the whole group was not resolved in most of these phylogenies [7]. Indeed, the taxonomic confusion of archigregarines and *Selenidium* in particular has been known for several years now and has been noted in the literature repeatedly [7,8,11,18,21–23,47,48]. *Selenidium* is typically unresolved in SSU rDNA phylogenies and has accordingly been described as a 'paraphyletic stem group' [7,11,17,47]. Similarly, previous analyses have repeatedly shown that *Selenidium* is hyper-diverse for a genus, encompassing a much wider range of phylotypes than equivalent categories within the apicomplexans or even specifically other gregarine subgroups [7,17,47]. One extreme example of how diverse taxa were lumped within the genus *Selenidium* can be seen in *Platyproteum vivax*, which was initially described as *Selenidium vivax*, but was later transferred to its own novel genus [48], and later found to not even be an apicomplexan but rather a squirmid [49].

With multigene data for all four archigregarine lineages, we now have ample evidence for the hyper-diversity of the genus and reasons to split *Selenidium* based on the four individual well-supported clades in rDNA and multigene phylogenetics (figures 2 and 3), which also correspond to their host-specificity. We propose to retain *Selenidium* to represent lineage Ag1,

as it includes the type species *Selenidium pendula* Giard 1884, representing the 'true' Selenidiidae or *Selenidium* [7,21]. We also propose to erect three new archigregarine genera to represent the already recognized major subgroups: *Metzidium* nov. gen. for Ag2, *Devanium* nov. gen. for Ag3 and *Lunidium* nov. gen. for Ag4.

The genus *Selenidioides* was established to distinguish *Selenidium* from species that were similar but lacked merogony [19]. Considering proving the absence of a life cycle stage is near impossible, the field has largely abandoned the usage of the genus *Selenidioides* [7,50]. Our SSU rDNA phylogeny reinforces this because several species nominally transferred to *Selenidioides* (*S. axiferens*, *S. hollandei* and *S. mesnili*) do not form a clade. There is no molecular data for the type species of this disputed genus, *S. caulleryi*, but it is difficult to imagine *Selenidioides* to be a viable genus based on existing data.

Support for each archigregarine lineage and thus distinct genera is strong based on phylogenies and their distinct host specificities. We collated morphological observations from our own and published data and could not find any set of morphological characters that are unique to any genus of archigregarines (electronic supplementary material, table S5). This is not to say none exist, but the current set of data does not show any clear pattern. Additional studies using high-quality light microscopy, scanning electron microscopy (SEM) and transmission electron microscopy (TEM) are needed to explore any potential morphological differences between these genera. For example, the occurrence, structure and organization of epicytic folds, and mucrons can be important morphological markers for archigregarines. Likewise observation of more than a single developmental stage is crucial in further refining taxonomic diagnoses. This is to say our current understanding of the morphological diversity of archigregarines does not reflect the diversity at the molecular level.

The genetic distances between SSU rDNA sequences of different archigregarine genera vary considerably, falling between 65% and 81% similarity (electronic supplementary material, table S6). Genetic distances within genera are lower but still vary considerably, underlining the divergent nature of archigregarine clades. Within *Selenidium* Ag1, for example, SSU rDNA sequences of different species range between 81% and 99% similarity, whereas they range from 79 to 93% in *Metzidium* Ag2.

4.3. Host specificity in archigregarine lineages

The four archigregarine subgroups each have a relatively high degree of host-specificity based on current data. Interestingly, the host specificity of *Selenidium* (Ag1) is less strict compared to the other archigregarine lineages—all its hosts are part of Sedentaria (Annelida), a group with over 5000 described species (as per World Polychaete Database, retrieved in May 2024). Ag1 includes specific subgroups of sedentarian hosts that also include hosts of *Devanium* (Ag3) and *Lunidium* (Ag4), and it is possible that a single host species can harbour multiple different archigregarine subgroups (see below). The other three archigregarine lineages are highly host specific. *Metzidium* (Ag2) is the only archigregarine lineage found in peanut worms, annelids that are distantly related to sedentarian polychaetes [48]. *Devanium* (Ag3) and *Lunidium* (Ag4) are to date only found in Cirratulidae and Terebellidae, respectively. *Selenidium* Ag1 also stands out because it has a high degree of both sequence and host diversity and considerably more described species (21 in *Selenidium* Ag1, 4 in *Metzidium* Ag2, 4 in *Devanium* Ag3 and 7 in *Lunidium* Ag4). Interestingly, Ag1 appears to be composed of at least two subclades, perhaps reflecting the larger number of taxa. A larger subclade that includes *Selenidium pendula*, *S. validusae*, *S. pherusae* and *S. capillus*, and a smaller one with *Selenidium serpulae* and *S. natalis* (figure 2). In our multigene tree, both of these clades are maximally supported (figure 3).

