Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal for Parasitology 34 (2004) 501–514
www.parasitology-online.com

The current status of the small subunit rRNA phylogeny of the coccidia (Sporozoa)☆

David A. Morrison[a,*], Set Bornstein[a], Per Thebo[a], Ulrich Wernery[b], Jörg Kinne[b], Jens G. Mattsson[a]

[a]*Department of Parasitology (SWEPAR), National Veterinary Institute and Swedish University of Agricultural Sciences, 751 89 Uppsala, Sweden*
[b]*Central Veterinary Research Laboratory, P.O. Box 597, Dubai, United Arab Emirates*

Received 16 September 2003; received in revised form 29 October 2003; accepted 10 November 2003

Abstract

There is no current comprehensive assessment of the molecular phylogeny of the coccidia, as all recently published papers either deal with subsets of the taxa or sequence data, or provide non-robust analyses. Here, we present a comprehensive and consistent phylogenetic analysis of the available data for the small-subunit ribosomal RNA gene sequence, including a number of taxa not previously studied, based on a Bayesian tree-building analysis and the covariotide model of evolution. The assumptions of the analysis have been rigorously tested, and the benefits and limitations highlighted. Our results provide support for a number of prior conclusions, including the monophyly of the families Sarcocystidae (cyst-forming coccidia) and Eimeriidae (oocyst-forming coccidia), but with bird-host *Isospora* species in the Eimeriidae and mammal-host species in the Sarcocystidae. However, it is clear that a number of previously reported relationships are dependent on the evolutionary model chosen, such as the placements of *Goussia janae*, *Lankesterella minimia* and *Caryospora bigenetica*. Our results also confirm the monophyly of the subfamilies Toxoplasmatinae and Sarcocystinae, but only some of the previously reported groups within these subfamilies are supported by our analysis. Similarly, only some of the previously reported groups within the Eimeriidae are supported by our analysis, and the genus *Eimeria* is clearly paraphyletic. There are unambiguous patterns of host–parasite relationship within the coccidia, as most of the well-supported groups have a consistent and restricted range of hosts, with the exception of the Toxoplasmatinae. Furthermore, the previously reported groups for which we found no support all have a diverse range of unrelated hosts, confirming that these are unlikely to be natural groups. The most interesting unaddressed questions may relate to *Isospora*, which has the fewest available sequences and host–parasite relationships apparently not as straightforward as elsewhere within the suborder.

Keywords: Coccidia; Eimeriorina; Sarcocystidae; Eimeriidae; Phylogeny; Bayesian phylogenetic analysis; 18S rRNA sequences; *Isospora*; Camel

1. Introduction

Isospora orlovi, Tsygankov (Sporozoa: Coccidea) has recently been identified as a cause of diphtheroid-haemorrhagic colitis in dromedary camels, *Camelus dromedarius* (Kinne et al., 2002), the first such report for any isosporan parasite. Like most coccidia, isosporans are intra-cellular parasites of vertebrates, and are known from all continents except Antarctica (Lindsay et al., 1997). Their current taxonomy is based on ultrastructural and life-cycle features (Lindsay et al., 1997), although there have been various recent attempts to elucidate their evolutionary history using the small-subunit ribosomal RNA gene (ssu rRNA) sequence (Carreno et al., 1998; Carreno and Barta, 1999; Franzen et al., 2000; Modrý et al., 2001). However, relatively few *Isospora* species have been sequenced, given the size of the genus, presumably because the lesser medical and veterinary importance of these species has focussed attention elsewhere within the order (Barta, 2001).

It has been suggested that the genus *Isospora* is not monophyletic, with bird-host species being associated with the members of the family Eimeriidae and mammal-host species being associated with those of the Sarcocystidae (Carreno and Barta, 1999). Furthermore, it has been suggested that *I. orlovi* might be a bird pseudoparasite (Péllerdy, 1965); and also the *I. orlovi* lesion found in

☆ Nucleotide sequence data reported in this paper are available in the EMBL/DDBJ databases under the accession number AY365026.

* Corresponding author. Tel.: +46-18-674161; fax: +46-18-309162.
E-mail address: david.morrison@vmm.slu.se (D.A. Morrison).

0020-7519/$30.00
doi:10.1016/j.ijpara.2003.11.006

the camels is reported to be similar to coccidiosis associated with *Eimeria* (Kinne et al., 2002). These mixed suggestions mean that determining the phylogenetic position of this species requires an analysis of the coccidia as a whole (i.e. suborder Eimeriorina sensu Lee et al., 2000, or suborder Eimeriina sensu Tenter and Johnson, 1997; Mehlhorn, 2001) in order to adequately test the possible evolutionary and taxonomic relationships.

Such an analysis thus allows us to provide a timely summary of the current knowledge of the phylogeny of the coccidia as a whole, as deduced from ssu rRNA sequences. This has not been attempted since the work of Morrison and Ellis (1997), and both rRNA gene sequencing projects and techniques for phylogenetic analysis have come a long way since then. Unfortunately, the recently published papers on coccidia deal either with subsets of the taxa or subsets of the sequence data, or they provide non-robust analyses—there has been no recent comprehensive analysis and assessment. One of the main contributions of the work presented here is thus the consistency of the analytical procedures used, which are then applied across the whole of the known coccidia (including several taxa not included in previous analyses). In the process of providing this assessment, we comment on the appropriate use of new evolutionary models, make original observations on the extent and nature of host–parasite relationships within the coccidia, and point out areas of potentially productive future research.

2. Materials and methods

2.1. Materials and sequencing

Oocysts of *I. orlovi* were obtained as described by Kinne et al. (2002), initially cracked with glass beads, and the DNA extracted using a standard phenol–chloroform protocol (Sambrook et al., 1989). The ssu rRNA gene was amplified with primers complementary to conserved regions of the ssu rRNA gene in eukaryotes. The forward primer (5′-AAC CTG GTT GAT CCT GCC AGT-3′) corresponds to nucleotides 1–21 in the *Saccharomyces cerevisiae* sequence and the reverse primer (5′-TGA TCC TTC TGC AGG TTC ACC TAC-3′) corresponds to nucleotides 1777–1795 (Sogin, 1990). A volume of 50 μl PCR mixture contained 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 20 pmol of each primer, 200 μM of each deoxynucleotide, 1 μl of template and 1 U of AmpliTaq DNA polymerase (Applied Biosystems). The amplifications were done in a PE 2400 from Applied Biosystems. After an initial 2 min incubation at 93 °C, the DNA was amplified for 30 cycles consisting of 45 s of denaturation at 94 °C for 30 s of annealing at 42 °C and extension for 2 min at 72 °C. The PCR ended with a final extension for 7 min at 72 °C. PCR products were verified on 1.5% agarose gels with 0.5 μg/ml ethidium bromide. The amplicons were purified over spin columns (QIAquick PCR purification kit, QIAgen) and finally eluted with 30 μl dH_2O. The ssu rRNA gene fragment was sequenced using the original PCR primer as well as primers complementary to conserved regions of the ssu rRNA gene (The European Ribosomal RNA database: http://oberon.fvms.ugent.be:8080/rrna/). BigDye chemistry (Applied Biosystems) was used for all the DNA sequencing reactions, and the samples were analysed on an ABI 3100 capillary DNA sequencer (Applied Biosystems). Both strands were sequenced in overlapping segments, and the sequence data were assembled using the Vector NTI suite (Informax).

2.2. Phylogenetic analyses

The ssu rRNA sequences were aligned using the rRNA secondary-structure information, following the strategy described by Morrison and Ellis (1997). Those coccidia sequences available in aligned form in the European Ribosomal Database (Wuyts et al., 2002) were accessed, plus several other members of the Apicomplexa to serve as an outgroup. Then those coccidia sequences available in the DDBJ/EMBL databases, plus our new sequence, were manually aligned against this prior alignment using MacClade version 4.05 (Maddison and Maddison, 2000). The alignments were then extensively checked by eye to ensure consistency and congruence among taxa with similar sequences.

Sequences were chosen for inclusion in the phylogenetic analysis if their length was >80% of the expected length of ~1800 bp (i.e. >1440 bp). Two shorter partial *Isospora* sequences (from *Isospora gryphoni*, AF080613, and *Isospora ohioensis*, AF029303) were also considered in some of the analyses. Where there were duplicate sequences of supposedly the same taxon (based on information in the sequence database or in relevant publications) these were merged into a consensus sequence if they differed by <10 alignment positions, using the standard IUPAC ambiguity codes for those nucleotide positions with more than one possible character-state in the consensus sequence. The DDBJ/EMBL accession numbers of the sequences included in the final alignment are listed in Fig. 1.

