

Evolution of Ruminant *Sarcocystis* (Sporozoa) Parasites Based on Small Subunit rDNA Sequences

O. J. M. Holmdahl,^{*,1} David A. Morrison,^{*} John T. Ellis,^{*} and Lam T. T. Huong[†]

^{*}Molecular Parasitology Unit, University of Technology, Sydney Westbourne Street, Gore Hill, New South Wales 2065, Australia; and [†]Department of Parasitology and Pathology, University of Agriculture and Forestry, Thu Duc, Ho Chi Minh City, Vietnam

Received August 18, 1997; revised May 19, 1998

We present an evolutionary analysis of 13 species of *Sarcocystis*, including 4 newly sequenced species with ruminants as their intermediate host, based on complete small subunit rDNA sequences. Those species with ruminants as their intermediate host form a well-supported clade, and there are at least two major clades within this group, one containing those species forming microcysts and with dogs as their definitive host and the other containing those species forming macrocysts and with cats as their definitive host. Those species with nonruminants as their intermediate host form the paraphyletic sister group to these clades. Most of the species have considerable genotypic differences (differing in more than 100 nucleotide positions), except for *S. buffalonis* and *S. hirsuta*. There is a large suite of genotypic differences indicating that those species infecting ruminant and nonruminant hosts have had very different evolutionary histories, and similarly for the felid- and canid-infecting species. Furthermore, the rDNA sequences that represent the different structural regions of the rRNA molecule have very different genotypic behavior within *Sarcocystis*. The evolution of these regions should be functionally constrained, and their differences can be explained in terms of the importance of the nucleotide sequences to their functions. © 1999 Academic Press

INTRODUCTION

Sarcocystis is a large genus of intracellular, cyst-forming coccidian parasites belonging to the protozoan (protistan) phylum Sporozoa (Apicomplexa) (Dubey *et al.*, 1989). The organisms have an obligatory two-host life cycle, involving a sexual generation in enteroepithelial cells of the definitive host, usually a carnivore, and asexual development in the tissues of the intermediate host, usually a herbivore. After oral up-

take of infective sporocysts shed by the definitive host in feces, the parasite will proliferate in the tissues of the intermediate host and finally establish itself as sarcocysts (sarcosporidia). The sarcocysts, containing cystozoites or bradyzoites, are found in muscle and nervous tissue in the intermediate host, which is then consumed by the definitive host.

Sarcocystis may be of considerable economic importance, because domesticated ruminants will act as an intermediate host for a wide range of species. For example, there are three species of *Sarcocystis* that infect cattle (*Bos taurus*), namely *S. cruzi*, *S. hirsuta*, and *S. hominis*, four species that infect sheep (*Ovis aries*), i.e., *S. arieticanis*, *S. gigantea*, *S. medusifformis*, and *S. tenella*, and three that infect goats (*Capra hircus*), i.e., *S. capracanis*, *S. hircicanis*, and *S. moulei*. Of the other domesticated ruminants, water buffalo (*Bubalus bubalis*) have three species, namely *S. fusiformis*, *S. leventei*, and a newly described species, *S. buffalonis* (Huong *et al.*, 1997), and camelids (camels, alpacas, llamas, etc.) and reindeer (*Rangifer tarandus tarandus*) also act as intermediate hosts for several species of *Sarcocystis*.

Two of the most pathogenic species of *Sarcocystis* are *S. cruzi* and *S. tenella*. These species form microscopic cysts in their intermediate host, and both have canines as the definitive host. *S. cruzi* and *S. tenella* can cause a variety of clinical symptoms in their respective intermediate hosts, from death and abortions to anemia, retarded growth, and reduced milk production. Species of *Sarcocystis* that form macroscopically visible cysts, like *S. fusiformis*, *S. buffalonis*, and *S. gigantea*, also may render meat unesthetic and thus lead to condemnation of infected carcasses (Dubey *et al.*, 1989; Ugglå and Buxton, 1990).

The identification and classification of *Sarcocystis* species has traditionally been based mainly on host range and the ultrastructural morphology of the cysts. The cyst-forming coccidia are usually placed in a single family, the Sarcocystidae, with two subfamilies, the Toxoplasmatinae and the Sarcocystinae (Levine, 1985). However, the phylogenetic relationships within the

¹ On leave from Department of Parasitology, National Veterinary Institute and Swedish University of Agricultural Sciences, Box 7073, S-750 07 Uppsala, Sweden.

genus *Sarcocystis* remain ambiguous, at least partly because of difficulties in assessing the homology of many of the phenotypic characters. It is possible that molecular data might alleviate these problems, and over the past decade several attempts have been made to infer the phylogeny from genotypical characters, predominantly small-subunit ribosomal DNA (ssu rDNA) sequence data (Ellis *et al.*, 1998). Unfortunately, there has been little consistency among the results of the molecular studies to date; for example, several analyses have questioned the monophyly of *Sarcocystis* (Barta *et al.*, 1991; Tenter *et al.*, 1992; Ellis *et al.*, 1994) while others do not (Ellis and Morrison, 1995; Morrison, 1996). Nevertheless, these studies suggest a strong correlation between the parasite phylogeny and that of their definitive hosts, with the species of the Sarcocystidae forming two clades based on the use of either canids or felids as their definitive hosts.

Here we present a phylogenetic analysis of 13 species of *Sarcocystis*, including 4 newly sequenced species with ruminants as their intermediate host (*S. aucheniae*, *S. buffalonis*, *S. cruzi*, and *S. hirsuta*), based on complete ssu rDNA sequences. We assess whether the groups based on their host-specificity are supported by the genotypic data, and the extent to which the apparent phenotypic relationships are reflected in the genotypes, with particular reference to the relationship of *S. buffalonis* and *S. hirsuta*, which are phenotypically very similar and have the same definitive host. We also consider the evolutionary patterns of character-state changes in the different structural regions of the ssu rRNA molecule.

MATERIALS AND METHODS

Parasite strains. Macroscopic sarcocysts of *S. buffalonis* and *S. hirsuta* were collected from naturally infected animals in an abattoir in Vietnam. The cysts were excised from skeletal muscles of carcasses from water buffalo (*S. buffalonis*) and cattle (*S. hirsuta*). Macrocyts of *S. aucheniae* were collected from naturally infected alpacas in Victoria, Australia. Cysts adjoining the picked ones (with similar morphological features) for these three species have been collected and subjected to ultrastructural studies for confirmation of species identification. Microcyts of *S. cruzi* were collected and further purified as described previously (Holmdahl *et al.*, 1993).