Cells of *Selenidium capillus* Ph213 and *Devanium cincinnus* Ph216 were isolated from different individuals of the same host, a cirratulid polychaete. While there is no close hit of the host COI to any sequence in public databases, the COIs from Ph213 and Ph216 are almost identical with 99.4% similarity. While uncommon, several other examples of divergent archigregarine species and even lineages occupying the same host are documented. Two species of *Selenidium* were found in the same species of slime feather duster worm, inhabiting different parts of the gut lumen [31]. *Lunidium terebellae* and *L. melongena* are likewise both found in the spaghetti worm *Thelepus japonicus* [22], with the former inhabiting the gut lumen, whereas the latter is found in the coelom. Christmas tree worms appear to be reservoirs of closely related species of *Selenidium*, with one study finding at least three different species in the intestines of *Spirobranchus giganteus* [8], in addition to *Selenidium natalis* nov. sp. SEL2980 (table 1). The symbiotic polychaete *Spirobranchus* itself represents a species complex [51], with different individuals showing genetic difference, which could lead to a different set of archigregarine symbionts, highlighting the need for increased sampling of marine invertebrates for gregarines.

4.4. Novel archigregarine species

Sampling increasingly larger numbers of marine invertebrates also increases the number of gregarines found. While there are over 50 described species of archigregarines [19], molecular data are only available for some, and in most cases, the SSU rRNA gene only. Many already described archigregarines have been sequenced [6,17,21] but also new ones described [7,8,20,22,31,48]. The number of both of these cases is expected to rise. We found several archigregarines in previously unsampled polychaete and sipunculid hosts that fit no existing description. In all of these cases, the SSU rDNA phylogeny shows their phylogenetic distance to related organisms, supporting our decision to assign them as new species.

The three *Lunidium* morphotypes SNEK, KNOB and SNEKD were all isolated from the same host individual, a *Eupolymnia* sp. terebellid worm from Curaçao. We propose new species for all three morphotypes, based on their morphological differences and phylogenetic distances: *Lunidium laculatum* SNEK, *L. shako* KNOB and *L. proboscidis* SNEKD (figures 1–3). The SSU sequence of *L. laculatum* is 93.6% similar to *L. proboscidis* SNEKD and only 90.2% similar to *L. shako*, indicating they should all be considered distinct species. *L. laculatum* has strong helical epicytic folds going across the cell, whereas *L. shako* has a conspicuous 'knobby' mucron, and *L. proboscidis* SNEKD lacks both the knobby mucron and does not show the helical striations (figure 1).

For each of these *Lunidium* species, we sequenced 2–3 cells separately and co-assembled them based on their morphology and SSU rDNA differences (electronic supplementary material, table S1) further providing support that these are novel species.

It should be noted that our observations of the morphology of all newly described species are relatively limited due to our sampling and imaging approach. In most cases, only a single cell to a handful were observed for a given species, and only at a singular life cycle stage (the trophozoite). This limits our understanding of the morphologies of these species, particularly in terms of cell size and observations of fine details like epicytic folds. Nevertheless, we consider them phylogenetically distinct from already described species and as such they merit novel descriptions, and we believe the standard of evidence is high, with imaging from light microscopy together with megabases of sequencing and thousands of genes.

5. Conclusions

Generating single-cell transcriptomes of 12 archigregarine and one blastogregarine species enabled us to generate the most comprehensive multigene dataset of archigregarines to date. While our 190-gene phylogenomic tree and additional phylogenomic analyses did not confidently resolve any branching order among archigregarines, together with existing SSU rDNA phylogeny data, it is now obvious that there are at least four subclades of archigregarines. This highlights the long-noted problem that all archigregarine diversity has been classified as a single genus, *Selenidium*, so we propose three additional genera: *Lunidium*, *Metzidium* and *Devanium*. These genera correspond to lineages identified in multiple published SSU rDNA phylogenetic studies, they are strongly supported in multigene phylogenies and they correspond to strong host-specific patterns. The early-diverging archigregarines and blastogregarines remain a crucial and intriguing group to study, and additional taxon sampling is probably needed to resolve their phylogeny. Another piece of the evolutionary puzzle is whether archigregarines have an apicoplast genome.