The final alignment contained 91 taxa (based on 138 rRNA sequences) and 2078 aligned nucleotide positions. In addition to the new sequence for *I. orlovi*, the sequences for the following *Eimeria* species have apparently not been phylogenetically analysed before: *Eimeria adenoeides*, *Eimeria catronensis*, *Eimeria ovinoidalis*, *Eimeria pilarensis*; and these species have not been previously discussed: *Eimeria ahsata*, *Eimeria crandallis*, *Eimeria faurei*, *Eimeria weybridgensis*. The beginning and end of some of the aligned sequences was truncated to match those that were less completely sequenced, so that the alignment as analysed covers positions 22–1765 of the 1795 bp *Toxoplasma gondii* ssu rRNA structure shown by Gagnon et al. (1996). At the time of the analysis there were a further

Fig. 1. Inferred phylogenetic tree for the coccidia, based on Bayesian maximum posterior probability. The branch lengths indicate the inferred amount of evolutionary change, according to the scale bar shown. The DDBJ/EMBL accession numbers for each taxon are listed, with these exceptions: *Cyclospora papionis*: AF061566, AF061567, AF061568, AF111187; *Eimeria tenella*: AF026388, U40264, U67121; *Neospora caninum*: AJ271354, L24380, U03069, U16159, U17345, U17346; *Sarcocystis hominis*: AF006470, AF006471, AF176942, AF176943, AF176944, AF176945; *Toxoplasma gondii*: L24381, L37415, M97703, U00458, U03070, U12138, X65508, X68523, X75429, X75430, X75453.

24 partial ssu rRNA sequences of coccidia in the databases that we did not use.

Phylogenetic relationships among all of the aligned ssu rRNA sequences were then examined (Swofford et al., 1996). The robustness of the phylogenetic analyses was assessed by trying various tree-building methods and several different evolutionary models. Neighbour-joining analysis was used to assess the various possible models

of nucleotide evolution, involving corrections both for multiple substitutions and for unequal rates of variation. This was done via hierarchical likelihood-ratio tests (Huelsenbeck and Crandall, 1997), using PAUP* version 4.0b10 (Swofford, D.L., 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods), Ver. 4. Sinauer Associates, Sunderland, MA) and MrModeltest version 1.1b (Nylander, J.A.A., 2002. Testing models of evolution—MrModeltest, Ver. 1.1b. Department of Systematic Zoology, Evolutionary Biology Centre, Uppsala University, Uppsala). Overall nucleotide compositional biases were tested via a goodness-of-fit χ^2-test, using PAUP*, while contingency χ^2-tests were applied to test each pair of taxa, using Tree-Puzzle 5.0 (Strimmer and von Haesler, 1996). Differences between lineages in among-site rate variation were assessed using the inequality test of Lockhart et al. (1998) and Steel et al. (2000), testing the Eimeriidae versus the Sarcocystidae, and also the approximate likelihood-ratio test of Huelsenbeck (2002), testing the overall phylogeny. Reversibility of the nucleotide-substitution model was tested using the goodness-of-fit χ^2-test of Waddell and Steel (1997), based on the final evolutionary tree and the average number of nucleotide substitutions inferred using MacClade. Stationarity of the nucleotide-substitution model can be tested using a triplet Markov analysis (Lin et al., 2002), but this usually requires longer sequences than we have here. So, as a heuristic alternative we carried out a goodness-of-fit χ^2-test comparing the observed substitutions for the Eimeriidae and the Sarcocystidae with the expected values for the whole data set, based on the final evolutionary tree and the average number of nucleotide substitutions inferred using MacClade.

Based on the final evolutionary model chosen by these tests, the data were then analysed via likelihood and Bayesian analysis using MrBayes version 3.0b4 (Ronquist and Huelsenbeck, 2003). The likelihood parameters converged after 70,000 generations, and so this was used as the burn-in period, followed by 1,000,000 further iterations that were sampled every 100 iterations to produce 10,000 trees. The final trees were drawn using TreeView version 1.6.6 (Page, 1996).

The relationship of the two partial *Isospora* sequences to the other *Isospora* sequences was assessed using a subset of the taxa consisting of: all of the *Isospora* sequences, all of the other taxa from the Toxoplasmatinae, plus *Sarcocystis dispersa* and *Sarcocystis mucosa* from the Sarcocystinae, and *Babesia microti* as the final outgroup. Maximum-likelihood analyses were run using PAUP*, with the default heuristic search strategy and 100 random-addition sequence replicates. The parameter values for the likelihood model were estimated from an initial heuristic search based on a best-guess for the final tree.

3. Results

The full-length ssu rRNA sequence of *I. orlovi* has been deposited in the DDBJ/EMBL databases, with accession number AY365026. The final sequence alignment used for the phylogenetic analyses, plus supporting documentation, is available online at http://hem.fyristorg.com/acacia/alignments.htm.

The final evolutionary model chosen from the comprehensive series of tests of the ssu rRNA sequence data was the most complex one currently available (i.e. all of the simpler models were rejected as inadequate). The statistics from the series of the likelihood-ratio tests were: (i) unequal base frequencies, $\delta = 79.1$; (ii) unequal transitions and transversions, $\delta = 482.1$; (iii) unequal transversion rates, $\delta = 158.3$; (iv) invariant sites, $\delta = 2611.3$; and (v) unequal variable sites, $\delta = 1939.2$. All of these are significant at $P < 0.001$. Thus, the most appropriate mathematical model for the phylogenetic analysis must allow the base frequencies to vary, all six substitution rates to vary (i.e. the general time-reversible substitution model, GTR), a proportion of the sites to be invariant, and the variable sites to vary with a gamma distribution—this is referred to here as the 'GTR + I + Γ' model (it is also known in the literature as a 'rates across sites' model).

Furthermore, both the Lockhart-Steel test ($z = 2.78$, $P = 0.008$) and the Huelsenbeck test ($\delta = 11.58$, $P < 0.001$) rejected the null hypothesis of equal evolutionary rates among lineages. Therefore, the most appropriate mathematical model must also allow the among-site rate variation to vary between different lineages (i.e. in addition to the GTR + I + Γ model)—this is referred to here as the 'covariotide' model. This form of analysis can currently only be done using Bayesian analysis, which was therefore used as our main inference tool for the phylogeny of the coccidia.

Fig. 1 shows the Bayesian maximum posterior probability (MAP) tree based on the final evolutionary model chosen from the various tests of the ssu rRNA sequence data (i.e. the covariotide model). This tree is a majority-rule consensus tree from the 10,000 trees sampled by the Bayesian analysis. The clade-credibility values from this analysis, which indicate the proportion of the 10,000 sampled trees that contained each of the branches (and which have recently been shown to be more accurate measures of branch support than are bootstrap values, Alfaro et al., 2003; Douady et al., 2003), are shown in Fig. 2. As an aid to the biological interpretation of the tree, it can be stated for these particular data that branches with clade credibilities <0.25 are not supported by any unequivocal nucleotide changes (i.e. there are no character-state changes that are unambiguously inferred to have occurred in the ancestor represented by that branch), while branches with values >0.90 are supported by at least four such nucleotide changes. The major clades within the coccidia that we

