

# Phylogeny of gregarines (Apicomplexa) as inferred from small-subunit rDNA and $\beta$ -tubulin

Brian S. Leander,<sup>1</sup> Richard E. Clopton<sup>2</sup> and Patrick J. Keeling<sup>1</sup>

## Correspondence

Patrick Keeling

pkeeling@interchange.ubc.ca

<sup>1</sup>Canadian Institute for Advanced Research, Program in Evolutionary Biology, Department of Botany, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

<sup>2</sup>Division of Science and Technology, Peru State College, Peru, NE 68421, USA

Gregarines are thought to be deep-branching apicomplexans. Accordingly, a robust inference of gregarine phylogeny is crucial to any interpretation of apicomplexan evolution, but molecular sequences from gregarines are restricted to a small number of small-subunit (SSU) rDNA sequences from derived taxa. This work examines the usefulness of SSU rDNA and  $\beta$ -tubulin sequences for inferring gregarine phylogeny. SSU rRNA genes from *Lecudina* (Mingazzini) sp., *Monocystis agilis* Stein, *Leidyana migrator* Clopton and *Gregarina polymorpha* Dufour, as well as the  $\beta$ -tubulin gene from *Leidyana migrator*, were sequenced. The results of phylogenetic analyses of alveolate taxa using both genes were consistent with an early origin of gregarines and the putative 'sister' relationship between gregarines and *Cryptosporidium*, but neither phylogeny was strongly supported. In addition, two SSU rDNA sequences from unidentified marine eukaryotes were found to branch among the gregarines: one was a sequence derived from the haemolymph parasite of the giant clam, *Tridacna crocea*, and the other was a sequence misattributed to the foraminiferan *Ammonium beccarii*. In all of our analyses, the SSU rDNA sequence from *Colpodella* sp. clustered weakly with the apicomplexans, which is consistent with ultrastructural data. Altogether, the exact position of gregarines with respect to *Cryptosporidium* and other apicomplexans remains to be confirmed, but the congruence of SSU rDNA and  $\beta$ -tubulin trees with one another and with morphological data does suggest that further sampling of molecular data will eventually put gregarine diversity into a phylogenetic context.

## INTRODUCTION

Apicomplexans are a diverse group of unicellular eukaryotes that parasitize the body cavities and tissues of metazoans. Some apicomplexans such as haemosporidians (e.g. *Plasmodium*, the causative agent of malaria), coccidians (e.g. *Toxoplasma* and *Eimeria*) and piroplasms (e.g. *Babesia*) are pathogens of humans and domesticated vertebrates, making these parasites economically important and focal organisms for medical research. There are approximately 4600 described species of apicomplexans (Ellis *et al.*, 1998), and probably more than 10 times that number yet to be discovered, but more is known about a handful of the medically important genera than about the rest combined.

Basing our understanding of apicomplexan biology and evolution solely on information from a few, potentially unrepresentative taxa may prove misleading.

Gregarines are a group of apicomplexans that are distinct from coccidians, haemosporidians and piroplasms and are thought to be the earliest lineage of apicomplexans. They are often found in marine worms and insects, where they are extremely abundant and diverse. Although gregarines currently do not have a recognized impact on human welfare, some species could potentially serve as disease agents for the biological control of insects (Levine, 1985), and it has been suggested that gregarines are closely related to the vertebrate pathogen *Cryptosporidium* (see Carreno *et al.*, 1999). Gregarines are certainly significant from an evolutionary perspective because of their suspected early diverging position within the Apicomplexa and their possession of several characteristics presumed to be plesiomorphic, such as an extracellular trophozoite stage, a monoxenous life-cycle and a prevailing presence in marine hosts (Cox, 1994; Théodoridés, 1984; Vivier & Desportes, 1990).

Dinoflagellates and ciliates have been shown to share a common ancestor with apicomplexans, collectively forming

Published online ahead of print on 26 July 2002 as DOI 10.1099/ijms.0.02284-0.

**Abbreviations:** ML, maximum likelihood for DNA; PML, protein maximum likelihood; SSU, small subunit.

The GenBank accession numbers for the small-subunit rDNA sequences from *Monocystis agilis*, *Lecudina* sp., *Gregarina polymorpha* and *Leidyana migrator* are respectively AF457127–AF457130. The GenBank accession number for the  $\beta$ -tubulin sequence from *Leidyana migrator* is AF457131.

the Alveolata (Fast *et al.*, 2001, 2002; Gajadhar *et al.*, 1991; Wolters, 1991). Accordingly, apicomplexans possess tubular mitochondrial cristae, micropores and a pellicle with three membranous layers subtended by microtubules, characteristics that define alveolates (Cavalier-Smith, 1993; Patterson, 1999). Flagella are absent in apicomplexans, except in the microgametes of some species. The phylum Apicomplexa is generally regarded as monophyletic, and a synapomorphy suggested for the group (from which the name is derived) is the presence of a microtubular apical complex, which consists of apical rings, a polar ring and a conoid (Barta *et al.*, 1991; Levine, 1970; Perkins, 1996; Vivier & Desportes, 1990). The apical complex is functionally linked to secretory organelles called rhoptries and micronemes, which together play a role in the attachment to the host and in most taxa, intracellular invasion (Black & Boothroyd, 2000). However, the presence or absence of an apical complex is rather ambiguous, since several predatory and parasitic alveolates (e.g. *Colpodella* and *Acrocoelus*) possess intermediate states for the apical complex and associated organelles, which makes their inclusion within the Apicomplexa *sensu stricto* controversial (Fernández *et al.*, 1999; Foissner & Foissner, 1984; Siddall *et al.*, 1997, 2001; Simpson & Patterson, 1996). For instance, the parasitic alveolate *Perkinsus* possesses an apical-complex-like apparatus, but molecular data show that *Perkinsus* is probably more closely related to modern dinoflagellates than to apicomplexans (Ellis *et al.*, 1998; Goggin & Barker, 1993; Reece *et al.*, 1997; Siddall *et al.*, 1997). Another feature that unites the Apicomplexa (perhaps less ambiguously) is their specific parasitic life history. This life history is often complex and appears to vary considerably within the group (Levine, 1985; Siddall, 1995; Vivier & Desportes, 1990) but, in general, consists of three alternating phases of development: merogony (asexual), gametogony (sexual) and sporogony (asexual). Gregarines display an interesting variety of these phases in the three subdivisions of the group: 'neogregarines' possess all three phases and trophozoites contain 'septa' between cell regions; 'archigregarines' are thought to possess all three phases (but merogony has not been indisputably observed) and here, trophozoites lack septa and are morphologically similar to the sporozoite stage; 'eugregarines' can be either septate or aseptate but are grouped together (perhaps artificially) based on the absence of merogony (Ellis *et al.*, 1998; Levine, 1971; Vivier & Desportes, 1990).