6. Formal taxonomic description

Apicomplexa (Levine 1970)
    Gregarinea (Bütschli 1882, stat. nov. Grassé 1953)
    Archigregarinorida (Grassé 1953)

6.1. *Lunidium* nov. gen. Lax & Keeling 2024

Description: A crown clade comprising *Lunidium terebellae*, *L. antevariabilis*, *L. spiralis*, *L. melongena* and *L. laculatum*. Archigregarines that infect the instestinal lumen of polychaetes of the marine invertebrate family Terebellidae Grube 1851 (Annelida, Sedentaria, Terebellida). This definition corresponds to lineage Ag4, outlined in [14].

    Type species: *Selenidium terebellae* Ray 1930 (=*L. terebellae*, comb. nov.).

    Etymology: Named after 'Luna', an ancient Roman moon deity, referring to the basionym *Selenidium* (Giard 1884), which was likely named after the moon deity 'Selene' in ancient Greek mythology.

    Transfer of existing species to Lunidium: *L. antevariabilis* (Rueckert & Horák 2017 [18]), *L. melongena* (Wakeman, Heintzelman & Leander 2014 [22]), *L. spiralis* (Rueckert & Horák 2017 [18]) and *L. terebellae* (Ray 1930 [22]) (basionym for all: *Selenidium*; all comb. nov.).

6.2. *Metzidium* nov. gen. Lax, Jacko-Reynolds & Keeling 2024

Description: A crown clade comprising *Metzidium pisinnus*, *M. orientale*, *M. pyroidea* and *M. perlucensae*. Archigregarines that infect the instestinal lumen of marine invertebrates of the class Sipuncula Rafinesque 1814 (Annelida). This definition corresponds to lineage Ag2, outlined in [14].

    Type species: *Selenidium pisinnus* (Rueckert & Leander 2009 [48]) (=*M. pisinnus*, comb. nov.).

    Etymology: Named after 'Mētztli' (Nahuatl) or 'Metzi', a moon deity in Aztec mythology, referring to the basionym *Selenidium* (Giard 1884), which was likely named after the moon deity 'Selene' in ancient Greek mythology.

    Transfer of existing species to Metzidium: *M. orientale* (Bogolepova 1953), *M. pisinnus* [48] and *M. pyroidea* (Wakeman 2019) (basionym for all: *Selenidium*; all comb. nov.).

6.3. *Devanium* nov. gen. Lax, Park, Na & Keeling 2024

Description: A crown clade comprising *Devanium planusae*, *D. fallax*, *D. robustum* and *D. cincinnus*. Archigregarines that infect the instestinal lumen of polychaetes of the marine invertebrate family Cirratulidae Carus 1863 (Annelida, Sedentaria, Terebellida). This definition corresponds to lineage Ag3, outlined in [14].

    Type species: *Selenidium planusae* Wakeman 2019 (=*D. planusae*, comb. nov.).

    Etymology: Named after 'Dziewanna' (Polish) or 'Devana', a Western Slavic moon deity of forests, hunting and the moon. Refers to the basionym. *Selenidium* (Giard 1884), which was likely named after the moon deity 'Selene' in ancient Greek mythology.

Transfer of existing species to Devanium: *D. fallax* (MacGregor and Thomasson 1965), *D. planusae* (Wakeman 2019) (basionym for all: *Selenidium*; all comb. nov.).

6.4. *Selenidium capillus* nov. sp. Lax, Park & Keeling 2024

Archigregarinorida Grassé 1953
  Selenidiidae Brasil 1907
  Selenidium Giard 1884
  Description: Trophozoite is vermiform, 216 µm long and 14 µm wide, with a pointed anterior end. The cell is translucent, and the nucleus is situated around the midpoint of the cell. Movement by bending and twisting.
  DNA sequence: SSU rRNA gene sequence, GenBank accession PP553619.
  Type locality: Hyacinthe Bay, Quadra Island, British Columbia, Canada.
  Type habitat: Marine.
  Type host: Cirratulidae sp. (Annelida, Polychaeta, Sedentaria, Terebellida, Cirratulidae).
  Location in host: Intestinal lumen.
  Type material: Cell depicted in figure 1*a*.
  Etymology: From Latin 'capillus' for 'hair', referring to the cirratulid host having tentacles, or 'hair'.