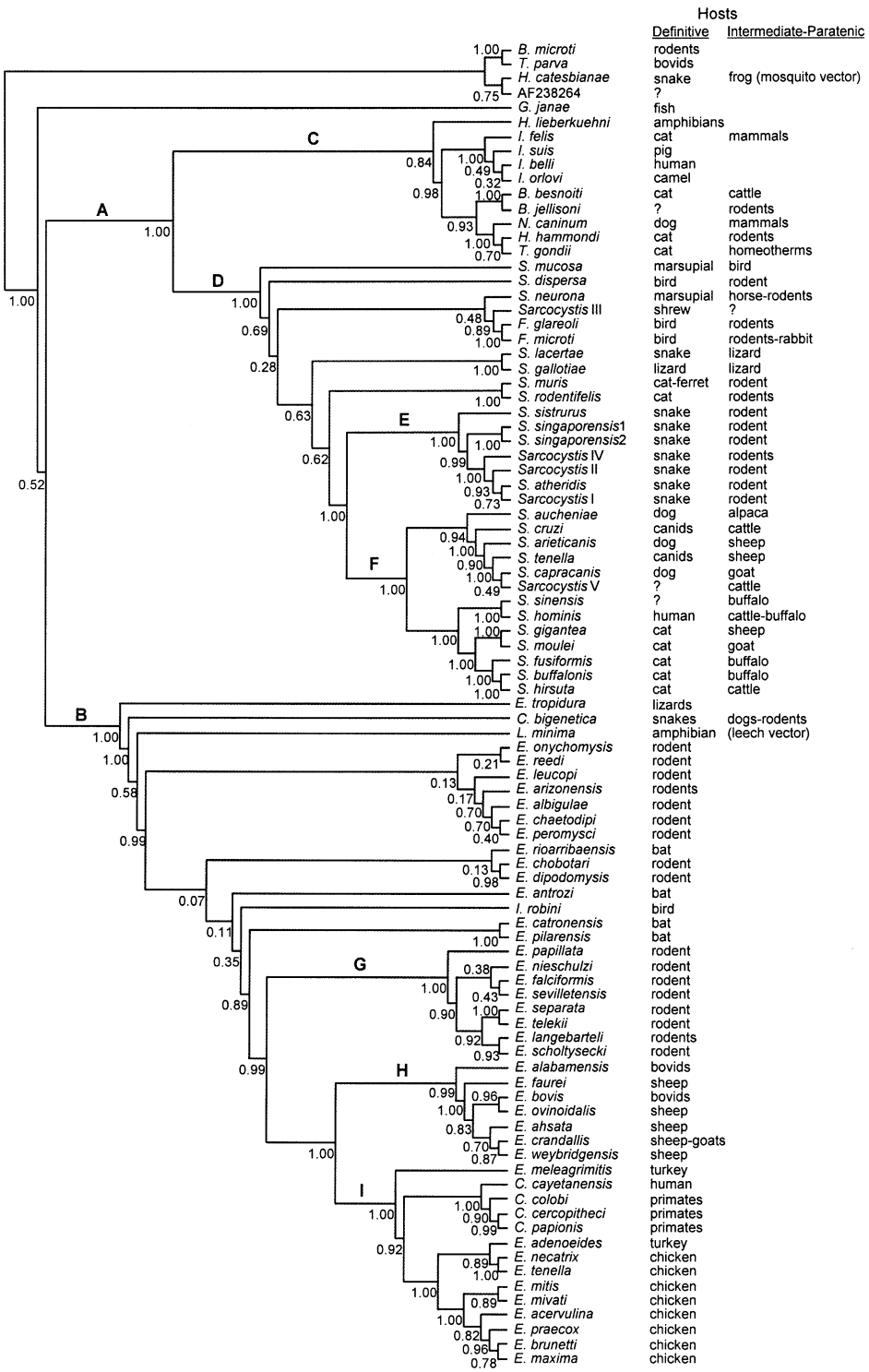

Fig. 2. Bayesian maximum posterior probability tree for the coccidia with the clade-credibility (or posterior probability) values shown for each branch of the tree, which indicate the proportion of the sampled trees containing that branch. Also indicated are the known hosts of each parasite taxon, including the definitive host (where sexual development of the parasite occurs), the intermediate host (with continuing asexual reproduction) and the paratenic host (with no development but where infectious stages can accumulate) if these are included in the known life cycle. The well-supported clades labelled A–I are discussed in the text. For genus names please refer to Fig. 1.

consider to be well-supported by the ssu rRNA data are indicated by the letters A–I in Fig. 2.

Starting from the root, the MAP tree shows two well-supported main clades, labelled A and B in Fig. 2. These clades correspond in most respects to the taxa sampled from the currently recognised families Sarcocystidae (A) and Eimeriidae (B). In the analysis shown, *Goussia janae* forms the sister to these two clades, but this relationship is only

weakly supported. Furthermore, this relationship occurs only if the covariotide model is used—if the non-covariotide model (i.e. GTR + I + Γ) is used then this species is the sister to group B, but once again only with weak support (clade credibility 0.70).

Within group A there are two well-supported groups, corresponding in many respects to the taxa sampled from the currently recognised subfamilies Toxoplasmatinae (C) and Sarcocystinae (D). However, members of *Isospora* have not traditionally been included within the Toxoplasmatinae. There are two well-supported clades within group D, labelled E and F, with the remainder of the species showing a very poorly supported set of relationships at the base of these two groups. If the non-covariotide model is used, then these basal taxa form a poorly supported monophyletic group.

Within group B there are three well-supported clades, labelled G, H and I in Fig. 2, with the remainder of the species showing sometimes poorly supported relationships at the base of these three groups. The placement of *Lankesterella minimia* and *Caryospora bigenetica* at the base of group B is well supported, but their relationship to each other is less clear—all of our analyses show them as paraphyletic with weak support, but an alternative weakly supported relationship is for them to form a monophyletic group. *Eimeria tropidura* is placed at the base of group B in all of the analyses.

The separate phylogenetic analysis of the relationships of the two partial *Isospora* sequences (using the GTR + I + Γ model only) produced the tree expected for the subset of taxa included, based on the tree shown in Figs. 1 and 2. In addition, it indicated an unambiguous sister relationship between *I. gryphoni* and *Isospora robini*, but an ambiguous placement of *I. ohioensis*, which formed a polychotomy with *Isospora suis* and the *Isospora belli–I. orlovi* clade.

Various tests were also carried out to assess some of the other assumptions of the chosen evolutionary model. The test of overall variation in base composition was non-significant, as was the comparison of the base composition of the Sarcocystidae and Eimeriidae. However, the pairwise contingency χ^2-tests for compositional biases indicated that the following taxa have unusual base compositions: *Hepatozoon catesbianae*, *Sarcocystis aucheniae*, *Sarcocystis buffalonis*, *Sarcocystis fusiformis*, *Sarcocystis gigantea*, *Sarcocystis hominis* and *Sarcocystis hirsuta*. Furthermore, the test of reversibility of the nucleotide-substitution model was significant ($\chi^2 = 61.49, P < 0.001$), with the T → A and T → C substitutions occurring at a much higher rate than the A → T and C → T substitutions, respectively. Finally, the test of stationarity of the nucleotide-substitution model was significant ($\chi^2 = 27.52$, $P = 0.004$), with the Eimeriidae having greater than expected C ↔ G and C ↔ T substitutions and fewer than expected A ↔ G and A ↔ T substitutions.

4. Discussion

A phylogenetic analysis is only as good as the steps taken to ensure the highest quality of data and to evaluate and use the most appropriate mathematical model for the data analysis. We presume that any discrepancies that exist between our analysis of the ssu rRNA data and the analyses of previous authors may be due to artefacts of the diversity of analysis techniques used by those authors, such as differences in taxon sampling, sequence alignment, partial sequence deletion, and the evolutionary model or tree-building algorithm used. We therefore emphasise that we have applied a consistent and comprehensive analysis philosophy across the whole of the coccidia sampled to date (including a number of sequences that have apparently not appeared in analyses before now), which should make our results comparable across the whole group (as well as providing improved taxon sampling, particularly within *Eimeria*). Our results are thus likely to be an accurate summary of the current status of the study of the phylogeny of the coccidia based on analysis of ssu rRNA gene sequence data, although they are unlikely to be perfect. The following points can be made from our summary. We start by considering the various aspects of phylogenetic analysis as they apply to the coccidia, before proceeding to a consideration of the biology of the coccidia themselves, especially *Isospora*.

4.1. Methodology

Sequence alignment should be phylogenetic not phenetic (Barta, 1997, 2001; Morrison and Ellis, 1997). Thus, we need to use alignment methods that try to maximise the evolutionary content of the resulting alignment, because the details of the alignment are known to greatly affect the results of phylogenetic analyses involving rRNA (Ellis and Morrison, 1995; Morrison and Ellis, 1997; Beebe et al., 2000; Hickson et al., 2000; Mugridge et al., 2000). To this end, it has been suggested that incorporating into the alignment procedure information on the structure and function of the molecule that the sequenced gene codes for will be of importance (Ellis and Morrison, 1995; Hickson et al., 2000), and this is explicitly the strategy that we have adopted here. We therefore believe that our alignments are robust and informative.

It seems to be frequently overlooked that all data analyses are based on some form of model, whether explicit or not, which specifies the assumptions that need to be met by the data in order for the results of the analyses to be reliable. In a phylogenetic analysis of sequence data the underlying model takes the form of assumptions about the process of nucleotide (or amino acid) substitution that will be used to infer the unknown events of the evolutionary history. Such assumptions include the relative frequencies of the bases, the rates of change among bases through time, the ability of bases at different alignment positions to

change, and whether the rates of change remain constant through time. We have extensively tested all of these aspects for the evolutionary models that we have used.

We chose to assess the possible models using hierarchical likelihood-ratio tests (Huelsenbeck and Crandall, 1997), which sequentially compare simpler models to more complex models in an attempt to find the simplest model that adequately fits the data. For our data, this procedure chose the most complex model available, which allows the base frequencies to vary, all six possible substitution rates to vary (GTR), a proportion of the sites to be invariant (I), and the variable sites to vary with a discrete gamma distribution (Γ). This leads to the GTR + I + Γ model.

Previous phylogenetic analyses of the coccidia have generally not followed this protocol of likelihood-ratio tests, but have simply chosen a restricted range of much simpler evolutionary models for comparison. The exceptions are the studies of Jirků et al. (2002) and Šlapeta et al. (2003), which followed the same protocol as we did but ended their testing by choosing a model that either does not allow all six possible substitution rates to vary or does not allow a proportion of the sites to be invariant (see also Perkins and Keller, 2001). Only Šlapeta et al. (2002a) have previously used the GTR + I + Γ model, but without explanation for its choice—they also used much greater values for their parameters than we have estimated here.

However, all of these models make the same basic assumption, that the model does not change along the evolutionary lineages (mathematically, the model is said to be stationary). In order to consider a more realistic model of evolution, we therefore considered it important to explicitly evaluate the covariotide model for the coccidia data, as this allows the rates of nucleotide substitution to vary between evolutionary lineages. This model has rarely been considered for the coccidia (Zhu et al., 2000), and we thus emphasise that our analyses indicate that the covariotide model is a superior one for phylogenetic analysis of the coccidia. This model can potentially change the shape of the tree (i.e. the branching order), and will almost certainly change the branch lengths and the values of some of the other model parameters, due to more effective estimates of saturation of nucleotide-substitutions (Galtier, 2001). For our data, all three of these effects were observed.