Gene sequence data suggest that apicomplexans and dinoflagellates are 'sister' groups (Carreno *et al.*, 1999; Fast *et al.*, 2002; Gajadhar *et al.*, 1991; Siddall *et al.*, 1997; Wolters, 1991). Morphologically, apicomplexans and dinoflagellates are quite different, but this seemingly large gap is shrinking as we begin to appreciate the actual diversity of alveolates, namely the combination of structural features in *Perkinsus*, *Parvilucera*, *Acrocoelus*, colpodellids, ellobiopsids, *Oxyrrhís* and parasitic dinoflagellates such as *Haplozoon* (Brugerolle & Mignot, 1979; Fernández *et al.*, 1999; Gajadhar *et al.*, 1991; Hoffman & Yancey, 1966; Norén *et al.*, 1999; Perkins, 1976,

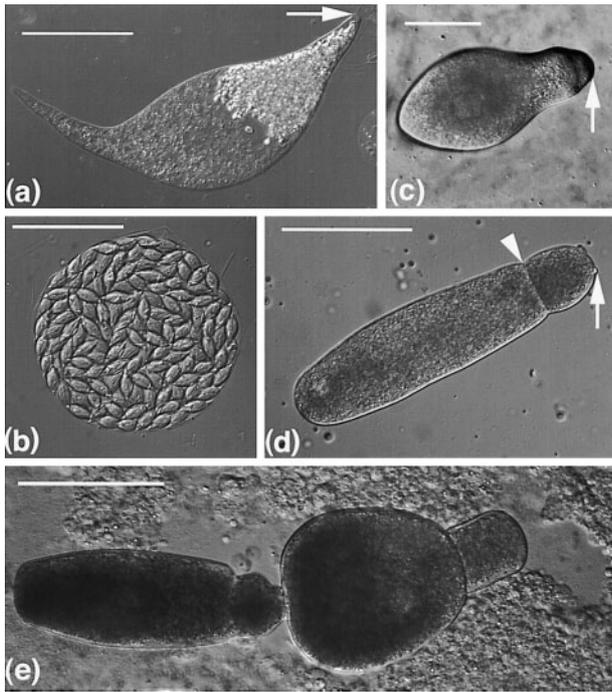
1996; Siddall *et al.*, 1997; Simpson & Patterson 1996; Whisler, 1990). Moreover, environmental PCR approaches are identifying potentially intermediate lineages that still need to be characterized at the morphological level (Diez *et al.*, 2001; López-García *et al.*, 2001; Moon-van der Staay *et al.*, 2001).

In terms of apicomplexans, the internal topology of the group is just beginning to emerge from comparisons of morphological characteristics and gene sequences (Barta, 1989; Ellis *et al.*, 1992, 1995, 1998; Ellis & Morrison, 1995; Escalante & Ayala, 1994; Hnida & Duszynski, 1999; Holmdahl *et al.*, 1999; Johnson *et al.*, 1990; Morgan *et al.*, 1999; Siddall, 1995; Siddall & Barta, 1992; Zhu *et al.*, 2000); however, most of this work has centred on representatives from three of the four major groups: coccidians, haemosporidians and piroplasms. Only two complete small-subunit (SSU) rDNA sequences from gregarines have been reported (Carreno *et al.*, 1999). We have sequenced four SSU rRNA genes and one  $\beta$ -tubulin gene from four different gregarines, including two aseptate eugregarines, in an attempt to examine the usefulness of SSU rDNA and  $\beta$ -tubulin for inferring the phylogeny of these poorly understood parasites. Phylogenetic studies of gregarines should help to provide the necessary framework for beginning to reconstruct the nature of the last common ancestor of dinoflagellates and apicomplexans, and for understanding the origin and early evolution of parasitism in apicomplexans.

## METHODS

**Organisms.** All stages in the life cycle of the aseptate eugregarine *Monocystis agilis* Stein were found in the seminiferous vesicles of earthworms (*Lumbricus terrestris*) purchased from Berry's Bait and Tackle (Fig. 1a, b); species identification was based on Meier (1956) and Levine (1988). Forty to fifty gametocysts filled with 'lemon-shaped' oocysts (Fig. 1b) were micropipetted from the seminiferous vesicles, washed three times with distilled water and then collected in a 1.5 ml Eppendorf tube. Trophozoites of an unidentified species of the aseptate eugregarine *Lecudina* Mingazzini (designated as *Lecudina* sp.) were identified in the gut of the marine polychaete *Nereis vexillosa*, collected at low tide in Stanley Park, Vancouver, BC, Canada (Fig. 1c); genus identifications were based on Levine (1976, 1988) and Kozloff (1983). Approximately 40 trophozoites were micropipetted from the worm's intestine, washed three times in filtered sea water and then deposited into an Eppendorf tube. The septate eugregarines *Leidyana migrator* Clopton and *Gregarina polymorpha* Dufour (Fig. 1d) were respectively collected from the intestines of the Madagascar hissing cockroach, *Gromphadorhina portentosa* (maintained at Peru State College, NE, USA), and the yellow mealworm, *Tenebrio molitor* (maintained at the University of British Columbia, BC, Canada) (Clopton, 1995; Clopton *et al.*, 1992). Approximately 40 trophozoites or gamonts from each were micropipetted, washed three times in saline solution (Belton & Grundfest, 1962) and then collected in Eppendorf tubes.

**Microscopy.** Cell isolations were done with a Meiji dissecting microscope and a Zeiss Axiovert 200 inverted microscope. Light micrographs were produced with a Zeiss Axioplan 2 imaging microscope connected to a Q-Imaging, Microimager II, black and white digital camera.



**Fig. 1.** Light micrographs of gregarines from which SSU rDNA sequences were obtained; arrows indicate the attachment end (i.e. mucron or epimerite) of the cell. (a)–(b) Trophozoites (a; bar, 50  $\mu\text{m}$ ) and spherical gametocysts filled with numerous 'lemon-shaped' oocysts (b; bar, 75  $\mu\text{m}$ ) of *M. agilis* (an aseptate eugregarine) were isolated from the seminiferous vesicles of the earthworm, *Lumbricus terrestris*. (c) A species of the aseptate eugregarine, *Lecudina* (bar, 70  $\mu\text{m}$ ), was isolated from the intestines of a marine polychaete, *Nereis vexillosa*. (d) The septate eugregarine *Gregarina polymorpha* (bar, 150  $\mu\text{m}$ ) possesses a transverse ectoplasmic septum between cell regions (arrowhead) and was isolated from the intestines of the yellow mealworm, *Tenebrio molitor*. (e) Associated gamonts of *Gregarina* occasionally had different morphologies that appeared to overlap the descriptions of three species in the intestines of *Tenebrio molitor* (bar, 165  $\mu\text{m}$ ). The promite (e; right-hand cell) possessed an overall shape reminiscent of *Gregarina niphandrodes* and the satellite (e; left-hand cell) was more similar to *Gregarina polymorpha* (d). This association was observed in the larval stage of *Tenebrio molitor*.

