PHYLOGENETIC ANALYSIS OF FIVE PROTOCTIST
PARASITES OF INSECTS

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

of the

Australian National University

Canberra

Molecular Evolution and Systematics Group

Research School of Biological Sciences

Jennifer Ann Ninham

May 1995

STATEMENT

This thesis contains no material which has previously been
submitted for an academic record at this or any other university
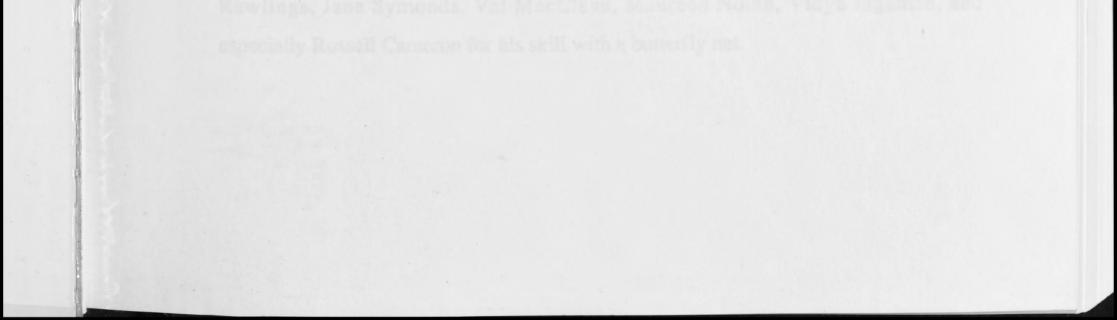and is the original work of the author, except where acknowledged.

Jennifer Ann Ninham

ACKNOWLEDGEMENTS

When I look back over the past four years, I am humbled by the number of people who have offered their help and friendship to me and I feel very privileged to have known and worked with them. The names below are the ones to whom I owe the greatest amount, there may be others who have helped in passing and I thank them all.

Firstly I would like to thank my supervisor Professor Adrian Gibbs for providing me with the opportunity to undertake this study. His understanding and willingness to be of help at any time, his guidance, sense of humour and continued good nature has made the time spent in his group a priviledge and pleasure.

I would also like to thank those people who supported me in the beginning of this undertaking without whom it could not have happened; Peter Stewart, Klaus Matthaei, Lex Beardsell and David Shaw.

For their encouragement and time, for always being there with patience, to discuss references, new ideas, computing, laboratory skills and especially for editing several chapters of this thesis, I would like to thank Paul Keese and Mary Skotnicki.

I would especially like to thank Anne MacKenzie for her expertise in the lab when without her continued patience, cloning and sequencing could not have happened.

As well, I would especially like to thank Georg Weiller for allowing me to use his DIPLOMO software prior to its initial publication, John Trueman, Jack Palmer, John Armstrong and David Sandilands for their continued patience and help with computing information and skills and Marjo Torronen who spent endless hours with me at the VAX helping with sequence alignments.

For other technical help I would like to thank Anne MacKenzie, Julia Playford, Marjo Torronen, Nelida Contreras, Denis Anderson, David Shaw, Mandy Reid, Val Rawlings, Jane Symonds, Val MacClean, Maureen Nolan, Vidya Jagadish, and especially Russell Cameron for his skill with a butterfly net.

TABLE OF CONTENTS

ABSTRACT

The work reported in this thesis is based on five single-celled parasites of insects belonging to the Kingdom Protoctista. Except for *Nosema apis*, which belongs to the Phylum Microspora, and was recovered from the cosmopolitan honey bee *Apis mellifera*, none of these parasites had been formally described. The latter comprised a microsporan-like organism recovered from the European wasp, *Vespula germanica*, and three gregarine-like organisms belonging to the Phylum Apicomplexa, and which were recovered from the Australian grasshopper, *Caledia captiva*, the Australian locust, *Chortoicetes terminifera*, and the Australian sugar-cane beetle, *Lepidiota frenchi*.

There were two aims. One was to describe the morphology and life cycle of the gregarine-like organism recovered from *Caledia captiva*. The other, was to use the sequences of the small subunit rRNA (SSU rRNA: 18S and 16S) in order to resolve the phylogenetic status of each of the five parasites. There are no previously published SSU rRNA sequences of gregarines which could indicate their evolutionary relationships within the apicomplexans, but morphological findings suggest that they are derived from a primitive class of apicomplexans, the Perkinsasida. The placement of the microsporans as a sister group to the eukaryotes is based on the published SSU rRNA sequence from a single microsporan, *Vairimorpha necatrix*. This organism has eukaryotic ultra-structure, but prokaryotic 70S ribosomes and 16S rRNA.

The morphology and life cycle of the gregarine-like organism from *Caledia captiva* was found to be similar to that of other cephaline gregarines. Variations in the size of sporonts was evident and this may be genetically determined.

Genomic DNA was extracted from the parasites and using synthetic oligonucleotide primers, the SSU rRNA gene (except for approximately 500 nucleotides at the 5' end) was amplified by the polymerase chain reaction (PCR), ligated into M13mp18/19*SmaI* and pBluescript SK(+)/*Eco*RV vectors, cloned into *E.coli* JM101 cells, and sequenced using a variety of different primers. Cloned fragments were first allocated to their relative positions using the 18S rDNA sequence from *Plasmodium falciparum* for comparison and then compiled into contiguous sequences using a computer program.

Secondary structure models of the SSU rRNAs of each of the five parasites were constructed using published secondary structures of those of closely related apicomplexans and microsporans as a guide. The SSU rRNA sequences of 25 other taxa from all kingdoms were obtained from public databases, edited and aligned with the gregarine and microsporan sequences. Based on the alignment and the secondary structure models, 715 conserved nucleotides of these sequences were selected. Different combinations of these conserved sequences were used to construct phylogenetic trees by distance and maximum parsimony methods and their topologies shown to be closely similar. The relationships of the gregarine and microsporan taxa were inferred using subsets of the data, the certainty of these tested by the T-PTP and bootstrap tests. The two methods gave similar results which suggested that these were not method-dependent.

Except for the anomalous *Giardia lamblia*, the wasp microsporan, *Nosema apis* and *Vairimorpha necatrix* were found to be a monophyletic clade which formed a sister group to the other eukaryotes. This was supported by the secondary structure models of the SSU rRNA which also showed that the newly-studied microsporans, like *Vairimorpha necatrix* had both eukaryotic and prokaryotic features. However the relative positions of the wasp microsporan, *Nosema apis* and *Vairimorpha necatrix* were not resolved with this data.

Within the apicomplexan monophyletic clade, the three gregarine-like parasites formed a monophyletic group. The parasite found in beetles was a sister group to a clade containing the two parasites found in the grasshopper and locust respectively. However the position of the gregarines within the apicomplexans was not resolved by the analysis. The secondary structure models of their SSU rRNAs were typically eukaryotic, but did not provide any other phylogenetic insights.

The work showed that while, in general, the phylogenetic position of the microsporan and gregarine-like organisms can be determined from their SSU rRNA sequences, the details of close relationships cannot, and it is likely that sequences from other genes will be needed to resolve them.

1

CHAPTER 1

GENERAL INTRODUCTION

1.1 Background

Phylogenetic studies based on both phenotypic and molecular data are thought to provide more accurate descriptions and interpretations of biological diversity than those derived from only one type of data (Moritz and Hillis, 1990). Within the Kingdom Protoctista (Margulis 1974 a, b; 1980; 1989), phenotypic analyses have resolved the single-celled, free-living eukaryotes into various phyla and other taxonomic sub-groups, but have not been able to establish sensible evolutionary relationships at higher taxonomic levels. The evolutionary pathways of parasitic forms are even more difficult to resolve because their specialisations often result in a reduction of their physical, ultra-structural and biochemical properties (Cavalier Smith, 1987). Thus the most reliable method is to compare parts of their genomic sequences, especially those shared by all taxa being compared, such as their small subunit ribosomal RNA (SSU rRNA: 18S and 16S) sequences (Woese, 1987; Olsen and Woese, 1993).

I report in this thesis studies of the phylogenetic relationships of five parasitic species belonging to two protoctistan phyla namely, Phylum Apicomplexa Levine, 1970 and Phylum Microspora (Canning, 1989). The apicomplexans were two septate gregarines and a pseudogregarine, and the microsporans were two *Nosema spp.*. Other taxa from these phyla which also appear in the analysis have already been evaluated phylogenetically by both phenotypic (Canning, 1989; Vivier and Desportes, 1989; Sprague *et al.*, 1992) and molecular (Johnson *et al.*, 1987, 1988, 1990; Johnson and Baverstock, 1989; Barta *et al.*, 1991; Gajadhar *et al.*, 1991; Ellis *et al.*, 1992; Gagnon *et al.*, 1993; Goggin and Barker, 1993) techniques. The following general descriptions of the two phyla include specific taxonomic information about the five parasitic species together with conclusions already published with the molecular data.

1.2 Phylum Apicomplexa Levine, 1970

This complex phylum contains parasites of invertebrates and vertebrates and has between 2400 and 4600 named species (Levine, 1988). All members of the phylum are characterised by having a life cycle with a sexual phase or gamogony in which male gametes have undulipodia, a sporogenic phase and sometimes a growth phase. The infective stage (sporozoite or zoite) has on its anterior end, a set of ultra-structural organelles collectively called the apical complex, by which it invades host cells. The apical complex consists of a polar ring connected with microtubules that extend under the cell surface as a sub-pellicular layer, two apical or conoidal rings, a cone shaped structure, the conoid, made up of several spirally arranged microtubules and two pedunculate secretory organelles called rhoptries (Levine, 1970; Vivier and Desportes, 1989). Although the conoid can be incomplete or absent, the microtubules and rhoptries seem to be of fundamental importance (Perkins, 1976); the latter probably contain agents which induce phagocytosis of the parasite by the host cell (Jones *et al.*, 1972; Trager, 1974).

These phenotypic traits define this phylum but provide no information about the relationships of the phylum with other taxa. There are no fossilised remains, but it is thought that the probable evolution of parasitic species may be deduced from their occurrence in, and possible co-evolution with, their hosts (Theodorides, 1984). The phylum is divided into four taxonomic classes. The class Perkinsasida Levine 1978 (Levine, 1988) is thought to be the most primitive one and has only one genus *Perkinsus* (Perkins, 1976; Levine, 1978) with four described species which parasitise marine molluscs and occur worldwide (Levine, 1988; Goggin and Barker, 1993). Members of the class Gregarinia (500 species) parasitise a range of marine invertebrates and freshwater and terrestrial arthropods, whereas those of the class Coccidia (more than 1600 species) parasitise both invertebrates and vertebrates. The class Haematozoa includes two taxonomic orders, the Haemosporidia (200 species) and Piroplasmida (about 100 species) (Vivier and Desportes, 1989), members of which parasitise vertebrates. It has been proposed that the dinoflagellates are the sister group to the apicomplexans because both groups have zygotic meioses and flagellated gametes (Sleigh, 1973 and Levine, 1988). Levine (1988) has also proposed that the original apicomplexans inhabited the guts of marine polychaetes. In these, primitive

dinoflagellates evolved into ancient members of the class Perkinsasida which in turn gave rise to primitive gregarines and coccidians. The gregarines subsequently entered water-associated and then terrestrial arthropods by way of marine and freshwater invertebrates whereas the coccidians evolved similarly to specialise in vertebrate hosts. During that time, coccidians gave rise to the haemosporids and piroplasmids which parasitise vertebrates using arthropod vectors.

1.2.1 Phenotypic descriptions

a. Class Perkinsasida Levine 1978

Members of this class have typical apicomplexan structures and cell division (Perkins, 1976) but differ from other members of the phylum. *Perkinsus marinus* (Perkins, 1976; Levine, 1978) has an incomplete conoid and forms bi-flagellated sporozoites in which the anterior flagellum contains mastigonemes (Levine, 1978). Such accessory structures have not been observed on other flagellated cells, which in other apicomplexans are not sporozoites but micro-gametes. Other structures are similar to those of other apicomplexans. The sporangium and its discharge tube are similar to analogous structures in gregarines, whereas the absence of a digestive vacuole associated with the conoid indicates affinities closer to the coccidians (Perkins, 1976).
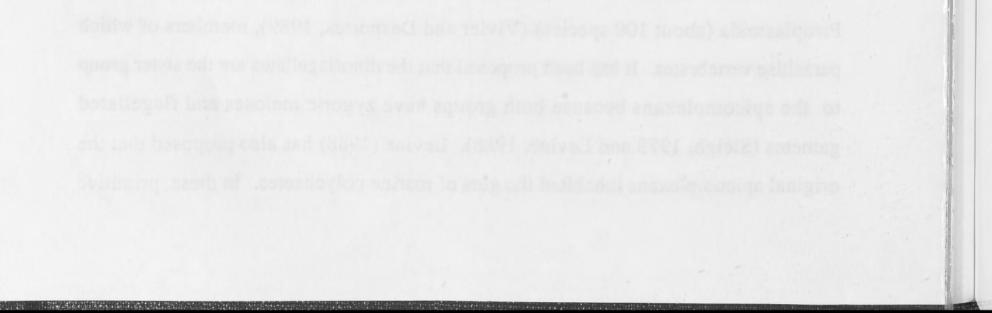
b. Class Gregarinia

Except for one order, the Neogregarinidae, members of the Gregarinia are extra-cellular gut parasites (Vivier and Desportes, 1989). Their apical complex is complete, but after entry of the sporozoite to the host alimentary canal, it becomes modified as a mucron or epimerite, which temporarily attaches each sporozoite to a gut epithelial cell. Cells of this stage develop into large extra-cellular cephalonts which subsequently detach from their epimerites as motile trophonts. Gamogony occurs within the gut inside a gametocyst. Trophic stages differentiate into approximately equal numbers of highly differentiated male and female gametes which undergo fertilisation. Zygotes form before sporogony, and the resulting gametocysts are shed into the environment with the host faeces. There they dehisce and release the spores into the environment. If they are ingested by a new host, the spores undergo meiosis to form haploid motile sporozoites (Vivier and Desportes, 1989).

page 3, 1.2, line 2

The sentence should read: "Contemporary findings (Levine, 1988) show that the gregarines subsequently entered water-associated and then terrestrial arthropods......"

The most recent taxonomies for septate gregarines, which are the subject of this thesis, are based on the morphology and cytology of all parasitic stages, aspects of the parasitic life cycle, the location of these in the host and the host's phyletic position. The following classification combines these taxonomies:

Phylum Apicomplexa Levine, 1970 (Levine, 1988; Vivier and Desportes, 1989)

Class Gregarina Dufour, 1828 (Vivier and Desportes, 1989)

Order Eugregarinida Leger, 1899 (Vivier and Desportes, 1989) or

Order Eugregarinorida Leger, 1900 (Levine, 1988)

Sub order Septatorina Lankester, 1885 (Levine, 1988)

Family Gregarinidae Labbé, 1899 (Levine, 1988; Vivier and Desportes, 1989) (septum present, trophonts develop in pairs, chains of cylindrical or ellipsoidal sporocysts released by several sporoducts) (Vivier and Desportes, 1989).

Genus Gregarina Dufour, 1828 (Levine, 1988; Vivier and Desportes, 1989)

c. Class Haematozoa

The haematozoans include the haemosporids *Plasmodium spp.,* which are the causal agents of malaria in birds, reptiles and mammals using mosquito vectors, and the piroplasmids *Babesia* and *Theileria spp.,* which use tick vectors to cause respectively, piroplasmosis and theileriosis in cattle (Vivier and Desportes, 1989).

Vivier and Desportes (1989) have reported the following features of the haematozoan life cycle. Gamogonous, syngamous and sporogonous stages occur in the guts of insect hosts which are themselves parasites of vertebrates. This has resulted in several adaptations. During gamogony the female gamont differentiates into a single macrogamete and the male gamont into several hundred microgametes. After syngamy, sporozoites penetrate the gut wall and migrate as ookinetes to the salivary glands from which they are subsequently injected into the vertebrate blood steam. Entry of sporozoites to the red blood cells is followed by haemotrophy which depends on a further adaptation whereby the apical complex is reduced to a rudimentary structure which lacks the conoid and conoidal rings. Once inside the red blood cells multiple fission occurs by merogony. The small size of merozoites and their subsequent ability to ingest and digest host red blood cells are adaptations to their development in erythrocytes.

page 5, 1.2.2 at the end of the paragraph ending "...(Lehninger, 1982)".
Addition of new material.

"In prokaryotes, there are one to several copies of rRNA genes either organised into a single operon separated by tRNA genes or dispersed throughout the genome (Morgan, 1982) but in most eukaryotes, the genes for 5.8S, 18S and 28S rRNA are present in several hundred tandemly repeated copies, each with internally and externally transcribed spacers and separated by non-transcribed spacers (Mandal, 1984). Within eukaryotic individuals, the rRNA genes may differ in total length (approximately 2 kilobases), but vary little in nucleotide sequence. In most multi-gene families, there is a higher genetic similarity between two genes from within a species than there is between genes from two different species (Dover and Tautz, 1986). This is the result of homogenisation of variant nucleotides within and between (Hancock and Dover, 1988) multiple gene copies which exist in both homologous and non-homologous chromosomes so that they evolve in concert (Arnheim *et al.,* 1980; Dover and Tautz, 1986). The relatively high frequency of inter-specific genetic similarity cannot be explained solely by drift or selection, because the time required to effect species homogeneity prior to fixation would be too long if selection had to wait for all loci within the gene cluster to mutate to the same condition (Dover *et al.,* 1982). In contrast to drift and selection however, concerted evolution acts two or more orders of magnitude faster than the rate of point mutation, and fast enough to produce species homogeneity (Dover *et al.,* 1982), thus indicating a novel force in evolution called molecular drive (Dover, 1982). Concerted evolution is mediated by a variety of non-reciprocal exchanges, which include transposition, slippage recognition and RNA-mediated transfers, but in particular, gene conversion and unequal exchange (Dover and Tautz, 1986). The last two result in duplication or elimination of many repeats at once and thus promote the relatively high frequency and spread of mutations through the family (molecular drive) and the population (fixation) over evolutionary time (Hillis and Dixon, 1991).