6.5. *Selenidium natalis* nov. sp. Lax, Jacko-Reynolds & Keeling 2024

Archigregarinorida Grassé 1953
  Selenidiidae Brasil 1907
  Selenidium Giard 1884
  Description: The trophozoite cell is >138 µm (cell ruptured during isolation) and 15 µm wide. The posterior forms into a pointed end, and the ovoid nucleus is situated around the midpoint of the cell. The cell is light brown in colour.
  DNA sequence: SSU rRNA gene sequence, GenBank accession PP553621.
  Type locality: Reef in front of CARMABI research station, Curaçao.
  Type habitat: Marine.
  Type host: *Spirobranchus giganteus* Pallas, 1766 (Annelida, Polychaeta, Sedentaria, Sabellida and Serpulidae).
  Location in host: Intestinal lumen.
  Type material: Cell depicted in figure 1*c*.
  Etymology: From Latin 'natalis' meaning 'of one's birth', commonly used as a meaning of Christmas, referring to the common name for the host, Christmas tree worm.

6.6. *Lunidium laculatum* nov. sp. Lax & Keeling 2024

Archigregarinorida Grassé 1953
  Lunidium Lax & Keeling 2024
  Description: The vermiform trophozoite is 115–123 µm long and 17 µm wide, with the ovoid nucleus situated around midpoint, closer to the anterior of the cell. Both anterior and posterior are narrowing to a blunt point. The cell is translucent and has clearly defined helical epicytic folds running across the surface, making the cell seem to have a cross-hatched pattern. Movement by bending and twisting.
  DNA sequence: SSU rRNA gene sequence, GenBank accession PP553614.
  Type locality: Reef in front of CARMABI research station, Curaçao.
  Type habitat: Marine.
  Type host: *Eupolymnia* sp. Verrill 1900 (Annelida, Polychaeta, Sedentaria, Terebellida, Terebellidae).
  Location in host: Intestinal lumen.
  Type material: Cell depicted in figure 1*e,f*.
  Etymology: From Latin 'laculatum' meaning 'checkered, checked, four-cornered', referring to epicytic folds in some cells rendering a checkered pattern visible with light microscopy.
  Notes: SSU rRNA gene sequence 93.6% similar to *L. proboscidis* and 90.2% similar to *L. shako*.

6.7. *Lunidium shako* nov. sp. Lax & Keeling 2024

Archigregarinorida Grassé 1953
  Lunidium Lax & Keeling 2024
  Description: Ttrophozoites are vermiform and measure 108–143 µm in length, 13–19.5 µm in width. The anteriors ends in a 'knob' or hat-like mucron (figure 1*i,j*), and the posterior ends in a blunt point. Distinct epicytic folds run across the whole length of the cell in a helical pattern. The oval nucleus is at the midpoint of the cell. Movement via bending and twisting.
  DNA sequence: SSU rRNA gene sequence, GenBank accession PP553616.
  Type locality: Reef in front of CARMABI research station, Curaçao.

Type habitat: Marine.

Type host: *Eupolymnia* sp. Verrill 1900 (Annelida, Polychaeta, Sedentaria, Terebellida, Terebellidae).

Location in host: Intestinal lumen.

Type material: Cell depicted in figure 1*i,j*.

Etymology: From 'shako' (from Hungarian 'csákó'), a cylindrical cap used in several militaries, which the mucron of this species resembles.

Notes: SSU rRNA gene sequence 90.2% similar to *Lunidium laculatum* and 89.4% similar to *Lunidium proboscidis.*

6.8. *Lunidium proboscidis* nov. sp. Lax & Keeling 2024

Archigregarinorida Grassé 1953

Lunidium Lax & Keeling 2024

Description: Trophozoites are vermiform, 110–170 μm long, 18–19 μm wide, with a blunt, almost square mucron resembling an elephant's trunk, and a widened central part of the cell. The round nucleus sits in the centre of the cell. Epicytic folds run longitudinally along the whole length of the cells. Movement through bending and twisting.

DNA sequence: SSU rRNA gene sequence, GenBank accession PP553617.

Type locality: Reef in front of CARMABI research station, Curaçao.