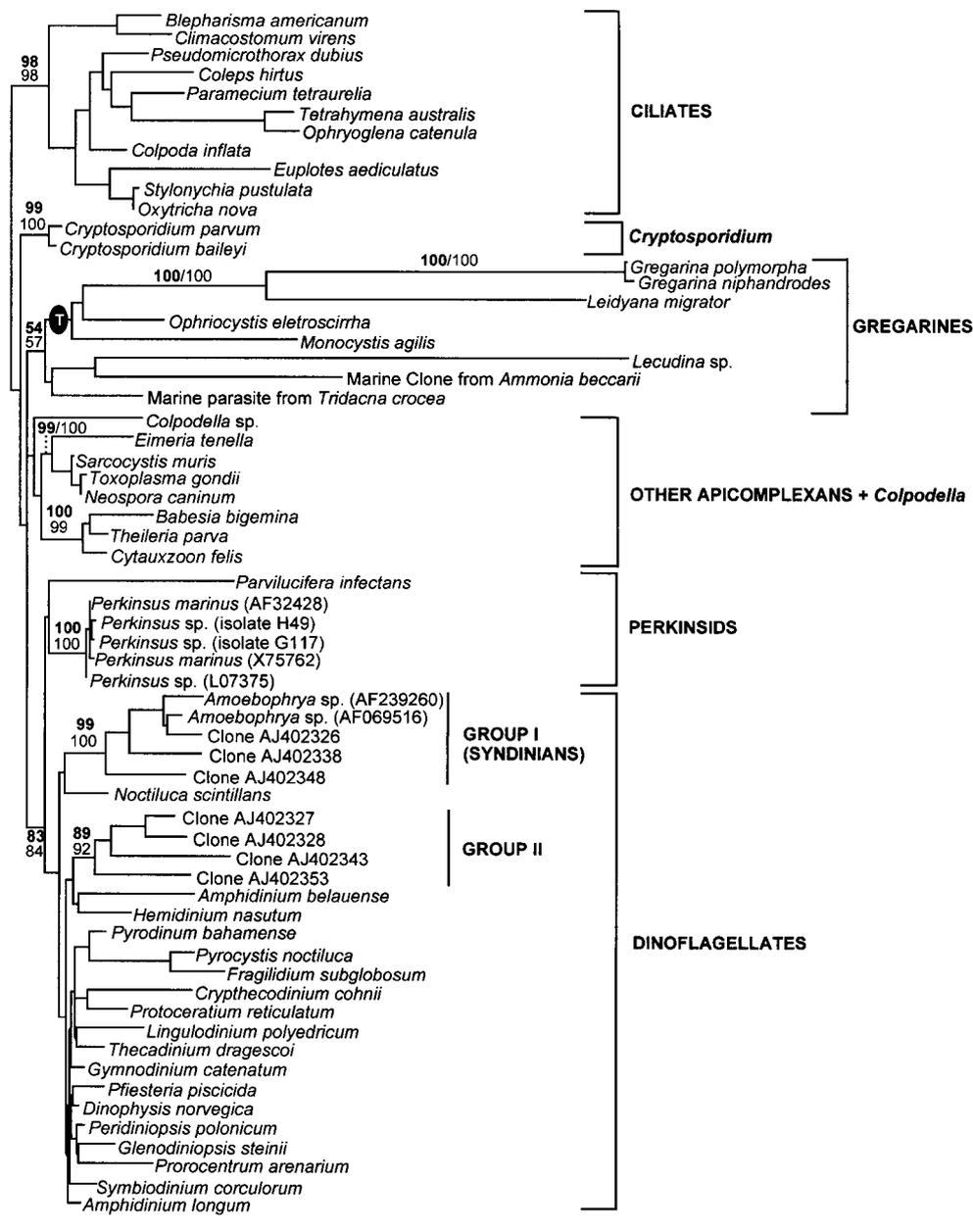
#### DNA extraction, PCR amplification, cloning and sequencing.

Genomic DNA was extracted from pellets of gregarine cells using a standard hexadecyltrimethylammonium bromide extraction protocol (Zolan & Pukhila, 1986). The SSU rRNA genes were amplified as a single fragment using universal eukaryotic primers 5'-CGAATTCAACCTGGTTGATCCTGCCAGT-3' and 5'-CCGG-ATCCTGATCCTTCTGCAGGTTACCTAC-3', except for *Gregarina polymorpha*, for which primers 5'-CGAATTCAACCTGGTTGATCCTGCCAGT-3' and 5'-GGCGGTGTGTACAARRGGCAGGG-3' were used. All amplifications consisted of an initial denaturing period (95°C for 2 min), 35 cycles of denaturing (92°C for 45 s), annealing (45°C for 45 s) and extension (72°C for 1.5 min) and a final extension period (72°C for 5 min). The  $\beta$ -tubulin sequence from *Leidyana migrator* was amplified as a single fragment using

the same PCR protocol with two pairs of nested primers: outside pair, 5'-GCCTGCAGGNCARTGYGGNAAYCA-3' and 5'-TCCTCGAGTRAAATTCATYTCRTCCAT-3'; inside pair, 5'-CAGATCGGCGCGAARTTYTGGGARAT-3' and 5'-CTCGTCC-ATGCCYTCNCCNGTRTACCA-3'. Products corresponding to the expected size were gel-isolated and cloned into the vector pCR2.1 using the TOPO TA cloning kit (Invitrogen). Two to four clones from each organism were sequenced with the ABI Big-Dye reaction mix using the vector primers and four internal primers oriented in both directions. New sequences were deposited in GenBank: *M. agilis* (AF457127), *Lecudina* sp. (AF457128), *Gregarina polymorpha* (AF457129), *Leidyana migrator* SSU rDNA (AF457130) and *Leidyana migrator*  $\beta$ -tubulin (AF457131).

**Alignments and phylogenetic analysis.** New SSU rDNA sequences were added to the alignment of Van de Peer *et al.* (1998) and modified manually using MacClade 4 (Maddison & Maddison, 2000). The number of taxa in this alignment was trimmed to include representatives of ciliates, dinoflagellates, apicomplexans, *Perkinsus* and other alveolates of uncertain affinity: *Colpodella* sp. (Siddall *et al.*, 2001), a marine clone from the giant clam, *Tridacna crocea* (GenBank accession no. AB000912), and a marine clone from the foraminiferan *Ammonia beccarii* (U07937). Substantially truncated sequences from gregarines were excluded from this alignment, namely *M. agilis* (AF129881), '*Pseudomonocystis lepidiota*' (L31843), '*Gregarina chortiocetes*' (L31841) and '*Gregarina caledia*' (L31799). Although alignments with a greater number of taxa were analysed (a 'tree of life' and the 62-taxon alignment; Fig. 2), we focused on an alignment consisting of 25 ingroup taxa, five outgroup taxa (ciliates) and including 1201 unambiguously aligned characters – the 30-taxon alignment. We compared the results from this alignment with one excluding five gregarine sequences with extremely fast substitution rates – the 25-taxon alignment. An inferred amino acid sequence of  $\beta$ -tubulin from *Leidyana migrator* was added to an existing alignment (Keeling *et al.*, 2000). This global alignment was analysed before being trimmed to an alveolate alignment consisting of 18 taxa and 354 unambiguously aligned amino acid characters. Unpublished  $\beta$ -tubulin sequences from *Perkinsus marinus*, *Oxyrrhis marina*, *Heterocapsa triquetra*, *Woloszynskia tenuissima*, '*Gymnodinium varians*' (name provisional) and *Gyrodinium instriatum* were provided by M. L. McEwan, J. F. Saldarriaga and N. M. Fast (University of British Columbia, Vancouver, BC, Canada). The SSU rDNA and  $\beta$ -tubulin amino acid sequence alignments are available upon request.

