Effects of Sequence Alignment and Structural Domains of Ribosomal DNA on Phylogeny Reconstruction for the Protozoan Family Sarcocystidae

N. B. Mugridge,* D. A. Morrison,* T. Jäkel,† A. R. Heckerth,‡ A. M. Tenter,‡ and A. M. Johnson,*

*Molecular Parasitology Unit, University of Technology, Sydney, New South Wales, Australia; †Division of Parasitology, University of Hohenheim, Stuttgart, Germany; and ‡Institut für Parasitologie, Tierärztliche Hochschule Hannover, Hannover, Germany

Finding correct species relationships using phylogeny reconstruction based on molecular data is dependent on several empirical and technical factors. These include the choice of DNA sequence from which phylogeny is to be inferred, the establishment of character homology within a sequence alignment, and the phylogeny algorithm used. Nevertheless, sequencing and phylogeny tools provide a way of testing certain hypotheses regarding the relationship among the organisms for which phenotypic characters demonstrate conflicting evolutionary information. The protozoan family Sarcocystidae is one such group for which molecular data have been applied phylogenetically to resolve questionable relationships. However, analyses carried out to date, particularly based on small-subunit ribosomal DNA, have not resolved all of the relationships within this family. Analysis of more than one gene is necessary in order to obtain a robust species signal, and some DNA sequences may not be appropriate in terms of their phylogenetic information content. With this in mind, we tested the informativeness of our chosen molecule, the large-subunit ribosomal DNA (lsu rDNA), by using subdivisions of the sequence in phylogenetic analysis through PAUP, fastDNAml, and neighbor joining. The segments of sequence applied correspond to areas of higher nucleotide variation in a secondary-structure alignment involving 21 taxa. We found that subdivision of the entire lsu rDNA is inappropriate for phylogenetic analysis of the Sarcocystidae. There are limited informative nucleotide sites in the lsu rDNA for certain clades, such as the one encompassing the subfamily Toxoplasmatae. Consequently, the removal of any segment of the alignment compromises the final tree topology. We also tested the effect of using two different alignment procedures (CLUSTAL W and the structure alignment using DCSE) and three different tree-building methods on the final tree topology. This work shows that congruence between different methods in the formation of cladases may be a feature of robust topology; however, a sequence alignment based on primary structure may not be comparing homologous nucleotides even though the expected topology is obtained. Our results support previous findings showing the paraphyly of the current genera Sarcocystis and Hammondia and again bring to question the relationships of Sarcocystis muris, Isospora felis, and Neospora caninum. In addition, results based on phylogenetic analysis of the structure alignment suggest that Sarcocystis zamani and Sarcocystis singaporensis, which have reptilian definitive hosts, are monophyletic with Sarcocystis species using mammalian definitive hosts if the genus Frenkelia is synonymized with Sarcocystis.

Introduction

For taxa that cannot be unequivocally classified based on phenotypic markers, DNA sequencing provides another marker for the investigation of phylogenetic relationships (Olsen and Woese 1993). The relative ease with which nucleotide sequences can be obtained today allows one to compare a large number of taxa based on nucleotide variations; hence, the phylogenetic informativeness of DNA now has the potential to outweigh that of the morphology or biology of an organism. However, the informativeness of the gene being used requires consideration if one is to make assumptions regarding species relationships (Olsen and Woese 1993). One way of ensuring a robust phylogeny in which the species signal, rather than the gene signal, emerges is to analyze numerous genes. Consequently, the DNA sequence chosen for the construction of a phylogeny is a very important factor, due to the varying phylogenetic content among genes. The phylogenetic informativeness of a gene can be a factor of the function of its product or a result of its length. In some cases, such as that of the stems and loops of the ribosomal RNA, different parts of the gene can have varying information contents due to evolutionary selection pressures on secondary structure and function (Sogin and Gundersen 1986; Olsen 1987, 1988; Olsen and Woese 1993).

The different methods available for building a phylogenetic tree can also affect the outcome of the analysis. Tree-building algorithms have different aims and assumptions. Therefore, congruence between tree topologies resulting from the use of varying algorithms can also be seen as a sign of a robust topology. The process of sequence alignment is also crucial, as it has been shown that nonhomologous alignment is the major cause of erroneous phylogenies (Morrison and Ellis 1997). Previous studies have shown that using a primary-structure alignment instead of a secondary-structure alignment can influence how well cladistic relationships within the Sarcocystidae are resolved (Ellis and Morrison 1995).

Ribosomal RNA is a useful molecule for phylogenetic analysis because it holds certain advantages for such studies. The conserved core regions that are interspersed among variable stretches of sequence assist in
the design of primers that can then be used to amplify even the more divergent areas. This mosaic structure also allows flexibility in experimental design of phylogenetic investigations so that closely related and less closely related organisms can be studied (Sogin and Gunderson 1986; Olsen 1987; Olsen and Woese 1993).