DNA extraction, PCR, and sequencing. Genomic DNAs from the four *Sarcocystis* species were extracted as previously described (Gajadhar *et al.*, 1991) and analyzed by agarose gel electrophoresis. In order to amplify the complete ssu rRNA gene, two overlapping fragments of this gene were amplified by polymerase chain reaction (PCR) with two pairs of primers complementary to conserved regions of the ssu rRNA; two fragments were needed to obtain reliably the complete

ssu rDNA gene sequence. The primers were AM and BM (Medlin *et al.*, 1988) and R18S and F18S (Holmdahl *et al.*, 1991). Both PCR mixtures (100 μ l each) contained 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM MgCl₂, approximately 0.5 μ g DNA, 1 μ M of each primer, and 200 μ M of each deoxynucleotide. The mixture was heated to 94°C for 5 min, and thereafter 0.5 units of *Taq* DNA polymerase (Bresatech, Australia) was added and the samples were amplified for 30 cycles. Each cycle consisted of 2 min of denaturation at 94°C, 2 min at the annealing temperature of 50°C, and 2 min of extension at 72°C. The final extension step was continued for an additional 4 min at 72°C. Each species and fragment were amplified three times, and the purified amplicons were pooled together. Amplicons were purified by a Qiagen PCR Purification Kit (Qiagen, USA) and run on a 1% agarose gel with a λ DNA/*Hind*III marker to estimate concentration and purity. No heterogeneity in the rRNA coding regions was detected (e.g., multiple bands).

The amplicons were sequenced by cycle sequencing and run on an ABI automated DNA sequencer (Supamac, Sydney, Australia). We used eight primers complementary to conserved regions of the ssu rDNA, which sequenced both strands for each species at least two times, to ensure accuracy in the sequence data. The primers were AM and BM (Medlin *et al.*, 1988), R18S and F18S (Holmdahl *et al.*, 1991), A, 8, and 19 (Sogin and Gunderson, 1987), and 2: AGGGTTCGATTCGGAGA. The sequence data have been deposited with GenBank (Table 1).

Data set. The data set used for the evolutionary analyses consisted of the complete ssu rDNA sequences of all Sarcocystidae lodged with GenBank (29 sequences covering 15 taxa; Table 1). Recent morphological and molecular data suggest that the subfamilies Toxoplasmatinae and Sarcocystinae are monophyletic (Morrison, 1996; Ellis *et al.*, 1997), and so the sister group to *Sarcocystis* is *Toxoplasma* + *Neospora*. Thus, these latter two species were used as a monophyletic outgroup to root the cladograms (see Smith, 1994).

Sequence alignment. The sequence alignment used was based on that described by Van de Peer *et al.* (1997), which defines the complete secondary structure of the ssu rRNA molecule. The alignment process is iterative, beginning with the juxtaposition of regions of extensive primary structural similarity and then refinement by invoking higher-order structural constraints (Gutell *et al.*, 1994); higher-order structures are inferred by comparative analysis, relying on the search for compensatory base substitutions or positional covariance (Gutell, 1996). This original alignment is available from the SSU rRNA Database (available at <http://www-rrna.uia.ac.be/>), maintained by Y. Van de Peer, P. De Rijk, and R. De Wachter (Departement Biochemie, Universiteit Antwerpen).

TABLE 1
The Species Used in the Phylogenetic Analyses

Species	GenBank Accession Nos. for small subunit rDNA sequences	Definitive host ^a	Sporocyst size ^a	Intermediate host ^a	Sarcocyst size ^a	Sarcocyst wall type ^a
<i>Neospora caninum</i>	L24380, U03069, U16159, U17346	Unknown	Unknown	Many	Micro	Thick smooth
<i>Sarcocystis arieticanis</i>	L24382	Dog	Large	Sheep	Micro	7
<i>Sarcocystis aucheniae</i>	AF017123	Dog	Large	Alpaca	Macro	20
<i>Sarcocystis buffalonis</i>	AF017121	Cat	Small	Water buffalo	Macro	10
<i>Sarcocystis capracanis</i>	L76472	Dog	Small + large	Goat	Micro	14
<i>Sarcocystis cruzi</i>	AF017120	Dog, fox, coyote	Large	Cattle	Micro	7
<i>Sarcocystis fusiformis</i>	U03071	Cat	Small	Water buffalo	Macro	20
<i>Sarcocystis gigantea</i>	L24384	Cat	Small	Sheep	Macro	20
<i>Sarcocystis hirsuta</i>	AF017122	Cat	Small	Cattle	Macro	10
<i>Sarcocystis moulei</i>	L76473	Cat	Small	Goat	Macro	20
<i>Sarcocystis muris</i>	M64244	Cat	Small	Mouse	Macro	1
<i>Sarcocystis neurona</i>	U07812, U33148 + U33149	Opossum	Small	Bird	Macro	10
<i>Sarcocystis tenella</i>	L24383	Dog, fox, coyote	Large	Sheep	Micro	14
<i>Sarcocystis species</i>	U97524	Snake	Unknown	Mouse	Macro	1
<i>Toxoplasma gondii</i>	M97703, X68523, X75453, X75429, X75430, U00458, U03070, L24381, X65508, U12138, L37415	Cat	Small	Many	Micro	1

Note. Newly sequenced taxa are in boldface type.

^a Data from Dubey *et al.* (1989), except for *S. buffalonis* and *S. hirsuta*, which are from Huong *et al.* (1997).

This alignment was then manually modified to eliminate minor inconsistencies between the different taxa. The final alignment is available at <http://www.science.uts.edu.au/~davidm/alignments.html>. The aligned sequence length was 2008 nucleotides. After alignment, consensus sequences were derived for those species for which there were several sequences available (11 for *T. gondii*, 4 for *N. caninum*, and 2 for *S. neurona*), using the MacClade 3.06 computer program (Maddison and Maddison, 1992). The standard IUPAC ambiguity codes were used for those few nucleotide positions with more than one possible character-state in the consensus sequence.

For the data analyses, all of the currently recognized helical regions within the ssu rRNA were identified from the alignment according to secondary structure (Van de Peer *et al.*, 1997). A separate data file was then created containing all of the helical (whether double- or single-stranded) positions (63.6% of the aligned sequence length) using the DCSE sequence editor (De Rijk and De Wachter, 1993). Following Ellis and Morrison (1995) and Morrison and Ellis (1997), who demonstrate that for the ssu rRNA of the Sporozoa most of the phylogenetically informative positions are in the helical regions, this data file was used to construct the phylogenetic trees (see Results for a further discussion of this point).

Cladistic analyses. Three tree-building methods were chosen as representative of the range of those available (Morrison, 1996), to examine whether the

placement of some of the taxa is sensitive to the cladistic analysis used. Neighbor-joining analyses were performed using the Phylip 3.57 package (Felsenstein, 1995), with distances calculated using the Kimura two-parameter model. Parsimony analyses were performed using the PAUP 3.1.1 program (Swofford, 1993), using the branch-and-bound algorithm. Maximum-likelihood analyses were performed using the Phylip 3.57 package (Felsenstein, 1995), using a transition:transversion ratio of 2:1, empirical base frequencies, one rate class for nucleotide substitutions across sites, and global branch rearrangements.