In addition, we tested the stationarity of the nucleotide composition among the coccidian lineages, and only detected some minor but statistically significant variation. Unfortunately, we also detected statistically significant variation when we tested both the stationarity of the nucleotide-substitution model and the reversibility of the substitution rates. Thus, our data do not accurately fit a time-reversible model, and even if they did do so, they still do not fit a model that assumes the same process of nucleotide substitutions in all of the evolutionary lineages. It is unclear just how large such variations can be before they affect the accuracy of phylogeny reconstruction (i.e. before the model assumptions are sufficiently violated so that inaccurate results are produced), and this may thus be the biggest limitation of our analyses.

In addition to the evolutionary model, for a worthwhile phylogenetic analysis the data set needs to have complete or near-complete sequences, and reasonable taxon sampling. Earlier studies of the coccidia often came to what is now considered to be a wrong conclusion, or the right conclusion for the wrong reason, because so few species were included in the sample, which then showed supposedly 'close' relationships among taxa solely because more closely related species were missing from the analysis. In parasitology taxon sampling is usually opportunistic (Barta, 2001), which clearly cannot create a statistically representative sample. Prior problems with taxon sampling are clearly shown by our analysis of the coccidia. For example, our analyses confirm the discussion of Holmdahl et al. (1999) concerning the relationships of ruminant-host *Sarcocystis* taxa from canid and felid definitive hosts, which had previously been misinterpreted due to small sample sizes (i.e. the host–parasite relationships turned out to be more complex).

The choice of an appropriate outgroup is also important for phylogenetic analysis because this determines the root of the tree, and it is thus involved in the tests of monophyly within the ingroup. For robust phylogenetic analysis the outgroup needs to consist of several members of the sister taxon to the ingroup, preferably the ones with relatively short branch lengths to the ingroup (Smith, 1994). As far as the coccidia is concerned several taxa have been commonly used as outgroups in previous studies. However, species of *Cryptosporidium* are phylogenetically too distant from the coccidia to be considered as the immediate sister group (Morrison and Ellis, 1997; Carreno et al., 1999; Zhu et al., 2000), and thus they were not included in our analyses even though they have traditionally been treated as part of the coccidia. Similarly, species of *Plasmodium* all have a long insertion in their ssu rRNA sequence that makes the branch length to other Apicomplexans too long to be useful (Morrison and Ellis, 1997; Xiao et al., 2002). We therefore used representatives of *Babesia* and *Theileria* (piroplasms) and *Hepatozoon* (haemogregarine) as the outgroup taxa, because other analyses of the Apicomplexa have suggested that these are the most closely related of the currently available sequenced taxa (Mathew et al., 2000; Barta, 2001; Barta et al., 2001; Perkins and Keller, 2001). However, in retrospect, the unusual base composition and the long terminal branch length indicate that *H. catesbianae* is not particularly useful as an outgroup in phylogenetic analyses of the coccidia, and we cannot recommend its continued use.

The choice of gene sequence(s) to be used for phylogeny reconstruction in the coccidia is somewhat problematic. Our analysis shows that at the base of the phylogenetic tree the ssu rRNA gene provides good phylogenetic signal, but does not do so equally within the Eimeriidae and Sarcocystidae. There is much less ssu rRNA sequence variation among taxa within the Eimeriidae than within the Sarcocystidae.

This means that sequence alignment is relatively straightforward within the Eimeriidae but that the relationships among the species are often hard to discern due to very short branch lengths (Fig. 1; cf. Zhao et al., 2001). Consequently, another gene sequence will be needed to further explore the species relationships within this family. For example, Zhao and Duszynski (2001b) and Zhao et al. (2001) have suggested using the plastid 23S gene, and Zhao and Duszynski (2001a) have suggested using the plastid ORF470, to supplement the ssu rRNA information.

Within the Sarcocystidae, on the other hand, there is a greater mixture of variabilities along the ssu rRNA sequence. Parts of the alignment are quite straightforward, but other parts, such as in the vicinity of helices E10-1 (domain 2), E21-1, E21-3, E21-5 (domain 4), 41 (domain 7) and 47 (domain 9) (all numbering according to Gagnon et al., 1996), are problematic. This has encouraged most people to delete these regions from their alignments, possibly losing phylogenetic information in the process (see below). Indeed, even within a species the ssu rRNA data can be quite variable, and even among strains from a single species (e.g. *Sarcocystis singaporensis*; Šlapeta et al., 2002a). The ssu rRNA sequence thus seems to be a good phylogenetic marker within this family. However, other genes still need to be examined, in order to distinguish gene phylogeny from species phylogeny (see below).

Although there are objective criteria for deleting regions of variable (or ambiguous) alignment in phylogenetic analyses (Castresana, 2000), deleting variable regions can be counter-productive for the robust reconstruction of phylogeny (Barta, 1997; Lutzoni et al., 2000). For ribosomal DNA these regions will usually be those parts of the sequence representing single-stranded loops between the helices, and there is no a priori reason to expect that phylogenetic signal will be absent from these regions (Morrison and Ellis, 1997; Beebe et al., 2000; Mugridge et al., 2000). Indeed, for our data there are clear synapomorphies within these regions, and these synapomorphies account for some of the differences in support for our groups compared to those found by other researchers. In other words, the phylogenetic signal in the variable regions supplements that in the regions of unambiguous alignment, but it may or may not be congruent with it. For example, the region consisting of the E21-1, E21-3 and E21-5 helices is a region of unambiguous alignment for the Eimeriidae and Toxoplasmatinae but is hypervariable for the Sarcocystinae. Deleting this region from any phylogenetic analysis would therefore reduce the amount of apparent support for the distinction of these taxonomic groups. Strategies such as the staggered sequence alignment discussed by Barta (1997) are thus essential for successful sequence alignment of ssu rRNA.

4.2. The phylogeny of the coccidia

Note, first, that the MAP tree (Fig. 2) is a summary of the support for the various clades shown, rather than being an 'optimal' tree from some specified criterion (such as maximum-likelihood). That is, it is basically a consensus tree that represents the combination of the best-supported groups of species. This analytical approach assumes that it is reasonable to suggest that the relationships among the well-supported groups represent the probable true phylogeny. Conversely, the poorly supported groups are unlikely to represent the true phylogeny, except by random chance, and therefore these parts of the tree require further investigation.

4.2.1. Phylogenetic patterns

Since this paper is essentially an overview, we have restricted our observations here to the main phylogenetic patterns rather than considering the details about exactly which species has which inferred relationships. Resolution of these details would require further detailed analyses of subsets of the taxa.

The two main clades resulting from our analysis correspond in most respects to the taxa sampled from the cyst-forming coccidia (A) and the intestinal oocyst-forming coccidia (B), which are basically the currently recognised families Sarcocystidae and Eimeriidae, respectively. The obvious exceptions to this simple pattern are the placement of *Goussia*, *Isospora* and *Lankesterella*.

Goussia has traditionally been placed in the Eimeriidae although other placements have been suggested (Jirků et al., 2002). The phylogenetic analysis of Jirků et al. (2002) supports possible placement within the Eimeriidae but only with weak support. Our analyses confirm that this placement has only weak support, and moreover indicate that this results from the use of a specific evolutionary model (i.e. the location was different for the covariotide and the $GTR + I + \Gamma$ models). Clearly the choice of evolutionary model for analysis is important here. Our analyses also indicate that a sister relationship to both the Eimeriidae and Sarcocystidae is a slightly more likely hypothesis for *Goussia* (Fig. 2). Further taxon sampling within suggested alternative family placements of *Goussia*, such as the Barroussiidae and Calyptosporidae, as well as further sampling within *Goussia* itself, will be needed in order to finally resolve this issue.

We also included in our analyses the rRNA sequence of an unidentified symbiont from a zooanthellate coral (accession number AF238264) because the analysis of Kuvardina et al. (2002) indicated that it might represent a sister species to the coccidia. However, none of our analyses support this placement of the sequence, instead indicating that it has a closer evolutionary relationship to *Hepatozoon*. However, the relative lengths of the branches (Fig. 1) indicate that even this is not a close relationship. More ssu rRNA sequences from the suborder Adeleorina will be needed to investigate this matter further.