Study of the phylogenetic informativeness of the large-subunit ribosomal DNA (lsu rDNA) helped us to resolve some unanswered questions regarding the Sarcocystidae. This family includes heteroxenous (two-host life cycle) coccidian parasites that have the ability to form tissue cysts in intermediate hosts. The family is usually divided into a range of different genera, such as Sarcocystis, Frenkelia, Toxoplasma, Besnoitia, Hammondia, Neospora, and Cystoisospora (Frenkel, Mehlhorn, and Heydorn 1987; Levine 1988; Current, Upton, and Long 1990; Tenter and Johnson 1997). Thus far, taxa within this family have been classified based on phenotypic characters, but recent results based on molecular data call for the revision of some of these genera. Phenotypically, the Sarcocystidae are distinguished from other coccidia by their heteroxenous life cycle, their ability to form tissue cysts in the intermediate host, and the morphology of their oocyst (disporous, tetrazoic). Division into two or three subfamilies is also based on phenotypic characters. The genera Sarcocystis and Frenkelia, which make up the subfamily Sarcocystinae, are obligatorily heteroxenous and develop only inside their hosts. The subfamily Toxoplasmatae includes the genera Toxoplasma, Besnoitia, Hammondia, and Neospora, which collectively consist of far fewer species than belong to the genus Sarcocystis alone. These species differ from the Sarcocystinae in that they have a different type of asexual development in the intermediate host, they have an additional asexual phase of proliferation preceding sexual development in the definitive host, and they sustain sporogony in the environment (Frenkel 1977; Levine 1988; Current, Upton, and Long 1990). A third, monogeneric, subfamily, the Cystoisosporinae, has been described based on the formation of monozoic cysts in optional intermediate hosts (Frenkel et al. 1979). However, the genus Cystoisospora has not been widely accepted and is most frequently synonymized with the genus Isospora in the family Eimeriidae.

Although the familial grouping of the Sarcocystidae has been widely accepted according to phenotypic characters, the biological and morphological data regarding current generic groups are sparse and conflicting (Tenter and Johnson 1997). Thus, in step with molecular biology advances, this has resulted in the use of other, more informative, markers, namely DNA sequences, for classification or reconstruction of phylogeny within the Sarcocystidae.

Phylogenetic research on the Sarcocystidae and other members of the phylum Apicomplexa has largely been based on the small-subunit ribosomal RNA (ssu rDNA) gene sequence (Tenter and Johnson 1997; Carreno et al. 1998; Votýpka et al. 1998; Carreno and Barta 1999; Dole El et al. 1999; Holmdahl et al. 1999; Jenkins et al. 1999). It is clear that the smaller size of the ssu rDNA makes it more popular when longer sequences are technically harder to obtain, and the helical regions of the ssu rDNA are phylogenetically more informative than the nonhelical regions (Ellis and Morrison 1995). Ssu rDNA failed to be sufficiently informative in some cases, as evidenced by conflicting phylogenies obtained for the Sarcocystidae (Jeffries et al. 1997; Tenter and Johnson 1997; Votýpka et al. 1998). To resolve some relationships and confirm others postulated from the ssu rDNA analyses, we recently published work on the entire lsu rDNA (Mugridge et al. 1999a, 1999b), based on the fact that this molecule is more informative than the ssu rDNA (Olsen 1987; Cavalier-Smith 1989). This work confirmed the paralogy of the genus Sarcocystis, unless Frenkelia was included, and the paralogy of the genus Hammondia, unless it was combined with Toxoplasma gondii and Neospora caninum. The results supported the hypothesis that the genus Frenkelia should be incorporated into the genus Sarcocystis (Odening 1998) and validated the biological evidence that showed N. caninum to be closer to Hammondia heydorni, and T. gondii to be closer to Hammondia hammondi (Mugridge et al. 1999a).

Here, we investigate the information content of the lsu rDNA gene through the subdivision of the nucleotide sequence and the effect of this on three tree-building methods, using members of the protozoan family Sarcocystidae as a model. We also include members of another family of coccidia, the Eimeriidae, which contains homoxenous (one-host life cycle) parasites, such as the genera Eimeria and Isospora. In addition, we address the importance of the alignment of sequences by observing the effect of using two methods, one of which we believe is superior in achieving homologous nucleotide alignment.

We present a detailed study on the informativeness of the lsu rDNA based on the analysis of particular domains and segments and their contribution in terms of phylogenetic information for the family Sarcocystidae. Using a larger data set of 21 taxa, we investigated the informativeness of lsu rDNA segments corresponding to secondary structures, which appeared to be hot spots for nucleotide variation. The informativeness of random segments of the molecule was also studied. Our data set served as a model for the analysis of congruence of results among different alignment strategies, alignment parameters, and tree-building methods and the sensitivity of these methods to the subdivision of data.