To assess the magnitude of the phylogenetic signal in the data set, we used relative apparent synapomorphy analysis (Lyons-Weiler *et al.*, 1996), as implemented in the RASA 2.1 computer program (Lyons-Weiler, 1997). The alternative tree topologies produced by the different cladistic analyses were evaluated by the Kishino-Hasegawa-Templeton test under the maximum-likelihood model (Kishino and Hasegawa, 1989), using the Phylip package. The robustness of each of the monophyletic groups represented on the final cladogram was assessed using bootstrapping with 200 replicates (Felsenstein, 1985).

The assessment of the distribution of nucleotide character-states and character-state changes on the phylogenetic tree was done using the MacClade 3.06 program. This reconstructs the character-states for each inferred ancestor on the cladogram (Swofford and Maddison, 1987); where there is more than one possible

most-parsimonious reconstruction, the numbers were averaged over all possibilities. Following Vawter and Brown (1993), the rDNA sequences were subdivided into functional classes based on the structure of the rRNA molecule, with three separate classes for helical bulges + loops (single-stranded regions), helical stems (double-stranded regions), and nonhelical regions. Analyses were performed separately for each of these classes; and separate analyses were also performed for some of the different clades on the phylogenetic tree. The frequencies of the nucleotide character-states and character-state changes were compared using log-likelihood ratio contingency χ^2 tests (Wilkinson, 1991); note that the data for these statistical tests will not be completely independent because they are constrained by the phylogeny, and thus the results of the tests need to be interpreted cautiously. The examination of the pattern of phenotypic character-states was done using the MacClade 3.06 program.

RESULTS

Phylogeny. The relative apparent synapomorphy analysis (RASA), which measures the natural cladistic hierarchy inherent in a data matrix, indicated that there is a statistically significant amount of phylogenetic signal in the total ssu rDNA sequence data set ($t = 7.28$ for the unrooted RASA; $P < 0.001$). However, the helical regions ($t = 8.14$) contain much more of this signal than do the nonhelical regions ($t = 5.41$), at least partly because most of the indels are in the nonhelical regions. In particular, the nonhelical regions contain relatively little phylogenetic signal about the relationship to the outgroup ($t = 2.88$ for the rooted RASA analysis is much reduced compared to $t = 5.76$ for the unrooted analysis of the ingroup). Thus, although there is considerable sequence variability in the nonhelical regions (see below), most of this variability is not phylogenetically informative. This conclusion confirms the results of Ellis and Morrison (1995) and Morrison and Ellis (1997), who used a different approach to the study of this information; and so the data from helical regions are the most suitable for reconstructing the evolutionary history of the *Sarcocystis* species.

The neighbor-joining tree (Fig. 1) shows that the *Sarcocystis* species form two main monophyletic groups, with *S. aucheniae* as their sister species. These relationships are strongly supported by the data, as indicated by the bootstrap analysis. However, the placement of *S. fusiformis* within its clade is not robust, nor are the relative positions of *S. arieticanis* and *S. cruzi* within their clade. The bootstrap analysis also well supports *S. muris* and *S. neurona* as the sister species to the rest of *Sarcocystis*, but there is little support for them as sister species to each other.

The optimal tree from the maximum-likelihood analysis differs from the neighbor-joining tree in that: (1) *S.*

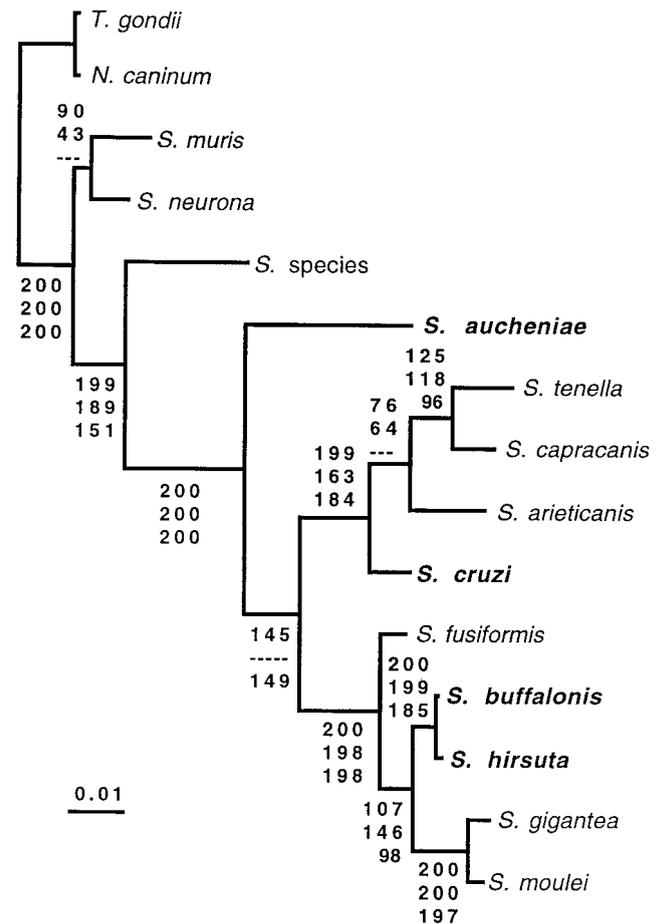


FIG. 1. Phylogenetic relationships among the *Sarcocystis* species as inferred from the helical regions of the ssu rDNA using the neighbor-joining tree-building method. Species names are as in Table 1 (newly sequenced taxa are in boldface type). The branch lengths are proportional to the amount of inferred evolutionary change, as shown by the scale bar. The numbers on the branches are the number of times that the branch was supported in 200 bootstrap replicates, based on (from top to bottom) the neighbor-joining, parsimony, and maximum-likelihood analyses; some branches do not have numbers for particular analyses because that branch does not appear in the tree produced from that analysis.

muris and *S. neurona* do not form a monophyletic group, but are paraphyletic sisters to the rest of the *Sarcocystis* species with *S. muris* as the basal species; and (2) *S. cruzi* is the sister to the *S. tenella* + *S. capracanis* clade. The parsimony analysis produced two equally optimal trees. The first of these differs from the neighbor-joining tree in that *S. aucheniae* is the sister to the *S. tenella* + *S. capracanis* + *S. arieticanis* + *S. cruzi* clade. The second parsimony tree is identical to the first tree, except that *S. muris* and *S. neurona* do not form a monophyletic group, but are paraphyletic sisters to the rest of the *Sarcocystis* species with *S. muris* as the basal species. The Kishino-Hasegawa-Templeton tests indicate that these three alternative tree topologies are not statistically significantly differ-

ent from the neighbor-joining tree ($z = 0.22, 0.08, 0.40$, respectively; $P > 0.050$ in all cases). Furthermore, the bootstrap analyses of the trees are approximately the same for the branches held in common (Fig. 1). Consequently, the neighbor-joining tree was used for the comparison with the phenotypic characters and for the assessment of character-state evolution.