Ribosomal RNA is considered to be a particularly useful molecule for phylogenetic analysis because different parts of the repeat unit evolve at different rates resulting in specific regions (domains) of the sequence which range between highly conserved (related to function) and highly variable (species specific). Although it is clear that for any species the entire rRNA sequence contains the greatest amount of

d. Class Coccidia

The coccidians include the *Eimeria spp.* which is the causal agent of coccidiosis particularly in domestic poultry, and the *Toxoplasma spp., Sarcocystis spp.* and *Cryptospororidium spp.,* which respectively cause toxoplasmosis, sarcosporidiosis and diarrhoea in domestic animals and livestock (Johnson *et al.,* 1990).

The coccids appear to form an intermediate evolutionary stage between the gregarines and haematozoans as they parasitise both invertebrates and vertebrate hosts, but have similar life styles in each. Like the gregarines, coccidians have complete apical complexes, but like the haematozoans (except *Babesia spp.*), gamogony produces a single female gamete and many male gametes (Vivier and Desportes, 1989).

1.2.2 18S rRNA sequence analysis

The SSU 18S rRNA gene has been studied in most eukaryote phyla (Gray *et al.,* 1984; Sogin *et al.,* 1986; Field *et al.,* 1988; Pace *et al.,* 1989) and almost exclusively in Protoctista, because it is one of the most slowly evolving genes found in living organisms, and thus has been used to reconstruct phylogenetic events from the Precambrian era (Hillis and Dixon, 1991). SSU rRNA genes are functionally equivalent, evolutionarily homologous and occur universally in the genomes of archaebacteria, eubacteria, mitochondria, plastids and the nucleus, and therefore provide local and global connections in molecular taxonomy (Gray *et al.,* 1984). Nuclear SSU rRNA forms part of the ribosome which is essential to protein biosynthesis and exists as two sub-units. The larger or 60S sub-unit (LSU) contains molecules of 5.8 rRNA, 28S rRNA and about 70 polypeptides. The smaller or 40S sub-unit (SSU) contains one molecule of 18S rRNA and about 30 polypeptides (Lehninger, 1982).

Analyses of the relationships of apicomplexan taxa inferred from 18S rRNA data usually contain only a small selection of protoctistan taxa. These include the phyla Ciliophora (Lynn and Small, 1981), Dinoflagellata (Taylor, 1989) and Microspora (Sprague *et al.,* 1992) and often a selection of multicellular taxa. Apicomplexan taxa are represented as follows:

a. The class Perkinsasida by *Perkinsus sp.* (Goggin and Barker, 1993).

phylogenetic information, the highly variable regions are usually deleted from any analyses, because compared to conserved regions, they carry a higher level of homoplastic nucleotides and accurate alignment with similar regions from other species is not possible (Hillis and Dixon, 1991). Moreover, it has been argued (Rothschild *et al.,* 1986) that molecular drive imposes limitations on phylogenetic analyses. This is because within the gene, rRNA sequence similarity among divergent domains is often higher than expected if the regions were evolving independently (Hancock and Dover, 1988), and between genes in the same organism, intragenomic variation is known (McCutchan *et al.,* 1988). Therefore in a phylogenetic analysis, the sequence positions within these regions cannot be treated as independent characters and subjective weighting needed in the analyses (Hillis and Dixon, 1991) would introduce bias. Alternatively, results of phylogenetic studies based on rDNA are usually consistent with those based on other sources of data that involve multiple comparisons (Hillis, 1987), so it is likely that methods of phylogenetic inference are sufficiently robust to handle the complexities of rDNA evolution (Hillis and Dixon, 1991)."

b. The class Coccidia by several *Eimeria spp.* (Barta *et al.*, 1991), *Toxoplasma spp.* (Johnson *et al.*, 1987, 1988; Johnson and Baverstock, 1989; Barta *et al.*, 1991 and Gagnon *et al.*, 1993), *Sarcocystis spp.* (Johnson *et al.*, 1988; Johnson and Baverstock, 1989; Barta *et al.*, 1991; Gajadhar *et al.*, 1991 and Gagnon *et al.*, 1993) and *Cryptosporidia spp.*, (Johnson *et al.*, 1990; Barta *et al.*, 1991 and Tenter *et al.*, 1992).

c. The orders Haemosporidia by *Plasmodium spp.* (referenced as in a. and b. above) and Piroplasmida by *Babesia spp.* (Ellis *et al.*, 1992; Goggin and Barker, 1993) and *Theileria spp.* (Barta *et al.*, 1991; Gajadhar *et al.*, 1991; Ellis *et al.*, 1992; Gagnon *et al.*, 1993; Goggin and Barker, 1993) (Figure 1).

Several studies made by both parsimony and distance methods placed the apicomplexans in a cluster quite separate from the other protoctistans and multicellular taxa. Within this cluster the coccidians *Sarcocystis, Toxoplasma* (Johnson *et al.*, 1988; Johnson and Baverstock, 1989; Gagnon *et al.*, 1993) and *Eimeria spp.* (Barta *et al.*, 1991) formed a second cluster that was a sister group to a third cluster formed by the two haematozoans, namely *Plasmodium* and *Theileria spp.* (Gagnon *et al.*, 1993), but the position of *Cryptosporidium spp.* varied; in some classifications they were monophyletic with the other three coccids (Tenter *et al.*, 1992) but in others formed a cluster with *Plasmodium spp.* (Johnson *et al.*, 1990; Barta *et al.*, 1991). Ellis *et al.* (1992) confirmed the close relatedness of *Babesia bovis* and *Babesia bigemina* which are known to share the same vectors and definitive hosts, and also showed that these were distant to *B. rodhaini*, which infects a different host and has *Theileria*-like meronts, implying that it might be more accurately classified as a member of that genus. In nearly all analyses, the *Plasmodium spp.* formed a sister group to the other apicomplexans, perhaps because they have evolved more rapidly (Wolters, 1991) which tends to make taxa appear to be separate sister groups (Felsenstein, 1978). The distinctiveness of *Plasmodium spp.* is supported by the presence of large insertions and deletions (indels) in the variable regions of their rRNA (Wolters, 1991), however it is difficult to determine their taxonomic positions with certainty as their genome has a very biased base composition (Hashimoto *et al.*, 1994).

In the only phylogenetic analyses of rRNA sequence data from the class Perkinsasida, Goggin and Barker (1993) used distance and parsimony methods to resolve the position of *Perkinsus sp.* They showed that this species forms a cluster with the dinoflagellates and the coccidians, which together form a sister group to a

second cluster containing haemosporids and piroplasmids. This supports the inference derived from morphological data (Perkins, 1976; Levine, 1978) that the Perkinsasida are a primitive apicomplexian group closely related to the dinoflagellates and the coccidians. No 18S rRNA sequences from the class Gregarinia have been reported.

1.3 Phylum Microspora

Microsporans are minute (diameter, 1-20µm), spore forming, obligate, intracellular parasites of hosts which range from vertebrates to species of Gregarinia (Canning, 1989). Certain species, namely *Nosema bombycis* and *Nosema apis* respectively, are economically important pathogens of silkworms (Ishihara and Hayashi, 1968) and honeybees (Bailey, 1963). Others like *Nosema acridophagus* and *N. cuneatum* are used as microbiological control agents for the grasshopper, *Melanoplus sanguinipes* (Erlandson *et al.*, 1985; Lockwood and Debrey, 1990). Cross infection by microsporans between insect and homeothermic vertebrate hosts does not usually occur. However, opportunistic infections by the microsporans *Encephalitozoon hellem* and *E. cuniculi* which parasitises laboratory rabbits (Wright and Craighead, 1922), have been recorded in humans infected with human immunodeficiency lentivirus (HIV) (Cali and Owen, 1988; Vossbrinck *et al.*, 1993; Zhu *et al.*, 1993), so that an understanding of phylogenetic relations is important in the diagnosis of these species.

Microsporans have not been reported from fossils, but ultrastructural and molecular studies have shown that extant forms may have ancestral character states. This is suggested by their having both eukaryotic and prokaryotic features (Vossbrinck *et al.*, 1987). Their eukaryotic features include the possession of a nucleus, an endo-membrane system and the separation of chromosomes on a mitotic spindle (Canning, 1989). However, like prokaryotes, they have 70S ribosomes with 16S and 23S rRNA (Ishihara and Hayashi, 1968; Curgy *et al.*, 1980). The 16S rRNA is shorter than the typically eukaryotic 18S rRNA sequence and the 5.8S rRNA, which in typical eukaryotes is cleaved from the 28S rRNA to form a separate molecule, is absent, although sequences complementary to the 5.8S rRNA are found in the 23S rRNA (Vossbrinck *et al.*, 1987).

page 8, 1.3.1

The second sentence should read: "Lack of these features are perhaps secondary reductions caused by their parasitic lifestyle (Cavalier Smith, 1987)".
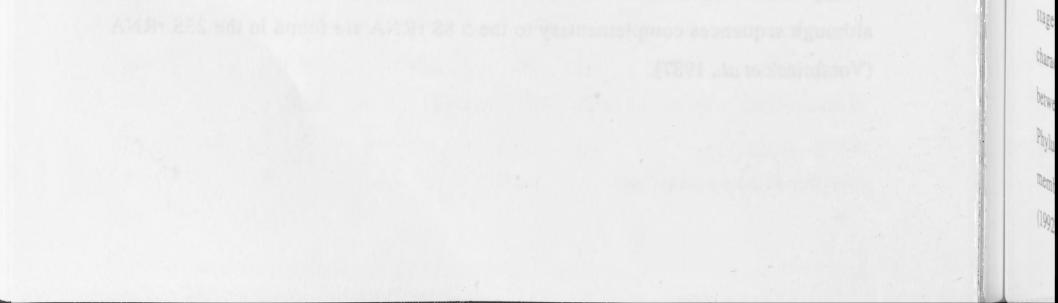
1.3.1 Phenotypic descriptions

Microsporans have no mitochondria, peroxisomes or Golgi dictyostelia. These features are perhaps secondary reductions caused by their parasitic life style (Cavalier Smith, 1987). The following description is reported by Tanada and Kaya (1993). Their life cycle has two phases; merogony and sporulation. Spores, which can survive outside their hosts, have an imperforate spore coat and are usually 3-6μm x 2-4μm, but may be up to 20μm in diameter. The interior of the spore, the sporoplasm, consists of a plasmalemma, cytoplasm, 1 or 2 nuclei, endoplasmic reticulum and ribosomes all enclosed in a coiled polar tube (diameter 0.1μm; length 100-400μm). The spore wall is enzymatically digested in the insect midgut and, as a result, the polar tube everts and penetrates the host gut cells, and the sporoplasm is injected through it into the host cells. Within the host cell, the sporoplasm develops into a rounded meront with one (unikaryon) or two (diplokaryon) nuclei. Meronts divide mitotically and the progeny infect other host cells. At a certain stage each meront develops into a sporont and this has a different cell membrane structure from the meront. Sporogony varies. In microsporans, such as in *Encephalitozoon* and *Nosema spp.*, the first nuclear division is followed by cytokinesis and this produces two potential sporoblasts. In *Vairimorpha necatrix*, cytokinesis does not follow nuclear division, and a pansporoblast or sporophorous vesicle is produced which later buds into unicellular sporoblasts (Larsson, 1986). However, polymorphic forms containing both free spores and spores in sporophorous vesicles have been found in both adult and larval forms of the same host (Canning, 1989). Sporoblasts develop into spores by the differentiation of the polar tube, the cell organelles and the spore coat. Spores are released when the host dies (Tanada and Kaya, 1993).

Even though phenotypic characters, such as morphology and ultrastructure, may be needed to confirm parasite identity, these vary greatly in the different life cycle stages of both the parasite and its host, so that any taxonomies based on these characters are too incomplete for identification and not sensitive enough to distinguish between species of the same genus (Vossbrinck *et al.*, 1993). Classification of the Phylum Microspora is therefore based on differences in the chromosome cycles of its members. Classification of the genera *Nosema spp.* is described by Sprague *et al.* (1992) as follows:

Phylum Microspora Sprague, 1969

Class Dihaplophasea (nuclei paired as a diplokaryon during part of the life cycle).

Order Dissociodihaplophasea (haplosis occurs by nuclear dissociation resulting in unpaired nuclei, spores are homosporous or heterosporous).

Super Family Nosematoidea Labbe 1899. (bi-nucleate homospores dissociate after the sporoplasm invades a new host individual, cycle of gamete production ends with plasmogamy and nuclear dissociation).

Family Nosematidae Labbe 1899. (all stages occur in host cell hyaloplasm).

Genus *Nosema*

1.3.2 16S rRNA sequence analysis

The evolutionary position of the microsporans was first indicated when a phylogenetic tree was calculated from a set of 16S rRNA sequences which included that of the microsporan *Vairimorpha necatrix*, a parasite of moths. These data indicated that the microsporans are a sister group to all other eukaryotes (Vossbrinck *et al.*, 1987).

In the past two years, two complete 16S rRNA sequences, those of *Nosema apis* and *Vavraia oncoperae* (isolated from hepialid lepidopterans) (Malone *et al.*, 1994) and four partial sequences, three isolated from human beings infected with HIV, namely *Encephalitozoon cuniculi* (Hartskeel *et al.*, 1993), a new species of microsporan (Vossbrinck *et al.*, 1993) and *Enterocytozoon bienusi* (Zhu *et al.*, 1993, submitted) and one from fish namely, *Ichthosporidium giganteum* (Vossbrinck et al., 1993) have been recovered. The first two have not been used to calculate phylogenetic trees, but Malone *et al.* (1994) found several parallels between phenotypic and genotypic data which could be taxonomically useful. *Nosema apis* and *Vairimorpha necatrix* have diplokarya in some life cycle stages which places them both in the Class Dihaplophasea, their 16S rRNAs are of similar lengths (1212 and 1209 nucleotides respectively) and they have 94% sequence identity, whereas *Vavraia oncoperae*, *Encephalitozoon cuniculi* and *Enterocytozoon bienusi* do not have diplokarya, and the first two have 16S rRNA of 1322 and 1299 nucleotides respectively and only 66.4% sequence identity. Vossbrinck *et al.* (1993) have calculated a molecular phylogeny based on DNA sequences coding for a segment spanning the LSU and SSU from the new microsporan, *Vairimorpha necatrix, V. lymantriae, Encephalitozoon cuniculi,* and *Ichthysporidium giganteum.* The results obtained support the close relationship of the

page 10, 1.4

The first sentence should read: ".....our understanding of the intra- and inter-taxonomic (between and within respectively) relationships of the Apicomplexa and Microspora".

two *Vairimorpha spp.* and the placement of the new microsporan in the genus *Encephalitozoon* as *E. hellem*. A separate work which compared the LSU rRNA sequences of several *Nosema* and *Vairimorpha spp.* (Baker *et al.*, 1994) supports their close relationship (Malone *et al.*, 1994) and places the *Nosema spp.*, which are composed of several apparently unrelated groups, as a sister group to *Vairimorpha spp.*

1.4 Conclusions and Aims of this thesis.

The available phylogenies based on SSU rRNA sequences and morphological data indicate that large gaps exist in our understanding of the intra and inter-taxonomic relationships of the Apicomplexa and Microspora. The conclusion that the primitive apicomplexan *Perkinsus sp.* forms a sister group to the other apicomplexans is deduced from a single SSU rRNA sequence from that taxon, as too is the conclusion that the microsporan *Vairimorpha necatrix* forms a sister group to all other eukaryotes. These would be more certainly accepted if confirmed by additional data from other closely related species. This could be done by constructing phylogenetic trees using the original SSU rRNA sequences from *Vairimorpha necatrix*, *Perkinsus sp.* and other apicomplexans and new SSU rRNA sequences from other microsporans and apicomplexans. Thus, in the work reported here, I decided:

1. To examine some aspects of the morphology and life cycle of a gregarine parasite from *Caledia captiva*, which, as it has not been formally described, is called "Gregarina caledia" in this thesis; see Materials and Methods.

2. To resolve the relative evolutionary positions of three previously undescribed apicomplexans "Gregarina caledia", "Gregarina chortoicetes" and "Pseudomonocystis lepidiota" and the microsporans *Nosema apis* and "Nosema vespula" using their SSU rRNA sequences.

Figure 1
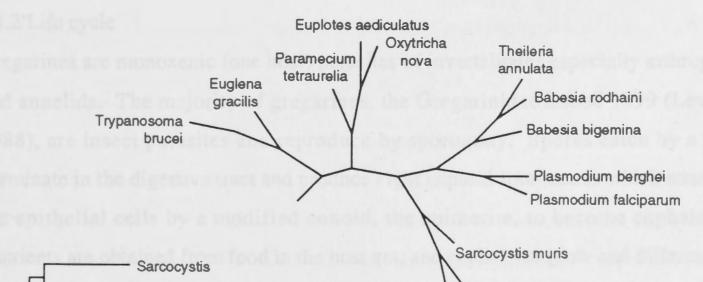
A selection of phylogenetic trees constructed by parsimony and distance methods from 18S rRNA sequences from apicomplexians, dinoflagellates, ciliophorans and other protoctistans. These trees have been modified by the removal of multicellular taxa and shortening of the branch lengths relative to those in the original trees.
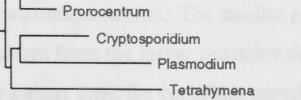
Trees made by Parsimony methods

Modified from Gagnon et al. (1993)

Modified from Ellis et al. (1992)

Modified from Barta et al. (1991)

Trees made by Distance methods

Modified from Johnson et al. (1990)

Modified from Goggin
and Barker (1993)

CHAPTER 2

A STUDY OF A GREGARINE PARASITE
OF THE GRASSHOPPER *Caledia captiva*

2.1 INTRODUCTION

2.1.1 Historical background

Sporozoites of single-celled organisms are found in the guts of all major groups of insects and form relationships with their hosts which range from that of parasite to commensal. Members of one taxon, the septate or cephaline gregarines of which there are 500 described species (Levine, 1988) were first reported in 1828 by Dufour, in the guts of coleopterous insects (Sprague, 1941). More recent investigations have described gregarines from a variety of commercially or medically important insect hosts and laboratory populations of orthopterans and coleopterans (Laird, 1959). This has sparked a renewed interest in their study as a potential source of parasites with which to control mosquitoes and other vectors (Walsh and Olsen, 1976; Canning, 1982; Beier and Craig, 1985; Levine, 1988).