In addition, our data set included two newly added species of Sarcocystis, Sarcocystis zamani and Sarcocystis singaporenensis, which use reptiles as definitive hosts (Beaver and Maleckar 1981). This paper reports the relationships of these species to Sarcocystis species which use mammals as definitive hosts and also shows the branching of the Toxoplasmatine species Besnoitia besnoiti and of Isospora felis (syn. Cystoisospora felis), which is often classified into the family Eimeriidae.

**Materials and Methods**

**Genomic DNA**

Genomic DNA of Sarcocystis arieticanis, Sarcocystis capracanis, Sarcocystis cruzi, Sarcocystis gigan-
Polymerase Chain Reaction Amplification, Cloning, and Sequencing

The 18-taxa list included the following taxa: Sarcocystis muis, Sarcocystis neurona, Sarcocystis muris, Sarcocystis miescheriana, Sarcocystis moulei, Sarcocystis zinckei, Besnoitia besnoiti, Frenkelia glareoli, Frenkelia microti, H. hammondi, T. gondii (ME49 strain), and I. felis. The shortest sequence was 3,212 bp (in I. felis), and the longest was 3,500 bp (in S. arietianis). Lsu rDNA sequences of N. caninum and Eimeria tenella were obtained from GenBank (accession numbers AF001946 and AF026388). The T. gondii RH strain sequence was a consensus of a GenBank submission (X75429) and the sequence published by Ellis et al. (1998). We aligned sequences according to their primary structure similarity using CLUSTAL W, version 1.5, with default gap values (Thompson, Higgins, and Gibson 1994) and according to their secondary structure using the DCSE program by De Rijk and De Wachter (1993), by way of observing the influence of two alignment strategies on the resulting trees. Both alignments were subject to three different tree-building methods: maximum parsimony (Farris 1970), neighbor joining (a distance method) (Saitou and Nei 1987), and maximum likelihood (Fukami and Tateno 1989). These methods were implemented in the computer programs PAUP, version 3.1.1 (Swofford 1990); neighbor in the PHYLIP version 3.5c, package (Felsenstein 1995); and fastDNAml, version 1.0.6 (Olsen et al. 1994), respectively. Maximum parsimony was carried out using the stepwise-addition option and a heuristic search method with 10 random starts. The Kimura two-parameter distance model was used for maximum likelihood and neighbor-joining analyses. Maximum likelihood used global rearrangements and a transition/transversion ratio of 2.0, and each analysis was repeated three times with the jumble option. Spatial variation in evolutionary rates along the sequence was also investigated for maximum likelihood using a heuristic procedure. The DNArates, version 1.0.3, program was used to infer the evolutionary rate at each sequence position for the full structure alignment, including the possibility of invariant sites. These rates were then used as nine possible rate categories using the weights and categories options in fastDNAml. This analysis changed the branch lengths of the resulting trees but not the branching order. Therefore, only the analyses using the equal-rates option are shown here.

A member of the genus Eimeria, E. tenella, which is closely related to the Sarcocystidae, was chosen as the outgroup for all of the analyses.

Informative Nucleotide Sites of the Lsu rDNA Molecule

To find the domains or stems of the Lsu rDNA molecule that were contributing most to informativeness, we subdivided the original full-length secondary-structure alignment into several shorter alignments, each including the individual sequence segments to be investigated through tree-building. First, we divided the structure alignment into two separate alignments, one including double-stranded, helical regions and the other including single-stranded, nonhelical regions. Other potentially informative domains or segments were found using the program MacClade, version 3.07. Trees from the secondary-structure alignment were examined using the MacClade, version 3.07, program under the branch reconstruction and character changes options. The first option reconstructs branch changes in a parsimonious manner and lists the sites that support each branch (Maddison and Maddison 1992). The latter option charts the minimum and maximum numbers of changes that could have occurred at each nucleotide site for a particular tree.

In this manner, segments seeming to contain a large number of sites involved in change and branch formation were taken as probable informative segments for phylogeny reconstruction within the Sarcocystidae. The structure alignment was then subdivided to produce alignments each containing one, or a combination, of the chosen segments. Each of these alignments was independently subjected to the three tree-building methods.

To test the informativeness of random segments of the Lsu rDNA, the entire sequence was divided in half without any consideration of secondary structure. The two resulting alignments thus represented the 5' end and the 3' end of the Lsu rDNA. Once again, these two regions were subjected to the three tree-building alignments.