Type habitat: Marine.

Type host: *Eupolymnia* sp. Verrill 1900 (Annelida, Polychaeta, Sedentaria, Terebellida, Terebellidae).

Location in host: Intestinal lumen.

Type material: Cell depicted in figure 1*g,h*.

Etymology: From Latin 'proboscis' for snout and the trunk of an elephant because the tip of the mucron resembles an elephant's trunk.

Notes: SSU rRNA gene sequence 93.6% similar to *L. laculatum* and 89.4% similar to *L. shako.*

6.9. *Metzidium perlucensae* nov. sp. Lax, Jacko-Reynolds & Keeling 2024

Archigregarinorida Grassé 1953

Metzidium Lax, Jacko-Reynolds & Keeling 2024

Description: Trophozoite cell is vermiform, 117 μm long and 23 μm wide at its widest point, the posterior ending in a point, the anterior in a rounded blunt point. Fine longitudinal striations cover the whole cell, and the elongated nucleus is situated around the midpoint of the cell, closer to the anterior end. The cell appears light brown.

DNA sequence: SSU rRNA gene sequence, GenBank accession PP553618.

Type locality: Reef in front of CARMABI research station, Curaçao.

Type habitat: Marine.

Type host: *Phascolosoma perlucens* Baird, 1868 (Annelida, Sipuncula, Phascolosomatidae).

Location in host: Intestinal lumen.

Type material: Cell depicted in figure 1*d*.

Etymology: From Latin 'perlucens' referring to the host *Phascolosoma perlucens* Baird, 1868.

6.10. *Devanium robustum* nov. sp. Lax, Na & Keeling 2024

Archigregarinorida Grassé 1953

Devanium Lax, Park & Keeling 2024

Description: Trophozoite cells measured are 104–215 μm in length and 10.5–20 μm in width. Cells are vermiform with a pointed posterior end and an anterior ending in a capitulum or 'knob-like' mucron. Faint longitudinal epicytic folds are running down the whole cell. The cells move by twisting and bending.

DNA sequence: SSU rRNA gene sequence, GenBank accession PP553613.

Type locality: Clover Point, Victoria, British Columbia, Canada.

Type habitat: Marine.

Type host: *Cirratulus robustus* Johnson, 1901 (Annelida, Polychaeta, Sedentaria, Terebellida, Cirratulidae).

Location in host: Intestinal lumen.

Type material: Cell depicted in figure 1*m*.

Etymology: From Latin 'robustum' meaning 'hard, solid', referring to the host species *Cirratulus robustus* Johnson, 1901.

6.11. *Devanium cincinnus* nov. sp. Lax, Park & Keeling 2024

Archigregarinorida Grassé 1953

Devanium Lax, Park & Keeling 2024

Description: Trophozoite is vermiform and measures 183 μm × 9.5 μm, with both anterior and posterior ending in sharp points. Faint longitudinal epicytic folds run along the whole cell. The round nucleus is located in the anterior quarter of the cell. Movement by bending and twisting.

DNA sequence: SSU rRNA gene sequence, GenBank accession PP553612.

Type locality: Hyacinthe Bay, Quadra Island, British Columbia, Canada.

Type habitat: Marine.

Type host: Cirratulidae sp. (Annelida, Polychaeta, Sedentaria, Terebellida, Cirratulidae).

Location in host: Intestinal lumen.

Type material: Cell depicted in figure 1*l*.

Etymology: From Latin 'cincinnus' meaning 'lock of hair', referring to the host having curly tentacles (Cirratulidae sp.).

6.12. *Siedleckia leitoscoloplosis* nov. sp. Lax, Park & Keeling 2024

Blastogregarinorina Chatton & Villeneuve 1936

  Siedleckiidae Chatton & Villeneuve 1936

  Siedleckia Caullery & Mesnil 1898

  Description: The trophozoite is vermiform with a club-like mucron and measures 70–84 μm × 7.5–8 μm. Several ovoid nuclei are stacked on top of each other throughout the cell body, and the mucron has a large vesicle with a granular interior. The cell is translucent and moves by bending and twisting.

DNA sequence: SSU rRNA gene sequence, GenBank accession PP553623.

Type locality: Hyacinthe Bay, Quadra Island, British Columbia, Canada.

Type habitat: Marine.