Lankesterella is traditionally placed in a separate family Lankesterellidae while *Caryospora* is placed within the Eimeriidae (Lee et al., 2000). However, the phylogenetic analyses of Barta et al. (2001) and Šlapeta et al. (2003)

suggest that these two taxa form a clade, but only with weak support. As with the placement of *Goussia* discussed above, our analyses confirm that this placement has only weak support, and moreover also indicate that it requires a specific evolutionary model to be used. Clearly the choice of evolutionary model for analysis is important here as well. Our analyses indicate that a paraphyletic sister relationship to (most of) the Eimeriidae is a slightly more likely evolutionary hypothesis for both genera (Fig. 2). Certainly, the ssu rRNA data indicate that if *Lankesterella* is placed taxonomically outside the Eimeriidae then *Caryospora* needs the same taxonomic treatment and so do *Eimeria* species such as *E. tropidura*. Further taxon sampling within the Lankesterellidae and the basal taxa of the Eimeriidae (especially other *Caryospora* species) is needed in order to finally resolve this issue.

Isospora is often placed in the family Eimeriidae (cf. Tenter and Johnson, 1997), although it has been pointed out that this placement is not supported by the ssu rRNA data (Carreno et al., 1998). Furthermore, it is now clear that *Isospora* is polyphyletic (Carreno and Barta, 1999; Franzen et al., 2000; Modrý et al., 2001), with the bird-host species being allied to the Eimeriidae and the mammal-host species allied to the Toxoplasmatinae. This situation is confirmed by our analyses, and additional support is provided by the inclusion of the *I. orlovi* sequence with the other mammal-host species (Fig. 2). Furthermore, our phylogenetic analyses provide robust support for the placement of *Hyaloklossia lieberkuehni* (*Isospora lieberkuehni*) as the sister to the Toxoplasmatinae clade rather than within the *Isospora* clade (Fig. 2), thus confirming the opinion of Modrý et al. (2001) that it should be placed in a separate genus.

Although our results agree with the basic phylogenetic pattern reported within the Sarcocystinae (our clade D) by previous workers (Jeffries et al., 1997; Tenter and Johnson, 1997; Votýpka et al., 1998; Doležel et al., 1999; Holmdahl et al., 1999; Yang et al., 2001; Šlapeta et al., 2001a, 2002a, 2003), the recognition of three clades within the subfamily (Doležel et al., 1999; Šlapeta et al., 2001a, 2002a, 2003) is not supported by our analyses. Two of the previously recognised groups are well supported (our clades E and F) but the third is not. This third 'group' only appears, with low support, when the GTR + I + Γ model is used, and not when the improved covariotide model is used. Similarly, although our results agree with the basic phylogenetic pattern reported within the Eimeriidae (our clade B) by previous workers (Barta et al., 1997; Šlapeta et al., 2001b; Zhao and Duszynski, 2001a,b; Zhao et al., 2001), the recognition of all four proposed clades is not supported. Three of the groups are well supported (our clades G–I) but the fourth is not; this fourth 'group' does not appear in any of our analyses. Our conclusions thus have close parallels in both the Sarcocystidae and Eimeriidae: several of the clades within the main genus of each family have been well sampled, and these form robust evolutionary groups, but the relationships among the remainder of the species will not be resolved without the sampling of many more sequences. These poorly placed species are shown at the base of clades B and D in Fig. 2. They have very short branch lengths (Fig. 1), which then have very weak branch support (Fig. 2), and so the evolutionary relationships among the taxa may not be determinable using ssu rRNA sequences.

It has previously been noted that recognition of the genus *Frenkelia* renders *Sarcocystis* paraphyletic (Votýpka et al., 1998; Doležel et al., 1999; Šlapeta et al., 2001a, 2002a, 2003) and that recognition of the genus *Cyclospora* renders *Eimeria* paraphyletic (Relman et al., 1996; Pieniazek and Herwaldt, 1997; Eberhard et al., 1999; Lopez et al., 1999; Shields and Olson, 2003; although cf. Barta, 2001), and this parallel pattern within the two families is confirmed by our more extensive analyses. It has been formally suggested that *Frenkelia* be synonymised with *Sarcocystis* (Votýpka et al., 1998; *Frenkelia microti* = *Sarcocystis buteonis*, *Frenkelia glareoli* = *Sarcocystis glareoli*), but it has only been informally suggested that *Cyclospora* be synonymised with *Eimeria* (Pieniazek and Herwaldt, 1997). Certainly, the ssu rRNA sequence data do not provide any support for the continued recognition of *Cyclospora* if the classification is to represent the evolutionary history of the organisms, but congruent support from other genes is still lacking, as is sampling of the other known *Cyclospora* species (e.g. those from snake, mole and rodent hosts). Most host types have both an eimeriid and a sarcocystid parasite, and so perhaps *Cyclospora* is simply the 'primate' version of *Eimeria*, there being no other *Eimeria* species recorded from humans for example.

Furthermore, the ssu rRNA data make it clear that the genus *Eimeria* as currently recognised is paraphyletic in several different ways. In addition to the placement of *Cyclospora* within the *Eimeria* clade, part of the genus *Isospora* is robustly placed there as well, and *E. tropidura* is placed as the sister to the entire clade (i.e. including *Caryospora*, *Cyclospora*, *Isospora* and *Lankesterella*). Thus, a major re-assessment of the taxonomy *Eimeria* is called for, with the likely outcome that it will be split into several smaller genera.

The phylogenetic relationships that we have found among the members of the Toxoplasmatinae sensu stricto (*Besnoitia*, *Hammondia*, *Neospora*, *Toxoplasma*) are as predicted from previous analyses of their ssu rRNA (Tenter and Johnson, 1997; Ellis et al., 2000). Importantly, however, the sequence differences among the species of the *Toxoplasma–Neospora–Hammondia* group provide the most glaring taxonomic anomaly in our data set. There are only 3–5 ssu rRNA nucleotide differences among the species of these separate genera, compared to 16 differences between the two species of *Besnoitia*, 20–24 differences between *Besnoitia* and these three genera, and 13–26 differences among the four sampled species of *Isospora*. If nothing else, it is obvious that the recognition of

Table 1
Percentage content of A + T nucleotides in various taxonomic groups shown in Fig. 2

Taxon	AT content (%)
Outgroup	55.0
Goussia janae	55.0
Clade C	54.5
Basal part of clade D	54.4
Clade E	55.9
Clade F	57.3
Basal part of clade B	53.5
Clade G	54.1
Clade H	53.1
Clade I	52.8

Toxoplasma, *Neospora* and *Hammondia* as separate genera is not based upon the distinctiveness of their rRNA sequences—in order to be comparable to the rest of the coccidia they would be separate species of a single genus based on this criterion. Indeed, studies of other gene sequences do not reveal large differences among these three genera either (Ellis et al., 1999; Jenkins et al., 1999; Mugridge et al., 1999; Šlapeta et al., 2002b), and such small genetic differences associated with large life-cycle differences may be a product of the clonal nature of *Toxoplasma* (Sibley, 2003). Nevertheless, it is high time that the taxonomy of this group was properly re-evaluated (Šlapeta et al., 2002b).

Finally, our statistical tests indicated that there is little overall variation in base composition among the ssu rRNA sequences when the phylogenetic relationships are ignored but that there are some statistically significant differences among individual taxa. In fact, there *are* consistent phylogenetic patterns in %AT content among the sequences (Table 1). Thus, the ancestor of the coccidia is inferred to have ca. 55% AT base composition, which has been maintained within the Toxoplasmatinae, but within the Sarcocystinae this composition has drifted towards increasing AT content through evolutionary time while within the Eimeriidae it has drifted towards decreasing AT content. Nevertheless, the coccidian ssu rRNA gene seems to be much less variable than is the same gene in *Plasmodium* or mycoplasmas, for example.

4.2.2. Host–parasite relationships

Most species of coccidia are fairly host-specific, and so host–parasite relationships are of particular interest in phylogenetic studies. There are clear patterns of host–parasite relationship within the coccidia, as shown in Fig. 2. Each of the well-supported groups of coccidia E–H has a consistent range of hosts associated with it. We interpret this pattern as confirmatory evidence that these are not arbitrary groupings but are likely to reflect the true evolutionary relationships among the taxa. This evidence is independent of the sequence data used to generate the tree.

The taxa in groups E and F are heteroxenous (dioxenous). The taxa in group E have a consistent relationship to both the definitive (snake) and intermediate (rodent) hosts, while the taxa in group F have a consistent relationship to the intermediate (bovid) hosts and form two well-supported subgroups with respect to the definitive (canid or felid) hosts. The taxa in groups G and H are monoxenous (homoxenous), and have a consistent relationship to their hosts (rodents and bovids, respectively). Perhaps the most remarkable thing about these consistent host–parasite relationships is that they cross continental boundaries, so that close evolutionary relationships reflect host-specificity rather than geographical co-location.

The taxa in group I are also monoxenous, but there are two groups of hosts, primates and galliform birds. With the exception of *Eimeria meleagrimitis*, these host relationships form two well-supported subgroups. The phylogenetic placement of *E. meleagrimitis* as the sister to these two subgroups is well supported (Fig. 2), and it occurs in all of our analyses. Therefore, the ssu rRNA data analysis strongly indicates that the primate-host eimeriid species are derived from a bird-host ancestor. This hypothesised origin has not been made explicit in previous analyses of *Cyclospora* with fewer *Eimeria* taxa, which have indicated *Cyclospora* as either the sister to the bird-host *Eimeria* (Relman et al., 1996; Pieniazek and Herwaldt, 1997; Eberhard et al., 1999; Lopez et al., 1999) or to the rodent-host *Eimeria* (Shields and Olson, 2003), although the pattern is shown in the phylogenetic tree of Jirků et al. (2002) with much less support than we have found here. Evidence from other genes will be needed to test this new hypothesis.