2.1.2 Life cycle

Gregarines are monoxenic (one host) parasites of invertebrates especially arthropods and annelids. The majority of gregarines, the Gregarinidae Labbé 1899 (Levine, 1988), are insect parasites and reproduce by sporogony. Spores eaten by a host germinate in the digestive tract and produce eight haploid sporozoites which attach to the epithelial cells by a modified conoid, the epimerite, to become cephalonts. Nutrients are obtained from food in the host gut, and cephalonts grow and differentiate into two compartments. The smaller anterior protomerite is separated by a septum of ectoplasm from the larger posterior deutomerite which contains the single nucleus. After a short time, the cephalont detaches from the host intestine and becomes a free-living, motile sporont. Sexual maturity is characterised by the fusion of two sporonts in an association called syzygy in which the deutomerite of the anterior sporont (the

page 12, 2.2

After the third sentence, add the following sentence: "The initial aim of studying variability between chromosomal races was not achieved, and the study was restricted to populations in New South Wales, Australia".

primite) fuses with the protomerite of the posterior sporont (the satellite). After an unknown length of time, the pair rotates to form a ball which becomes encysted as a gametocyst and passes out in host faeces. Within the gametocyst, the nucleus of each sporont becomes a gamont which undergoes gametogenesis and a similar number of male and female iso-gametes is formed. These pair in syngamy and divide to form thousands of diploid spores which are expelled in chains through everted spore ducts which form on the gametocyst surface. Each spore undergoes meiosis to produce 8 haploid sporozoites, which if eaten by new hosts, germinate into potential cephalonts in the gut (Kudo, 1954; Levine, 1988; Vivier and Desportes, 1989).

Vavra (1969) first proposed that gregarines were species specific, however the single species *Gregarina garnhami*, appears to infect two species of locust, *Locusta migratoria* and *Schistocerca gregaria* (Vivier and Desportes, 1989).

2.1.3 Pathogenicity

Gregarine infections cause loss of appetite, general sluggishness and, occasionally, a colour change of the insect host, but most are only detected by histological examinations. In the locust host *Schistocerca gregaria* Forsk, the gregarine parasite *Gregarina garnhami* has been shown to destroy large areas of caecal epithelium, however damaged epithelial cells are either continually replaced, or may survive with little regeneration (McCray *et al.*, 1970). In the midgut, sporonts cause little damage to the gut epithelium but interfere with the absorption of food material by forming localized barriers between the digested food material and the epithelial cells (Canning, 1956).

2.2 AIM

The initial reason for studying gregarines was to look for variability in those which inhabited the Australian grasshopper *Caledia captiva*. This insect exists in a coastal band in northern and eastern Australia as two parapatrically associated chromosomal races which, in southeast Queensland, are separated by a hybrid zone. It was thought that these parasites may have exhibited co-evolution with their parental and hybrid hosts (Moran, 1979). A study of the morphology and life cycle was made of the

13

gregarine-like organisms recovered from local populations of *Caledia captiva* at Uriarra and Araluen in New South Wales.

The aim of this study was to study a gregarine parasite "Gregarina caledia" obtained from a wild population of the grasshopper *Caledia captiva* Orthoptera; Acrididae (Shaw *et al.*, 1976). Stages of the life cycle of "Gregarina caledia" in laboratory-reared *Caledia captiva* were described.

2.3 MATERIALS AND METHODS

2.3.1 Collection of grasshopper egg-pods

Grasshoppers captured in the field (Section 2.4) were placed in cages measuring 250 x 300 x 280 mm^3, constructed of wood and screen wire with a 350 ml plastic pot containing sand in their floor, for oviposition. Damp, autoclaved sand in the container was replaced every day, and the egg pods collected by immersing each container in water so that they floated out of the sand. Pods were surface sterilised in 3-4% sodium hypochlorite solution for 5 minutes and washed 5 times in water, then transferred to new containers of damp autoclaved sand, covered loosely and incubated at 30°C and 90% relative humidity. The hatchlings, which usually appeared 28 to 31 days after the pod was laid, were caged separately (see below). Five hatchlings were initially placed in each cage.

2.3.2 Rearing of infection-free grasshoppers

Infection-free grasshoppers were reared in two types of infection-proof cages. In the initial experiments, nymphs or grasshoppers were confined separately in 250 ml clear plastic containers, each stoppered with a foam plastic bung. These were kept in an incubator at 30°C. Humidity and food were provided by a daily supply of glasshouse-grown, fresh wheat shoots. The insects were kept from 7.00am to 11.00pm, at 30°C with lights (50 watt silver backed bulbs (Phillips)) and from 11.00pm to 7.00 am, at 20°C without lights. These cages were used in the second study reported in this chapter (Section 2.5).

Nymphs and grasshoppers were monitored daily for mortality, moulting and gametocyst production, so that re-infection by spores from dehiscing gametocysts was avoided. The gametocysts usually dehisced two or three days after faeces were voided

by the insects. Every day the grass was removed and the nymph was tipped into a new container, and examined under a dissecting microscope. These cages proved to be ideal for rearing and handling single nymphs and grasshoppers; each hatchling survived and developed into an adult in about 60 days. It was however difficult to check fully for gametocysts. Faeces, which were sometimes transferred along with grasshoppers into a new cage, sometimes carried gametocysts. These ultimately dehisced, grasshoppers became reinfected, and new gametocysts appeared after 10 days. Therefore, before each experiment, faeces from grasshoppers were monitored daily for 15 days and any infected grasshoppers were discarded.

Larger cages for up to 8 grasshoppers were also used for the experiments described in Chapter 3 to keep gregarine-infected *Caledia captiva* and "Chortoicetes terminifera". Genomic DNA of these parasites was extracted for use in experiments described in Chapters 3, 4 and 5.

2.3.3 Inoculation of infection-free grasshoppers.

Spores were suspended in sterile distilled water and transferred with a Gilson pipette from the slides or Petri dish lids, to a Microfuge tube. Each grasshopper was held between the thumb and the forefinger and squeezed gently until it regurgitated a drop of undigested matter, then a 10µl drop of water containing the spores was transferred to the mouth and mouthparts using a 25µl glass capillary (Microcap; Drummond Scientific Instruments).

2.3.4 Collection of the sporont and gametocyst stages of the parasites.

Grasshoppers were anaesthetised in gaseous CO_2. The gut was removed by cutting off the last three abdominal segments, grasping the head with forceps and pulling it from the body (Harry, 1969). Each intestine was transferred to sterile insect saline (0.7% NaCl, 0.02% $CaCl_2$) or to sterile distilled water and teased open under a dissecting microscope. The liberated sporonts and gametocysts were collected and washed several times in sterile distilled water. Gametocysts were recovered from faeces that had been softened by soaking for a few minutes in distilled water, by teasing them away from the faeces with forceps.

2.3.5 Storage of sporonts and gametocysts

Sporonts and gametocysts were collected in sterile water and kept at -70°C in Microfuge tubes (Eppendorf).

Gametocysts to be used as a source of spores for infecting grasshoppers were kept on labelled glass slides in sterile plastic disposable Petri dishes. Stacks of these were placed in tightly lidded plastic lunch boxes with moist blotting paper to provide humidity, while minimising the possibility of fungal infection. Gametocysts were kept at 26°C, and examined daily. Spore chains from dehisced gametocysts formed characteristic patterns on the slides or on the underside of the Petri dish lids.

2.3.6 Measurement of sporonts and gametocysts of the parasite

All sporonts and gametocysts were measured at magnifications of x125 and x1250 using a light microscope (Zeiss) fitted with a calibrated micrometer eyepiece.

2.3.7 Photography

Cephalont and sporont stages of the parasites were photographed either using reflected light with Kodacolor Gold 100 (GA135-24) in a camera mounted above a macroscope (Makroscope 400, Wild Leitz), or under transmitted light using Kodak Ektachrome (EPY 135-36) film in a camera mounted above a light microscope (Zeiss). Spores were photographed in reflected light under oil immersion at a maximum magnification of x1200.

2.3.8 Formulae used for measuring developing cephalonts

Bush (1928); Allegre (1948); Canning (1956) and Laird (1959) found that the ratios of the sizes of different parts of sporonts and gametocysts were sufficiently constant to be useful for distinguishing between taxa. The volumes of developing cephalonts were calculated using the formulae:

Volume of a sphere = $4/3\pi r^3$

Volume of a cylinder = $\pi r^2 h$

Volume of an ellipsoid = $4/3\pi a^2 b$, where a and b are the smaller and larger radii respectively.

2.4   A STUDY OF "Gregarina caledia" OBTAINED FROM A WILD POPULATION OF *Caledia captiva*

2.4.1 Introduction

The vegetative and reproductive stages of "Gregarina caledia" were observed and measured, and the parasite load and site of infection was investigated in *Caledia captiva*.

2.4.2 Materials and methods

In April 1990, 44 grasshoppers from the South East Australian taxon of *Caledia captiva* and at various stages of maturity were collected at Araluen, NSW. When caught, 7 grasshoppers (6 males and 1 female) were chosen at random, killed immediately and examined for parasites. None appeared to be infected with gregarines. The remaining grasshoppers were placed in wire cages in the humidifying cabinet and stressed by being given a minimum of food in order to activate latent infections, as suggested by Appel (1983) who reported a five fold increase in the number of gametocysts of *Gregarina rhyparobiae* when its cockroach hosts were starved for 12 days.

2.4.3 Results

After two days of starvation, 9 grasshoppers (7 females and 2 males) were dissected. In all of these, the mid guts and most hind guts were packed with vegetative sporonts of "Gregarina caledia" ranging in numbers from 17 to 330, with a mean of 137 per grasshopper. Most sporonts were not syzygous, and were 180 to 240μm long and 60 to 70μm wide (mean 222μm x 65μm). Each sporont was cream to pale-yellow coloured and barrel-shaped, and had a smaller anterior protomerite and a larger posterior deutomerite separated by a septum of ectoplasm which appeared to be confluent with the ectoplasm which surrounded the entire organism. The protomerite was roughly spherical and at its anterior end had a flattened scar which marked the position of the earlier epimerite. The deutomerite was bacilliform, wider than the protomerite and rounded at the posterior end. The endoplasm was granular especially in the centre of the organism. The nucleus was refractile, filled with granules of variable size and its position varied. Ratios of protomerite length : deutomerite length

pages 17, 20, 21 and in the legends corresponding to plates 2.0a-2.0e and 2.0f-2.0i:

Replace Figure 2.0 by Figure 2. Thus Figure 2a-2e and 2f-2i.

ranged from 1 : 3.0 to 1 : 3.5. The primite and satellite of most syzygous pairs were of a similar length, but a few had a primite that was slightly larger than the satellite, their mean total lengths being 260μm and 255μm respectively (Figure 2.0a).

One to three spherical gametocysts, each surrounded by a gelatinous layer varying in thickness from 18μm to 144μm were found in the hind gut of each of five grasshoppers infected ; eight in all. Of these, four were in the stage where two homogenous gamonts could still be distinguished (Figure 2.0b), and four had progressed to the stage where a single mass of tissue filled the gametocyst coat. Gametocyst diameters, minus the gelatinous layer, ranged from 168μm to 360μm with a mean of 259μm. Gametocysts which had been allowed to develop on glass microscope slides developed from the two gamont stage until the orange-coloured basal discs of inverted spore ducts appeared, approximately 3 days later (Figure 2.0c). The number of basal discs on the side of the gametocyst facing the microscope lens, varied from 5 to 10, and their diameter measured 18μm to 24μm. The surface of each gametocyst remained unaltered for the next 2 days while the spore ducts grew toward the inner mass (Canning, 1956), after which they everted through the gametocyst wall. Spore chains consisting of ellipsoidal spores embedded in a sticky matrix (Figure 2.0d), were extruded from the spore ducts and formed a sinous pattern of fine threads which stuck to the glass slides and to the underside of the Petri dish lid. Diameters of spores measured 12μm x 6μm which included a thick surrounding coat and which according to Harry (1965), formed approximately two thirds of the total mass (Figure 2.0e).

After six days of starvation, 4 females were dissected. All were infected, and the number of very small sporonts in the midgut ranged from 50 to 700, with a mean of 285 per host. In the most heavily infected insects, sporonts were visible through the walls of the midgut, and in one instance had pushed through the junction between the foregut and the crop and protruded into the haemocoele. However in these insects, the small size of the sporonts and the lack of syzygy suggested that there had been secondary infections, perhaps from dehisced gametocysts trapped on the wire of the cage. No gametocysts were detected *in vivo*. Seven days later, a single female was dissected. Its crop contained 450 sporonts, midgut 1250 sporonts and 4 gametocysts and hindgut 53 sporonts and 3 gametocysts. These numbers are fewer than those reported by Harry (1970) who found in 18 laboratory reared desert locusts

(*Schistocerca gregaria*) infected with the gregarine *Gregarina garnhami*, that 10 had >10,000 gregarines, 1 had >5000 and 7 had >1000.

44% of the individuals in a field population of *Caledia captiva* collected at Araluen, NSW, were infected, mostly in the midgut. After re-infection in the laboratory, the sporonts were also found in the crop, hindgut and haemocoele, but without any obvious pathological effects on their hosts.

2.4.4 Discussion

The life cycle of "Gregarina caledia" appears to be similar to that of other septate gregarines of arthropods (Kudo, 1954; Canning, 1956). The general body shape, the size of sporonts, gametocysts and basal discs, and the timing of various stages of development are summarised in Table 2.1 below. It is unfortunate that the sizes of gametes and the exact number of basal discs in "Gregarina caledia" are not known, but they are probably variable (Allegre, 1948; Harry, 1970). This shows them to be similar in nearly all aspects except the size of the spores, to those described by Canning (1956) and Harry (1965) for *Gregarina garnhami* which is parasitic in the locust *Schistocerca gregaria* Forsk. The variable size of the unassociated sporonts was possibly related to the time of infection and germination of spores, or genetic variation in the host or the parasite, factors which were not determined in these experiments. Similar ratios of protomerite length : deutomerite length of the gregarine from *Caledia captiva* might be taken to suggest that these gregarines were related, however, similar ratios between these and those found for *Gregarina garnhami* indicate that ratios of measurement and absolute dimensions have no taxonomic significance by themselves, as has already been realised by others (Bush, 1928; Allegre 1948).

2.5 A STUDY OF THE LIFE CYCLE OF "Gregarina caledia" IN THE GRASSHOPPER *Caledia captiva*

2.5.1 Introduction

The field and laboratory studies of *Caledia captiva* nymphs artificially infected with spores of "Gregarina caledia" reported above (Section 2.4), did not define the developmental stages of the parasite *in vivo* nor explain the size polymorphism of

sporonts. These stages were investigated in an experiment by observing both fresh tissue and fixed sections collected from the time of infection of the the grasshopper, until about 12 days later when gametocysts are extruded on its faeces. Dehiscence of these gametocysts *in vitro* was recorded and described below.

2.5.2 Materials and methods

Twenty four third instar infection-free nymphs were infected with gregarine spores on the same day. These nymphs had been hatched in the laboratory from surface-sterilised pods obtained from adult females collected at Araluen, NSW. The spores came from dehisced gametocysts dissected from the gut of grasshoppers captured at Araluen and passaged in laboratory-reared nymphs from Uriarra, NSW. One nymph was kept as an uninfected control. Each nymph was separately confined in a 250 ml disposable clear plastic container and kept in an incubator at $30^{\circ}C$. Each day for 12 days 2 nymphs were anaesthetised in CO_2, and from each, the gut was removed. One gut was placed in sterile insect saline and immediately checked for the presence of any parasites, and these were photographed. The other gut was fixed immediately in 10% formal saline until needed. After 12 days when all the nymphs had been dissected, the 12 fixed guts were washed in 0.85% saline, and dehydrated in an alcohol series (50%, 70%, 95%, 100%). The tissue was cleared in xylene, infiltrated with and embedded in wax. Serial sections 10 to 15μm thick were cut using a microtome (Leitz), stained with Ehrlich's haematoxylin (Carleton, 1980) and mounted in Histomount (National Diagnostics). Faeces were monitored daily for gametocysts.

2.5.3 Results

Of the 24 infected nymphs, 2 died of unknown causes, 2 had gametocysts present and none were found in 6. To complete the daily sequence, 10 more nymphs were infected with spores from the original source and dissected, as above, on 1 to 6 days after infection. All of this group became infected and none died. The uninfected nymph remained free of infection.

a. Stages in the life cycle (Figures 2.1 and 2.2)

No parasites were seen in histological sections from the guts collected in first two days after infection, but sporozoites are too small to be seen easily in sections thicker than 5μm (Harry, 1970). Spores of "Gregarina caledia" were ellipsoidal, with shorter and longer radii of 3μm and 6μm respectively, so their volume is 220μm^3 of which about two thirds (147μm^3) was the spore coat (Harry, 1970). Thus each of the 8 sporozoites in the germinal area of the spore had a volume of about 9μm^3 (Figure 2.1). The spore coats of ingested spores are thought to be removed by enzymes in the foregut of the host, and hence released the 8 cephalonts (Sprague, 1941). Samples taken from the midgut after two days showed gregarines in different stages of development and volume which had increased over a range of 12 fold to 370 fold range since germination (Figure 2.1). One smaller, ellipsoidal gregarine with a differentiated epimerite, was not attached to a host cell and may have been the earliest stage in development of the sporozoite. The larger gregarines had longer axial ratios, were attached to the brush border of the intestine and had each differentiated into a protomerite and deutomerite by a pale septum of ectoplasm. These large differences in volume may reflect genetic polymorphism as all were infected on the same day but, as the parasites were growing exponentially, the differences may have resulted from small differences in the actual time of germination of the sporozoites. Cephalonts sampled on the third day had a greater axial ratio and had increased in volume by 550 fold since infection (Figure 2.1). Samples taken on day 4, showed that several changes had occurred. The cephalont had become a free swimming sporont by detaching from its epimerite and leaving as evidence, a small scar on the anterior surface of the protomerite in the region of the ectoplasm (Figure 2.0f). The sporont was now cylindrical and had a clear outer ectoplasm surrounding a granular endoplasm. The shorter anterior protomerite was separated from a larger posterior deutomerite, by a septum of ectoplasm which caused a constriction between the two compartments. There was a nucleus in the centre of the deutomerite and the volume of the total organism (the sum of the elliptical protomerite and the cylindrical deutomerite) showed an increase in volume of 7.5 fold from the cephalonts of the previous day (Figure 2.1). In samples taken on days 7 to 11, it was difficult to measure the sporonts accurately because of their alignment in the gut, however they showed daily increases in volume and on the eighth day one sample was 10^7 fold larger than a spore nucleus (Figure 2.1).