Distances for both the SSU rDNA and the  $\beta$ -tubulin datasets were calculated with TREE-PUZZLE 5.0 using the Hasegawa–Kishino–Yano model and WAG substitution matrix, respectively (Strimmer & Von Haeseler, 1996). Site-to-site variation was modelled on a gamma distribution with invariable sites and eight variable rate categories. Nucleotide frequencies, transition/transversion ratios (1.72 for the 30-taxon alignment and 1.63 for the 25-taxon alignment) and the shape parameter (0.37 for the 30-taxon alignment and 0.30 for the 25-taxon alignment) were estimated from the DNA dataset; a shape parameter (0.81) was also estimated from the protein dataset. Distance trees were constructed with weighted neighbour-joining using WEIGHBOR (Bruno *et al.*, 2000) and with Fitch–Margoliash analyses using FITCH (Felsenstein, 1993), the latter with global rearrangements and 10 input-order jumbles. One hundred bootstrap datasets were constructed using SEQBOOT (Felsenstein, 1993). Distances for each bootstrap dataset were generated with the gamma shape parameters (and transition/transversion ratios in the case of DNA analyses) estimated from the original dataset using the shell script PUZZLEBOOT (M. Holder and A. Roger; <http://www.tree-puzzle.de>). Bootstrap distance trees were constructed with the same two algorithms; however, in the Fitch–Margoliash analyses, the number of input jumbles was reduced to five.



**Fig. 2.** Gamma-corrected ML tree ( $-\ln L = 14\,085$ ) inferred from an alignment containing 62 SSU rDNA sequences and 1201 sites, with ciliates as the outgroup, showing the general relationships of major alveolate groups (bracketed to the right). Numbers at the relevant nodes correspond to bootstrap proportions for Fitch–Margoliash (bold) and weighted neighbour-joining (roman). Gregarine branch lengths were very long. Thus, to minimize the effects of long-branch attraction, we excluded sequences from the haemosporidian *Plasmodium*, which also has very long SSU rDNA branches in phylogenetic analyses (Escalante & Ayala, 1994; Zhu *et al.*, 2000). The ML tree recovered two major clades of gregarines: parasites from terrestrial hosts (see black oval) and parasites from marine hosts. Dinoflagellate sequences labelled as group I and group II were identified by environmental PCR approaches (Diez *et al.*, 2001; López-García *et al.*, 2001). The transition/transversion ratio was 1.78; the scale bar indicates 0.1 substitutions (corrected) per site.

Maximum-likelihood trees with correction for site-to-site rate variation were also constructed for both the DNA and protein datasets [maximum likelihood for DNA (ML) and protein maximum likelihood (PML), respectively]. For the DNA datasets, ML trees from the original dataset and 100 resampled bootstrap datasets were inferred using

PAUP\* 4.0 (Swofford, 1999) under a Hasegawa–Kishino–Yano model and several input-order jumbles: ten for the original dataset and three for the bootstrap datasets (using corrections for site-to-site rate variation from the original dataset). For the protein dataset, ProML (Felsenstein, 1993) was used with global rearrangements and five

input-order jumbles. Site-to-site rate variation was modelled using the 'R' option with nine categories of substitution rates (eight rates plus invariable sites), using the probabilities and frequencies estimated by TREE-PUZZLE. ProML bootstrapping was also performed for 100 resampled datasets (from SEQBOOT) using the same settings but with only two input-order jumbles.

## RESULTS AND DISCUSSION

### Pairwise sequence comparisons

Two of our SSU rDNA sequences were quite similar to existing GenBank entries. Our sequence from *M. agilis* (accession no. AF457127) is 6.2% different from the partial sequence of *M. agilis* (AF129881) published previously (Carreno *et al.*, 1999); there were 83 differences in 1337 aligned base positions. Not surprisingly, the two sequences grouped together with 100% bootstrap support (distance and ML analysis) when they are included in the same alignment (data not shown). Although both isolates came from the same host, *Lumbricus terrestris*, there is currently not enough information to clarify whether these two sequences came from one described species, *M. agilis*, or, perhaps, from different species within the genus (Fig. 1a, b).

There are four similar species of eugregarines in the genus *Gregarina* described as inhabiting the intestines of *Tenebrio molitor* (Clopton *et al.*, 1991). SSU rRNA genes have now been sequenced from two of these species: *Gregarina polymorpha* and *Gregarina niphandrodes* (Carreno *et al.*, 1999). Our sequence from *Gregarina polymorpha* was less than 1.0% different from the previously published sequence from *Gregarina niphandrodes*; there were 15 differences in 1539 aligned positions. This sequence similarity plus the continuum of morphological shapes and sizes of ontogenetic stages in *Gregarina* within *Tenebrio* could suggest that some of the species designations therein (e.g. *Gregarina polymorpha*, *Gregarina niphandrodes* and *Gregarina cuneata*) warrant re-evaluation. This is also supported by our observations of associated gamonts with shapes that overlap the species descriptions for *Gregarina polymorpha* and *Gregarina niphandrodes* (Fig. 1e) (Clopton & Janovy, 1993; Clopton *et al.*, 1991, 1992). Moreover, we observed gamonts that appear to fit the description of *Gregarina niphandrodes* in the larval stage of the host (Fig. 1e), which is inconsistent with reports that this gregarine is restricted to adult *Tenebrio molitor* (Clopton & Janovy, 1993; Clopton *et al.*, 1991). However, broad morphological variation as an ontogenetic effect is characteristic of gregarine populations, so it is difficult to make an unambiguous identification of a population or individual that is inconsistent with the host restrictions defined previously (Clopton *et al.*, 1991; Clopton & Janovy, 1993). Complete SSU rDNA sequences from morphotypes attributable to *Gregarina cuneata* and *Gregarina steini* should shed additional light onto this taxonomic issue. It is interesting to note, however, that some species encompass isolates that are more than 6% dissimilar in SSU rDNA sequence (*M. agilis*) while, in other cases, different species are less than 1% dissimilar

at the gene sequence level (*Gregarina polymorpha* and *Gregarina niphandrodes*).

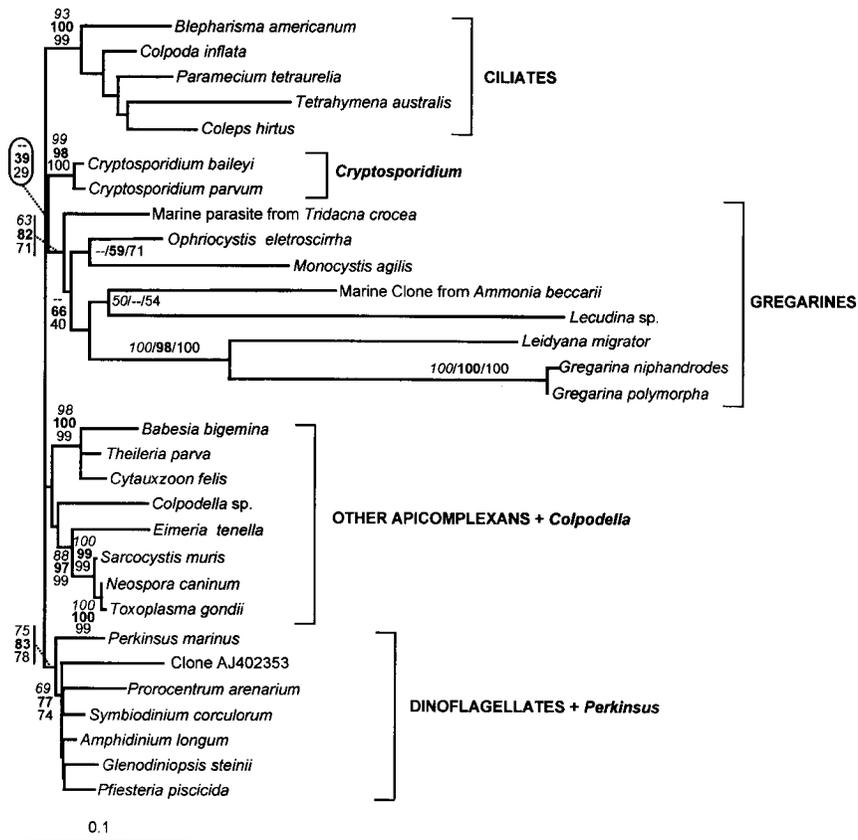
### Phylogeny of gregarines as inferred from SSU rDNA and $\beta$ -tubulin