In contrast to these patterns, the taxa in the well-supported group C do not show clear host–parasite relationships. The amphibian-hosted coccidian species is the sister to the mammal-hosted species (Fig. 2), as expected, but within the mammal-hosted species the relationships are mixed. This is not necessarily unexpected within *Isospora* due to the poor taxon sampling (see below), but most of the Toxoplasmatinae sensu stricto have been sampled and therefore the contrasting pattern is very marked. At this stage we can offer no simple explanation for this antithetical pattern, nor have previous commentators (e.g. Tenter and Johnson, 1997; Doležel et al., 1999; Šlapeta et al., 2001a, 2003).

As noted in the previous section, one ‘evolutionary group’ identified by previous authors within *Eimeria* and within *Sarcocystis* does not stand up to close scrutiny. In both cases it is a coccidian group with a mixture of hosts, and in our analyses both of these break up into smaller groups with more-consistent internal host relationships. For example, the grouping of bat- and rodent-hosted *Eimeria* species suggested by Zhao and Duszynski (2001a, 2001b) and Zhao et al. (2001), which mixes two different hosts, is not supported in our analyses, nor is the grouping of miscellaneous hosts in group ‘B’ of Doležel et al. (1999) and Šlapeta et al. (2001a, 2002a, 2003). We interpret this

discordance between the coccidian groups and their hosts as confirmatory evidence that these previously recognised groups were not natural evolutionary groupings, and we predict that each subgroup will in fact form host-related groups when taxon sampling of these groups is increased.

Furthermore, it is notable (Fig. 2) that the sister taxon to the rest of the coccidia is a fish-host species, and that the basal species within the Toxoplasmatinae and the Eimeriidae are reptile- or amphibian-host species, while the crown species are mammal- or bird-host species. These patterns suggest a co-evolutionary relationship between the coccidia and their hosts, given that poikilotherms are usually considered to be basal to homeotherms. If so, then the diversification of the coccidia occurred at the same time as the diversification of the vertebrate hosts themselves.

However, the general patterns of host–parasite relationship *within* the mammal- and bird-host species are unlikely to be the result of host–parasite co-evolution, and so we have not performed any formal coevolutionary analyses. Given the relative distributions of the parasites on the hosts, the overall evolutionary history is much more likely to have been 'invasion' followed by speciation in the new host type. However, there may be host–parasite coevolutionary patterns within closely related groups of parasite species, and this hypothesis could be tested by further sampling of ssu rRNA sequences within these groups, as has been done for other Apicomplexan taxa such as *Cryptosporidium* (Xiao et al., 2002).

Finally, there are two species of *Sarcocystis* for which the definitive hosts are as yet unknown (Fig. 2). One of the many uses for knowledge about phylogenetic relationships is the ability to make predictions about unknown information such as this, and consequently it is important to explicitly consider this situation. The unnamed *Sarcocystis* species reported by Yang et al. (2001), shown as *Sarcocystis* V in Figs. 1 and 2, is robustly included in a clade that has canids as the definitive host, and so the clear prediction is that this species will have the same host. Similarly, *Sarcocystis sinensis* is robustly included in a clade that has felids (or in one case humans) as their definitive host, and so the clear prediction is that this species will have the same host.

4.3. Isospora *and* I. orlovi

One of the original motivations for our analysis of the coccidia was an examination of the phylogenetic relationship of *I. orlovi* to the other *Isospora* species. Our phylogenetic analysis provides independent confirmation of two points. First, the organism isolated from the camels is an *Isospora* species rather than a *Sarcocystis* species, because it robustly forms part of a clade with other *Isospora* species. Second, it is not an avian-host species (i.e. a pseudoparasite) but is a true camel parasite, as it robustly forms part of a clade with the other mammal-host *Isospora* species rather than with the bird-host *Isospora* species. These conclusions are congruent with the ultrastructural observation that the *I. orlovi* sporocysts lack stieda bodies (Kinne et al., 2002), which is also a feature of the other mammal-host *Isospora* but not the bird-host species (Carreno and Barta, 1999).

However, other aspects of the phylogenetic placement of *I. orlovi* are still ambiguous, because the relationships among the mammal-host *Isospora* species are not well supported in our analyses. This ambiguity in placement is easily explained. Based on the ssu rRNA data, *I. belli* is equally related genetically to both *I. suis* (13 nucleotide differences) and *I. orlovi* (14 nucleotide differences) but *I. suis* and *I. orlovi* are much less closely related to each other (21 nucleotide differences). *Isospora felis*, on the other hand, is equally unrelated to all of the other species (19–26 differences). Therefore, in the phylogenetic trees either *I. orlovi* or *I. suis* can appear as the sister to *I. belli* or as the sister to the whole group of *Isospora* species.

One possible way of resolving this part of the phylogeny will be to acquire sequences from more of the closely related *Isospora* taxa, as information about the relationships of these taxa to other species should stabilise the reconstructed pattern. We therefore attempted to further evaluate the situation within this group of species by including in an analysis the available partial ssu rRNA sequence (1392 bp) for *I. ohioensis*. However, the result of this analysis was not useful, as the only portion of the sequence available does not even provide sufficient phylogenetic information to form a dichotomously branching tree. This species thus needs to be sequenced fully.

Nevertheless, in all of our analyses those mammal-host *Isospora* species that are believed to be monoxenous (*I. belli*, *I. orlovi*, *I. suis*) form a clade, even if the facultatively heteroxenous species (*I. felis*, *I. ohioensis*) do not (Fig. 2). However, based on our analyses it seems that the artiodactyl-host species of *Isospora* (*I. suis* and *I orlovi*) do not form a monophyletic group, unlike the situation for the other monoxenous coccidia where species with similar hosts are sisters. This pattern of disrupted host–parasite relationships parallels that observed within the Toxoplasmatinae sensu stricto, suggesting that it might be a real biological pattern rather than an artefact of poor taxon sampling (which is possible since there are only two taxa sampled to date). Alternatively, the development of *I. orlovi* is reported to be quite different from that of most other *Isospora* species (Kinne et al., 2002), developing and causing haemorrhagic enteritis within the colon itself. It is possible that evolutionary patterns associated with this different developmental strategy may confound any host-related phylogenetic pattern.

The bird-host *Isospora* species (*I. gryphoni*, *I. robini*) are robustly placed within the Eimeriidae in our analyses, which is the conventional taxonomic placement (Lee et al., 2000). The analysis including the available partial ssu rRNA sequence (582 bp) for *I. gryphoni* places it as the sister to *I. robini* even using such a small sequence length.

This placement renders *Eimeria* paraphyletic, just as does *Cyclospora*. Unfortunately, the precise placement of these *Isospora* species in relation to *Eimeria* is not well resolved in our analyses (cf. Barta, 2001), and so no positive predictions can yet be made.

The evolutionary history of the taxa currently included in *Isospora* thus needs a thorough investigation, as it is clearly not as simple as the current taxonomy (based on a few ultrastructural and life-cycle features) suggests, and the taxonomy needs a major re-assessment. For further details of some of the historical confusions regarding isosporan taxonomy see Carreno et al. (1998) and Carreno and Barta (1999).

4.4. The future

The remaining 'big' questions concerning the coccidia may relate to the relationships within the genus *Isospora*. Despite their importance, including parasitising domestic animal hosts, this genus has the fewest available ssu rRNA sequences for such a large group—for example, there are only five sequences from ca. 250 species, compared to 30 sequences from ca. 150 species for *Sarcocystis* and 40 sequences from ca. 1200 species for *Eimeria*. So, it remains the most poorly investigated group within the coccidia—even the common species of domesticated dogs and cats have not all been sampled (see Lindsay et al., 1997), let alone those from other host taxa. Furthermore, the host–parasite relationships are clearly not as straightforward as they apparently are elsewhere within the suborder, and so the patterns may be biologically revealing. Similarly, the patterns of homoxeny/heteroxeny, including the existence of paratenic hosts, are clearly not as simple as elsewhere within the suborder, and elucidation of the evolutionary patterns of these life cycles will also be biologically interesting. The apparent relationships (Lindsay et al., 1997; Carreno and Barta, 1999) between ultrastructural features, such as the presence or absence of stieda bodies, and life-history features, such as the number and type of hosts, also remain to be investigated in more detail.

Members of the genera *Sarcocystis* and/or *Eimeria* also parasitise many other host species, such as domesticated pigs, goats, horses, rabbits, geese and ducks, as well as numerous other vertebrates. Clearly, if the host–parasite relationships of the coccidia are going to be evaluated in any further detail then these other hosts need to be included in the analysis. While it is useful to collect further samples of coccidia from types of hosts for which samples already exist, this is not as useful as sampling species from as-yet-unsampled host types.