Results

Phylogenetic Analysis of Full-Length Lsu rDNA

Sequences were aligned by two alignment strategies in order to observe the effect on tree building. An example tree with bootstraps and branch lengths based on full-length Lsu rDNA and the maximum-likelihood tree-building method is shown in figure 1.
FIG. 1.—Phylogenetic relationships among the Sarcocystidae as inferred from the structure alignment of the full-length lsu rDNA and the maximum-likelihood tree-building method for 21 taxa. Neospora caninum and the Toxoplasma gondii strains split the genus Hammondia; Isospora felis is a sister taxon to the subfamily Sarcocystinae; Sarcocystis muris is a sister taxon to Sarcocystis neurona, Frenkelia glareoli, and Frenkelia microti; Sarcocystis zamani and Sarcocystis singaporensis are placed within the genus Sarcocystis. The branch lengths are proportional to the amount of inferred evolutionary change, and the numbers on the branches are the numbers of times that the branch was supported in 100 bootstrap replicates.

The importance of the alignment step in phylogeny is well established (Morrison and Ellis 1997). Any errors in alignment will lead to potential errors in the trees. To corroborate the importance of the resultant alignment, we varied the gap opening and gap extension values in the CLUSTAL W alignment (Wheeler 1995; Morrison and Ellis 1997). This was done following the method of Wheeler (1995), whereby the values increase logarithmically (gap extension range of 0.031–4, gap opening range of 0.5–64), and using the experimental plan of Morrison and Ellis (1997), which tests all orthogonal combinations of values (72 in total). Results (not shown) for this experiment clearly demonstrated that alignment parameters could be influential. Eight different topologies were obtained from the 72 alignments, consisting of varying gap opening and extension penalty values, using PAUP as the tree-building method, and seven different topologies were obtained using fastDNAml.

The trees obtained from the full-length CLUSTAL W sequence alignment using maximum parsimony, maximum likelihood, and neighbor joining mostly supported the same phylogenetic groupings. The tree given by maximum likelihood is shown in figure 2. The group including the Toxoplasmatinae as inferred from the structure alignment of the full-length lsu rDNA and the maximum-likelihood tree-building method for 21 taxa. Neospora caninum and the Toxoplasma gondii strains split the genus Hammondia; Isospora felis is a sister taxon to the subfamily Sarcocystinae; Sarcocystis muris is a sister taxon to Sarcocystis neurona, Frenkelia glareoli, and Frenkelia microti; Sarcocystis zamani and Sarcocystis singaporensis are placed within the genus Sarcocystis. The branch lengths are proportional to the amount of inferred evolutionary change, and the numbers on the branches are the numbers of times that the branch was supported in 100 bootstrap replicates.
The genus Sarcocystis, too, demonstrates paraphyly of the genus Hammondia is supported. Tree-building analysis on the secondary-structure alignment of the full lsu rDNA sequences was more sensitive to the tree-building method used than was the CLUSTAL W alignment. Figure 1 shows the tree topology given by maximum likelihood using the secondary-structure alignment. It differs from the maximum likelihood tree based on the CLUSTAL W alignment in the placement of I. felis, H. heymdorni and N. caninum, S. muris, and S. zamani and S. singaporenisis. In figure 1, I. felis splits the subfamilies Sarcocystinae and Toxoplasmatinae. Neospora caninum splits the genus Hammondia, as in the tree from the CLUSTAL W alignment using neighbor joining. Sarcocystis muris does not form a monophyletic clade with the Frenkelia species and S. neurona, but splits the latter from the rest of the Sarcocystis species. Sarcocystis zamani and S. singaporenisis prove to be part of the genus Sarcocystis by branching after the Frenkelia species/S. neurona clade and S. muris. Neighbor joining, when applied to the secondary-structure alignment, gave the same tree given by this method using the CLUSTAL W alignment, which is also essentially the same as that given by the maximum-likelihood method and the CLUSTAL W alignment (fig. 2). Maximum parsimony, when applied to the secondary-structure alignment, resulted in eight equally parsimonious trees with a consensus showing polytomies within the Toxoplasmatinae, and in the placement of I. felis.

Informative Sites of the lsu rDNA
Phylogenetic Analysis of Helical and Nonhelical Regions

The results obtained by analysis of helical and nonhelical regions differed from each other in the inferred branching of S. muris, I. felis, and the Toxoplasmatine taxa. Variations in topology were found among tree-building algorithms and between the two alignments. Figure 3 summarizes the differences in topology between trees obtained from helical and nonhelical alignments using the three tree-building methods.