On day 7 after infection, gametocysts appeared for the first time on faeces collected from both nymphs, (Figure 2.0g). One nymph produced 11 gametocysts, the other 3, and these were measured. Diameters ranged from 228μm to 312μm with a mean of 270μm and standard deviation of 59.4. All gametocysts had orange-coloured basal disc patches, which may have been present before extrusion and which from the earlier studies (Section 2.4.3) indicated that the two gamont stage must have occurred on about day 5, and that syzygy had probably occurred on day 4, some time after detachment of the sporont from the host epithelium. Although I inspected many hundreds of dissections, I did not see the early formation of the gametocyst as described by both Sprague (1941) and Canning (1956). I found individuals of different sizes in pairs, but usually such partners were of similar size, with the primite being only slightly larger than the satellite. As there is a variety of different sized individuals available for mating, the mechanism by which similarly sized individuals are attracted to each other is not known. In one unusual pairing, the primite had two smaller satellites (Figure 2.0h). Samples taken on days 10 and 11 after infection, had sporonts between the peritrophic membrane of the gut and the food mass, possibly interfering with food absorption (Canning, 1956) (Figure 2.0i). Their volumes were variable and suggested that the volume of the individual on the eighth day may have been unusually large (Figure 2.1). A relatively small gametocyst (diameter 42 μm) sampled from the mid-gut caecae on the tenth day suggests that syzygy may occur late and that there is variation in the onset of sexual maturity. Sporonts collected over 12 days after infection had axial ratios of approximately 3 : 1 and that this was constant after day 4. The lengths of the protomerite : deutomerite were approximately 1 : 4 indicating that the deutomerite had increased in length, as had been reported by other workers (Canning, 1956).

b. Gametocysts

Gametocysts recovered on day 5 from the midgut of one of the fresh nymphs had two sporonts with no internal septa, so that they appeared as two dark masses separated by a clear zone (Figure 2.0b). These were a pale yellow to gold in colour and showed a variation in diameter, with a mean of 279μm. The thickness of the gelatinous layer around each gametocyst varied from 24μ to 100μ, which Canning (1956) suggests may depend on the duration of the rotation period.

Gametocysts collected from faeces were each kept on separate glass microscope slides in humidity chambers at 26°C. Each was observed daily from its first appearance on the faeces as a two gamont stage on day 5 to its dehiscence 5 days later (day 10). It appeared that development *in vitro* and *in vivo* were similar.

After one day (day 6) the gametocyst had a very dark interior, but around the periphery, it was translucent, and gametes could be seen to be moving. After two days (day 7) orange coloured patches, which are the basal discs of spore ducts, formed over the surface of the cyst. The total number was not counted, but 4-9 were on the side facing the lens. There is disageement about whether the number of spore ducts is proportional to the size of the gametocyst (Sprague, 1941), is species specific (Ellis, 1913) or is variable (Allegre, 1941; Harry, 1970). Measuring approximately 24μm in diameter, these patches were the basal discs of inverted spore tubes which had developed towards the centre of the gametocyst (Sprague, 1941; Canning, 1956,). After three days (day 8), the inner mass of gametes and spores appeared to shrink to approximately half its former volume, and on the fifth day (day 9) each spore duct began to evert pushing through the basal disc to form an exterior tube. After five days (day 10) the gametocyst dehisced. Spore chains were ejected with force and some stuck in a sinous pattern on the underside of the Petri dish lid at least one centimetre away, which is approximately 30 times the mean diameter of the gametocyst itself. When gametocysts dehisced in the cage, the spore chains formed large white coils each containing thousands of spores. Sprague (1941), estimated that spore chains from cysts of *Gregarina blattarum* were 87 millimetres in length and contained approximately 10,000 spores.

Some gametocysts were accidently squashed or, if conditions were too dry, burst prematurely to release cytoplasm containing uniformly sized, spherical bodies which measured 3μm in diameter and were probably gametes (Figure 2.0e). Droplets resembling oil globules were found beside many dehisced gametocysts. Sprague (1941) suggested that these may be a lubricant to minimise friction between the spore duct and the spores, and also act as an anti-wetting agent to protect the spores. However, this is inconsistent with the observation that spores stick to surfaces making it difficult to remove them from a slide with a pipette, which in the environment would prevent them from being washed away.

page 23, add a new section

"2.6.1a Mixed gregarine infection in field captured *Caledia captiva*.

Despite the proposal that gregarines are host specific (Vavra, 1969), it is feasible that the original field captured grasshoppers harboured mixed infections of "Gregarina caledia". This possibility was not tested as it was not technically possible to infect a grasshopper with a single spore".

page 23, 2.6.2

The third sentence should begin "During the first 8 days, the volume of the parasite increased in size 10^7 fold......".

2.6 GENERAL DISCUSSION

Although I attempted to observe and record day by day changes in the parasite, I was only sampling the process and my daily random selection of nymphs for their associated parasites is obviously an imperfect way to study the process. Therefore, my results can only be viewed as general indicators of the changes.

2.6.1 Infection efficiency

In laboratory tests, 75% of grasshoppers were infected compared to 44% of those observed in field collections. It is possible that the large numbers of parasites observed in laboratory populations of *Caledia captiva* simply resulted from multiple re-infections. This could occur during the transfer of grasshoppers to new containers, by the simultaneous transfer of gametocysts on faecal pellets, which then produced spore chains inside the cages.

2.6.2 Life cycle of "Gregarina caledia"

The two most prominent features of these results were the large growth rate of the parasite and variation in size between days 2 and 11 of development. Although few parasites were observed and statistical analyses were not done, a general picture of growth and development of "Gregarina caledia" in *Caledia captiva* was established. In each of the first 8 days, the volume of the parasite increased in size 10^7 fold from a spore of approximately $9\mu m^3$, to a sexually mature sporont with a maximum volume of $10^8\mu m^3$. The large growth rate, coupled with differences in germination and attachment times of sporozoites, could have accounted for the size variation. Small differences in germination and attachment times could have caused the large size differences seen up to day 8. However if all sporonts were genetically similar, it is possible that by days 10 or 11 all sporonts would have attained a similar size. The size range of $10^6\mu m^3$ to $10^7\mu m^3$ on days 10 and 11 is probably not great enough to deduce that the size polymorphism results from genetic variation, or by differences in germination and attachment times. The large unassociated sporonts seen on days 8 to 12, may not have ever entered syzygy, and in turn, may have become a loss to parasite numbers and their evolution.

page 24, 2.6.3. The ninth line ending "(this study)". Add a new sentence: "These ratios show that from the time of differentiation of a sexually immature sporont from a cephalont to a sexually mature sporont and finally to an un-associated sporont, all the gregarines follow a pattern of allometric growth whereby their form exhibits a constant ratio (Thompson, 1968) between the protomerite and the total length."

2.6.3 Affinities with other gregarines

Ratios of length : width and of protomerite : deutomerite was almost the same in samples collected between day 2 and day 11 and indicated that the shape of this species did not change as it grew in size. Furthermore the ratio of protomerite : total length of the sporont is similar to that of other gregarines from different hosts such as *Gregarina garnhami* from *Schistocerca gregaria* which has a ratio of 1:4 (Harry, 1965), *Gregarina oblonga* from *Locusta migratoria*, ratio, 1:4 (Canning, 1956), an unknown *Gregarina spp.* from *Anacrydium aegyptium*, ratio, 1:4 (Canning, 1956) and *Gregarina caledia* from *Caledia captiva*, ratio, 1:3.6 to 1:4.2 (this study). This confirms that these ratios are of little value for taxonomic comparisons.

2.7 CONCLUSIONS and FUTURE WORK

The form, size and timing of each life cycle stage of "Gregarina caledia" found in the grasshopper *Caledia captiva* appear to be very similar to those of *Gregarina garnhami* which parasitises the locust *Schistocerca gregaria* (Canning, 1956; Harry, 1965). To determine whether the two are related, their 18S rRNA sequences could be compared. To resolve the large size variation in the sporonts of "Gregarina caledia", it may be possible to separate the large and small variants and to examine genomic DNA from groups of these by RAPD (Randomly Amplified Polymorphic DNA) PCR (Polymerase Chain Reaction) (Williams *et al.,* 1993).

The reason for the size polymorphism in individual vegetative forms cannot be determined from these observations, but it must be influenced by genetic variation, environmental differences and in germination or attachment time of these.
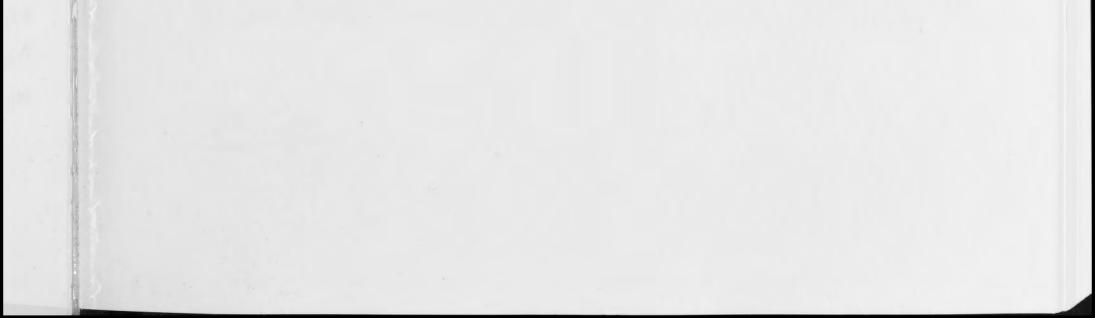
Figure 2.0a

Two sporonts of "Gregarina caledia" in syzygy. The primite (PR) is usually larger than the satellite (S). The protomerite (P), deutomerite (D) and nucleus (N) are well differentiated. Scale bar = 90μm

Figure 2.0b

Gametocysts of "Gregarina caledia" showing the two gamonts within the gametocyst wall surrounded by a gelatinous layer. Scale bar = 160μm

Figure 2.0c
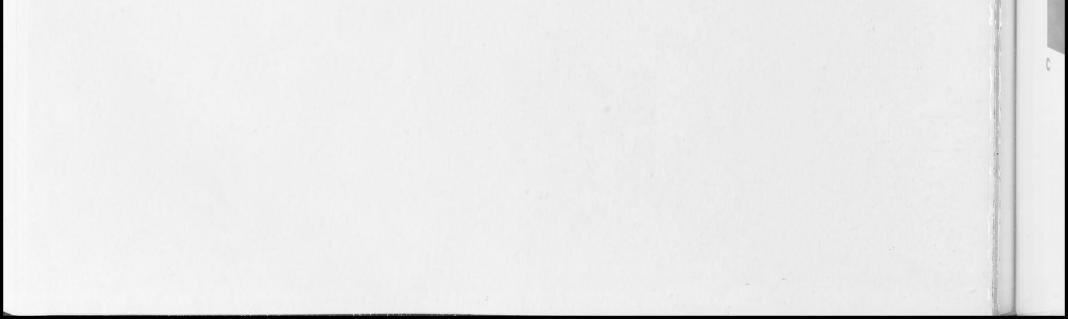
Gametocysts of "Gregarina caledia" showing orange patches (OP) that cover the inverted spore ducts. Scale bar = 125μm
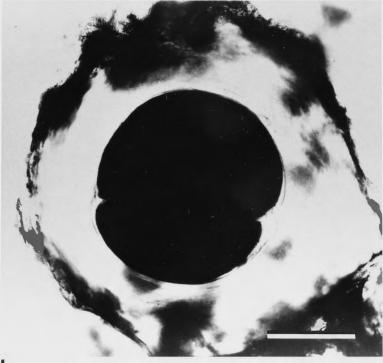
Figure 2.0d
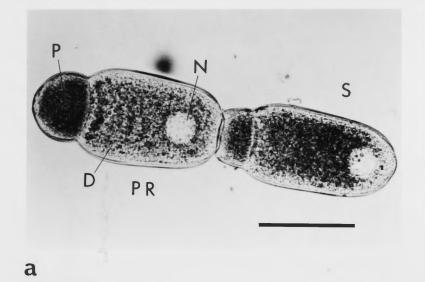
Dehisced gametocyst showing the extruded spore chain (arrow). The everted spore duct is not visible. Scale bar = 125μm

Figure 2.0e

Spores (SP) forming a spore-chain and gametes (G). Scale bar = 12μm
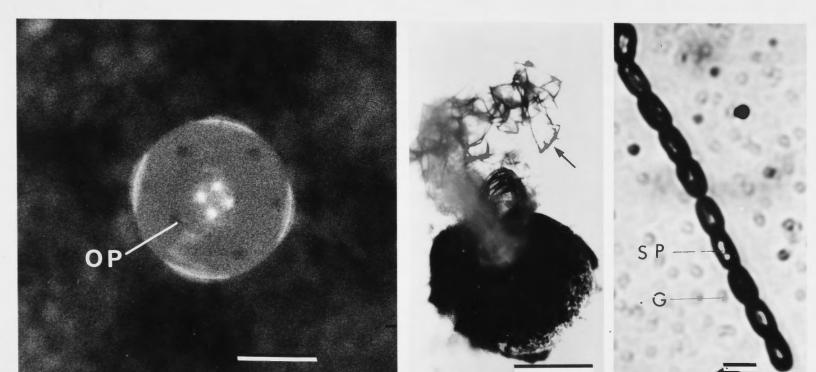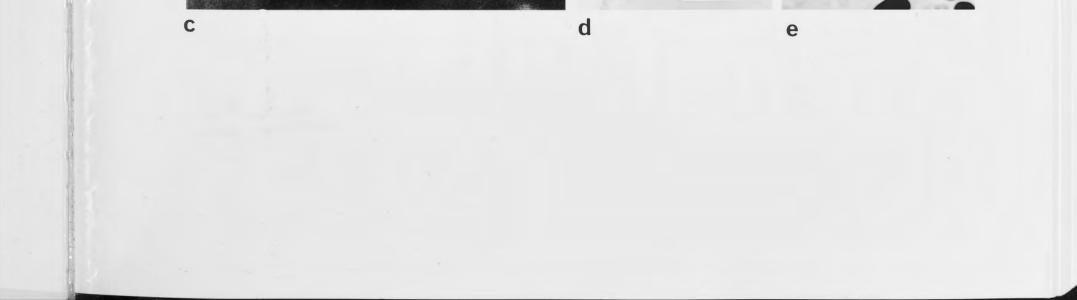
Table 2.1 Column 1, Row 2

The following two lines should read:

"Sporonts: n = 100   (l) 180μm to 264μm, mean 210μm......."

"gametocysts: n = 45   (d) 168μm to 360μm, mean 259μm......."

Sizes of free living and parasitic stages
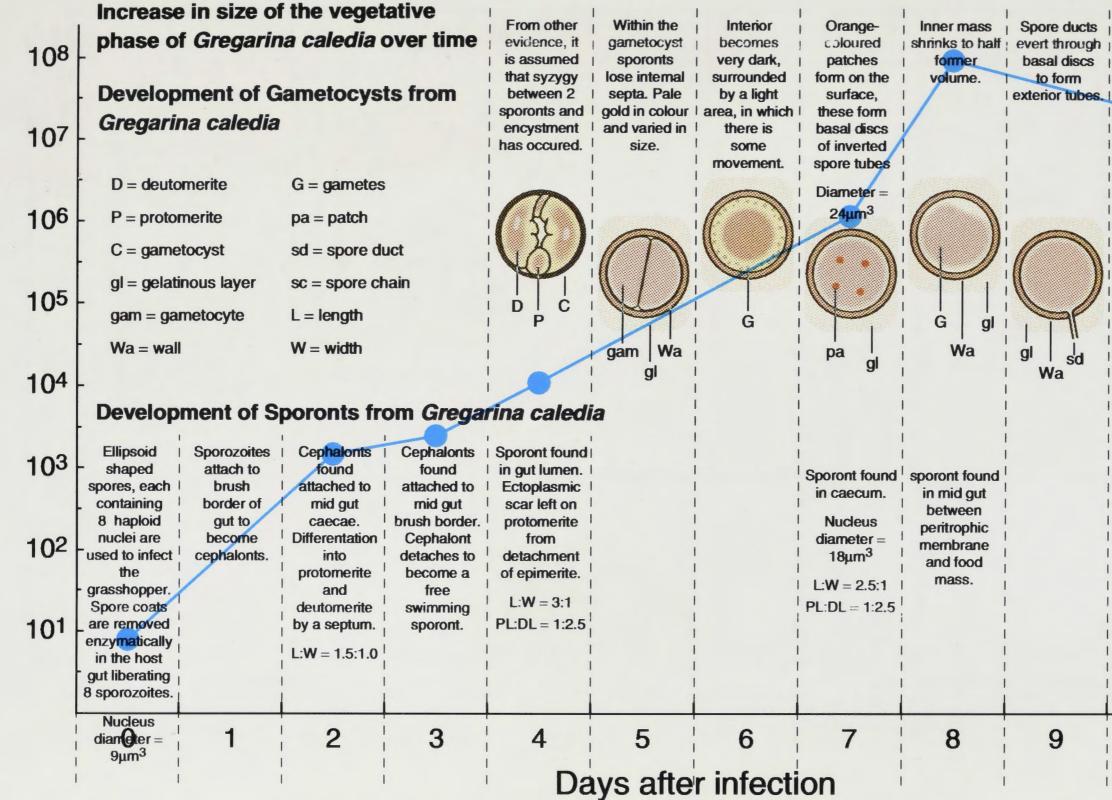
Stages	"Gregarina caledia" in a natural population of *Caledia captiva*	*Gregarina garnhami* in a laboratory population of *Schistocerca gregaria* (Canning, 1956).
sporont: (l)	180μm to 264μm, mean 210μm	up to 554μm, mostly 250μm
gametocyst:(d)	168μm to 360μm, mean 259μm	114μm to 470μm,
basal discs:		
eg. 1. number	5 to 10	8
gametocyst (d)	unknown	unknown
eg. 2. number	18 - 24	20
gametocyst (d)	250-400μm	250μm
eg. 3. number	unknown	35
gametocyst (d)	unknown	>500μm
(dt)	3 days	3 days
spores: (l x d)	12 x 6	6.5 x 4 to 7 x 4
gametes: (d)	3.0	3.5
ratios of sporont parts:	LP : LD = 1 : 2.5 to 1 : 3.5 LP : TL = 1 : 3.6 to 1 : 4.75	LP : LD = 1 : 3 LP : TL = 1 : 3.7 to 1 : 4.4

Table 2.1. Sizes and numbers of free living and parasitic stages of "Gregarina caledia" from the grasshopper *Caledia captiva* and *Gregarina garnhami* from the locust *Schistocerca gregaria*. (l) = length, (d) = diameter, (dt) = development time for two-gamont-gametocysts to form basal discs, LP= length of protomerite, LD = length of deutomerite, TL = total length

Figure 2.1

A diagram representing the life cycle of "Gregarina caledia" in the grasshopper *Caledia captiva*. Reproductive stages are described from *in vitro* development of gametocysts and vegetative stages are described from histological sections. The graph shows the mean logarithmic increase in volume of sporonts with time from days 0-11.