We analysed alignments with a large number of taxa (a 'tree of life' and the 62-taxon alignment of alveolates; Fig. 2) but focused on an alignment consisting of 25 ingroup taxa and five outgroup taxa (ciliates) – the 30-taxon alignment (Fig. 3a). We compared the 30-taxon alignment with one that excluded five gregarine sequences with extremely fast substitution rates – the 25-taxon alignment (Fig. 3b). In all of our analyses of SSU rDNA, gregarine (and gregarine-like) sequences formed a moderately to well-supported clade, but without any obvious sister group (Figs 2 and 3). A sister relationship between dinoflagellates (plus *Perkinsus*) and apicomplexans was rarely recovered, and SSU rDNA did not produce a well-supported apicomplexan clade. In general, gregarine sequences showed an accelerated rate of evolutionary change, which makes them prone to long-branch attraction and tends to decrease statistical support for nodes in the topology; bootstrap proportions were markedly lower than those in published phylogenetic analyses of representative alveolates (Carreno *et al.*, 1999; Siddall *et al.*, 2001; Zhu *et al.*, 2000).

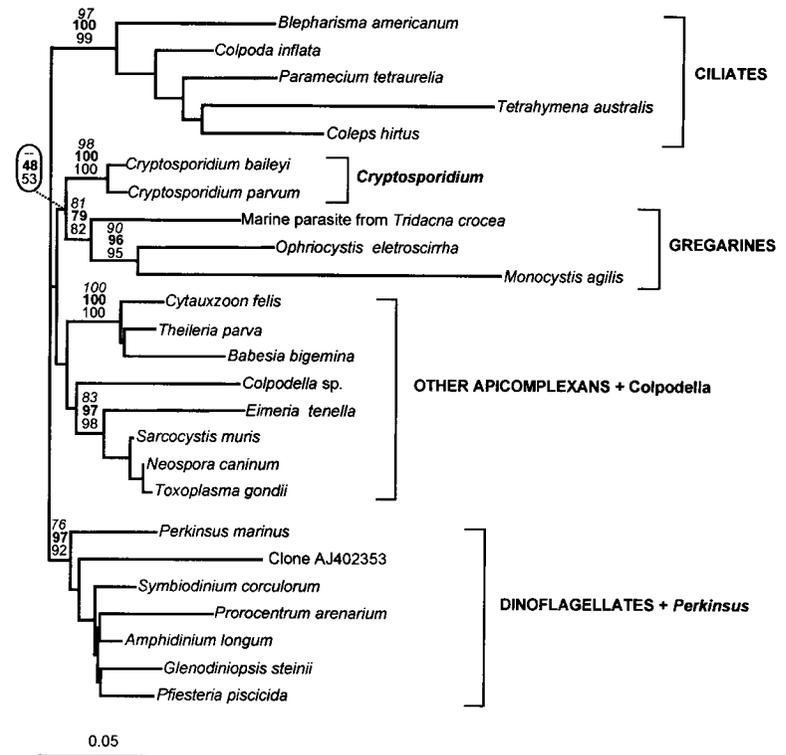
Previous studies have suggested that gregarines and the vertebrate pathogen *Cryptosporidium* are sister groups that diverge early within the Apicomplexa (Barta, 1989; Carreno *et al.*, 1999; Zhu *et al.*, 2000). Our distance analyses of SSU rDNA weakly supported this hypothesis (Figs 2 and 3), and bootstrap proportions in support of this relationship increased when the five longest gregarine branches were removed from the analysis (Fig. 3b). Distance and PML analyses of  $\beta$ -tubulin sequences also consistently recovered the sister relationship between the gregarine *Leidyana migrator* and *Cryptosporidium* with very weak statistical support (42–46% support; Fig. 4). However, ML analysis of SSU rDNA never recovered a sister relationship between gregarines and *Cryptosporidium*, and  $\beta$ -tubulin sequences from both *Cryptosporidium parvum* and *Leidyana migrator* are very divergent in comparison with the other alveolate sequences in the analysis, so we were not able to rule out long-branch attraction as the primary cause of this relationship in  $\beta$ -tubulin trees. Nonetheless, all analyses of  $\beta$ -tubulin sequences did recover a monophyletic Apicomplexa with weak to moderate statistical support (Fig. 4). Overall, our analyses do not provide strong support either for or against the hypothesis that *Cryptosporidium* and gregarines are sister groups within the Apicomplexa, yet it is clear that both lineages are distinct from other apicomplexans and diverged prior to the origins of coccidians, haemosporidians and piroplasmids.

The internal topology of the gregarine clade was only partially resolvable with SSU rDNA data (Figs 2 and 3). The septate eugregarines (*Leidyana migrator*, *Gregarina niphandrodes* and *Gregarina polymorpha*) formed a

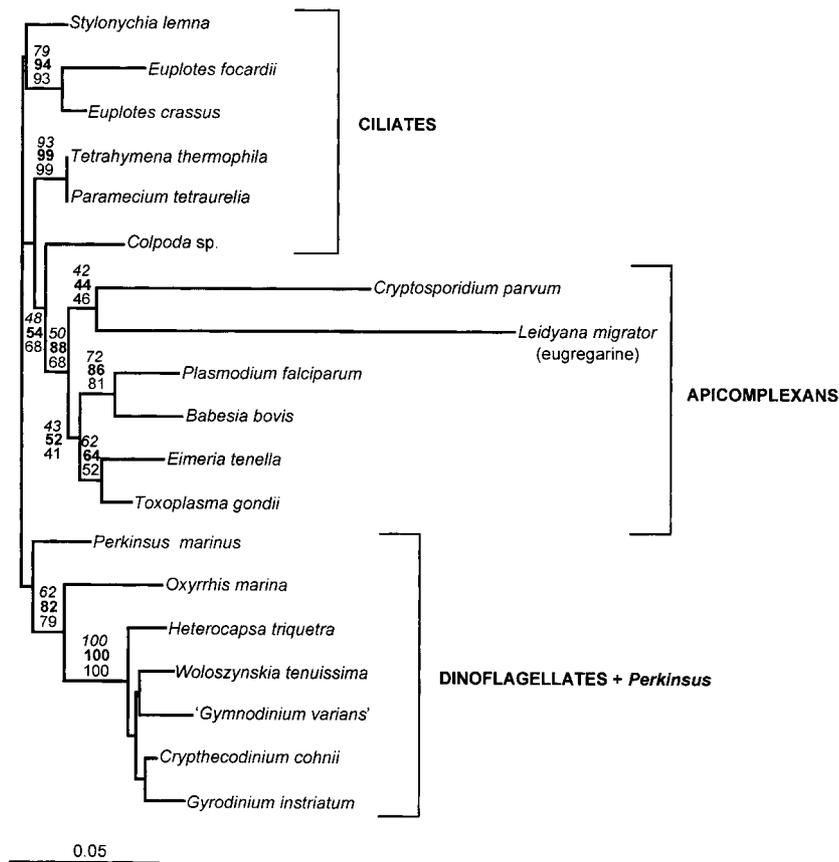
(a) Gamma-corrected distance tree using SSU rDNA