Type host: *Leitoscoloplos pugettensis* Pettibone, 1957 (Annelida, Polychaeta, Sedentaria, Orbiniidae).

Location in host: Intestinal lumen.

Type material: Cell depicted in figure 1o.

Etymology: From Greek 'leitoscoloplos' referring to the host *Leitoscoloplos pugettensis* Pettibone, 1957.

References

1. Kwong WK, Del Campo J, Mathur V, Vermeij MJA, Keeling PJ. 2019 A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes. *Nature* **568**, 103–107. (doi:10.1038/s41586-019-1072-z)

2. Saffo MB, McCoy AM, Rieken C, Slamovits CH. 2010 Nephromyces, a beneficial apicomplexan symbiont in marine animals. *Proc. Natl Acad. Sci. USA* **107**, 16190–16195. (doi:10.1073/pnas.1002335107)

3. Rueckert S, Leander BS. 2008 Morphology and phylogenetic position of two novel marine gregarines (Apicomplexa, Eugregarinorida) from the intestines of north-eastern Pacific ascidians. *Zool. Scr.* **37**, 637–645. (doi:10.1111/j.1463-6409.2008.00346.x)

4. Rueckert S, Wakeman KC, Jenke-Kodama H, Leander BS. 2015 Molecular systematics of marine gregarine apicomplexans from Pacific tunicates, with descriptions of five novel species of *Lankesteria*. *Int. J. Syst. Evol. Microbiol.* **65**, 2598–2614. (doi:10.1099/ijs.0.000300)

5. Rueckert S, Simdyanov TG, Aleoshin VV, Leander BS. 2011 Identification of a divergent environmental DNA sequence clade using the phylogeny of gregarine parasites (Apicomplexa) from crustacean hosts. *PLoS One* **6**, e18163. (doi:10.1371/journal.pone.0018163)

6. Rueckert S, Glasinovich N, Diez ME, Cremonte F, Vázquez N. 2018 Morphology and molecular systematic of marine gregarines (Apicomplexa) from Southwestern Atlantic spionid polychaetes. *J. Invertebr. Pathol.* **159**, 49–60. (doi:10.1016/j.jip.2018.10.010)

7. Rueckert S, Glasinovich N, Diez ME, Cremonte F, Vázquez N. 2018 Morphology and molecular systematic of marine gregarines (Apicomplexa) from Southwestern Atlantic spionid polychaetes. *J. Invertebr. Pathol.* **159**, 49–60. (doi:10.1016/j.protis.2018.06.004)

8. Wakeman KC, Leander BS. 2013 Molecular phylogeny of marine gregarine parasites (Apicomplexa) from tube-forming polychaetes (Sabellariidae, Cirratulidae and Serpulidae), including descriptions of two new species of *Selenidium*. *J. Eukaryot. Microbiol.* **60**, 514–525. (doi:10.1111/jeu.12059)

9. Na I, Campos C, Lax G, Kwong WK, Keeling PJ. 2024 Phylogenomics reveals adeleorina are an ancient and distinct subgroup of Apicomplexa. *Mol. Phylogenet. Evol*. **195**, 108060. (doi:10.1016/j.ympev.2024.108060)

10. Janouškovec J, Paskerova GG, Miroliubova TS, Mikhailov KV, Birley T, Aleoshin VV, Simdyanov TG. 2019 Apicomplexan-like parasites are polyphyletic and widely but selectively dependent on cryptic plastid organelles. *Elife* **8**, e49662. (doi:10.7554/eLife.49662)

11. Leander BS. 2008 Marine gregarines: evolutionary prelude to the apicomplexan radiation? *Trends Parasitol*. **24**, 60–67. (doi:10.1016/j.pt.2007.11.005)

12. Rueckert S, Betts EL, Tsaousis AD. 2019 The symbiotic spectrum: where do the gregarines fit? *Trends Parasitol*. **35**, 687–694. (doi:10.1016/j.pt.2019.06.013)

13. Archibald JM, Simpson AG, Slamovits CH. 2017 *Handbook of the protists*, pp. 567–624, 2nd edition. Cham, Switzerland: Springer International Publishing. (doi:10.1007/978-3-319-28149-0)

14. Simdyanov TG, Paskerova GG, Valigurová A, Diakin A, Kováčiková M, Schrével J, Guillou L, Dobrovolskij AA, Aleoshin VV. 2018 First ultrastructural and molecular phylogenetic evidence from the blastogregarines, an early branching lineage of Plesiomorphic Apicomplexa. *Protist* **169**, 697–726. (doi:10.1016/j.protis.2018.04.006)