Informative Domains and Segments

The trees obtained using maximum parsimony, neighbor joining, and maximum likelihood on the structure alignment were analyzed under the branch reconstruction and character change options in MacClade, version 3.07. From this analysis, we observed the sites which were involved in the formation of each branch on a tree. All trees showed the same hot spots for nucleotide change under the character change options. Figure 4 shows part of a chart giving the amount of character

---

Fig. 2.—Phylogenetic relationships among the Sarcocystidae as inferred from the CLUSTAL W alignment of the full-length lsu rDNA. Hammondia heydorni and Neospora caninum are on sister branches; Isospora felis is a sister taxon to the subfamily Toxoplasmatinae; Sarcocystis muris is a part of a monophyletic group including Sarcocystis neurona, Frenkelia glareoli, and Frenkelia microti; the Sarcocystis species with reptilian definitive hosts (Sarcocystis zamani and Sarcocystis singaporenisis) form a sister clade to the Sarcocystis species with mammalian definitive hosts and the Frenkelia species.

N. caninum appear on sister branches. Besnoitia besnoitii is the immediate relative to this group, followed by I. felis. The genus Sarcocystis, too, demonstrates paraphyly as previously shown (Mugridge et al. 1999b). Frenkelia microti and F. glareoli demonstrate a close relationship with S. neurona and S. muris by forming the monophyletic group {S. muris [S. neurona (F. glareoli, F. microti)]}, which is a sister clade to the other Sarcocystis species with mammalian definitive hosts. Other branchings within the genus Sarcocystis include the monophyletic clades {S. cruzi [S. arieticanis (S. tenella, S. capracaenisis)]} and {S. miescheriana (S. moulei, S. gigantea)}. The Sarcocystis species with reptilian definitive hosts (S. zamani and S. singaporenisis) form a monophyletic clade which is a sister group to the rest of the genus Sarcocystis and the Frenkelia species.

The tree topology resulting from neighbor joining is like that in figure 2 except that N. caninum and H. heymdorni form a monophyletic clade rather than being sister branches; this results in the genus Hammondia being split by N. caninum in addition to the two T. gondii strains. Maximum parsimony gave four equally parsimonious trees where the only varying branches were those in the Toxoplasmatinae. This method found conflicting information within the clade [H. hammondii (T. gondii RH, T. gondii ME49)] and conflicting information regarding whether the nearest sister taxon to this clade is N. caninum or H. heymdorni; however, the paraphyly of the genus Hammondia is supported.
Fig. 3.—Summary of results obtained using (A) the helical regions and (B) the nonhelical regions of the structure alignment in conjunction with maximum parsimony, maximum likelihood, and neighbor joining. The only varying placements involve Sarcocystis muris, Isospora felis, and the Toxoplasmatinae. The tree diagrams are of sections of full trees featuring groupings of interest.

In total, 12 phylogenetic analyses were performed (three cladistic methods for each of four sequence subsets). The D2 domain gave trees supporting the major groupings found for the full-length analyses using maximum parsimony and neighbor joining, with the exception that S. cruzi was placed as a sister taxon to S. tenella instead of S. capracanis (fig. 5). The result from maximum likelihood based on the D2 region shows this and also changes the placement of S. zamani and S. singaporensis to form a clade with S. miescheriana, S. moulei, and S. gigantea (fig. 6). With respect to the Toxoplasmatinae, the D2 segment does not deviate from results given by the full-length analyses. The three trees from the D2 region all place I. felis as a sister taxon to the subfamily Sarcocystinae.
Adding D3 to the D2 domain did not change the above results for any of the tree-building methods, except that neighbor joining placed S. cruzi and S. caprarae as monophyletic sisters, and maximum likelihood did not place S. zamani and S. sangaporensis in a clade with S. mouleli, S. miescheriana, and S. gigantea.

The other two segments from bases 1426–1904 and 2047–2720 gave trees that should not be counted as phylogenetically informative for the Sarcocystidae because of groupings that intersperse the genus Sarcocystis with members of the Toxoplasmatinae.

Results from the truncation of the full-length lsu rDNA alignment into two alignments containing either the 5′ or the 3′ end gave results more in accordance with those from the full-length alignment. The maximum-likelihood trees from the 5′ and 3′ ends, respectively, are shown in figures 7 and 8. Maximum parsimony and neighbor joining also gave results in congruence with the full-length results, except for the neighbor-joining tree of the 3′ end, which split the T. gondii strains with H. hammondi.

Discussion

Informative Regions of the lsu rDNA
Secondary-Structure Regions

Our investigation of the use of partial lsu rDNA sequences to study the phylogeny of the Sarcocystidae revealed that the entire length of the gene is needed to obtain full phylogenetic informativeness. In order to analyze the influence of different structural segments of the alignment on the final phylogeny, all trees were compared based on tree topology only. Previous work on the phylogeny of the Sarcocystidae using ssu rDNA proposed that helical regions, which correspond to double helices, are more informative than the nonhelical areas (Morrison and Ellis 1997). Nonhelical areas, which are single-stranded regions forming globular domains, have a higher rate of change than helical areas (Olsen and Woese 1993). Therefore, changes in the more conserved helical regions that make up the backbone of the molecule could be construed to be more significant in the evolution of these taxa than changes in the faster-evolving nonhelical areas. However, these conserved areas, although useful in the alignment of sequences, may sometimes contain little phylogenetic information for closely related taxa. Our investigations into the amount...
Fig. 5.—Phylogenetic relationships among the Sarcocystidae as inferred from maximum-parsimony and neighbor-joining analyses of the structure alignment of the D2 domain. Deviation from full-length corresponding analyses is included.