Increase in size of the vegetative phase of *Gregarina caledia* over time

Development of Gametocysts from *Gregarina caledia*

Mean logarithmic volume (μm³)

10^8

10^7

10^6

10^5

10^4

10^3

10^2

10^1

D = deutomerite G = gametes
P = protomerite pa = patch
C = gametocyst sd = spore duct
gl = gelatinous layer sc = spore chain
gam = gametocyte L = length
Wa = wall W = width

Development of Sporonts from *Gregarina caledia*

Ellipsoid shaped spores, each containing 8 haploid nuclei are used to infect the grasshopper. Spore coats are removed enzymatically in the host gut liberating 8 sporozoites.

Nucleus diameter = 9μm³

Sporozoites attach to brush border of gut to become cephalonts.

Cephalonts found attached to mid gut caecae. Differentation into protomerite and deutomerite by a septum.

L:W = 1.5:1.0

Cephalonts found attached to mid gut brush border. Cephalont detaches to become a free swimming sporont.

Sporont found in gut lumen. Ectoplasmic scar left on protomerite from detachment of epimerite.

L:W = 3:1
PL:DL = 1:2.5

From other evidence, it is assumed that syzygy between 2 sporonts and encystment has occured.

D C
P

Within the gametocyst sporonts lose internal septa. Pale gold in colour and varied in size.

gam Wa
gl

Interior becomes very dark, surrounded by a light area, in which there is some movement.

G

Orange-coloured patches form on the surface, these form basal discs of inverted spore tubes

Diameter = 24μm³

pa gl

Sporont found in caecum.

Nucleus diameter = 18μm³

L:W = 2.5:1
PL:DL = 1:2.5

Inner mass shrinks to half former volume.

G gl
Wa

sporont found in mid gut between peritrophic membrane and food mass.

Spore ducts evert through basal discs to form exterior tubes.

gl Wa sd

Sporont found in mid gut caecum.

Dehiscence of gametocyst spore chains ejected through spore ducts.

gl
sc

Sporonts found in mid gut.

L:W = 3:1
PL:DL = 1:4

Days after infection

0 1 2 3 4 5 6 7 8 9 10 11

Figure 2.2

A diagram summarising the life cycle of "Gregarina caledia" in the grasshopper *Caledia captiva*. Vegetative stages were recorded in histological and fresh specimens. Reproductive stages from gametocysts were kept in a humid chamber and developed *in vitro*.

DAY 1   Germination

Within host gut each spore germinates into 8 haploid sporozoites.

Nucleus
Wall } SPORE

Sporozoite

DAY 2   Attachment

Epimerite forms on sporozoite which attaches to gut epithelium as a cephalont.

Epimerite

Nucleus

Brush border

Sporozoite

Cephalont

DAY 3   Differentiation

The cephalont further differentiates into the protomerite and deutomerite by the laying down of a septum of ectoplasm.

Septum

Nucleus

Epimerite

Protomerite

Deutomerite

Epicyte

Grasshopper eats spore adhered to grass

DAY 4   Detachment and Syzygy

The cephalont detaches from the host epithelium at the junction of the epimerite to form a freely mobile sporont. A scar is left in the position of the epimerite.

Two sporonts of approximately the same size enter syzygy. The anterior primite is usually larger than the posterior satellite. The association is freely mobile.

Scar

Sporont

Primite

Satellite

DAY 10   Dehiscence and Sporulation

Spore chains pass through spore tubes to exterior environment.

WITHIN HOST
WITHIN ENVIRONMENT

Cyst wall

Everted spore duct

Gelatin layer

DAY 9   Eversion of Spore Ducts

Meiosis occurs within each spore to form 8 haploid nuclei.

DAY 5   Encystment

The syzygid pair rotates to form a ball which becomes encapsulated as a gametocyst. Internal structures disappear and each sporont becomes a gamont.

Gamont

Sporont

Wall of cyst

Cyst wall

Basal disc

Gelatinous layer

Gametes

DAY 8 or 9   Shrinkage

The inner mass of the gametocyst shrinks.

DAY 7 or 8   Syngamy

Isogametes enter syngamy. Orange-coloured basal discs of inverted spore tubes appear on external gametocyst wall.

DAY 6   Gametogenesis

Interior becomes dark.

Figure 2.0f

Early free-swimming sporont of "Gregarina caledia". Note the epimerite scar on the protomerite (arrow). Scale bar = 12.5μm

Figure 2.0g

Gametocysts of "Gregarina caledia" on extruded faeces. Scale bar = 500μm
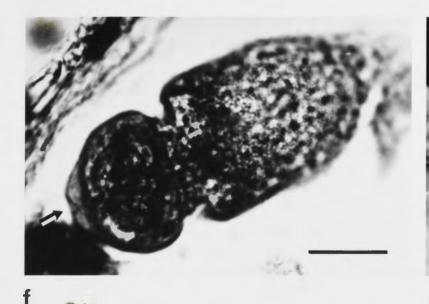
Figure 2.0h

Variation in the sizes of sporonts. Note the unusual occurrence of the two satellites attached to a single large primite (arrow). Scale bar = 30μm
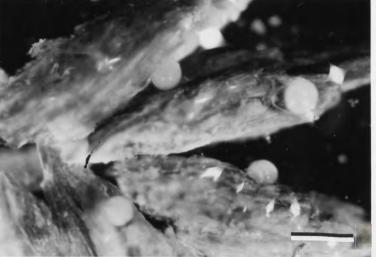
Figure 2.0i

Section of grasshopper mid-gut showing sporonts (arrowed) lying inside the peritrophic space on Day 11 after infection. Scale bar = 180μm

f

g

h

i
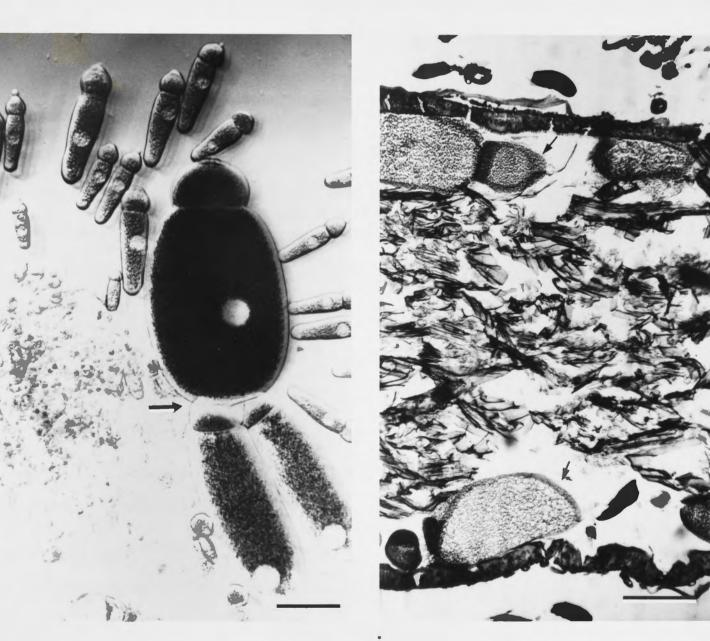
page 25, 3.1

3.1 Sources and isolation of five protoctist parasites: is relabelled as 3.1a
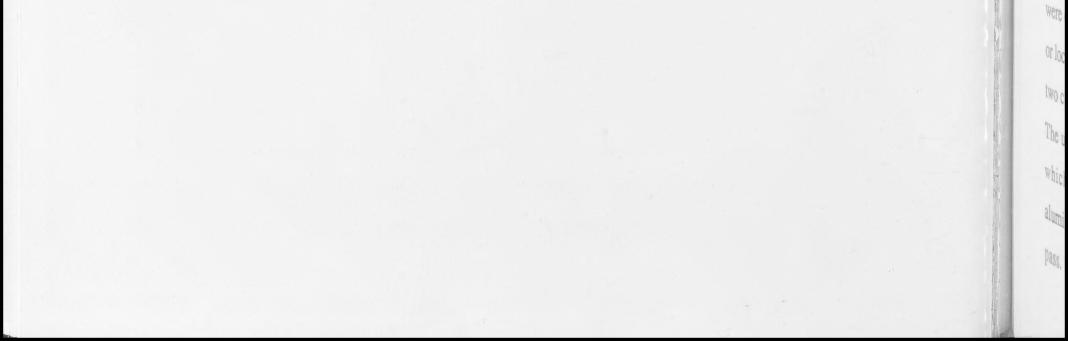
CHAPTER 3

GENERAL MATERIALS AND METHODS

3.1 Sources and isolation of five protoctist parasites

a. "Gregarina caledia"

"Gregarina caledia" was isolated from the Australian grasshopper *Caledia captiva*, Insecta: Orthoptera: Acridoidea: Acrididae (Key, 1979) and was first found in the early 1980's in New South Wales and Queensland, by Dr David D. Shaw (personal communication). In February 1991, 27 specimens of *Caledia captiva* were netted at Brooyar Forest $152^\circ35$'E, $26^\circ00$'S approximately 25km west of Gympie, Queensland, Australia.

Living grasshoppers were killed as described in section 2.3.2. The gut was slit open with scissors, allowing the sporonts and gametocysts of the parasites to be flushed out in water. The sporonts and gametocysts were counted under a dissecting microscope, washed 5 times in sterile, distilled water, placed in 1ml screw top plastic tubes and stored in liquid nitrogen. A total of 608 sporonts and 1 gametocyst were recovered.

b. "Gregarina chortoicetes"

"Gregarina chortoicetes" was isolated from the Australian plague locust, *Chortoicetes terminifera*, Insecta: Orthoptera: Acridoidea: Acrididae (Key, 1979) collected in Wyandra, Queensland, and kindly provided by Dr Peter Christian, Division of Entomology, CSIRO, Canberra, Australia.

Locusts were maintained in sterile cages made of polythene. These cages were constructed as designed by Harry (1969), to hold up to 8 adult grassgrasshoppers or locusts and to eliminate cross infection by *Gregarina spp*. Each cage consisted of two cylinders of rigid polythene tubing of 10 cm diameter, one on top of the other. The upper cylinder (30 cm high) formed a cage of volume $2357cm^3$, the base of which was formed by a 10 cm diameter circle of perforated (4.5 mm holes) aluminium, kept in place by plastic sealant, and through which faeces could easily pass. The top consisted of a standard plastic lid with a 2cm hole, into which was

placed a foam-plastic bung. The lower cylinder (10 cm high) sat in a large Petri dish in which the faeces collected. This confined the faeces and kept the grasshoppers out of range of spores that ejected from dehisced gametocysts in the Petri dish. A bunch of grass tied with string was lowered through the hole in the lid, and kept in place by the bung. This acted both as a food supply and as a source of humidity. Each day the cage was transferred to a clean lower cylinder and the entire structure placed in a clean Petri dish. Using forceps, the gametocysts were gently teased away from the faeces and washed 10 times in distilled water. After each use, all cage parts were washed in a concentrated solution of an iodine-based disinfectant, Turco-dine (Turco Industries, Australia). Up to 8 uninfected grasshoppers or 5 infected grasshoppers could be successfully maintained in each of these cages for periods of 30 and 20 days, respectively. During this time uninfected grasshoppers remained free of infection.

c. "Pseudomonocystis lepidiota"

Gametocysts of "Pseudomonocystis lepidiota" (Drs Denis Anderson and Ray Akhurst, personal communication) were isolated from the body cavity of the sugar cane beetle, *Lepidiota frenchi*, Insecta: Coleoptera: Scarabaeoidea: Scarabaeidae (Britten, 1979) found at Skeleton Creek, Todd Farm, White Rock, Cairns, Queensland. These were kindly donated by Drs Anderson and Akhurst, CSIRO Division of Entomology, Canberra, Australia.

Gametocysts were much larger than those of the other two gregarines and measured 1-2mm in diameter. These were washed in sterile distilled water 10 times and stored in a 1 ml screw topped plastic tube at -20°C.

d. "Nosema vespula"

Spores of "Nosema vespula" initially recovered from larval stages of the European wasp *Vespula germanica,* Insecta: Hymenoptera: Apocrita: Vespoidea: Vespidae (Riek, 1979) were kindly provided by Dr Denis Anderson, who found them in a nest of the European wasp in New South Wales, Australia. These wasps were first seen in Tasmania in the early 1980's (Denis Anderson, personal communication).

Larvae were removed from the comb and ground in distilled water with a mortar and pestle. The slurry was initially coarse-filtered 5 times through 3 layers of muslin, then through filter paper (Whatman Number 54) and finally separated from

page 27

"Heliothis armigera" is renamed to *"Helicoperva armigera"*.

host tissue by centrifugation through a discontinuous gradient of Percoll for 30 minutes at 4000rpm and 15°C in a Beckman L8-70M ultracentrifuge with a SW28 rotor. The Percoll gradient had been prepared by sequentially layering 7 ml of 100%, 8ml of 75%, 8 ml of 50%, 8 ml of 25% into each centrifuge tube, and overlaying with approximately 1 ml of spore suspension. Purified "Nosema vespula" spores were visible as a white band in the 75% Percoll, and as a white pellet among the brown debris in the base of the tube. Spores were removed with a Pasteur pipette, and counted in a Neubauer counting chamber. 2.3×10^8 spores ml^{-1} were obtained, and purified spores were stored in 1 ml screw top polythene vials at -20°C.

To increase spore numbers sufficiently for use in molecular biology techniques, these were passaged in larva of *Heliothis armigera* (kindly donated by Denis Anderson). Third instar larvae of *Heliothis armigera* which had been reared on a specific diet (Appendix) were separately housed in 35 ml lidded plastic cups (Lily, Australia) and starved for three hours. Each was provided with a 0.125 mm^3 cube of *Heliothis* food, to which 10μl of a 10^5 spores ml^{-1} dilution in distilled water had been added, and left overnight. From this time onward, food was provided *ad libitum* with daily monitoring. Lethargic individuals found to contain spores were stored at -20°C. Infected third instar larvae usually died 6 days after infection, and did not survive the fourth instar.

Five *Heliothis* larvae infected with "Nosema vespula" spores were ground in 30 ml sterile distilled water using a mortar and pestle. The slurry was filtered through 3 layers of muslin to remove the larger pieces (3 times) then through Miracloth (Calbiochem Corporation, USA) (3 times) and finally 6 layers of facial tissue. The filtrate was centrifuged twice in 40 ml centrifuge tubes in a bench centrifuge for 10 minutes at 12,000 rpm, the pellet being resuspended in distilled water each time. The final pellet was resuspended in 1 ml distilled water. A discontinuous density gradient of Percoll was prepared as before. Five ml of spore suspension were layered on top and centrifuged at 12,000 rpm in a Sorvall OTD, 75B-2 ultra-centrifuge using a SW 28 rotor. A white pellet of spores without any visible host tissue formed in the base of the tube. All top layers were removed with a Pasteur pipette and discarded, after which the spores were transferred to a fresh tube, from which the Percoll was removed by centrifuging 3 times in sterile distilled water at 12,000 rpm for 20 minutes. The spores were counted in a Neubauer Counting Chamber, the undiluted

page 28, addition of a new section, 3.1b:

"3.1b Naming of formally undescribed species

Because the identities of workers who made the first sightings and collections of the formally undescribed 3 apicomplexans namely, "Gregarina caledia", "Gregarina chortoicetes" and "Pseudomonocystis lepidiota" and the single microsporan "Nosema vespula" are not known, the convention used in this thesis is partly based on morphology of the parasite and the name of its specific host and is partly a matter of convenience for identification. It is referred to as a Genus species. For the first two apicomplexans, the generic name "Gregarina" was chosen by me, because the morphology of their vegetative forms both resemble those of the septate gregarine *Gregarina garnhami* (Harry, 1965; David Shaw, personal communication). The specific names "caledia" and "chortiocetes" are the generic names of their respective hosts *Caledia captiva* (Shaw, personal communication) and *Chortoicetes terminifera* (Key, 1979). The generic name "Pseudomonocystis" was suggested for the third apicomplexan by Drs Denis Anderson and Ray Akhurst, as the morphology of the gametocysts partially resembled those of members of the the Monocystidae Butschli, 1882 (Margulis *et al.,* 1989). The specific name "lepidiota" was chosen by me to define this species in terms of its host *Lepidiota frenchi* (Britten, 1979). The Genus name of the microsporan "Nosema vespula" was chosen because it has a morphology and life cycle resembling those of other *Nosema spp.* (Denis Anderson, personal communication), and its specific name was chosen to define it in terms of its host *Vespula germanica* (Denis Anderson, personal communication)."

page 28, 3.2

In the third sentence "maximum speed" should be read as "8000*g*."

concentration being 2×10^8 ml^{-1}. The final pellet was resuspended in 500µl sterile distilled water, and stored at -20^0 C.

e. *Nosema apis*

Genomic DNA extracted from spores of *Nosema apis*, isolated in New Zealand from the honey bee, *Apis mellifera*, Insecta: Hymenoptera: Apocrita: Apoidea: Apidae: (Riek, 1979) was kindly donated by Dr Louise Malone, (Hort+Research, Mt Albert Research Centre, Auckland, New Zealand).