15. Mahé F *et al*. 2017 Parasites dominate hyperdiverse soil protist communities in neotropical rainforests. *Nat. Ecol. Evol*. **1**, 0091. (doi:10.1038/s41559-017-0091)

16. Simdyanov TG, Guillou L, Diakin AY, Mikhailov KV, Schrével J, Aleoshin VV. 2017 A new view on the morphology and phylogeny of eugregarines suggested by the evidence from the gregarine Ancora sagittata (Leuckart, 1860) Labbé, 1899 (Apicomplexa: Eugregarinida). *PeerJ* **5**, e3354. (doi:10.7717/peerj.3354)

17. Wakeman KC. 2020 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* **171**, 125710. (doi:10.1016/j.protis.2019.125710)

18. Rueckert S, Horák A. 2017 Archigregarines of the english channel revisited: new molecular data on selenidium species including early described and new species and the uncertainties of phylogenetic relationships. *PLoS One* **12**, e0187430. (doi:10.1371/journal.pone.0187430)

19. Levine ND. 1971 Taxonomy of the Archigregarinorida and Selenidiidae (Protozoa, Apicomplexa). *J. Protozool*. **18**, 704–717. (doi:10.1111/j.1550-7408.1971.tb03401.x)

20. Wakeman KC, Leander BS. 2012 Molecular phylogeny of pacific archigregarines (Apicomplexa), including descriptions of *Veloxidium leptosynaptae* n. gen., n. sp., from the sea cucumber *Leptosynapta clarki* (Echinodermata), and two new species of *Selenidium*. *J. Eukaryot. Microbiol*. **167**, 339–368. (doi:10.1111/j.1550-7408.2012.00616.x)

21. Schrével J, Valigurová A, Prensier G, Chambouvet A, Florent I, Guillou L. 2016 Ultrastructure of *Selenidium* pendula, the type species of archigregarines, and phylogenetic relations to other marine apicomplexa. *Protist* **167**, 339–368. (doi:10.1016/j.protis.2016.06.001)

22. Wakeman J, Heintzelman A, Leander G. 2014 Comparative ultrastructure and molecular phylogeny of *Selenidium melongena* n. sp. and *S. terebellae* Ray 1930 demonstrate niche partitioning in marine gregarine parasites (Apicomplexa). *Protist* **165**, 493–511. (doi:10.1016/j.protis.2014.05.007)

23. Mathur V, Salomaki ED, Wakeman KC, Na I, Kwong WK, Kolisko M, Keeling PJ. 2023 Reconstruction of plastid proteomes of Apicomplexans and close relatives reveals the major evolutionary outcomes of cryptic plastids. *Mol. Biol. Evol*. **40**, msad002. (doi:10.1093/molbev/msad002)

24. Lax G, Simpson AGB. 2020 The molecular diversity of phagotrophic euglenids examined using single-cell methods. *Protist* **171**, 125757. (doi:10.1016/j.protis.2020.125757)

25. Rueckert S, Villette PMAH, Leander BS. 2011 Species boundaries in gregarine apicomplexan parasites: a case study-comparison of morphometric and molecular variability in *Lecudina* cf. *tuzetae* (Eugregarinorida, Lecudinidae). *J. Eukaryot. Microbiol*. **58**, 275–283. (doi:10.1111/j.1550-7408.2011.00553.x)

26. Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. 2014 Full-length RNA-seq from single cells using smart-seq2. *Nat. Protoc*. **9**, 171–181. (doi:10.1038/nprot.2014.006)

27. Song L, Florea L. 2015 Rcorrector: efficient and accurate error correction for illumina RNA-seq reads. *Gigascience* **4**, 48. (doi:10.1186/s13742-015-0089-y)

28. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170/-/dc1)

29. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. 2019 RnaSPAdes: a de novo transcriptome assembler and its application to RNA-seq data. *Gigascience* **8**, giz100. (doi:10.1093/gigascience/giz100)

30. Grabherr MG *et al*. 2011 Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol*. **29**, 644–652. (doi:10.1038/nbt.1883)