The phenotypic characteristics correlate well with the genotypic phylogeny (Table 1, Fig. 2). Those species with a ruminant as the intermediate host form a monophyletic group, as do those species with dogs or cats as the definitive host. Those species with small sporocysts form a clade, as do those with microscopic sarcocysts. Finally, some of the different sarcocyst wall types form smaller clades nested within the clades formed by the sarcocyst sizes.

Character and character-state evolution. The *Sarcocystis* species varied at 414 (20.6%) of the 2008 aligned nucleotide positions. Most of the species differed unambiguously from one another at more than 100 of these positions, with the notable exceptions of *S. buffalonis* and *S. hirsuta* (13 positional differences), *S. gigantea*

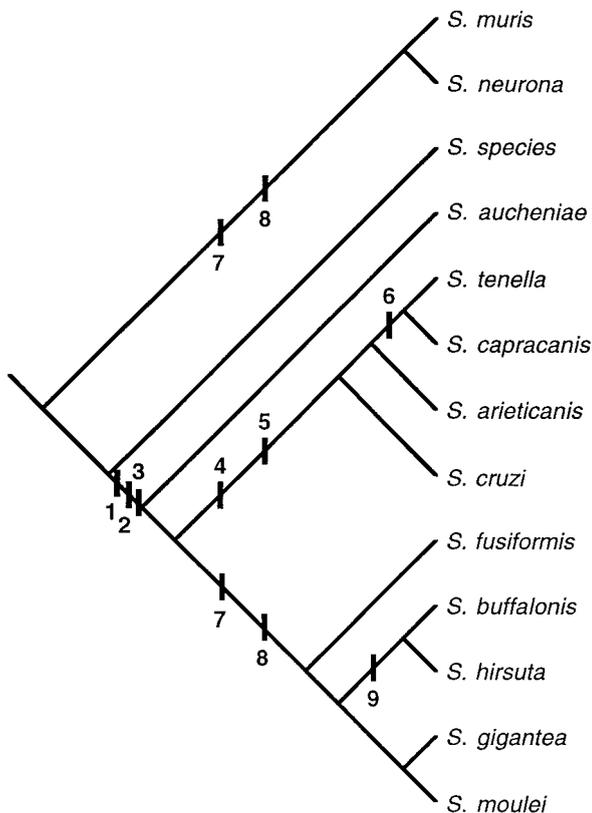


FIG. 2. Distribution of phenotypic characteristics among the *Sarcocystis* species. Species names and characters are as in Table 1. 1, intermediate host a ruminant; 2, definitive host a dog; 3, sarcocyst wall type 20; 4, sarcocyst wall type 7; 5, microscopic sarcocyst; 6, sarcocyst wall type 14; 7, definitive host a cat; 8, sporocyst small; 9, sarcocyst wall type 10.

TABLE 2
Percentage Frequency of Nucleotides in *Sarcocystis* and Its Subsets

Nucleotide	Region of rRNA				
	Overall	Helical- double	Helical- single	Nonhelical	
<i>Sarcocystis</i>					
A	27.6	20.3	34.9	35.5	
C	18.4	22.0	14.3	14.9	
G	25.1	29.7	21.3	19.9	
T	28.8	28.1	29.5	29.7	
GC	43.5	51.7	35.7	34.8	
Species with a ruminant as intermediate host					
A	28.0	20.6	35.2	35.8	
C	18.1	21.8	13.9	14.5	
G	24.9	29.3	21.3	19.8	
T	29.0	28.3	29.6	29.8	
GC	43.0	51.1	35.2	34.3	
Species with a nonruminant as intermediate host					
A	26.4	19.2	33.6	34.4	
C	19.5	22.7	15.8	16.1	
G	25.9	30.8	21.4	20.1	
T	28.2	27.2	29.2	29.4	
GC	45.4	53.6	37.2	36.2	
Species with a dog as definitive host ^a					
A	27.7	20.1	34.9	36.0	
C	18.5	22.0	14.5	15.0	
G	25.3	30.0	21.2	20.0	
T	28.5	27.9	29.3	29.0	
GC	43.8	52.0	35.7	35.1	
Species with a cat as definitive host					
A	28.2	20.9	35.5	35.6	
C	17.8	21.5	13.6	14.2	
G	24.6	28.9	21.4	19.7	
T	29.4	28.7	29.5	30.4	
GC	42.4	50.4	35.0	34.0	
Total number of nucleotide positions		2008	971	307	730

^a With the exception of *S. aucheniae*, which is not part of that clade.

and *S. moulei* (47 differences), and *S. muris* and *S. neurona* (49 differences).

Within *Sarcocystis*, the nucleotide frequencies are very different in the double-stranded regions compared to the single-stranded regions ($G = 404.69$; $P < 0.001$) and the nonhelical regions ($G = 737.61$; $P < 0.001$), with fewer A and more G and C (Table 2). Furthermore, the nucleotide frequencies were different in those species with ruminants as their intermediate host compared to those with nonruminants as their intermediate host ($G = 10.89$; $P = 0.012$), with fewer C and more A (Table 2). There were no differences in nucleotide frequencies between those species with dogs as opposed to cats as their definitive host ($G = 3.69$; $P = 0.298$) (Table 2).

TABLE 3
Percentage Frequency of Variable Nucleotide
Positions in *Sarcocystis* and Its Subsets

Nucleotide	Overall	Region of rRNA		
		Helical- double	Helical- single	Nonhelical
<i>Sarcocystis</i>	16.4	11.3	12.7	24.7
Species with a ruminant as intermediate host	13.1	9.0	10.4	19.7
Species with a nonrumi- nant as intermediate host	5.2	5.0	3.9	6.0
Species with a dog as definitive host ^a	4.9	4.0	2.6	7.1
Species with a cat as definitive host	5.8	2.7	3.6	10.8
Total number of nucleo- tide positions	2008	971	307	730

^a With the exception of *S. aucheniae*, which is not part of that clade.

There are more variable nucleotide positions for those species with ruminants as their intermediate host than for those with nonruminants as their intermediate host ($G = 76.94$; $P < 0.001$), and this is true in the double-stranded regions ($G = 11.56$; $P = 0.001$), the single-stranded regions ($G = 10.14$; $P = 0.001$), and the nonhelical regions ($G = 63.91$; $P < 0.001$) (Table 3). The number of variable nucleotide positions for those species with cats as their definitive host and that for those with dogs as their definitive host are the same ($G = 1.42$; $P = 0.233$), but this is also true only in the double-stranded regions ($G = 2.71$; $P = 0.100$) and the single-stranded regions ($G = 0.49$; $P = 0.484$) (Table 3); in the nonhelical regions there are more variable nucleotide positions for those species with cats as their definitive host than for those with dogs as their definitive host ($G = 6.15$; $P = 0.013$).