Fig. 6.—Phylogenetic relationships among the Sarcocystidae as inferred from maximum-likelihood analysis of the structure alignment of the D2 domain.

of nucleotide substitutions throughout the 18S rDNA molecule for the Sarcocystidae support the idea that globular domains can have more changes, as evidenced by the fact that the areas of highest variation observed by us were present in the globular domains. As seen in figure 4, the D2 domain is the area with the most change. The D2 region of the 18S rDNA is only 305 bp long in the *T. gondii* RH strain but has nevertheless been used to infer phylogenies on a small number of taxa belonging to the Sarcocystidae (Ellis et al. 1999). One of the aims of this work was to determine whether this site alone could be informative enough for the Sarcocystidae. It was observed that many of the changes involved in the formation of tree branches are indeed part of the D2 domain but that informative sites involved in resolving branches for very similar species or strains, such as *H. hammondii* and the *T. gondii* RH and ME49 strains, also come from other areas of the molecule, including the 3’ end. For example, there are only three nucleotide variations supporting the branch grouping of *H. hammondii, T. gondii* RH, and *T. gondii* ME49, and only one of these is from the D2 domain. Therefore, it is evident that in such cases the entire 18S rDNA sequence is more informative. Furthermore, from figures 5 and 6, it can be seen that the D2 domain gave trees which departed from two major groupings given by the full-length analyses by placing *S. cruzi* as a sister taxon to *S. tenella* instead of *S. capracanis* (fig. 5) and placing *S. zamani* and *S. singaporensis* with the *S. miescheriana, S. moulei, and S. gigantea* clade (fig. 6). The considerable number of site changes not within the D2 domain supporting the various groupings within the genus Sarcocystis suggests that the entire length of the gene, as opposed to the D2 domain alone, gives a more robust result. The fact that the addition of the D3 domain to the D2 domain did not add to the informativeness, and that the other truncated regions of change gave questionable results, also supports this view.

**Helical and Nonhelical Regions**

Our results from the use of helical and nonhelical parts of the secondary alignment for tree building also demonstrated that subdivisions of the full-length gene could have an effect on tree topology. In figure 3, it can be seen that the placement of *S. muris* remained constant for the two alignment subdivisions for the maximum-parsimony and neighbor-joining tree-building methods but that maximum likelihood had a different placement
for this species depending on the alignment. The branching of *I. felis* did not stay constant for the different tree-building methods for each alignment, and furthermore, subdividing the full sequence into helical and nonhelical sections significantly reduced the resolution of the Toxoplasmatinae (fig. 3).

Previously, we reported that the nonhelical regions of the *lsu rDNA* are as informative as the helical areas (Mugridge et al. 1999b). This remains the case in this study, as neither the helical nor the nonhelical alignments appear more informative than the other in conjunction with any of the tree-building algorithms. However, we note, not unexpectedly, that adding to the number of species (only nine were used in the previous study) decreases the congruence of topologies obtained from the different tree-building methods. For this work, we added more Sarcocystis species and all current genera of the Toxoplasmatinae; however, the incongruities remain with the placement of *S. muris*, as well as *I. felis*. The paraphyly of *T. gondii* shown in the nonhelical alignment is a clear result of the removal of informative sites for the Toxoplasmatinae, which comprise only 4 nt.

Random Choice of Nucleotide Sites

Dividing the full-length secondary alignment in half to produce 5' and 3' alignments also proved to be unsatisfactory, as the 3' alignment split the *T. gondii* strains using the maximum-likelihood and neighbor-joining trees. Minor branch swappings diverging from the full-length alignment results were also present for these two alignments within the genus Sarcocystis. This result also demonstrates that neither the 5' nor the 3' end of the gene is more informative than the other in the final analysis.