Furthermore, there are several remaining small families (e.g. Aggregatidae, Calyptosporiidae, Elleipsisomatidae, Selenococcidiidae, Spirocystidae) and genera (especially within the Eimeriidae) within which no species have yet had their ssu rRNA sequenced. These groups have different biological characteristics that led to their distinction in the first place, such as the number of sporozoites per sporocyst and the number of sporocysts per oocyst. The evolutionary relationships implied by these ultrastructural features need testing, and the relationships of these taxa thus require further genetic investigation.

Finally, we still need another gene or two at the suborder level in order to produce a species phylogeny that could be used for the purpose of producing a robust taxonomy (Tenter et al., 2002). The ribosomal internal transcribed spacer 1 (ITS1) sequence may be useful for closely related species (Hnida and Duszynski, 1999; Ellis et al., 2000; Šlapeta et al., 2002b,c), but probably not otherwise. The evolutionary information in the nuclear large-subunit (lsu) ribosomal RNA gene is apparently too similar to the ssu rRNA to provide independent phylogenetic evidence (Mugridge et al., 1999, 2000), although both the plastid lsu rRNA (Zhao et al., 2001) and ssu rRNA (Oborník et al., 2002) gene sequences have both been used successfully. Protein-coding regions should therefore be the most useful for phylogenetic purposes, as independent evidence, including possibly: the plastid open reading frame ORF470 (Zhao and Duszynski, 2001a), and the genes for actin (Xiao et al., 2002), cytosolic 70-kDa heat-shock protein (Xiao et al., 2002; Zhu et al., 2000), α- or β-tubulin (Zhu et al., 2000), and RNA polymerase II or even DNA polymerase α.

Acknowledgements

This study was in part funded by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) and in part by the Swedish International Development Cooperation Agency (SIDA).

References

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.

Barta, J.R., 1997. Investigating phylogenetic relationships within Apicomplexa using sequence data: the search for homology. Methods 13, 81–88.

Barta, J.R., 2001. Molecular approaches for inferring evolutionary relationships among protistan parasites. Vet. Parasitol. 101, 175–186.

Barta, J.R., Martin, D.S., Liberator, P.A., Dashkevicz, M., Anderson, J.W., Feighner, S.D., Elbrecht, A., Perkins-Barrow, A., Jenkins, M.C., Danforth, H.D., Ruff, M.D., Profous-Juchelka, H., 1997. Phylogenetic relationships among eight *Eimeria* species infecting domestic fowl inferred using complete small subunit ribosomal DNA sequences. J. Parasitol. 83, 262–271.

Barta, J.R., Martin, D.S., Carreno, R.A., Siddall, M.E., Profous-Juchelka, H., Hozza, M., Powles, M.A., Sundermann, C., 2001. Molecular phylogeny of the other tissue coccidia: *Lankesterella* and *Caryospora*. J. Parasitol. 87, 121–127.

Beebe, N.W., Cooper, R.D., Morrison, D.A., Ellis, J.T., 2000. Subset partitioning of the ribosomal DNA small subunit and its effects on the phylogeny of the *Anopheles punctulatus* group. Insect Mol. Biol. 9, 515–520.

Carreno, R.A., Barta, J.R., 1999. An eimeriid origin of isosporoid coccidia with stieda bodies as shown by phylogenetic analysis of small subunit ribosomal RNA gene sequences. J. Parasitol. 85, 77–83.

Carreno, R.A., Schnitzler, B.E., Jeffries, A.C., Tenter, A.M., Johnson, A.M., Barta, J.R., 1998. Phylogenetic analysis of coccidia based on 18S rDNA sequence comparison indicates that *Isospora* is most closely related to *Toxoplasma* and *Neospora*. J. Eukaryot. Microbiol. 45, 184–188.

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.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Doležel, D., Koudela, B., Jirků, M., Hypša, V., Oborník, M., Votýpka, J., Modrý, D., Šlapeta, J.R., Lukeš, J., 1999. Phylogenetic analysis of *Sarcocystis* spp. of mammals and reptiles supports the coevolution of *Sarcocystis* spp. with their final hosts. Int. J. Parasitol. 29, 795–798.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.

Eberhard, M.L., da Silva, A.J., Lilley, B.G., Pieniazek, N.J., 1999. Morphologic and molecular characterization of new *Cyclospora* species from Ethiopian monkeys: *C. cercopitheci* sp.n., *C. colobi* sp.n., and *C. papionis* sp.n. Emerg. Infect. Dis. 5, 651–658.

Ellis, J., Morrison, D.A., 1995. Effects of sequence alignment on the phylogeny of *Sarcocystis* deduced from 18S rDNA sequences. Parasitol. Res. 81, 696–699.

Ellis, J.T., Morrison, D.A., Liddell, S., Jenkins, M.C., Mohammed, O.B., Ryce, C., Dubey, J.P., 1999. The genus *Hammondia* is paraphyletic. Parasitology 118, 357–362.

Ellis, J.T., Holmdahl, O.J.M., Ryce, C., Njenga, J.M., Harper, P.A.W., Morrison, D.A., 2000. Molecular phylogeny of *Besnoitia* and the genetic relationships among *Besnoitia* of cattle, wildebeest and goats. Protist 151, 329–336.

Franzen, C., Müller, A., Bialek, R., Diehl, V., Salzberger, B., Fätkenheuer, G., 2000. Taxonomic position of the human intestinal protozoan parasite *Isospora belli* as based on ribosomal RNA sequences. Parasitol. Res. 86, 669–676.

Gagnon, S., Bourbeau, D., Levesque, R.C., 1996. Secondary structures and features of the 18S, 5.8S and 26S ribosomal RNAs from the Apicomplexan parasite *Toxoplasma gondii*. Gene 173, 129–135.

Galtier, N., 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. Mol. Biol. Evol. 18, 866–873.

Hickson, R.E., Simon, C., Perrey, S.W., 2000. The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. Mol. Biol. Evol. 17, 530–539.

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.

Holmdahl, O.J.M., Morrison, D.A., Ellis, J.T., Huong, L.T.T., 1999. Evolution of ruminant *Sarcocystis* (Sporozoa) parasites based on small subunit rDNA sequences. Mol. Phylogenet. Evol. 11, 27–37.

Huelsenbeck, J.P., 2002. Testing a covariotide model of DNA substitution. Mol. Biol. Evol. 19, 698–707.

Huelsenbeck, J.P., Crandall, K.A., 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst. 28, 437–466.

Jeffries, A.C., Schnitzler, B., Heydorn, A.O., Johnson, A.M., Tenter, A.M., 1997. Identification of synapomorphic characters in the genus *Sarcocystis* based on 18S rDNA sequence comparison. J. Eukaryot. Microbiol. 44, 388–392.

Jenkins, M.C., Ellis, J.T., Liddell, S., Ryce, C., Munday, B.L., Morrison, D.A., Dubey, J.P., 1999. The relationship of *Hammondia hammondi* and *Sarcocystis mucosa* to other heteroxenous cyst-forming coccidia as inferred by phylogenetic analysis of the 18S SSU ribosomal DNA sequence. Parasitology 119, 135–142.

Jirků, M., Modrý, D., Šlapeta, J.R., Koudela, B., Lukeš, J., 2002. The phylogeny of *Goussia* and *Choleoeimeria* (Apicomplexa: Eimeriorina) and the evolution of excystation structures in coccidia. Protist 153, 379–390.

Kinne, J., Ali, M., Wernery, U., Dubey, J.P., 2002. Clinical large intestinal coccidiosis in camels (*Camelus dromedarius*) in the United Arab Emirates: description of lesions, endogenous stages, and redescription of *Isospora orlovi*, Tsygankov, 1950 oocysts. J. Parasitol. 88, 548–552.

Kuvardina, O.N., Leander, B.S., Aleshin, V.V., Myl'Nikov, A.P., Keeling, P.J., Simdyanov, T.G., 2002. The phylogeny of colpodellids (Alveolata) using small subunit rRNA gene sequences suggests they are the free-living sister group to Apicomplexans. J. Eukaryot. Microbiol. 49, 498–504.

Lee, J.J., Leedale, G.F., Bradbury, P. (Eds.), 2000. Illustrated Guide to the Protozoa, 2nd ed., Society of Protozoologists, Lawrence, KA.

Lin, Y.-H., McLenachan, P.A., Gore, A.R., Phillips, M.J., Ota, R., Hendy, M.D., Penny, D., 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Mol. Biol. Evol. 19, 2060–2070.

Lindsay, D.S., Dubey, J.P., Blagburn, B.L., 1997. Biology of *Isospora* spp. from humans, nonhuman primates, and domestic animals. Clin. Microbiol. Rev. 10, 19–34.

Lockhart, P.J., Steel, M.A., Barbrook, A.C., Huson, D.H., Charleston, M.A., Howe, C.J., 1998. A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. Mol. Biol. Evol. 15, 1183–1188.