Within *Sarcocystis* as a whole, there are twice as many variable nucleotide positions in the nonhelical regions than in the helical regions ($G = 55.70$; $P < 0.001$) (Table 3). This is also true for those species with ruminants as their intermediate host ($G = 43.18$; $P < 0.001$), but not for those with nonruminants as their intermediate host ($G = 2.14$; $P = 0.343$) (Table 3). It is also true for those species with dogs ($G = 12.71$; $P = 0.002$) as well as cats ($G = 51.83$; $P < 0.001$) as their definitive host (Table 3).

Within *Sarcocystis*, there are an average of two character-state changes per variable nucleotide position, with more changes in the double-stranded regions (2.28) and the nonhelical regions (2.02) than in the single-stranded regions (1.64) ($G = 16.94$; $P = 0.031$). This is also true for those species with ruminants as their intermediate host ($G = 15.48$; $P = 0.017$), but not for those with nonruminants as their intermediate host

($G = 2.59$; $P = 0.274$). This is also true for those species with dogs as their definitive host ($G = 14.38$; $P = 0.001$), but not for those with cats as their definitive host ($G = 0.45$; $P = 0.799$).

The rates of nucleotide changes for those species with ruminants or nonruminants as their intermediate host are the same ($G = 7.49$; $P = 0.187$), and this is true in the double-stranded regions ($G = 7.24$; $P = 0.203$), the single-stranded regions ($G = 5.71$; $P = 0.335$), and the nonhelical regions ($G = 2.95$; $P = 0.708$) (Table 4). The rates of nucleotide changes are very different for those species with dogs as their definitive host compared to those with cats as their definitive host ($G = 27.58$; $P < 0.001$), with fewer $A \leftrightarrow T$ and more $A \leftrightarrow C$ (Table 4). This is also true in the double-stranded regions ($G = 10.94$; $P = 0.053$), the single-stranded regions ($G = 9.90$; $P = 0.042$), and the nonhelical regions ($G = 11.82$; $P = 0.037$) (Table 4).

Within *Sarcocystis* as a whole, the rates of nucleotide changes are very different in the nonhelical regions compared to the double-stranded regions ($G = 47.04$; $P < 0.001$) and the single-stranded regions ($G = 10.70$; $P = 0.058$), with fewer $A \leftrightarrow G$ and more $A \leftrightarrow T$ (Table 4). This is also true for those species with ruminants as their intermediate host ($G = 26.50$; $P < 0.001$; $G = 11.91$; $P = 0.036$; respectively), but not for those with nonruminants as their intermediate host ($G = 9.23$; $P = 0.100$; $G = 1.74$; $P = 0.883$; respectively) (Table 4). This is also true for those species with cats as their definitive host ($G = 14.24$; $P = 0.014$; $G = 23.30$; $P < 0.001$; respectively), but not for those with dogs as their definitive host ($G = 6.87$; $P = 0.231$; $G = 5.36$; $P = 0.374$; respectively) (Table 4).

Within *Sarcocystis*, the average transition:transversion ratio is 1.54, with a higher value in the double-stranded regions (2.31) than in the single-stranded (1.69) and nonhelical regions (1.13) ($G = 18.21$; $P < 0.001$). This is also true for those species with nonruminants as their intermediate host (ti:tv = 3.40, 1.43, and 1.43, respectively; $G = 5.59$; $P = 0.061$), but for those species with ruminants as their intermediate host the value for the single-stranded regions (2.08) is as high as that for the double-stranded regions (2.11) compared to the nonhelical regions (1.13) ($G = 10.578$; $P = 0.005$). For those species with dogs as their definitive host, there are no differences in the transition:transversion ratios (ti:tv = 1.37, 1.20, and 1.03, respectively; $G = 0.67$; $P = 0.714$), but for those species with cats as their definitive host the value for the single-stranded regions (22.00) is much higher than the more usual values for the double-stranded regions (2.84) and the nonhelical regions (1.16) ($G = 17.16$; $P < 0.001$).

DISCUSSION

The different tree-building procedures investigated in this study yielded the same basic structure for the

TABLE 4
Percentage Frequency of Nucleotide Changes
in *Sarcocystis* and Its Subsets

Nucleotide change	Region of rRNA			
	Overall	Helical-double	Helical-single	Nonhelical
<i>Sarcocystis</i>				
A ↔ C	9.4	10.4	12.5	8.3
A ↔ G	21.7	28.7	26.6	16.0
A ↔ T	22.6	10.4	17.2	32.0
C ↔ G	8.6	9.2	12.5	7.4
C ↔ T	21.8	25.1	18.8	20.4
G ↔ T	15.9	16.3	12.5	16.0
Total No. changes	678	251	64	363
Species with a ruminant as intermediate host				
A ↔ C	9.9	10.9	14.3	9.0
A ↔ G	22.2	27.3	35.7	16.1
A ↔ T	23.6	12.1	16.7	32.2
C ↔ G	9.1	10.9	9.5	8.2
C ↔ T	20.7	23.6	14.3	19.9
G ↔ T	14.6	15.2	9.5	14.6
Total No. changes	474	165	42	267
Species with a nonruminant as intermediate host				
A ↔ C	9.6	11.1	8.3	8.2
A ↔ G	20.9	31.5	8.3	14.3
A ↔ T	17.4	9.3	16.7	24.5
C ↔ G	5.2	1.9	16.7	6.1
C ↔ T	30.4	18.5	33.3	28.6
G ↔ T	16.5	13.0	16.7	18.4
Total No. changes	115	54	12	49
Species with a dog as definitive host ^a				
A ↔ C	16.2	16.1	12.5	16.7
A ↔ G	19.17	22.6	25.0	16.7
A ↔ T	14.0	6.5	12.5	21.2
C ↔ G	16.2	17.7	37.5	12.1
C ↔ T	17.6	17.7	12.5	18.2
G ↔ T	16.9	19.4	0.0	15.2
Total No. changes	136	62	8	66
Species with a cat as definitive host				
A ↔ C	4.4	3.3	8.3	4.3
A ↔ G	29.6	46.7	75.0	18.3
A ↔ T	30.4	13.3	0.0	39.8
C ↔ G	6.7	6.7	0.0	7.5
C ↔ T	16.3	10.0	16.7	18.3
G ↔ T	12.6	20.0	0.0	11.8
Total No. changes	135	30	12	93

^a With the exception of *S. aucheniae*, which is not part of that clade.

estimate of the phylogenetic relationship among the members of *Sarcocystis*, which presumably represents the underlying phylogenetic signal. This indicates that there are phylogenetically informative regions of the ssu rDNA that are relatively robust to different strategies for constructing cladograms. In particular, those species with ruminants as their intermediate host form a well-supported clade, and there are two major clades

within this group, one containing those species forming microcysts and with dogs as their definitive host (*S. arieticanis*, *S. capracanis*, *S. cruzi*, and *S. tenella*) and the other containing those species forming macrocysts and with cats as their definitive host (*S. buffalonis*, *S. fusiformis*, *S. gigantea*, *S. hirsuta*, and *S. moulei*).