Alignments

Recent work on the effect of sequence alignment on the phylogeny of the genus Sarcocystis showed that the resulting trees can be influenced by the choice of a primary- or secondary-structure alignment (Ellis and Morrison 1995). This work addressed the issue by investigating the phylogeny of the family Sarcocystidae using two alignment strategies and three tree-building methods in order to observe the effect this had on the various groupings in the family. The DCSE alignment of De Rijk and De Wachter (1993) was chosen for its
consideration of the secondary structure of the ribosomal RNA molecule. Since homology between nucleotides of the aligned sequences is imperative for a true phylogenetic reconstruction, we believe that adhering to the biological constraints of the molecule during the alignment process is more likely to achieve this aim. The CLUSTAL W alignment algorithm (Thompson, Higgins, and Gibson 1994) aims at achieving the highest nucleotide identity among sequences; hence, in some parts of the alignment, nucleotides from different sequences may be positionally aligned because of their similarity rather than their true homology. Any nonhomologous alignment will lead to an erroneous phylogeny.

From the results obtained using the two full-length alignments with the three tree-building programs, we found that all of the trees derived from the CLUSTAL W alignment had identical topologies except for variations in the placement of *N. caninum* as a sister taxon to the genus Hammondia or as a sister taxon to *H. heydorni*, thereby splitting the genus Hammondia. The trees resulting from the application of the structure alignment gave inconsistencies among tree-building methods. Again, the three differing groups involved *S. muris*, *I. felis*, and the Toxoplasmataceae. In trees derived by the maximum-parsimony and maximum-likelihood methods, which are both character-based methods, *S. muris* splits a clade consisting of *S. neurona* and the Frenkelia species from the rest of the Sarcocystis species. In contrast, neighbor joining, which deals with distance data, placed *S. muris* in a clade with *S. neurona* and the two Frenkelia species.

The branching of *I. felis* differed for the three tree-building methods, with maximum parsimony and neighbor joining placing it as the sister taxon to all species after *E. tenella*, while maximum likelihood placed it as the sister taxon to the subfamily Sarcocystinae. With regard to this last group, the three methods gave the same topology, except that in the case of maximum parsimony, the clade involving the *T. gondii* ME49 and RH strains and *H. hammondii* could not be resolved.

The question that arises is whether the result obtained from the CLUSTAL W alignment is the true evolutionary topology of the group under analysis. The congruence in results within the CLUSTAL W analyses could be linked to the use of the LSU rDNA molecule. This molecule is about double the length of the SSU rDNA, and this increase in the available number of nucleotides for analysis may render any differences between alignment strategies in this case irrelevant. The structure alignment contained many more gaps and areas in which the alignment was not intuitively obvious, but in keeping to secondary structure, this should in theory be a more informative alignment, and the few differences found among the tree-building methods could be the effect of these methods having different goals in their tree-building processes. Areas in the alignment that contain many gaps could be removed from the alignment with the aim of reducing inconsistencies among trees. With respect to the LSU rDNA structure alignment, most of the areas that have a larger amount of gaps correspond to the nonhelical regions of the molecule. The results here pertaining to the helical segments of the alignment (fig. 3) demonstrate that removing the regions with more gaps does not aid the congruency of results among different tree-building methods. Furthermore, excluding such regions in this case means the removal of areas that contain informative nucleotide sites, and this is particularly important for the genera Hammondia, Neospora, and Toxoplasma, which have a very small number of informative sites in the LSU rDNA.

Using the three tree-building methods, however, did show that the majority of the groups within the inferred phylogeny are very robust. For example, the major groupings within the subfamily Sarcocystinae remained the same in all of the full-length trees. The placement of *S. muris* was dependent on the tree-building method as to whether it branched in a clade with *S. neurona* and the two Frenkelia species or split the latter from the rest of the Sarcocystis species. The other variation is due to the different alignments and concerns *S. singaporensis* and *S. zamani*, which branch before the other Sarcocystis species in trees from the CLUSTAL W alignment but branch after the Frenkelia species/S. neurona clade and *S. muris* in trees derived using the structure alignment.

Congruence among trees with respect to the Toxoplasmatidae is also high, with the only exception being that maximum parsimony provides eight equally parsimonious trees because it cannot resolve which are closest among *T. gondii* RH, *T. gondii* ME49, and *H. hammondii*. The other methods agree that the *T. gondii* strains are closest, and in the case of parsimony, it is probably just a case of the LSU rDNA not having sufficient characters to resolve this particular group using this method. All of the full-length results, however, support our previous findings that *N. caninum* and *H. heydorni* are closely related, thus splitting the genus Hammondia. It can be seen in figure 1 that the local rearrangements in this group are due to the short branch lengths between the genera Hammondia, Neospora, and Toxoplasma, with even a branch length of 0 between *H. hammondii* and the *T. gondii* strains.

Applying different alignment and tree-building methods provides an opportunity to observe how the data behave under different assumptions and aims. Literature on phylogeny algorithms remains focused on the advantages and pitfalls of various methods, despite the fact that as a taxonomic and evolutionary tool, phylogeny is becoming increasingly popular. Different methods may or may not extract varying information from a gene, and reducing the number of data (the number of taxa or the length of a sequence) may add to the variation, as evidenced by our results using partial sequences. We have shown that using the entire LSU rDNA gives a more informative topology of the Sarcocystidae, and the use of different tree-building methods and even two alignment algorithms has revealed that most of the groupings within this family are robust.