(b) Gamma-corrected distance tree using SSU rDNA: Excluding long gregarine branches



**Fig. 3.** Gamma-corrected distance trees, based on 30 (a) and 25 (b) SSU rDNA sequences, showing the position of gregarines within the Alveolata. Numbers at the relevant nodes correspond to bootstrap proportions for maximum likelihood (italics), Fitch–Margoliash (bold) and weighted neighbour-joining (roman). In (a), all full-length or nearly full-length gregarine sequences are shown; they have long branch lengths and their phylogenetic position within the alveolates is left unresolved. Gregarine sequences along with sequences from both an unidentified marine parasite clone from the giant clam, *Tridacna crocea*, and a marine clone misattributed to the foraminiferan *A. beccarii* cluster together with moderate bootstrap support (63/82/71). Although the tree topologies from the distance analyses show a sister relationship between gregarines and *Cryptosporidium*, the bootstrap proportions (in the oval symbol) are very low. Transition/transversion ratio, 1.72. When the five longest branches are removed from the analysis (b), the bootstrap proportions supporting a sister relationship between gregarines and *Cryptosporidium* (in the oval symbol) are increased but remain weak. Transition/transversion ratio, 1.63. The scale bars for (a) and (b) respectively indicate 0.1 and 0.05 substitutions (corrected) per site.



**Fig. 4.** Weighted neighbour-joining tree derived from 19 gamma-corrected  $\beta$ -tubulin sequences showing the phylogenetic position of the septate eugregarine *Leidyana migrator* within the Apicomplexa. Numbers at the nodes correspond to bootstrap proportions for PML (italics), Fitch–Margoliash (bold) and weighted neighbour-joining (roman). The phylum Apicomplexa is monophyletic with weak to moderate bootstrap support (50/88/68), and all three analyses show a weak (42/44/46) sister relationship between the long branches of *Cryptosporidium* and *Leidyana*. The scale bar indicates 0.05 substitutions (corrected) per site.

highly supported clade without any obvious sister group (98–100% support; Fig. 3). *Ophriocystis elektroscirrha* (also spelled *Ophryocystis*) is classified as a neogregarine (McLaughlin & Myers, 1970); however, this taxon did not cluster strongly with the septate eugregarines in distance analyses. Instead, *Ophriocystis elektroscirrha* tended to group with the aseptate eugregarine *M. agilis*, which is consistent with previous results using three gregarine sequences (Carreno *et al.*, 1999). However, ML analysis of the 62-taxon SSU rDNA matrix (Fig. 2) produced two separate clades of gregarines: one consisted of gregarines from terrestrial hosts and one consisted of gregarines from marine hosts. The ML analyses of this dataset also placed *Ophriocystis elektroscirrha* as the sister group to septate eugregarines, which supports the hypothesis that all septate gregarines form a monophyletic group and that the septa evolved once in their common ancestor (Fig. 2).

The marine aseptate eugregarine *Lecudina* sp. tended to cluster with two enigmatic SSU rDNA sequences in GenBank, AB000912 and U07937. An SSU rDNA sequence from an unidentified parasite in the haemolymph of the giant clam, *Tridacna crocea* (AB000912), has previously been shown with a distance analysis to diverge with weak bootstrap support between apicomplexans and dinoflagellates (Nakayama *et al.*, 1998). The tiny sporozoites of this parasite (about 9  $\mu$ m in length) were described by using light

microscopy and transmission electron microscopy and were shown to invade the blood cells (haemocytes) of the host. At the time, these authors did not have access to any gregarine sequences, but they suspected that this parasite might be related to gregarines (Nakayama *et al.*, 1998). In all of our analyses, the sequence from this parasite branched among the gregarine sequences with moderate to high bootstrap support (Figs 2 and 3). We infer that this parasite is indeed a gregarine, but a better understanding of its ontogenetic and structural characteristics is needed before we can make sense of its phylogenetic position and incorporate it into a classification scheme.

The second enigmatic SSU rDNA sequence (accession no. U07937) was published as coming from the benthic foraminiferan *A. beccarii* (Wray *et al.*, 1995). This sequence clusters within the gregarine clade and often as the sister to *Lecudina* sp. (Figs 2 and 3). It was demonstrated previously that the SSU rDNA sequence from *A. beccarii* was misattributed to the foraminiferan and probably came from DNA originating from food vacuoles or endobiotic eukaryotes (Pawlowski *et al.*, 1996). Our analyses indicate that the sequence misattributed to *A. beccarii* came from a gregarine with unknown characteristics. We suspect that the foraminiferan host either ingested gregarine oocysts from the benthos or, perhaps more likely, harboured a gregarine parasite. Despite this uncertainty, our analyses provide circumstantial evidence that apicomplexans known to

infect foraminiferans such as *Trophosphaera planorbulinae* (Levine, 1985; Pawlowski *et al.*, 1996) are closely related to gregarines.

### The early evolution of apicomplexans and alveolates

The specific characteristics of the common ancestor of all alveolates and the common ancestor of dinoflagellates and apicomplexans remain unresolved issues. The modern lineages with known gene sequences that best represent the features in these ancestors are *Perkinsus*, *Parvilucifera*, *Oxyrrhis* and *Colpodella*. Sequence evidence indicates that *Oxyrrhis*, *Parvilucifera* and *Perkinsus* are more closely related to dinoflagellates than to other alveolates (Ellis *et al.*, 1998; Goggin & Barker, 1993; Norén *et al.*, 1999; Reece *et al.*, 1997; Siddall *et al.*, 1997; P. J. Keeling, unpublished data), but the phylogenetic position of *Colpodella* is much less clear. *Colpodella* sp. consistently clusters with the apicomplexans in our analyses (Figs 2 and 3), which differs from previous analyses of combined SSU rDNA and actin sequences that suggest *Colpodella* is a sister to ciliates (Siddall *et al.*, 2001). Our results are, however, consistent with published trees generated from only SSU rDNA sequences using non-alveolate outgroups (Siddall *et al.*, 2001) and with interpretations of comparative morphology (Simpson & Patterson, 1996). Moreover, there is suspicion that the actin sequence from *Colpodella* was derived from its food, *Bodo*, which would make inferences involving this sequence unreliable (J. Saldarriaga, personal communication). Thus, there does not appear to be a strong relationship between *Colpodella* and ciliates, and current evidence suggests that the common ancestor of apicomplexans and dinoflagellates possessed many of the character states present in *Colpodella*, *Acrocoelus*, perkinsids and *Oxyrrhis* (Brugerolle & Mignot, 1979; Dodge & Crawford, 1971; Fernández *et al.*, 1999; Foissner & Foissner, 1984; Perkins, 1976, 1996; Simpson & Patterson, 1996). The inferred characteristics of the common ancestor of all alveolates remain an open question, but perhaps the modern lineage that best represents these features is the predatory flagellate *Colponema* (Mignot & Brugerolle, 1975).