However, it is clear that the placement of some of the taxa is sensitive to the tree-building procedure used. There are, for example, a number of possible rearrangements within the two major clades, and the relationship between *S. muris* and *S. neuropa* is not clear. These latter two species have very similar ssu rDNA sequences, differing in less than 50 nucleotide positions; however, it seems that these differences are capable of different phylogenetic interpretations.

The placement of *S. aucheniae* on the cladogram is also not particularly robust. However, phenotypically it is different from those other species with ruminants as their intermediate host, because it forms macrocysts but has the dog as its definitive host. Consequently, its placement as the sister species to the other two clades is reasonable. This species may, in fact, be part of a third clade, consisting of those other species with the same combination of phenotypic characteristics, such as *S. alceslatrans* (intermediate host moose), *S. hircicanis* (goat), and *S. sibirica* (roe deer) (see Dubey *et al.*, 1989). There are no species forming microcysts that have been reported to have cats as their definitive host, but they would be a valuable addition to the data set should they prove to exist.

Those species with nonruminants as their intermediate host do not form a monophyletic group. Although the relationship between *S. muris* (intermediate host mouse) and *S. neuropa* (bird) is ambiguous, their paraphyletic relationship with *S. species* (mouse) is not. This situation could be further investigated by sampling more species with an intermediate host of either rodents (such as *S. cymruensis*, *S. dispersa*, or *S. sebeki*) or birds (such as *S. horvathi* or *S. riyeli*).

Both *S. aucheniae* and *S. species* have relatively long branches on the cladograms, indicating that they have many unique character-state changes on the tree. These long branches will not have affected the structure of the phylogenetic tree (e.g., via long-branch attraction), because they are separated by an equally long branch. These branches may be shortened by the inclusion of other taxa that branch off from these long branches, such as the other species listed above.

The fact that our analyses indicate that the clades are based on combinations of phenotypic characteristics indicates that previous rDNA analyses of the Sarcocystidae (e.g., Barta *et al.*, 1991; Tenter *et al.*, 1992; Ellis *et al.*, 1994) were limited by their sampling of the species. It appears that there are not just two groups within the family, based on canid versus felid definitive hosts, but that the relationship is more

complex, involving other features such as intermediate host and sarcocyst size.

There is a close agreement between the genotypic groups and their phenotypic characteristics, including definitive and intermediate hosts, sporocyst and sarcocyst size, and sarcocyst wall type. Thus, it should be possible to predict the definitive host for those *Sarcocystis* species where it is unknown, based on the intermediate host and its sarcocyst size. While the extent to which our conclusions can be generalized to other taxa and genes is unknown, the robust nature of most of our clades allows us to predict that other *Sarcocystis* species with the same phenotypic characteristics should fall into the same genotypic groups.

The close concordance between the genotypic and phenotypic data sets also allows us to infer the evolutionary transformation series for the phenotypic characters. For example, the cladogram indicates the following as the evolutionary order for each of the character states within the Sarcocystidae:

Definitive host:	cat → dog → cat
Intermediate host:	many → nonruminant → ruminant
Sporocyst size:	small → large → small
Sarcocyst size:	micro → macro → micro
Sarcocyst wall type:	thick smooth ← 1 → 20 → 7 → 14
	$\begin{array}{ccc} \downarrow & & \downarrow \\ 10 & & 10 \end{array}$

Most of these transformation series involve reversals of states through time, for example, definitive host-switching and cyst size. Furthermore, the intermediate hosts have become more specialized through time.

We expect that the putative secondary-structure model that we employed is likely to have produced a multiple-sequence alignment that is close to the true alignment, in the sense of having aligned homologous nucleotides (Kjer, 1995; Hickson *et al.*, 1996; Morrison and Ellis, 1997), since the higher-order structures inferred from comparative analyses are now quite refined (Gutell *et al.*, 1994; Gutell, 1996). Comparative sequence analysis provides a powerful way of identifying functionally important elements in a molecular structure, based on the principle that during evolution the structure of the RNA has been better conserved than has the corresponding DNA sequence, and therefore even quite divergent sequences must still produce a ribosome with the same three-dimensional structure and the same molecular function (Gutell *et al.*, 1994). Our sequence alignment therefore represents our preferred set of hypotheses concerning the homology (i.e., evolutionary similarity) of the ssu rDNA sequences, and it is thus possible to consider the genotypic characters and character-state changes in detail.

Most of the species have considerable genotypic

differences based on their ssu rDNA sequences (differing in more than 100 nucleotide positions), except for *S. buffalonis* and *S. hirsuta*. Nevertheless, the fact that these latter two species differ unambiguously in 13 nucleotide positions supports the characterization of *S. buffalonis* as a newly described species that infects water buffaloes (Huong *et al.*, 1997). However, the genotypic similarity of these species is also reflected in their close phenotypic resemblance.

Most of the characters and character-state changes that we investigated differ significantly between those species that have ruminants as their intermediate host and those that have nonruminants. For example, compared to the nonruminant species the ruminant species have (1) fewer A nucleotides and more C; (2) more variable nucleotide positions; (3) a different number of variable nucleotide positions in the helical relative to the nonhelical regions; (4) fewer character-state changes per variable nucleotide position; (5) a different set of character-state changes in the nonhelical regions relative to the helical regions; and (6) equal transition:transversion ratios in the double- and single-stranded regions of the helices. This suite of genotypic differences suggests that these two groups have had very different evolutionary histories. However, we can find no particularly convincing rationalization for the origins of the patterns of differences that we have observed.

Similarly, most of the characters and character-state changes that we investigated differ significantly between those clades that have dogs as their definitive host and those that have cats. For example, compared to dog species the cat species have (1) more variable nucleotide positions in the nonhelical regions; (2) an equal number of variable nucleotide positions in the helical regions relative to the nonhelical regions; (3) more A ↔ T and fewer A ↔ C character-state changes; (4) a different set of character-state changes in the nonhelical regions relative to the helical regions; and (5) a different transition:transversion ratio in the helical relative to the nonhelical regions. This suite of genotypic differences suggests that these two clades have also had very different evolutionary histories. There seems to be no reason to expect that there should be any special explanation for these patterns of differences, given that many of the phenotypic character-state changes involve reversals.