3.2 Isolation of DNA from protoctistans

DNA was extracted from all protoctistan material by the mechanical disruption method of Undeen and Cockburn (1989). After washing and pelleting, the protoctistan material was resuspended in 200µl of mechanical disruption buffer (100mM NaCl, 200mM sucrose, 30mM Tris-HCl pH 8.0, 10 mM EDTA) with an equal quantity of glass beads (Sigma, USA) (0.45-0.5mm diameter). DNA was liberated from the spores by vortexing for 40 seconds at maximum speed. For microsporans, the percentage of broken spores was estimated using a Neubauer blood counting chamber, as 19% or 2×10^7 of the total count of 10^8 spores. For each protoctistan sample, proteins were denatured from 100µl of homogenate with an equal volume of the same mechanical disruption buffer containing Proteinase K (100µgml^{-1}) plus 0.25 volumes of DNA extraction buffer (2.5% SDS, 250mM EDTA, 500mM Tris-HCl pH 9.2) and incubated at 55°C for one hour. In a second incubation at 4°C for 60 minutes, the SDS and protein were precipitated with 1M potassium acetate and sedimented by centrifugation at 8000*g* for 20 minutes. DNA in the supernatant was concentrated by ethanol precipitation, air dried at 37°C and incubated at room temperature for 60 minutes in 20µl dissolving buffer (10mM Tris-HCl pH 8.0, 1mM EDTA 1mg ml^{-1} RNAase A). The solution was made up to 100µl, and was shaken with an equal volume of phenol/chloroform, the DNA concentrated by ethanol precipitation and dissolved in 50µl water. The DNA was further purified by agarose gel electrophoresis. Bands were excised and the DNA electroeluted. DNA was concentrated by ethanol precipitation, resuspended in 20µl water and stored at -20°C (Sambrook *et al.*, 1989). In the gregarines, the amount of isolated DNA was very small and 2µl of a 20µl sample of

resuspended DNA was not visible when fractionated by electrophoresis in agarose and stained with ethidium bromide ($10\mu g\ ml^{-1}$).

3.3 Selection of primers for PCR

Four primers for PCR (numbers 1, 2, 3A and 3B) were commercially synthesised in an ABI DNA synthesizer. The sequences were selected from the 18S rDNA sequence of the apicomplexian *Plasmodium falciparum* (McCutchan *et al.,* 1988). A fifth primer (number 4) was constructed later for use as a sequencing primer based on a nucleotide sequence from the apicomplexian "Gregarina caledia", a clone of which had already been generated using two of the PCR primers. These primers were chosen so that each contained more than 50% G + C with G or C nucleotides at the 5' and 3'ends, that their 3' sequences would not hybridize, that they did not contain poly-purines or poly-pyrimidines and that they each contained about 25 nucleotides. Primers of about this length and composition have optimal annealing temperatures of 60°C or above, and are more robust for priming DNA synthesis from difficult samples (Taylor, 1991).

Primer sequences were as follows:

Primer 1. Positions 581-603 complementary to the *P. falciparum* non-coding strand.

5' GCG GTA ATT CCA GCT CCA ATA GC 3' (23 nucleotides).

Primer 2. Positions 2089-2067 complementary to the *P. falciparum* coding strand.

5' CCG CAG GTT CAC CTA CGG AAA CC 3' (23 nucleotides).

Primer 3A. Positions 1275-1299 complementary to the *P. falciparum* non-coding strand.

5' GCG GCT TAA TTT GAC TCA ACA CGG G 3' (25 nucleotides).

Primer 3B. Positions 1299-1275 complementary to the *P. falciparum* coding strand.

5' CCC GTG TTG AGT CAA ATT AAG CCG C 3' (25 nucleotides).

Primer 4. Positions 462-484 complementary to the "Gregarina caledia" non-coding strand.

5' GGA TCG GAG GCC GTA ACA ATA CG 3' (23 nucleotides).

In all protoctistans, nucleotides concordant with those from positions 1-580 in the apicomplexan *Plasmodium falciparum* SSU rRNA gene were not sequenced; therefore each sequence is approximately 66% of the total of that gene.

Primer concentrations were calculated from the absorbance at OD260 and kinased before use as described by Sambrook *et al.* (1989).

3.4 Gene amplification using PCR.

Gene amplification of the five protoctistan rRNA gene sequences was done using a modification of the PCR methods of Mullis and Faloona (1987) and Saiki *et al.*, (1988). Each 100μl reaction mixture contained 100 picomoles of each of the amplification primers, 200μl of each dNTP, 1 x PCR buffer (Perkin Elmer Cetus, USA), 1.5mM $MgCl_2$, 2.5 Units of *Amplitaq* DNA polymerase (Perkin Elmer Cetus, USA), 100ng of DNA template and 100μl paraffin oil overlay. PCR amplifications were done in an Intelligent Heating Block (Cherlyn Electronics, UK) as described by Medlin *et al.* (1988) for 35 cycles of 90°C for 2.5 minutes, 30°C for 4.25 minutes and 65°C for 7.5 minutes followed by a single primer extension at 72°C for 7 minutes. PCR products were purified by agarose gel electrophoresis, followed by electroelution of the bands of interest.

3.5 Procedures for molecular cloning

Vector DNA was prepared by linearisation of either phage M13mp18 or mp19 with restriction endonuclease *Sma*I, or plasmid pBLUESCRIPT SK (+/-) (Stratagene, USA) with restriction endonuclease *Eco*RV. Eluted PCR products were end-filled, and then ligated into vector DNA which had been treated with calf-intestinal-phosphatase.

Ligation mixtures were transformed into supercompetent *E.coli* K12 JM101 cells (Messing, 1983; Hanahan, 1983), and clones with inserts were selected by ampicillin resistance and blue/white colour on X-gal/IPTG plates as appropriate.

Single-stranded DNA for sequencing was prepared from M13 clones, and double stranded DNA was prepared from pBLUESCRIPT clones by the methods of Holmes and Quigley (1981), Titus (1991) and Sambrook *et al.* (1989).

Both single-stranded DNA and double-stranded DNA from recombinant clones were sequenced manually using Sequenase (USB) and ^{35}S-dATP; the reaction

page 31, 3.6

Replace the existing section with the following:

"3.6 Resolution of variable nucleotides

In otherwise identical fragments from each of the five protoctistan taxa, between 0.1 and 0.8% of nucleotides were variable. It is possible that the original PCR products were mixtures, amplified from two or more apicomplexan or microsporan individuals each with different genomes, in which case heterogeneity is real. Alternatively, the original PCR products were homogeneous and the changes were incorporated during amplification by the *Amplitaq* DNA polymerase, or during sequencing by T7 DNA polymerase . Although such minor levels of variation are unlikely to have a significant impact on the final interpretation of relationships, it was important to determine whether these were true heterogeneities and therefore phylogenetically important. To resolve this ambiguity, a minimum of 4 extra clones covering the region were sequenced. Final descisions were based on statistical certainty, clarity of the sequencing gels and comparisons with nucleotides in the same positions from a range of other taxa. In all cases, these appeared to be sequencing errors and not true heterogeneities."
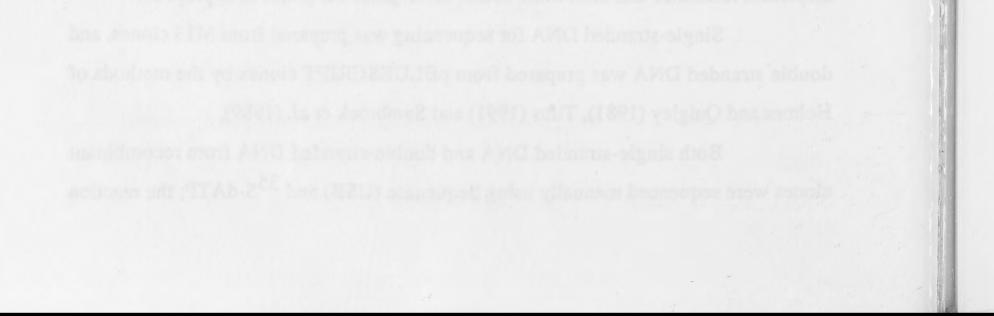
products were fractionated by electrophoresis through a 6% polyacrylamide gel and exposed to Kodak XAR film.

3.6 Resolution of variable nucleotides

Nucleotide differences were observed in otherwise identical fragments from each of the five protoctistan taxa. In all cases these appeared to be sequencing errors and not true heterogeneities. To resolve any ambiguity which occurred in the sequence, a minimum of 4 extra clones covering the region were sequenced.

3.7 Computer handling of nucleotide sequences

Using as a comparison, the 18S rDNA sequence of *Plasmodium falciparum* (McCutchan *et al.*, 1988), sequenced fragments from various clones were allocated to their relative positions by the Shotgun Sequencing program (Staden, 1982; 1987) and SEQ library programs of the Research School of Biological Sciences at the Australian National University and the University of Wisconsin Genetic Computing Group (UWGCG) versions 6 and 7 (Devereux *et al.*, 1984).

Overlapping clones formed contigs and are shown in Figure 3.1 a, b, c, d, and e. These were collated into consensus SSU rDNA sequences for the 5 organisms and are shown in Figures 3.2, 3.3, 3.4, 3.5 and 3.6.

Figure 3.1 a-e
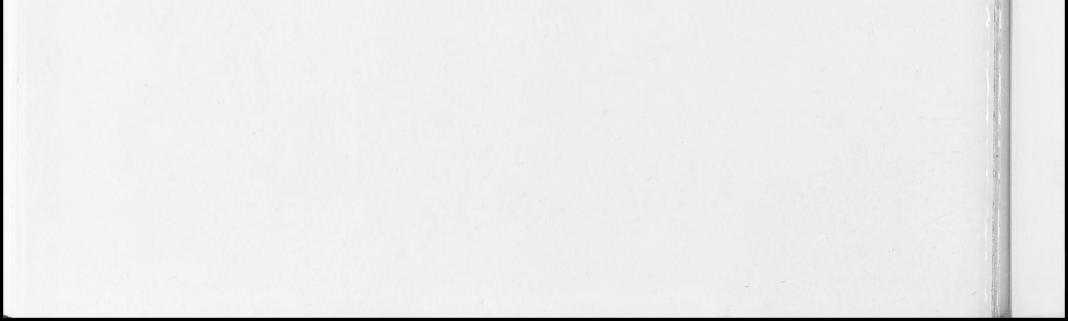
Maps showing the overlapping clones used to determine the 18S rDNA
sequences from the apicomplexians (a, b, and c) and 16S rDNA sequences from
the microsporans (d and e). The scale bar indicates the position of each clone in
the sequence. The filled squares indicate the positions of the sequence used to
prepare synthetic oligonucleotide primers P1, P2, P3 and P4. Each arrow and its
direction represents an 18S or 16S rDNA insert and its orientation in either
M13mp18/19*Sma*I or pBluescript SK(+)*Eco*RV vectors.
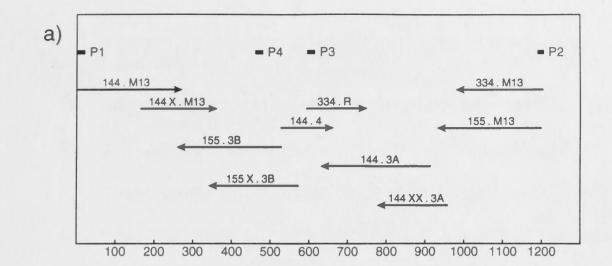
a. "Gregarina caledia:"
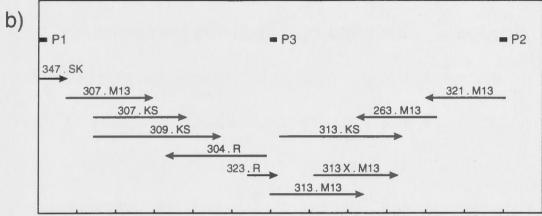
b. "Gregarina chortoicetes"

c. "Pseudomonocystis lepidiota"

d. "Nosema vespula"

e. *Nosema apis*

a)

P1   ■ P4   ■ P3   ■ P2

144 . M13 →
144 X . M13 →
334 . R →
334 . M13 ←
144 . 4 →
155 . M13 ←
155 . 3B ←
144 . 3A ←
155 X . 3B ←
144 XX . 3A ←

100 200 300 400 500 600 700 800 900 1000 1100 1200

b)

P1   ■ P3   ■ P2

347 . SK →
307 . M13 →
321 . M13 ←
307 . KS →
263 . M13 ←
309 . KS →
313 . KS →
304 . R ←
323 . R ←
313 X . M13 →
313 . M13 →

100 200 300 400 500 600 700 800 900 1000 1100 1200

c)

P1   ■ P3   ■ P2

68 . R ←
91 . M13 →
75 . R →
83 . M13 →
67 . M13 ←
77 . M13 →
63 . M13 ←
83 . R ←
68 . M13 →
67 . R →

100 200 300 400 500 600 700 800 900 1000 1100 1200

d)

P1   ■ P3   ■ P2

32 . M13 →
30 . R ←
40 . R →
42 . R →
42 SS (2) . M13 ←
42 SS (1) . M13 ←
48 . R ←

100 200 300 400 500 600 700 800 900 1000

e)

P1   ■ P3   ■ P2

409 . T7 →
404 . T7 ←
457 . KS ←
409 . T3 ←
404 . T3 ←
511 . SK ←
506 . SK →
511 . KS ←
506 . KS ←
506 X . KS ←
511 X . SK ←

100 200 300 400 500 600 700 800 900 1000

Figure 3.2

18S rDNA sequence of "Gregarina caledia". Position 1 corresponds to position 581 in the SSU rRNA gene sequence of the apicomplexan *Plasmodium falciparum* (McCutchan *et al.*, 1988).

"Gregarina caledia"
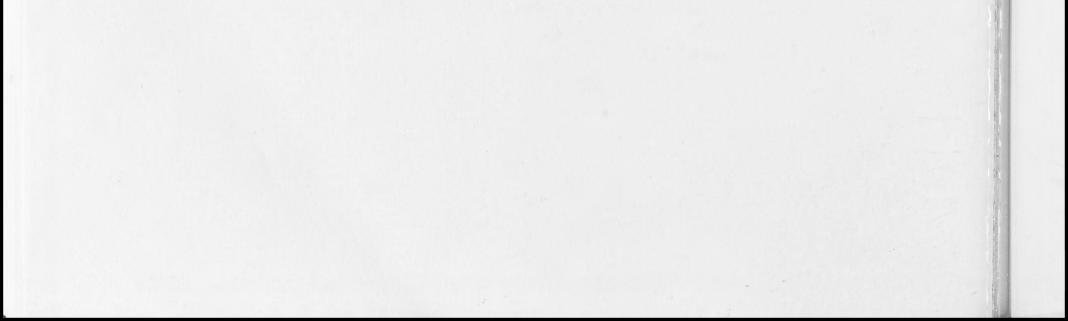
```
   1  GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAAUUGCUG CAGUUAAAGC

  51  GUCCGUAGUU GAAUUUCGUU CGCACGGGUA GGAUGCCGAU UGAUUCUUUG

 101  GAGUCAUCGU GCGCUUUCCC CAGUGAACUU GAGGGAAAUA UCCAGCCUGC

 151  UGGAUGUUCU CCUCCGUUAC UUUGAGCAAA UUGGAGUGCU CCAACCAGGC

 201  UUACGCUUGA ACAGCUCAGC AUGGAAUAAC AAGAUAGGAC UUUGGUUCUU

 251  CUUGUUGGUG UCAUGAACCA AAAGUAAUGG UUGAUAAGGA CAUACGGGGG

 301  CAUUUGUACU UGCUGGAGAG AGGUGAAAUU CUAAGACCCA GCAAAGACAA

 351  ACAACUGCGA AAGCAUUUGC CCAGUGUGUA CCUAUUAAUC AAGGACGAAA

 401  GUUGGGGGAU CGAAGACGCU UAGAUACCGU CGUAGUCCCA ACUAUAAACU

 451  AUGCCGACUG AGGAUCGGAG GCCGUAACAA UACGACUCUU CGGCACUCCA

 501  AGAGAAAUCU AAGUCUCUGG GCCCUGGGGG GAGUAUGGUC GCAAGGCUGA

 551  AACUUAAAGG AAUUGACGGA AGGGCACCAC CAGGAGUGAG CUGCGGCUUA

 601  AUUUGACUCA ACACGGGGAA CCUCACCAGG CCCGGACAUA GUCAUGAUUG

 651  ACAGAUCGAG AGUUCCUUCU CGAUUCUAUG GGUGGUGGUG CAUGGCCGUU

 701  CUUAGUCGGU GAGGUGACUU GUCUGGUUAA UUCCGAUAAC GGACGAGACC

 751  UCGACCUGCU AACUAGCAGA CACUUACAGA UAGUAAGUGU UUAGCUUCUU

 801  AGAGGGACUU UGUGAGUCUA UCACAAGGAA GUUCGAGCCU AUAACAGGUC

 851  UGUGAUGCCC UUAGAUGGCC UGGGCUGCAC GUGCGCUACA AUGACAGAGC

 901  CAGCGAGUAU UACCCUUCUU CCGUUCGGAG UGGGCAAUCU UUUGAAACUC

 951  UGUCGUGAUA GGGAUUGACC CUUGCAAUUA UGGGUCAUGA ACGAGGAAUU

1001  CCUAGUAAGA ACAAGUCAUC ACCUUGUGCU GAUUACGUCC CUGCCCUUUG

1051  UACACACCGC CCGUCGCUUC AACCGAUUGG AUGAUCCGGC AAACUGUACA

1101  GACAUUUGGA AUUACCUGGA GUUCCUCUAG GCUUUCGAGC GGAAGUACCG

1151  UGAGCCUUAU CAUCUAGAGG AUGAAGAAGU CGUAACACGG UUUCCGUAGG

1201  UGAACCUGCG
```

Figure 3.3

18S rDNA sequence of "Gregarina chortoicetes". Position 1 corresponds to position 581 in the SSU rRNA gene sequence of the apicomplexan *Plasmodium falciparum* (McCutchan *et al.*, 1988).