**Relationships Within the Sarcocystidae**

The findings from this work support our previous results regarding the genus Frenkelia and the subfamily
Toxoplasmata (Mugridge et al. 1999a, 1999b). All the trees in this study using lsu rDNA sequences confirm that the current genus Sarcocystis is split by the genus Frenkelia. This has also been found in a study using ssu rDNA sequences (Votýpka et al. 1998), and it has already been suggested that Frenkelia should be synonymized with Sarcocystis based on phenotypic characters (Tadros and Laarmann 1976; Odening 1998; Mugridge et al. 1999b), which would render the current subfamily Sarcocystinae monogenic. New findings obtained here show that the clades [S. miescheriana (S. moulei, S. gigantea), {S. cruzi [S. arieticanis (S. tenella, S. capracanis)]}, [S. neurona (F. glareoli, F. microti)], and (S. zamani, S. singaporensis), within the subfamily Sarcocystinae, appear to be very robust when exposed to varying alignment methods and parameters and different tree-building methods. The Sarcocystis species with reptilian definitive hosts (S. zamani and S. singaporensis) are shown to form a lineage that is separated from the other Sarcocystis species, which use mammals as definitive hosts, and from the Frenkelia species, which use birds as definitive hosts. In contrast, as found in analyses using ssu rDNA sequences (Jeffries et al. 1997; Votýpka et al. 1998), the placement of S. muris also remained the main variant within the Sarcocystinae based on the lsu rDNA data.

The current study also confirmed previous analyses based on lsu rDNA and ITS1 sequences (Ellis et al. 1999; Mugridge et al. 1999a), which showed the genus Hammondia to be paraphyletic and showed species placed in the genera Toxoplasma, Hammondia, and Neospora to be very closely related. As described previously, this finding, together with the broad range of phenotypic characters that are shared by these taxa, calls for a revision of these taxa within the subfamily Toxoplasmatinae (Mugridge et al. 1999a). In contrast, the branching of B. besnoiti as the nearest sister to these taxa is a robust result in this study.

Isospora felis (syn. C. felis), which has the ability to produce monozoic tissue cysts in optional intermediate hosts, is sometimes placed in the genus Cystoisospora, which is then placed in the monogenic subfamily Cystoisopora of the Sarcocystidae (Frenkel et al. 1979; Frenkel, Mehllhorn, and Heydorn 1987). However, it is most frequently placed in the genus Isospora, which is a member of the family Eimeriidae and contains several coccidia with differences in life cycle and morphology (Levine 1988; Current, Upton, and Long 1990; Tenter and Johnson 1997). In our present work, all of the full-length sequence results from the CLUSTAL W alignment, and from the neighbor-joining and maximum-parsimony methods using the structure alignment, place I. felis as the sister taxon to a monophyletic group comprising the subfamilies Toxoplasmatinae and Sarcocystinae (fig. 2), while maximum likelihood in conjunction with the structure alignment suggested that I. felis splits the Sarcocystinae from the Toxoplasmatinae (fig. 1). A recent phylogenetic analysis using ssu rDNA sequences of several Isospora species showed the genus Isospora to be paraphyletic and identified different lineages within this genus which are in accordance with morphologic similarities among their life cycle stages (Carreno and Barta 1999). That study suggests that a monophyletic group of I. felis and homoxenous Isospora species of mammals is a sister taxon to the Toxoplasmatinae, whereas Isospora species of birds are monophyletic with species of Eimeria. Further studies involving lsu rDNA sequences of more species of Isospora will be needed to validate the different lineages within the family Sarcocystidae and its subdivision into subfamilies and genera.

**Supplementary Material**

Nucleotide sequence data reported in this paper are available in the GenBank, EMBL, and DDBJ databases under accession numbers AF044250 (S. arieticanis), AF076899 (S. tenella), AF076902 (S. miescheriana), AF076903 (S. cruzi), AF092927 (S. neurona), AF012883 (S. muris), AF012884 (S. moulei), AF012885 (S. capracanis), U85706 (S. gigantea), AF076900 (B. besnoiti), AF044251 (F. glareoli), AF044252 (F. microti), AF076901 (T. gondii ME49), AF101077 (H. hammondii), AF159240 (H. haydorni), U85705 (I. felis), AF237616 (S. zamani), and AF237617 (S. singaporensis).

**LITERATURE CITED**


Felsenstein, J. 1995. PHYLIP (phylogenetic inference package). Distributed by the author, Department of Genetics, University of Washington, Seattle.


William Taylor, reviewing editor

Accepted August 9, 2000