Lopez, F.A., Manglicmot, J., Schmidt, T.M., Yeh, C., Smith, H.V., Relman, D.A., 1999. Molecular characterization of *Cyclospora*-like organisms from baboons. J. Infect. Dis. 179, 670–676.

Lutzoni, F., Wagner, P., Reeb, V., Zoller, S., 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst. Biol. 49, 628–651.

Maddison, D.R., Maddison, W.P., 2000. MacClade: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, MA.

Mathew, J.S., van den Bussche, R.A., Ewing, S.A., Malayer, J.R., Latha, B.R., Panciera, R.J., 2000. Phylogenetic relationships of *Hepatozoon* (Apicomplexa: Adeleorina) based on molecular, morphologic, and life-cycle characters. J. Parasitol. 86, 366–372.

Mehlhorn, H. (Ed.), 2001. Encyclopaedic Reference of Parasitology, 2nd ed., Springer, Berlin.

Modrý, D., Šlapeta, J.R., Jirků, M., Oborník, M., Lukeš, J., Koudela, B., 2001. Phylogenetic position of a renal coccidium of the European green frogs, *Isospora lieberkuehni* Labbé, 1894 (Apicomplexa: Sarcocystidae) and its taxonomic implications. Int. J. Syst. Evol. Microbiol. 51, 767–772.

Morrison, D.A., Ellis, J.T., 1997. Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of Apicomplexa. Mol. Biol. Evol. 14, 428–441.

Mugridge, N.B., Morrison, D.A., Heckeroth, A.R., Johnson, A.M., Tenter, A.M., 1999. Phylogenetic analysis based on full-length large subunit ribosomal RNA gene sequence comparison reveals that *Neospora caninum* is more closely related to *Hammondia heydorni* than to *Toxoplasma gondii*. Int. J. Parasitol. 29, 1545–1556.

Mugridge, N.B., Morrison, D.A., Jäkel, T., Heckeroth, A.R., Tenter, A.M., Johnson, A.M., 2000. Effects of sequence alignment and structural domains of ribosomal DNA on phylogeny reconstruction for the protozoan family Sarcocystidae. Mol. Biol. Evol. 17, 1842–1853.

Oborník, M., Jirků, M., Šlapeta, J.R., Modrý, D., Koudela, B., Lukeš, J., 2002. Notes on coccidian phylogeny, based on the apicoplast small subunit ribosomal DNA. Parasitol. Res. 88, 360–363.

Page, R.D.M., 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Applic. Biosci. 12, 357–358.

Péllerdy, L.P., 1965. Coccidia and Coccidiosis. Akadémiai Kiadó, Budapest.

Perkins, S.L., Keller, A.K., 2001. Phylogeny of nuclear small subunit rRNA genes of hemogregarines amplified with specific primers. J. Parasitol. 87, 870–876.

Pieniazek, N.J., Herwaldt, B.L., 1997. Reevaluating the molecular taxonomy: is human-associated *Cyclospora* a mammalian *Eimeria* species? Emerg. Infect. Dis. 3, 381–383.

Relman, D.A., Schmidt, T.M., Gajadhar, A., Sogin, M., Cross, J., Yoder, K., Sethabutr, O., Echeverria, P., 1996. Molecular phylogenetic analysis of *Cyclospora*, the human intestinal pathogen, suggests that it is closely related to *Eimeria* species. J. Infect. Dis. 173, 440–445.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shields, J.M., Olson, B.H., 2003. *Cyclospora cayetanensis*: a review of an emerging parasitic coccidium. Int. J. Parasitol. 33, 371–391.

Sibley, L.D., 2003. Recent origins among ancient parasites. Vet. Parasitol. 115, 185–198.

Šlapeta, J.R., Modrý, D., Votýpka, J., Jirků, M., Koudela, B., Lukeš, J., 2001a. Multiple origin of the dihomoxenous life cycle in sarcosporidia. Int. J. Parasitol. 31, 413–417.

Šlapeta, J.R., Modrý, D., Votýpka, J., Jirků, M., Oborník, M., Lukeš, J., Koudela, B., 2001b. *Eimeria telekii* n.sp., (Apicomplexa: Coccidia) from *Lemniscomys striatus* (Rodentia: Muridae): morphology, pathology and phylogeny. Parasitology 122, 133–143.

Šlapeta, J.R., Kyselová, I., Richardson, A.O., Modrý, D., Lukeš, J., 2002a. Phylogeny and sequence variability in the *Sarcocystis singaporensis* Zaman and Colley, (1975) 1976 ssrDNA. Parasitol. Res. 88, 810–815.

Šlapeta, J.R., Modrý, D., Kyselová, I., Hořejš, R., Lukeš, J., Koudela, B., 2002b. Dog shedding oocysts of *Neospora caninum*: PCR diagnosis and molecular phylogenetic approach. Vet. Parasitol. 109, 157–167.

Šlapeta, J.R., Koudela, B., Votýpka, J., Modrý, D., Hořejš, R., Lukeš, J., 2002c. Coprodiagnosis of *Hammondia heydorni* in dogs by PCR based of ITS 1 rRNA: differentiation from morphologically indistinguishable oocysts of *Neospora caninum*. Vet. J. 163, 147–154.

Šlapeta, J.R., Modrý, D., Votýpka, J., Jirků, M., Lukeš, J., Koudela, B., 2003. Evolutionary relationships among cyst-forming coccidia *Sarcocystis* spp. (Alveolata: Apicomplexa: Coccidea) in endemic African tree vipers and perspective for evolution of heteroxenous life cycle. Mol. Phylogenet. Evol. 27, 464–475.

Smith, A.B., 1994. Rooting molecular trees: problems and strategies. Biol. J. Linn. Soc. 51, 279–292.

Sogin, M.L., 1990. Amplification of ribosomal RNA genes for molecular evolution studies. In: Innes, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego, CA, pp. 307–322.

Steel, M., Huson, D., Lockhart, P.J., 2000. Invariable sites models and their use in phylogeny reconstruction. Syst. Biol. 49, 225–232.

Strimmer, K., von Haesler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, 2nd ed., Sinauer Associates, Sunderland, MA, pp. 407–514.

Tenter, A.M., Johnson, A.M., 1997. Phylogeny of the tissue cyst-forming coccidia. Adv. Parasitol. 39, 69–139.

Tenter, A.M., Barta, J.R., Beveridge, I., Duszynski, D.W., Mehlhorn, H., Morrison, D.A., Thompson, R.C.A., Conrad, P.A., 2002. The conceptual basis for a new classification of the coccidia. Int. J. Parasitol. 32, 595–616.

Votýpka, J., Hypša, V., Jirků, M., Flegr, J., Vávra, J., Lukeš, J., 1998. Molecular phylogenetic relatedness of *Frenkelia* spp. (Protozoa, Apicomplexa) to *Sarcocystis falcatula* Stiles, 1893: is the genus *Sarcocystis* paraphyletic? J. Eukaryot. Microbiol. 45, 137–141.

Waddell, P.J., Steel, M.A., 1997. General time-reversible distances with unequal rates across sites: mixing Γ and inverse gaussian distributions with invariant sites. Mol. Phylogenet. Evol. 8, 398–414.

Wuyts, J., Van de Peer, Y., Winkelmans, T., De Wachter, R., 2002. The European database on small subunit ribosomal RNA. Nucleic Acids Res. 30, 183–185.

Xiao, L., Sulaiman, I.M., Ryan, U.M., Zhou, L., Atwill, E.R., Tischler, M.L., Zhang, X., Fayer, R., Lal, A.A., 2002. Host adaptation and host–parasite co-evolution in *Cryptosporidium*: implications for taxonomy and public health. Int. J. Parasitol. 32, 1773–1785.

Yang, Z.-Q., Zuo, Y.-X., Yao, Y.-G., Chen, X.-W., Yang, G.-C., Zhang, Y.-P., 2001. Analysis of the 18S rRNA genes of *Sarcocystis* species suggests that the morphologically similar organisms from cattle and water buffalo should be considered the same species. Mol. Biochem. Parasitol. 115, 283–288.

Zhao, X., Duszynski, D.W., 2001a. Phylogenetic relationships among rodent *Eimeria* species determined by plastid ORF470 and nuclear 18S rDNA sequences. Int. J. Parasitol. 31, 715–719.

Zhao, X., Duszynski, D.W., 2001b. Molecular phylogenies suggest the oocyst residuum can be used to distinguish two independent lineages of *Eimeria* spp in rodents. Parasitol. Res. 87, 638–643.

Zhao, X., Duszynski, D.W., Loker, E.S., 2001. Phylogenetic position of *Eimeria antrozoi*, a bat coccidium (Apicomplexa: Eimeriidae) and its relationship to morphologically similar *Eimeria* spp. from bats and rodents based on nuclear 18S and plastid 23S rDNA sequences. J. Parasitol. 87, 1120–1123.

Zhu, G., Keithly, J.S., Philippe, H., 2000. What is the phylogenetic position of *Cryptosporidium*? Int. J. Syst. Evol. Microbiol. 50, 1673–1681.