Gregarina chortoicetes
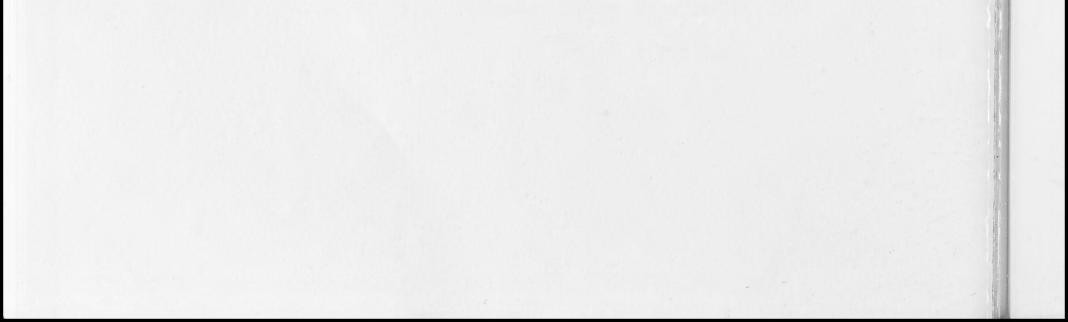
```
   1  GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAAUUGCUG CAGUUAAAGC

  51  GUCCGUAGUU GGAUUUCGUU CGCACGGGUA GGAUGCCGAU UGAGUCUUCG

 101  GAUUCAUCGU GCGCUUUCCC UAGUGAACUU GAGGGAAACU UUCAGCUUGC

 151  UGAAUGUUCU CCUCCGUUAC UUUGAGCAAA UUGGAGUGCU CCAACCAGGC

 201  UUACGCUUGA ACAGCUCAGC AUGGAAUAAC AAGAUAGGAC UUUGGUUCUU

 251  CUUGUUGGUG UCAUGAACCA AAAGUAAUGG UUGAUAAGGA CAUACGGGGG

 301  CAUUUGUACU UGCUGGAGAG AGGUGAAAUU CUAAGACCCA GCAAAGACAA

 351  ACAACUGCGA AAGACUUUGC CCAGUGUGUA CCUGUUAAUC AAGGACGAAA

 401  GUUGGGGGAU CGAAGACGAU UAGAUACCGU CGUAGUCCCA ACUAUAAACU

 451  AUGCCGACUG AGGAUCGGAG GCCGUAACAA UACGACUCUU CGGCACUCCA

 501  AGAGAAAUCU AAGUCUCUGG GCCCUGGGGG GAGUAUGGUC GCAAGGCUGA

 551  AACUUAAAGG AAUUGACGGA AGGGCACCAC CAGGAGUGGA GUGCGGCUUA

 601  AUUUGACUCA ACACGGGGAA CCUCACCAGG CCCGGACAUA GUCAUGAUUG

 651  ACAGUUCGAG AGUUCUUUCU CGAUUCUAUG GGUGGUGGUG CAUGGCCGUU

 701  CUUAGUCGGU GAGGUGACUU GUCUGGUUAA UUCCGAUAAC GGACGAGACC

 751  UCGACCUGCU AACUAGCAAA CAUUUACAGA CCGUAAGUGU UUAGCUUCUU

 801  AGAGGGACUU UGUGAGUCUA UCACAAGGAA GUUCGAGCCU AUAACAGGUC

 851  UGUGAUGCCC UUAGAUGGCC UGGGCUGCAC GUGCGCUACA AUGACAGAGC

 901  CAGCGAGUAU UACCCUUCUU CCGUUCGGAG UGGGCAAUCU UUUGAAACUC

 951  UGUCGUGAUA GGGAUUGACC CUUGCAAUUA UGGGUCAUGA ACGAGGAAUU

1001  CCUAGUAAGG ACAAGUCAUC ACCUUGUGCU GAUUACGUCC CUGCCCUUUG

1051  UACACACCGC CCGUCGCUUC AACCGAUUGG AUGAUCCGGC AAACUGUACA

1101  GACAUUUGAA ACCACCUGGA GUUCCUCUAG GCUUUCGAGU GGAAGUACCG

1151  UGAGCCUUAU CAUCUAGAGG AUGAAGAAGU CGUAACACGG UUUCCGUAGG

1201  UGAACCUGCG
```

Figure 3.4

18S rDNA sequence of "Pseudomonocystis lepidiota". Position 1 corresponds to position 581 in the SSU rRNA gene sequence of the apicomplexan *Plasmodium falciparum* (McCutchan *et al.*, 1988).

Pseudomonocystis lepidiota

```
   1  GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAGUUGUUG CAGUUAAAAA

  51  GCUCGUAGUU GAACUUAUGU UGGCGAAAGU CAGUUGCCUU CAUUAGGUUU

 101  AUUGUACUUU CAACAACAUU GGUCUACUGC UUGGCAACUU CACUGCUGUC

 151  AAUCCAGUAG AGCUGUUACU UUGAGUAAAU UAGAGUGUUU CAAGCAGGCG

 201  CAGUGCCUUG AAUACCCAGC AUGGAAUAAC AAAUAAGGAC UUCGGUUCUU

 251  UUUUGUUGGU UUGGAGGCCG AAGUAAUGAU UAACAGGAAC AGUUGGGGGC

 301  AUUCGAAUUU GGUAGCUAGA GGUGAAAUUC UUAGAUUUAC CAAAGACGGA

 351  CUACUGCGAA AGACUCUGCC AAGGAUGUUU UCAUUAAUCA AGAACGAAAG

 401  UUAGGGGAUC GAAGACGAUC AGAUACCGUC GUAGUCUUAA CUAUAACGAU

 451  GCCAACUAGA GAUUGGAAAU UGUCACUUAU UGACAUUUUC AGCACCUUAU

 501  GAGAAAUCAA AGUCUUUGGG UUCUGGGGGG AGUAUGGCCG CAAGUCUGAA

 551  ACUUAAAGGA AUUGACGGAA GGGCACCACC AGGAGUGAGC UGCGGCUUAA

 601  UUUGACUCAA CACGGGAAAA CUCACCAGGU CCAGACAUAG GAAGGAUUGA

 651  CAGAUUGAGA GCUCUUUCUU AAUUCUAUGG GUGGUGGUGC AUGGCCGUUC

 701  UUAGUUGGUG GAGUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU

 751  UAACCUGCUA AAUAGACACC AAGGUCAUAA CCUUGGCUGU GCUUCUUAGA

 801  GGGACUUUGC GUAUCUAACG CACGGAAGUU UAAGGCAAUA ACAGGUCUGU

 851  GAUGCCCUUA GAUGUCCUGG GCUGCACGCG CGCUACACUG AUGCACUCAG

 901  CGAGUAUUCC AUGACCUGUU AAGGUUGGGU AAUCUUGUGA AUGUGCAUCG

 951  UGAUUGGGCU AGAUGAUGUA AUUAUUCAUC UUUAACGAGG AAUUCCUAGU

1001  AAGUACAAGU CAUUAGUUGU GCUGAUUACG UCCCUGCCCU UUGUACACAC

1051  CGCCCGUCGC CUCAAUCGAC UGGAUGAUCC GGUGAAUGGC UCAGACUGGG

1101  AUGCAGGUGG AAACAUCUAC GUCUUGGGAA GUUCUGUGAA CCAAAUCAUC

1151  UGAAGAAUGA GAAAGUCGUA ACAUGGUUUC CGUAGGUGAA CCUGCG
```
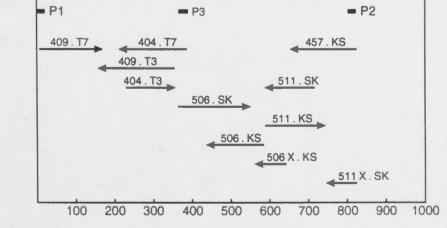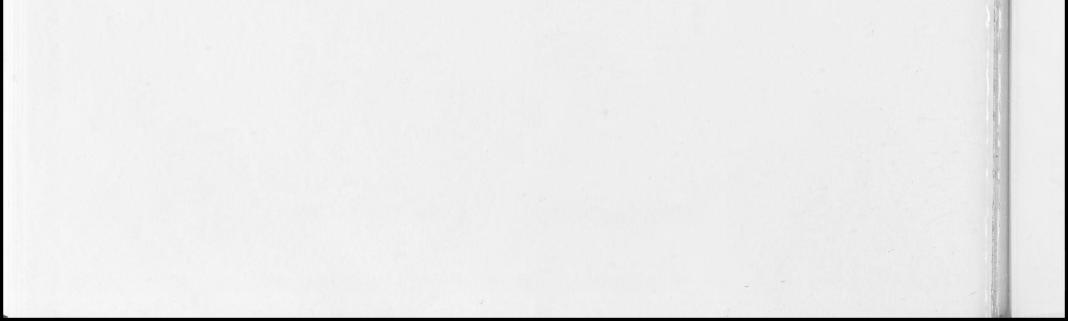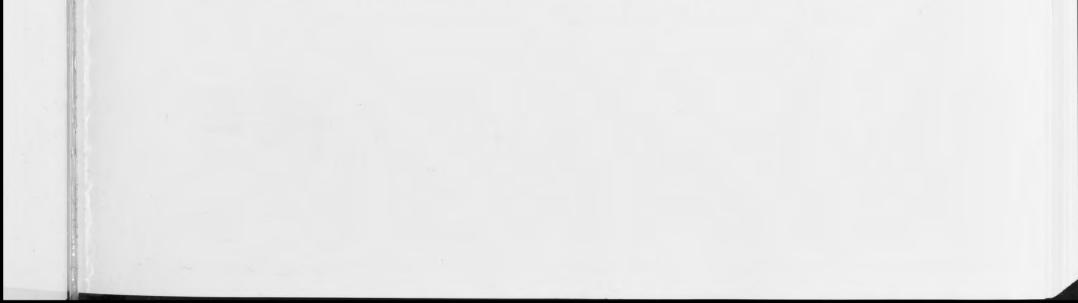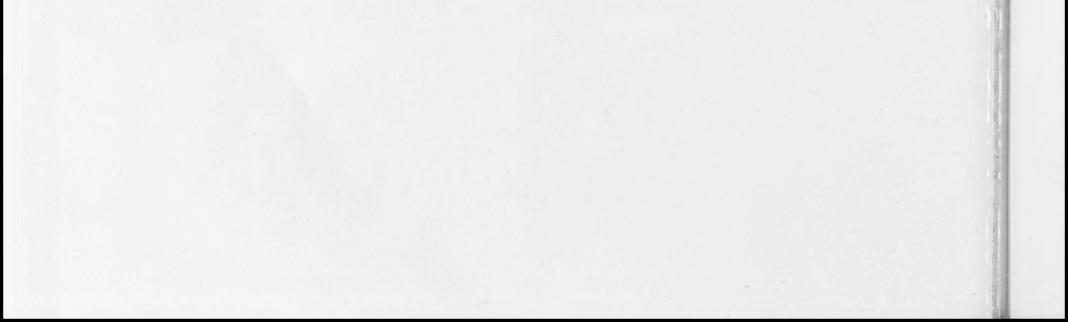
Figure 3.5

16S rDNA sequence of "Nosema vespula". Position 1 corresponds to position 581 in the SSU rRNA gene sequence of the apicomplexan *Plasmodium falciparum* (McCutchan *et al.*, 1988).

Nosema vespula
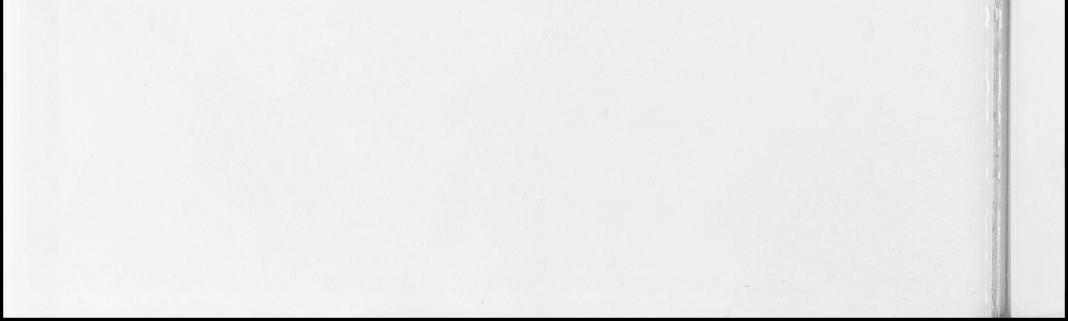
```
  1  GCGGUAAUUC CAGCUCCAAU AGCGUGUAUG AUGAUUGAUG CAGUUAAAAA

 51  GUCCGUAGUU UAUAUUUAAG AAGCAAUAUG AGGUGUACUG UAUAGUUGGG

101  AGAAAGAUGA AAUGUAACGA CCCUGACUGG ACGAACAGAA GCGAAAGCUG

151  UACACUUGUA UGUAUUUUUU GAACAAGGAC GUAAGCUGGA GGAGCGAAGA

201  UGAUUAGAUA CCAUUGUAGU UCCAGCAGUA AACUAUGCCG ACGAUGUGAU

251  AUGAUAUAUU UUGUAUUACA UAAUAGAAAU UAGAGUUUUU UGGCACUGGG

301  GAUAGUAUGA UCGCAAGAUU GAAAAUUAAA GAAAUUGACG GAAGAAUACC

351  ACAAGGAGUG GAUUGUGCGG CUUAAUUUGA CUCAACACGG GGUAACUUAC

401  CAAUAUUUUA UUAUUUGAGA CGAUUUUUAA UCAGAGAAUG AUAAUAGUGG

451  UGCAUGGCCG UUUUCAAUGG AUGCUGUGAA GUUUUGAUUA AUUUCAACAA

501  GACGUGAGAC CCUUUUAUUU AUAGACAGAC ACAAUCAGUG UAGGAAGGAA

551  AGGAUUAAAA CAGGUCCGUU AUGCCCUCAG ACAUUUUGGG CUGCACGCGC

601  AAUACAAUAG AUAUAUAAUC UUUAUGGGAU AAUAUUUUGU AAGAGAUAUU

651  UGAACUUGGA AUUGCUAGUA AAUUUUAUUA AAUAAGUAGA AUUGAAUGUG

701  UCCCUGUUCU UUGUACACAC CGCCCGUCGC UAUCUAAGAU GAUAUAUGUU

751  GUGAAAUUAG UGAAAACUAC UUGAACAAUA UGUAUUAGAU CUGAUAUAAG

801  UCGUAACAUG GUUUCCGUAG GUGAACCUGC G
```

Figure 3.6

16S rDNA sequence of *Nosema apis*. Position 1 corresponds to position 581 in the SSU rRNA gene sequence of the apicomplexan *Plasmodium falciparum* (McCutchan *et al.*, 1988).

Nosema apis

```
  1   GCGGUAAUUC CAGCUCCAAU AGCGUGUAUG AUGAUUGAUG CAGUUAAAAA

 51   GUCCGUAGUU UAUUAUUAAG AAGCAAUAUG AUGUGUACUG UAUAGUUGGG

101   AGAGAGAUGA AAUGUGACGA CCCUGACUGG ACGAACUGAA GCGAAAGCUG

151   UACACUUGUA UGUAUUUUUU GAACAGGGAC GUAAGCUGGA GGAUCGAAGA

201   UGAUUAGAUA CCAUUGUAGU UCCAGCAGUA AACUAUGCCG ACGAUGUGAU

251   AUGAGAUGUU GUAUUACAUU AUAGAAAUUA GAGUUUUUUG GCUCUGGGGA

301   UAGUAUGAUC GCAAGAUUGA AAAUUAAAGA AAUUGACGGA AGAAUACCAC

351   AAGGAGUGGA UUGUGCGGCU UAAUUUGACU CAACACGGGG UAACUUACCA

401   AUAUUUUAUU GUUCCGGCCG AGGAUAUGAU CUGAGGAUGA UAAUAGUGGU

451   GCAUGGCCGU UUUCAAUGGA UGCUGUGAAG UUUUGAUUAA UUUCAAAAGA

501   CGUGAGACCC UUUAUUAGAC UGACACUAUU AGUGUAGGAA GGAAAGGACU

551   AAAACAGGUC AGUUAUGCCC UCUGAGCAUU UUGGGCAGCA CGCGCAAUAC

601   AAUAGACUUU AAUCUUUAUG GGAUAAUAUU UUGUAAGAGA UAUUUGAACU

651   UGGAAUUGCU AGUAAAUUUU AUCAAAUAAG UAGAAUUGAA UGUGUCCCUG

701   UUCUUUGUAC ACACCGCCCG UCGCUAUCUA AGAUGAUAUG UGUUGUGAAA

751   UUAGUGCAAG CUACUUGAAC AAUAUGUAUU AGAUCUGAUA UAAGUCGUAA

801   CAUGGUUUCC GUAGGUGAAC CUGCGG
```

page 32, 4.1

add to the end of the first paragraph:

"Thermodynamic stability

The first such model based on thermodynamic stabilities was proposed by Fresco *et al.* (1960) who showed that single-stranded rRNA folds back on itself forming secondary double-stranded structures stabilised by hydrogen bonds between complementary bases. Simple, combinatorial computer algorithms using thermodynamics to predict

CHAPTER 4

SECONDARY STRUCTURE OF SSU rRNA

4.1 INTRODUCTION

Until two decades ago, ribosomal function was thought to be directed by its structural proteins. However in 1973, these beliefs were confounded by the suggestion that the basic functional units of the ribosome were the RNAs (Woese, 1973). Woese's views have since been vindicated with a series of dramatic discoveries, in which the rRNA itself has assumed a new role of supreme importance as a molecule with a wide range of highly conserved, functional activities (Dahlberg, 1989). Thus maintenance of correct biochemical and biophysical interactions within the ribosome, and higher order interactions between the ribosome and the other intracellular components, determine the functional integrity of the rRNA molecule, which in turn depends on conservation of its secondary and tertiary structure. The secondary structure of SSU rRNA is more highly conserved than the nucleotide sequence and therefore underpins accurate sequence alignments used as the basis of most phylogenetic comparisons. Factors such as the great variations in base composition between the SSU rRNA from different organisms appear to have little influence on secondary structure, but can cause considerable difficulty in assessing sequence similarity. Therefore secondary structure models of SSU rRNA are usually constructed as a means to help with the alignment process, of gaining a greater insight into the evolution of this molecule, and to delineate regions that evolve rapidly. Because of its size and complexity, the techniques of chemical and enzymatic nucleases, NMR spectroscopy and X-ray crystallography have not been able to resolve SSU rRNA structures directly. Thus predictions based on thermodynamic stability (Williams and Tinoco Jr., 1986; Zuker, 1989; Zuker *et al.*, 1991) and phylogenetic comparison (Gutell *et al.*, 1985; Pace *et al.*, 1989; Larsen, 1992) have been the primary means of determining rRNA structure.