The rDNA sequences that represent the different structural regions of the rRNA molecule have very different genotypic behavior throughout *Sarcocystis*. These structural regions are not arbitrary, as the double-stranded helical regions (48.4% of the positions) contain all of the paired bases within the helices, the single-stranded helical regions (15.3%) contain all of the unpaired bases within the helices (whether bulges along the stems or loops at the ends), and the nonhelical regions (36.4%) contain the remainder of the single-

stranded regions. The helical regions are assumed to form first and thus define the secondary structure of the RNA, and tertiary contacts between the secondary-structure regions then fold the RNA into its three-dimensional structure (Westhof and Michel, 1994). Consequently, the evolutionary behavior of these groups should be functionally constrained, and their differences should be open to explanation in terms of the importance of the nucleotide sequences to the functions of the various structural classes (Vawter and Brown, 1993).

First, the nucleotide frequencies are very different in the double-stranded regions compared to the single-stranded and nonhelical regions, with fewer A and more G+C. This GC-richness in the stems might be expected on the basis that G-C pairs have a lower free energy value than do A-U or G-U pairs (Turner *et al.*, 1988; Zuker, 1989). Furthermore, Gutell *et al.* (1985) have suggested that the bulges and loops might be more A-rich than the stems because A is the least polar of the bases and might thus facilitate hydrophobic interactions with proteins.

Second, there are twice as many variable nucleotide positions in the nonhelical regions as in the helical regions. This result appears to contradict the studies by Gutell *et al.* (1985) and Van de Peer *et al.* (1996b), who found that more of the universally conserved positions are located in the single-stranded regions (both helical and nonhelical). These single-stranded regions are presumed to be conserved because they are involved in ribosomal subunit association (involving tertiary-structure contacts) as well as protein binding and the initiation of protein synthesis (involving sequence-specific recognition sites) (Varani and Pardi, 1994). It is possible that different patterns of sequence conservation occur in different phylogenetic groups, as a result of different structural constraints (Gutell, 1996). For example, our results are consistent with the data of Morrison and Ellis (1997), who examined the Sporozoa as a whole. The possible difference in this phylum thus deserves further attention.

Third, there are more character-state changes per variable nucleotide position in the double-stranded regions and nonhelical regions than in the single-stranded regions. The helical stems are assumed to be involved in the formation of the secondary structure of the ribosome, helping to position the single-stranded regions correctly for site-specific contacts (Varani and Pardi, 1994). Consequently, substitutions in the stems can occur freely within the constraint that the base-pairing is maintained, as the base-pairing is presumed to be involved mainly in maintaining the helix rather than in protein synthesis (Gutell, 1996). The large number of substitutions in the nonhelical regions may be related to the phenomenon discussed in the previous paragraph.

The method that we used to estimate the number of

substitutions (i.e., character-state changes) is an improvement over methods that merely tally the differences between the taxa when taken pairwise (Vawter and Brown, 1993). However, our method is likely to have underestimated the actual number of changes that have occurred (van de Peer *et al.*, 1996b), because it cannot detect superimposed substitutions. Thus it is an estimate of the minimum number of substitutions, and if these potential underestimates are randomly distributed among the regions, then the rank-order of our estimates will still be correct. Alternative methods for assessing nucleotide-site variability have been proposed (e.g., van de Peer *et al.*, 1993, 1996a, 1996b), based on corrected pairwise distances, which may be more effective than the procedure used here. Nevertheless, distance measures also lose phylogenetic information (Steel *et al.*, 1988) and may therefore have their own limitations, and methods that take into account the phylogenetic relationships may produce more accurate estimates (Yang, 1996).

Fourth, the rates of nucleotide changes are very different in the nonhelical regions compared to the helical regions, with fewer A \leftrightarrow G and more A \leftrightarrow T. The high rate of A \leftrightarrow G and C \leftrightarrow T changes in the stems (Table 4) is likely to be related to the maintenance of base-pairing in the RNA molecule (Vawter and Brown, 1993). In RNA the G-U pairing is stable (unlike the G-T pairing in DNA) (Westhof and Michel, 1994), and therefore certain single-nucleotide changes allow the stem structure to be maintained (e.g., A-U \rightarrow G-U \rightarrow G-C).

Finally, the transition:transversion ratio is highest in the double-stranded regions, followed by the single-stranded regions and then the nonhelical regions. This is presumably a consequence of the maintenance of base-pairing in the stems, since the transitions (A \leftrightarrow G and C \leftrightarrow T) can maintain the base-pairing while the transversions cannot (i.e., they convert a stem position into a bulge). These compensatory changes should thus be more frequent than expected by chance. The average transition:transversion ratio in the helical regions is 2.16, thus justifying our use of a ratio of 2 in the cladistic analyses.

The method that we used to estimate the relative number of transitions and transversions is likely to have underestimated the actual transition:transversion ratio (Wakeley, 1996), because it cannot detect superimposed substitutions. For example, using the method of Yang and Kumar (1996) to correct the parsimony counts for multiple substitutions produces an overall transition:transversion ratio of 1.76, as opposed to our estimate of 1.54. However, if these potential underestimates are randomly distributed among the regions, then the rank-order of our estimates will still be correct.

In conclusion, we have determined the complete nucleotide sequence of the ssu rRNA gene for four

species of *Sarcocystis*: *S. aucheniae*, *S. buffalonis*, *S. cruzi*, and *S. hirsuta*. This has allowed a detailed phylogenetic analysis of the genus. This revealed that those species with ruminants as their intermediate host form a well-supported clade, and that there are at least two major clades within this group, one containing those species forming microcysts and with dogs as their definitive host, and the other containing those species forming macrocysts and with cats as their definitive host. Those species with nonruminants as their intermediate host form the paraphyletic sister group to these clades. There is a large suite of genotypic differences that indicate that those species infecting ruminant and nonruminant hosts have had very different evolutionary histories and also that the felid- and canid-infecting species have had very different evolutionary histories. Furthermore, within *Sarcocystis*, the rDNA sequences that represent the different structural regions of the rRNA molecule have very different genotypic behavior.

ACKNOWLEDGMENTS

We thank J. Parsons and P. O'Donoghue for materials and information relating to *S. aucheniae*, J. Barta for providing the *S. sp.* sequence, R. De Wachter for providing the original rDNA structure alignment, A. Uggla for helpful discussions and comments concerning the work, and C. Rice for technical support. This work was supported by grants from the Svedberg Foundation and the Australian Research Council.