### ACKNOWLEDGEMENTS

We wish to thank N. M. Fast, J. F. Saldarriaga and J. M. Archibald for technical advice and J. M. Archibald for comments on the manuscript. We are grateful to M. L. McEwan, J. F. Saldarriaga and N. M. Fast (University of British Columbia) for providing us with  $\beta$ -tubulin sequences and to M. E. Siddall (American Museum of Natural History) for providing us with an SSU rDNA sequence from *Colpodella* sp. B.S.L. was supported by the National Science Foundation (Postdoctoral Research Fellowship in Microbial Biology, USA). P. J. K. is a Scholar of the Canadian Institute for Advanced Research and the Michael Smith Foundation for Health Research, and was supported by a grant from the Canadian Institutes for Health Research (MOP-42517).

### REFERENCES

- Barta, J. R. (1989).** Phylogenetic analysis of the class Sporozoa (phylum Apicomplexa Levine, 1970): evidence for the independent evolution of heteroxenous life cycles. *J Parasitol* **75**, 195–206.
- Barta, J. R., Jenkins, M. C. & Danforth, H. D. (1991).** Evolutionary relationships of avian *Eimeria* species among other Apicomplexan protozoa: monophyly of the apicomplexa is supported. *Mol Biol Evol* **8**, 345–355.
- Belton, P. & Grundfest, H. (1962).** Potassium activation and K<sup>+</sup> spikes in muscle fibers of mealworm larva (*Tenebrio molitor*). *Am J Physiol* **203**, 588–594.
- Black, M. W. & Boothroyd, J. C. (2000).** Lytic cycle of *Toxoplasma gondii*. *Microbiol Mol Biol Rev* **64**, 607–623.
- Brugerolle, G. & Mignot, J. P. (1979).** Observations sur le cycle l'ultrastructure et la position systematique de *Spiromonas perforans* (*Bodo perforans* Hoolande 1938), flagelle parasite de *Chilomonas paramecium*: ses relations avec les dinoflagelles et sporozoaires. *Protistologica* **15**, 183–196.
- Bruno, W. J., Socci, N. D. & Halpern, A. L. (2000).** Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol* **17**, 189–197.
- Carreno, R. A., Martin, D. S. & Barta, J. R. (1999).** *Cryptosporidium* is more closely related to the gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites inferred using small-subunit ribosomal RNA gene sequences. *Parasitol Res* **85**, 899–904.
- Cavalier-Smith, T. (1993).** Kingdom Protozoa and its 18 phyla. *Microbiol Rev* **57**, 953–994.
- Clopton, R. E. (1995).** *Leidyana migrator* n. sp. (Apicomplexa, Eugregarinida, Leidyaniidae) from the Madagascar hissing cockroach, *Gromphadorhina portentosa* (Insecta, Blattodea). *Invertebr Biol* **114**, 271–278.
- Clopton, R. E. & Janovy, J. (1993).** Developmental niche structure in the gregarine assemblage parasitizing *Tenebrio molitor*. *J Parasitol* **79**, 701–709.
- Clopton, R. E., Percival, T. J. & Janovy, J. (1991).** *Gregarine niphandroides*, n. sp. (Apicomplexa: Eugregarinorida) from adult *Tenebrio molitor* (L.) with oocyst descriptions of other gregarine parasites of the yellow mealworm. *J Protozool* **38**, 472–479.
- Clopton, R. E., Janovy, J., Jr & Percival, T. J. (1992).** Host stadium specificity in the gregarine assemblage parasitizing *Tenebrio molitor*. *J Parasitol* **78**, 334–337.
- Cox, F. E. G. (1994).** The evolutionary expansion of the Sporozoa. *Int J Parasitol* **24**, 1301–1316.
- Díez, B., Pedrós-Alió, C. & Massana, R. (2001).** Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl Environ Microbiol* **67**, 2932–2941.
- Dodge, J. D. & Crawford, R. M. (1971).** Fine structure of the dinoflagellate *Oxyrrhis marina*. I. The general structure of the cell. *Protistologica* **7**, 339–409.
- Ellis, J. & Morrison, D. (1995).** Effects of sequence alignment on the phylogeny of *Sarcocystis* deduced from 18S rDNA sequences. *Parasitol Res* **81**, 696–699.
- Ellis, J., Hefford, C., Baverstock, P. R., Dalrymple, B. P. & Johnson, A. M. (1992).** Ribosomal DNA sequence comparison of *Babesia* and *Theileria*. *Mol Biochem Parasitol* **54**, 87–95.
- Ellis, T. J., Luton, K., Baverstock, P. R., Whitworth, G., Tenter, A. M. & Johnson, A. M. (1995).** Phylogenetic relationships between *Toxoplasma* and *Sarcocystis* deduced from a comparison of 18S rDNA sequences. *Parasitology* **110**, 521–528.