31. Park E, Leander B. 2024 Coinfection of slime feather duster worms (Annelida, *Myxicola*) by different gregarine apicomplexans (*Selenidium*) and astome ciliates reflects spatial niche partitioning and host specificity. *Parasitology* **151**, 400–411. (doi:10.1017/S0031182024000209)

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

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

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

35. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019 RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455. (doi:10.1093/bioinformatics/btz305)

36. Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, Keeling PJ. 2016 Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B* **283**, 20152802–20152810. (doi:10.1098/rspb.2015.2802)

37. Roure B, Rodriguez-Ezpeleta N, Philippe H. 2007 SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol. Biol*. **7 Suppl 1**, S2. (doi:10.1186/1471-2148-7-S1-S2)

38. Minh BQ, Nguyen MAT, von Haeseler A. 2013 Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol*. **30**, 1188–1195. (doi:10.1093/molbev/mst024)

39. Wang HC, Li K, Susko E, Roger AJ. 2008 A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. *BMC Evol. Biol*. **8**, 331–13. (doi:10.1186/1471-2148-8-331)

40. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020 IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol*. **37**, 1530–1534. (doi:10.1093/molbev/msaa015)

41. Tice AK *et al*. 2021 Phylofisher: a phylogenomic package for resolving eukaryotic relationships. *PLoS Biol*. **19**, e3001365. (doi:10.1371/journal.pbio.3001365)

42. Minh BQ, Hahn MW, Lanfear R. 2020 New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol*. **37**, 2727–2733. (doi:10.1093/molbev/msaa106)

43. Mo YK, Lanfear R, Hahn MW, Minh BQ. 2023 Updated site concordance factors minimize effects of homoplasy and taxon sampling. *Bioinformatics* **39**, btac741. (doi:10.1093/bioinformatics/btac741)

44. Salomaki ED, Terpis KX, Rueckert S, Kotyk M, Varadínová ZK, Čepička I, Lane CE, Kolisko M. 2021 Gregarine single-cell transcriptomics reveals differential mitochondrial remodeling and adaptation in apicomplexans. *BMC Biol*. **19**, 77. (doi:10.1186/s12915-021-01007-2)

45. Mathur V, Wakeman KC, Keeling PJ. 2021 Parallel functional reduction in the mitochondria of apicomplexan parasites. *Curr. Biol*. **31**, 2920–2928.(doi:10.1016/j.cub.2021.04.028)

46. Hernandez AM, Ryan JF. 2021 Six-state amino acid recoding is not an effective strategy to offset compositional heterogeneity and saturation in phylogenetic analyses. *Syst. Biol*. **70**, 1200–1212. (doi:10.1093/sysbio/syab027)

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

48. Rueckert S, Leander BS. 2009 Molecular phylogeny and surface morphology of marine archigregarines (Apicomplexa), *Selenidium* spp., *Filipodium phascolosomae* n. sp., and *Platyproteum* n. g. and comb. from north-eastern Pacific peanut worms (Sipuncula). *J. Eukaryot. Microbiol.* **56**, 428–439. (doi:10.1111/j.1550-7408.2009.00422.x)

49. Mathur V, Kolísko M, Hehenberger E, Irwin NAT, Leander BS, Kristmundsson Á, Freeman MA, Keeling PJ. 2019 Multiple independent origins of apicomplexan-like parasites. *Curr. Biol*. **29**, 2936–2941.(doi:10.1016/j.cub.2019.07.019)

50. Leander BS. 2006 Ultrastructure of the archigregarine *Selenidium vivax* (Apicomplexa) – a dynamic parasite of sipunculid worms (host: *Phascolosoma agassizii*). *Mar. Biol. Res.* **2**, 178–190. (doi:10.1080/17451000600724395)

51. Perry O, Sapir Y, Perry G, Ten Hove H, Fine M. 2018 Substrate selection of christmas tree worms (*Spirobranchus* spp.) in the Gulf of Eilat, Red Sea. *J. Mar. Biol. Ass.* **98**, 791–799. (doi:10.1017/S0025315416002022)

52. Kissling WD *et al*. 2015 Data from: Establishing macroecological trait datasets: digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. Dryad Digital Repository. (doi:10.5061/dryad.6cd0v)

53. Lax G, Park E, Na I, Jacko-Reynolds V, Kwong WK, House CSE *et al*. 2024 Data from: Phylogenomic diversity of archigregarine apicomplexans. Figshare. (doi:10.6084/m9.figshare.c.7456278)