REFERENCES

- Barta, J. R., Jenkins, M. C., and Danforth, H.D. (1991). Evolutionary relationships of avian *Eimeria* species among other apicomplexan protozoa: Monophyly of the Apicomplexa is supported. *Mol. Biol. Evol.* **8**: 345–355.
- De Rijk, P., and De Wachter, R. (1993). DCSE, an interactive tool for sequence alignment and secondary structure research. *Comp. Appl. Biosci.* **9**: 735–740.
- Dubey, J. P., Speer, C. A., and Fayer, R. (1989). "Sarcocystosis of Animals and Man," CRC Press, Boca Raton, FL.
- Ellis, J., Luton, K., Baverstock, P. R., Brindley, P. J., Nimmo, K. A., and Johnson, A. M. (1994). The phylogeny of *Neospora caninum*. *Mol. Biochem. Parasitol.* **64**: 303–311.
- Ellis, J., and Morrison, D. (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., and Jeffries, A. C. (1998). The phylum Apicomplexa: An update on the molecular phylogeny. In "Evolutionary Relationships among Protozoa" (G. Coombs and A. Warren, Eds.), Chapman & Hall, London.
- Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783–791.
- Felsenstein, J. (1995). "PHYLIP (Phylogeny Inference Package), Version 3.57c," University of Washington, Seattle.
- Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Camona, E. V., and Sogin, M. L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Cryptosporidium parvum* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol. Biochem. Parasitol.* **45**: 147–154.
- Gutell, R. R. (1996). Comparative sequence analysis and the structure of 16S and 23S rRNA. In "Ribosomal RNA" (R. A. Zimmermann and A. E. Dahlberg, Eds.), pp. 111–128, CRC Press, Boca Raton, FL.
- Gutell, R. R., Larsen, N., and Woese, C. R. (1994). Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* **58**: 10–26.
- Gutell, R. R., Weiser, B., Woese, C. R., and Noller, H. F. (1985). Comparative anatomy of 16S-like ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **32**: 155–216.
- Hickson, R. E., Simon, C., Cooper, A., Spicer, G. S., Sullivan, J., and Penny, D. (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Mol. Biol. Evol.* **13**: 150–169.
- Holmdahl, O. J. M., Mattsson, J., Uggla, A., and Johansson, K. E. (1991). The phylogeny of *Neospora caninum* and *Toxoplasma gondii* based on ribosomal RNA sequences. *FEMS Microbiol. Lett.* **194**: 187–192.
- Holmdahl, O. J. M., Mattsson, J. G., Uggla, A., and Johansson, K.-E. (1993). Oligonucleotide probes complementary to variable regions of 18S rRNA from *Sarcocystis* species. *Mol. Cell. Probes* **7**: 481–486.
- Huong, L. T. T., Dubey, J. P., Nikkilä, T., and Uggla, A. (1997). *Sarcocystis buffalonis* n. sp. (Protozoa: Sarcocystidae) from the water buffalo (*Bubalus bubalis*) in Vietnam. *J. Parasitol.* **83**: 471–474.
- Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**: 170–179.
- Kjer, K. M. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. *Mol. Phylogenet. Evol.* **4**: 314–330.
- Levine, N. D. (1985). Phylum II. Apicomplexa Levine 1970. In "An Illustrated Guide to the Protozoa" (J. J. Lee, S. H. Hunter and E. C. Bovee, Eds.), pp. 322–374, Society of Protozoologists, Lawrence.
- Lyons-Weiler, J. (1997). "Manual for RASA 2.1," University of Nevada, Reno.
- Lyons-Weiler, J., Hoelzer, G. A., and Tausch, R. J. (1996). Relative apparent synapomorphy analysis (RASA). I. The statistical measurement of phylogenetic signal. *Mol. Biol. Evol.* **13**: 749–757.
- Maddison, W. P., and Maddison, D. R. (1992). "MacClade: Analysis of Phylogeny and Character Evolution," Sinauer Associates, Sunderland, MA.
- Medlin, L., Elwood, H. J., Stickel, S., and Sogin, M. L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**: 491–499.
- Morrison, D. A. (1996). Phylogenetic tree-building. *Int. J. Parasitol.* **26**: 589–617.
- Morrison, D. A., and 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.
- Smith, A. B. (1994). Rooting molecular trees: Problems and strategies. *Biol. J. Linn. Soc.* **51**: 279–292.
- Sogin, M. L., and Gunderson, J. H. (1987). Structural diversity of eukaryotic small subunit ribosomal RNAs. Evolutionary implications. *Ann. N. Y. Acad. Sci.* **503**: 125–139.
- Steel, M. A., Hendy, M. D., and Penny, D. (1988). Loss of information in genetic distances. *Nature* **336**: 118.
- Swofford, D. L. (1993). "PAUP: Phylogenetic Analysis Using Parsimony," Smithsonian Institution, Washington, DC.
- Swofford, D. L., and Maddison, W. P. (1987). Reconstructing ancestral states under Wagner parsimony. *Math. Biosci.* **87**: 199–229.
- Tenter, A. M., Baverstock, P. R., and Johnson, A. M. (1992). Phyloge-

- netic relationships of *Sarcocystis* species from sheep, goats, cattle and mice based on ribosomal RNA sequences. *Int. J. Parasitol.* **22**: 503–513.
- Turner, D. H., Sugimoto, N., and Frier, S. M. (1988). RNA structure prediction. *Annu. Rev. Biophys. Biophys. Chem.* **17**: 167–192.
- Uggla, A., and Buxton, D. (1990). Immune responses against *Toxoplasma* and *Sarcocystis* infections in ruminants: Diagnosis and prospects for vaccination. *Rev. Sci. Tech. Off. Int. Epiz.* **9**: 441–462.
- Van de Peer, Y., Neefs, J.-M., De Rijk, P., and De Wachter, R. (1993). Reconstructing evolution from eukaryotic small ribosomal subunit RNA sequences: Calibration of the molecular clock. *J. Mol. Evol.* **37**: 221–232.
- Van de Peer, Y., Van der Auwera, G., and De Wachter, R. (1996a). The evolution of Stramenopiles and Alveolates as derived by “substitution rate calibration” of small ribosomal subunit RNA. *J. Mol. Evol.* **42**: 201–210.
- Van de Peer, Y., Chapelle, S., and De Wachter, R. (1996b). A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Res.* **24**: 3381–3391.
- Van de Peer, Y., Jansen, J., De Rijk, P., and De Wachter, R. (1997). Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res.* **25**: 111–116.
- Varani, G., and Pardi, A. (1994). Structure of RNA. In “RNA–Protein Interactions” (K. Nagai and I. W. Mattaj, Eds.), pp. 1–24, IRL Press, Oxford.
- Vawter, L., and Brown, W. M. (1993). Rates and patterns of base change in the small subunit ribosomal RNA gene. *Genetics* **134**: 597–608.
- Wakeley, J. (1996). The excess of transitions among nucleotide substitutions: New methods of estimating transition bias underscore its significance. *Trends Ecol. Evol.* **11**: 158–163.
- Westhof, E., and Michel, F. (1994). Prediction and experimental investigation of RNA secondary and tertiary foldings. In “RNA–Protein Interactions” (K. Nagai and I. W. Mattaj, Eds.), pp. 25–51, IRL Press, Oxford.
- Wilkinson, L. (1991). “SYSTAT: The System for Statistics,” Systat Inc., Evanston, IL.
- Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* **11**: 367–372.
- Yang, Z. and Kumar, S. (1996). Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. *Mol. Biol. Evol.* **13**: 650–659.
- Zuker, M. (1989). Computer prediction of RNA structure. *Methods Enzymol.* **180**: 262–288.