- Ellis, J. T., Morrison, D. A. & Jeffries, A. C. (1998). The phylum Apicomplexa: an update on the molecular phylogeny. In *Evolutionary Relationships Among Protozoa*, pp. 255–274. Edited by G. H. Coombs, K. Vickerman, M. A. Sleigh & A. Warren. Boston: Kluwer.
- Escalante, A. A. & Ayala, F. J. (1994). Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc Natl Acad Sci U S A* **91**, 11373–11377.
- Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* **18**, 418–426.
- Fast, N. M., Xue, L., Bingham, S. & Keeling, P. J. (2002). Re-examining alveolate evolution using multiple protein molecular phylogenies. *J Eukaryot Microbiol* **49**, 30–37.
- Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package). Distributed by the author. Department of Genetics, University of Washington, Seattle, USA.
- Fernández, I., Pardos, F., Benito, J. & Arroyo, N. L. (1999). *Acrocoelus glossobalani* gen. nov. et sp. nov., a protistan flagellate from the gut of the enteropneust *Glossobalanus minutus*. *Eur J Protistol* **35**, 55–65.
- Foissner, W. & Foissner, I. (1984). First record of an ectoparasitic flagellate of ciliates: an ultrastructural investigation of the morphology and the mode of attachment of *Spiromonas gonderi* nov. spec. (Zoomastigophora, Spiromonadidae) invading the pellicle of ciliates of the genus *Colpoda* (Ciliophora, Colpodidae). *Protistologica* **20**, 635–648.
- Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Carmona, E. V. & Sogin, M. L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Cryptosporidium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol Biochem Parasitol* **45**, 147–154.
- Goggin, C. L. & Barker, S. C. (1993). Phylogenetic position of the genus *Perkinsus* (Protista, Apicomplexa) based on small subunit ribosomal RNA. *Mol Biochem Parasitol* **60**, 65–70.
- Hnida, J. A. & Duszynski, D. W. (1999). Taxonomy and systematics of some *Eimeria* species of murid rodents as determined by the ITS1 region of the ribosomal gene complex. *Parasitology* **119**, 349–357.
- Hoffman, E. G. & Yancey, R. M. (1966). Ellobiopsidae of Alaskan coastal waters. *Pacific Sci* **20**, 70–78.
- Holmdahl, O. J., Morrison, D. A., Ellis, J. T. & Huong, L. T. (1999). Evolution of ruminant *Sarcocystis* (Sporozoa) parasites based on small subunit rDNA sequences. *Mol Phylogenet Evol* **11**, 27–37.
- Johnson, A. M., Fielke, R., Lumb, R. & Baverstock, P. R. (1990). Phylogenetic relationships of *Cryptosporidium* determined by ribosomal RNA sequence comparison. *Int J Parasitol* **20**, 141–147.
- Keeling, P. J., Luker, M. A. & Palmer, J. D. (2000). Evidence from beta-tubulin phylogenies that microsporidia evolved from within the fungi. *Mol Biol Evol* **17**, 23–31.
- Kozloff, E. N. (1983). *Seashore Life of the Northern Pacific Coast*, 2nd edn. Seattle: University of Washington Press.
- Levine, N. D. (1970). Taxonomy of the Sporozoa. *J Parasitol* **56**, 208–209.
- Levine, N. D. (1971). Taxonomy of the Archigregarinorida and Selenidiidae (Protozoa, Apicomplexa). *J Protozool* **18**, 704–717.
- Levine, N. D. (1976). Revision and checklist of the species of the aseptate gregarine genus *Lecudina*. *Trans Am Microsc Soc* **95**, 695–702.
- Levine, N. D. (1985). Phylum Apicomplexa Levine 1970. In *An Illustrated Guide to the Protozoa*, pp. 322–374. Edited by J. J. Lee, S. H. Hutner & E. C. Bovee. Lawrence, KS: Society of Protozoologists.
- Levine, N. D. (1988). *The Protozoan Phylum Apicomplexa*. Boca Raton, FL: CRC Press.
- López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C. & Moreira, D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**, 603–607.
- Maddison, D. R. & Maddison, W. P. (2000). *MacClade*. Sunderland, MA: Sinauer.
- McLaughlin, R. E. & Myers, J. (1970). *Ophryocystis elektroscirrha* sp. n., a neogregarine pathogen of the monarch butterfly *D. gillippus berenice* Cramer. *J Protozool* **17**, 300–305.
- Meier, M. (1956). Die Monocystideenfauna der Oligochaten von Erlangen und Umgebung. *Arch Protistenkd* **101**, 335–400.
- Mignot, J. P. & Brugerolle, G. (1975). Etude ultrastructurale de flagelle phagotrophe *Colponema loxodes* Stein. *Protistologica* **11**, 429–444.
- Moon-van der Staay, S. Y., De Wachter, R. & Vaulot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**, 607–610.
- Morgan, U. M., Monis, P. T., Fayer, R., Deplazes, P. & Thompson, R. C. (1999). Phylogenetic relationships among isolates of *Cryptosporidium*: evidence for several new species. *J Parasitol* **85**, 1126–1133.
- Nakayama, K., Nishijima, M. & Maruyama, T. (1998). Parasitism by a protozoan in the hemolymph of the giant clam, *Tridacna crocea*. *J Invertebr Pathol* **71**, 193–198.
- Norén, F., Moestrup, Ø. & Rehnstam-Holm, A.-S. (1999). *Parvilucifera infectans* Norén et Moestrup gen. et sp. nov. (Perkinsozoa phylum nov.): a parasitic flagellate capable of killing toxic microalgae. *Eur J Protistol* **35**, 233–254.
- Patterson, D. (1999). The diversity of eukaryotes. *Am Nat* **154**, 96–124.
- Pawlowski, J., Bolivar, I., Fahrni, J. F., Cavalier-Smith, T. & Gouy, M. (1996). Early origin of Foraminifera suggested by SSU rRNA gene sequences. *Mol Biol Evol* **13**, 445–450.
- Perkins, F. O. (1976). Zoospores of the oyster pathogen, *Dermocystidium marinum*. I. Fine structure of the conoid and other sporozoan-like organelles. *J Parasitol* **62**, 959–974.
- Perkins, F. O. (1996). The structure of *Perkinsus marinus* (Mackin, Owen and Collier, 1950) Levine, 1978 with comments on the taxonomy and phylogeny of *Perkinsus* spp. *J Shellfish Res* **15**, 67–87.
- Reece, K. S., Siddall, M. E., Burrenson, E. M. & Graves, J. E. (1997). Phylogenetic analysis of *Perkinsus* based on actin gene sequences. *J Parasitol* **83**, 417–423.
- Siddall, M. E. (1995). Phylogeny of adeleid blood parasites with a partial systematic revision of the haemogregarine complex. *J Eukaryot Microbiol* **42**, 116–125.
- Siddall, M. E. & Barta, J. R. (1992). Phylogeny of *Plasmodium* species: estimation and inference. *J Parasitol* **78**, 567–568.
- Siddall, M. E., Reece, K. S., Graves, J. E. & Burrenson, E. M. (1997). ‘Total evidence’ refutes the inclusion of *Perkinsus* species in the phylum Apicomplexa. *Parasitology* **115**, 165–176.
- Siddall, M. E., Reece, K. S., Nerad, T. A. & Burrenson, E. M. (2001). Molecular determination of the phylogenetic position of a species in the genus *Colpodella* (Alveolata). *Am Mus Novit* **3314**, 1–10.
- Simpson, A. G. B. & Patterson, D. J. (1996). Ultrastructure and identification of the predatory flagellate *Colpodella pugnax* Cienkowski (Apicomplexa) with a description of *Colpodella turpis* n. sp. and a review of the genus. *Syst Parasitol* **33**, 187–198.
- Strimmer, K. & Von Haeseler, A. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol Biol Evol* **13**, 964–969.
- Swofford, D. L. (1999). PAUP\* 4.0. Phylogenetic analysis using parsimony (and other methods). Sunderland, MA: Sinauer.

- Théodoridés, J. (1984).** The phylogeny of the Gregarina. *Orig Life* **13**, 339–342.
- Van de Peer, Y., Caers, A., De Rijk, P. & De Wachter, R. (1998).** Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res* **26**, 179–182.
- Vivier, E. & Desportes, I. (1990).** Phylum Apicomplexa. In *The Handbook of Protoctista*, pp. 549–573. Edited by L. Margulis, J. O. Corliss, M. Melkonian & D. J. Chapman. Boston: Jones and Bartlett.
- Whisler, H. C. (1990).** Incertae Sedis: Ellobiopsida. In *The Handbook of Protoctista*, pp. 715–716. Edited by L. Margulis, J. O. Corliss, M. Melkonian & D. J. Chapman. Boston: Jones and Bartlett.
- Wolters, J. (1991).** The troublesome parasites – molecular and morphological evidence that Apicomplexa belong to the dinoflagellate–ciliate clade. *Biosystems* **25**, 75–83.
- Wray, C. G., Langer, M. R., DeSalle, R., Lee, J. J. & Lipps, J. H. (1995).** Origin of the foraminifera. *Proc Natl Acad Sci U S A* **92**, 141–145.
- Zhu, G., Keithly, J. S. & Philippe, H. (2000).** What is the phylogenetic position of *Cryptosporidium*? *Int J Syst Evol Microbiol* **50**, 1673–1681.
- Zolan, M. E. & Pukkila, P. J. (1986).** Inheritance of DNA methylation in *Coprinus cinereus*. *Mol Cell Biol* **6**, 195–200.