The first basic secondary structures elucidated for prokayotic 16S rRNA (Woese *et al.*, 1980) and eukaryotic 18S rRNA (Brimacombe, 1980) showed several similarities. More recent compilations include sequences from a large number of

secondary structures in short RNA sequences, were first introduced by Pipas and McMahon (1975). These were extended for longer sequences by Studnika *et al.* (1978) and further refined into dynamic programming algorithms by Nussinov and Jacobsen (1980) and Zuker and Steigler (1981). Using these algorithms, optimal secondary structures are found from the huge number of possible foldings (10^{102} foldings in a sequence of 400 nucleotides) (Zuker *et al.* 1991) by calculating the free energies for all possible base pairings, and choosing the set with the lowest free energy. Ambiguities inherent in experimental data and in the possibility that molecular structures are not in equilibrium, are compensated by algorithmic generation of sub- optimal foldings with free energies close to the minimum (Williams and Tinoco Jr., 1986; Zuker, 1989; Zuker *et al.*, 1991).

Folded RNA is not completely double-stranded, base paired regions are separated by single-stranded regions consisting of isolated unpaired bases, loops, and variously sized bulges (Tinoco Jr. *et al.,*1990). Folded conformations of RNA molecules are largely stabilised by the interactions that form double-stranded helices or stems. Duplexes of antiparallel strands are stabilised both by hydrogen bonding between the strands and by intra- and inter-strand stacking of the aromatic bases. Predictions for forming a correctly folded molecule from a single strand are basically the sum of the free energies of the constituent elements which is influenced, not only by the interactions between these elements, but also by the thermodynamic contributions of secondary and tertiary interactions. Free energies calculated using the nearest neighbour model (Jaeger *et al.,* 1989) are easier to quantify in short duplexes than in single stranded loops. Recently, loop and bulge solution structures that correlate with thermodynamic data have been resolved using NMR spectroscopy, which indicates that loops can provide additional stability to the 3 dimensional conformation (Cheong *et al.,*1990; Heus and Pardi, 1991; Varani *et al.,* 1991 and Wolters, 1992). Despite the limitations of energy minimizing algorithms (Zuker and Steigler 1981), they do correctly predict about 70% and 90% of the secondary structure within optimal and sub-optimal structures respectively, when compared to predicted phylogenetic models (Jaeger *et al.,* 1989).

Phylogenetic comparison

Phylogenetic comparison is still the most useful tool for definition of rRNA structure. Used independently to clarify folding patterns (Gutell *et al.,* 1985), or to confirm and supplement data inferred by energy minimisation (Jaeger *et al.,* 1989), it is regarded as

prokaryotic and eukaryotic organisms (Huysmans and De Wachter, 1986; Dams *et al.*, 1988; Neefs *et al.*, 1991; De Rijk *et al.*, 1992) and have resulted in the definition of two possible secondary structure models, one applicable to SSU rRNAs from eubacteria, archaebacteria, plastids and mitochondria, and the other which applies to eukaryotic cytoplasmic SSU rRNA (De Rijk *et al.*, 1992). These models show the primary sequence as forming five structural elements namely, helices, hairpin loops, internal loops, bulges and junction loops (Neefs *et al.*, 1991; De Rijk *et al.*, 1992; Figure 4). All models have 48 helices (universal helices) some of which are grouped into variable or V regions of which the V4 region, which lies between helices 21 and 22 and contains the conserved eukaryotic helix E21-5, is the most variable. Variable sequences are thought to be either remnants of mobile elements that inserted into ribosomal genes, or the remnants of linkers that connected different functional elements during the evolutionary assembly of the ribosome and which have been subsequently eliminated from all but nuclear encoded eukaryotic ribosomes (Clark *et al.*, 1984; Clark, 1987). In bacteria and organelles, the V4 region consists of a single hairpin, but in eukaryotes it forms up to 9 helices, none of which are homologous to the prokaryotic form and which often contain pseudoknots (the pairing of a single-stranded region to the loop sequence in a hairpin loop; Pleij *et al.*, 1986). Exceptions are *Giardia lamblia* which has a reduced V4 region and *Vairimorpha necatrix* (Vossbrinck *et al.*, 1987) which lacks helices 10, 11 and 44 and the entire V4 region. (Figures 4.1 to 4.5).

This chapter records my attempts to fit the SSU rRNA molecules of the microsporans "Nosema vespula" and *Nosema apis*, and the apicomplexans "Gregarina caledia", "Gregarina chortoicetes" and "Pseudomonocystis lepidiota" into the two types of established secondary structure models.

4.2 MATERIALS AND METHODS

For both microsporan and apicomplexan SSU rRNA sequences, phylogenetic comparisons using primary and secondary structures of closely related taxa, were used in combination with free energy minimisation techniques incorporated in the computer program FOLD (Zucker and Steigler, 1981).

having only slightly less resolving power than X-ray crystallography or NMR spectroscopy (Larsen 1992). Phylogenetic comparisons assume that function is dependent on structure, and that mutations which do not alter function will be preserved, therefore despite base changes in rRNA sequences between organisms, the preservation of the same structure is evidence for its existence. Aligned sequences from several taxa are searched for regions capable of base pairing maintained by Watson-Crick (A-U, C-G) and non-canonical (G-U) rules. A helix is usually considered to exist if two or more covariations are found in it (Pace *et al.*, 1989). Phylogenetic methods are limited in that they do not provide information about regions of secondary structure that contain conserved nucleotides, and may predict fewer helices than actually exist. (Chastain and Tinoco, Jr., 1991). "

page 33, 4.1

the second paragraph should read:

"This chapter records the fitting of SSU rRNA molecules of the microsporans...."

page 34, a. Microsporans
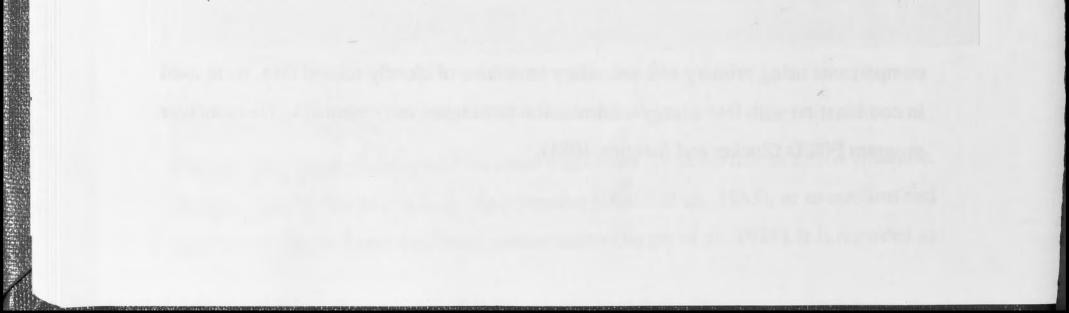
the second sentence should read:

"Nucleotide sequences of "Nosema vespula" and *Nosema apis* and the sequence of *Vairimorpha necatrix* were aligned using the computer program PILEUP (Devereux *et al.*, 1984)."

page 34, b. Apicomplexans

the second line should read:

"the causative organism of human tertian malaria, *Plasmodium falciparum* (McCutchan *et al.,* 1988) and the causative organism of mouse malaria *P. berghei* (Gunderson *et al.,* 1987) were edited at the 5' and 3' ends......"

a. Microsporans

Secondary structures of the two 16S rRNAs from "Nosema vespula" and *Nosema apis* were based on the secondary structure model proposed for *Vairimorpha necatrix* (Neefs *et al.*, 1991). Nucleotide sequences of "Nosema vespula" and *Nosema apis* and an edited sequence of *Vairimorpha necatrix* were aligned using the computer program PILEUP (Devereaux *et al.*, 1984). Nucleotide differences which existed in double-stranded areas of each species, and which were found to cause changes to the relative helical structures, were revised from 5' to 3' in each complete helix, using the program FOLD (Zuker and Steigler, 1981).

b. Apicomplexans

The 18S rRNA primary sequences of *Drosophila melanogaster* (Tautz *et al.*, 1987), *Plasmodium falciparum* (McCutchan *et al.*, 1988) and *P. berghei* (Gunderson *et al.*, 1987) were edited at the 5' and 3' ends to remove those bases which were absent from the three apicomplexan and two microsporan sequences determined in Chapter 3. The sequences were aligned using PILEUP (Devereaux *et al.*, 1984) and then compared visually with the alignment by Dams *et al.* (1988) and corrections made accordingly using the computer program LINEUP (Devereaux *et al.*, 1984) as shown in Chapter 5.

The secondary structure of *P. falciparum* 18S rRNA was derived from a correspondence between its primary sequences (McCutchan *et al.*, 1988) and the deduced secondary structure of *Drosophila melanogaster* 18S rRNA (Neefs *et al.*, 1991). Regions of ambiguity were resolved as for microsporans, using FOLD (Zuker and Steigler, 1981). The resultant structure was used as a template to determine the secondary structure for the apicomplexans. Pseudoknots (Pleij *et al.*, 1986) at E21-7 were aligned by eye, based on those predicted in *D. melanogaster* (Neefs *et al.*, 1991).

4.3 RESULTS and DISCUSSION

a. Microsporans

The sequences used to construct the models were respectively 826, 831 and 828 nucleotides long for *Nosema apis,* "Nosema vespula" and *Vairimorpha necatrix* (Vossbrinck *et al.*, 1987; Figures 4.1 and 4.2). The secondary structure models for *Nosema apis* and "Nosema vespula" were identical to each other and to that for

Vairimorpha necatrix (De Rijk *et al.*, 1992) except in helix 41 and helix 35 and its junction loop with helix 36 (Table 4.1). In *Nosema apis*, the hairpin in helix 41 was reduced to a single nucleotide, whereas in each microsporan sequence, helix 35 had a different arrangement of internal loops and bulges and the junction loop varied in size. The variability of helix 35 within and between microsporan genera suggests that it may prove to be a useful taxonomic marker within this phylum. Typically, all microsporan models lacked helix 44 and the entire V4 region (helices 21 and 22), even though a helix within this region is present in a secondary structure model created from a recently described sequence of the microsporan *Encephalitozoon cuniculi* (Hartskeel *et al.*, 1993).

Stem 20 in the secondary structure model for *Vairimorpha necatrix* 16S rRNA (Neefs *et al.*, 1991) may be incorrect. The published representation of this taxon (Neefs *et al.*, 1991) shows a non-canonical base pair C-C. If drawn in the same representation as the other taxa , the pairing becomes canonical (Figure 4.6).

b. Apicomplexans

The secondary structure models deduced from 18S rRNA primary nucleotide sequences were 1210 nucleotides each for "Gregarina caledia" and "Gregarina chortoicetes" and 1196 nucleotides for "Pseudomonocystis lepidiota". The models were typically eukaryotic, with all helices present including the variable helices 41, 43, 44 and 47 and a V4 region of more than 200 nucleotides which includes helices E21-1 to E21-8 and a pseudoknot at E21-7 and E21-8 (Figures 4.3, 4.4 and 4.5). Models for "Gregarina caledia" and "Gregarina chortoicetes" were almost identical to each other and varied only slightly from "Pseudomonocystis lepidiota" with changes in the sizes and occurrence of some internal loops, the largest difference between them being in helix 35 (Table 4.1). In figures 4.3, 4.4 and 4.5, helix 35 has been drawn as a eubacterial-like double hairpin with increased stability (Freier *et al.*, 1986; Figure 4.7). Helix 35 in other eukaryotes could also be accommodated as a double hairpin (Figure 4.7).

Other secondary structure elements of the microsporans conform to a eukaryotic pattern, except in at least two features. In helix 32, all microsporans have a standard Watson-Crick base pair (G-C, Figure 4.8) equivalent to that found in all eubacteria and archaebacteria. However all the other eukaryotes except "Gregarina

chortoicetes" have a non-canonical base-pair or single nucleotide bulge in the equivalent position. In the centre of helix 36, microsporans have a three nucleotide residue internal loop (Figure 4.9) identical in size and similar in sequence to that of eubacteria. However, all other eukaryotes have a single nucleotide bulge (adenine) in the equivalent position.

4.4 SUMMARY

A summary of secondary structure components shows that these may offer important insights into phylogenetic relationships. Although in microsporans and apicomplexans these components share almost 90% identity with prokaryotes and eukaryotes, the microsporans typically lack the V4 region and helix 44, but resemble the eukaryotes in helix 35 (Figures 4.1 and 4.2) and the prokaryotes in helices 32 (Figure 4.8) and 36 (Figure 4.9) implying that this taxon could be intermediate between the prokaryotes and eukaryotes. Alternatively, apicomplexans resemble eukaryotes in the V4 region and helix 41 (Figures 4.1 to 4.5; Table 4.1), but may resemble prokaryotes in helix 35 (Figure 4.7). Furthermore, the uncertain phylogenetic position of the zoomastigote *Giardia lamblia* as a sister group to all eukaryotes (Sogin *et al.,* 1989; Hashimoto *et al.,* 1994), is accentuated by the eukaryote-like structural components in helices 32 (Figure 4.8) and 36 (Figure 4.9).
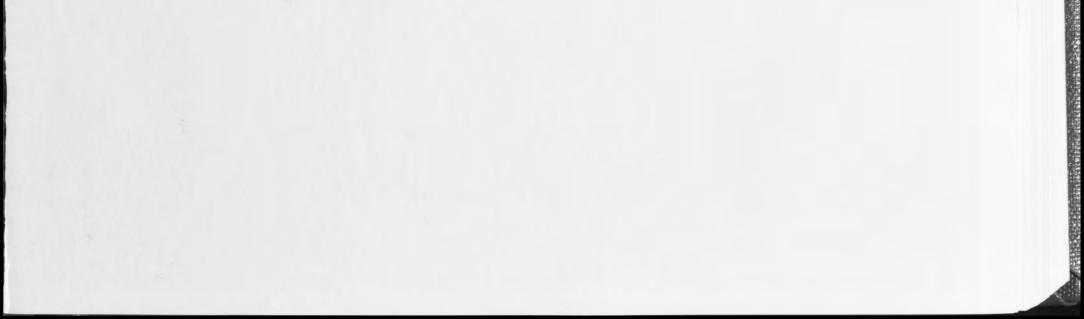
Figure 4

A model representing the components of secondary structure of SSU rRNA.

Definitions:
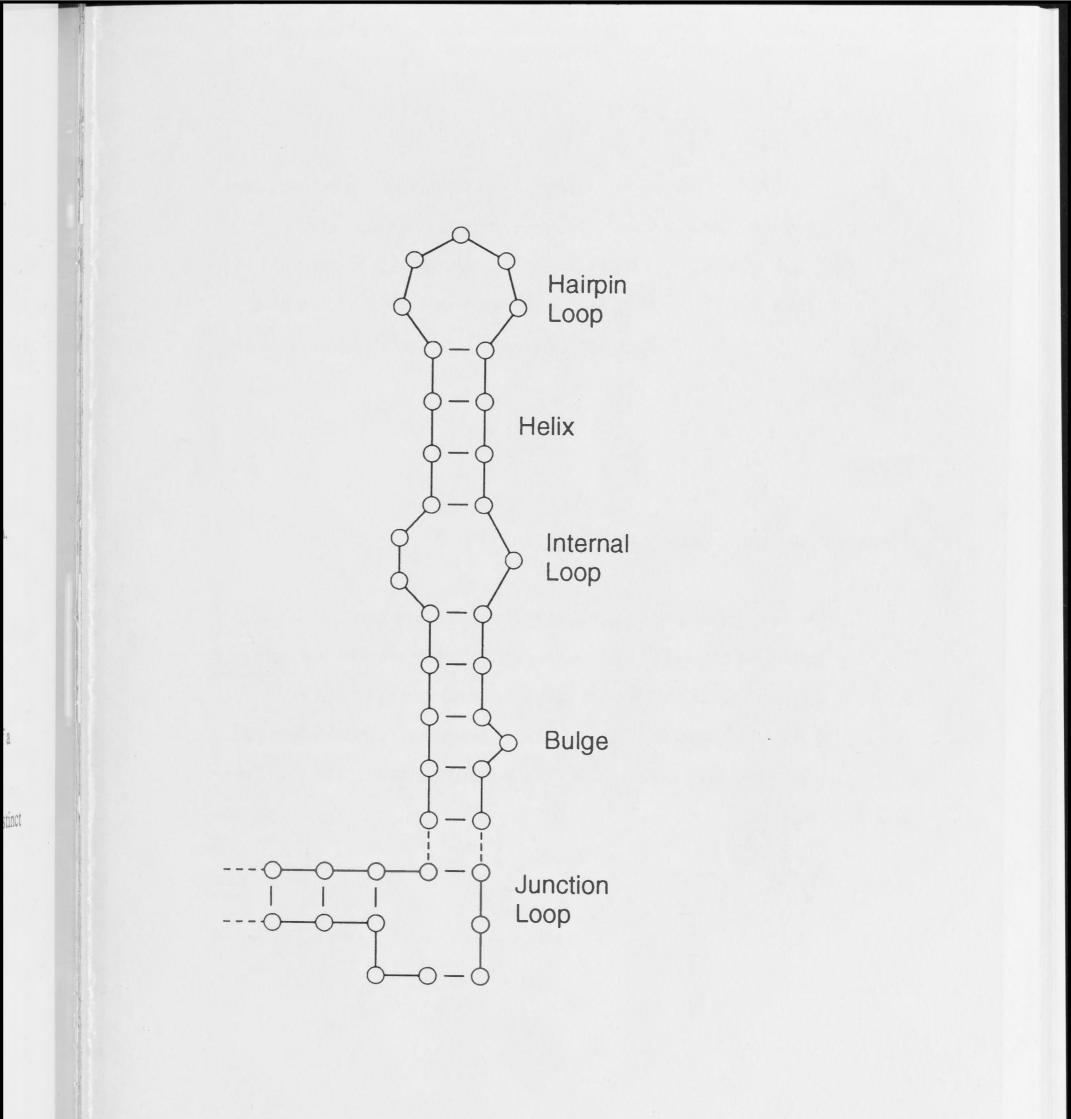
1. Helices are base-paired regions which form a stem.

2. Hairpin loops terminate a single stem.

3. Internal loops are non base-paired regions within a stem.

4. Bulges are non base-paired nucleotide(s) that occur in only one strand of a stem.

5. Junction loops are non based-paired regions that connect two or more distinct stems.

Hairpin Loop

Helix

Internal Loop

Bulge

Junction Loop
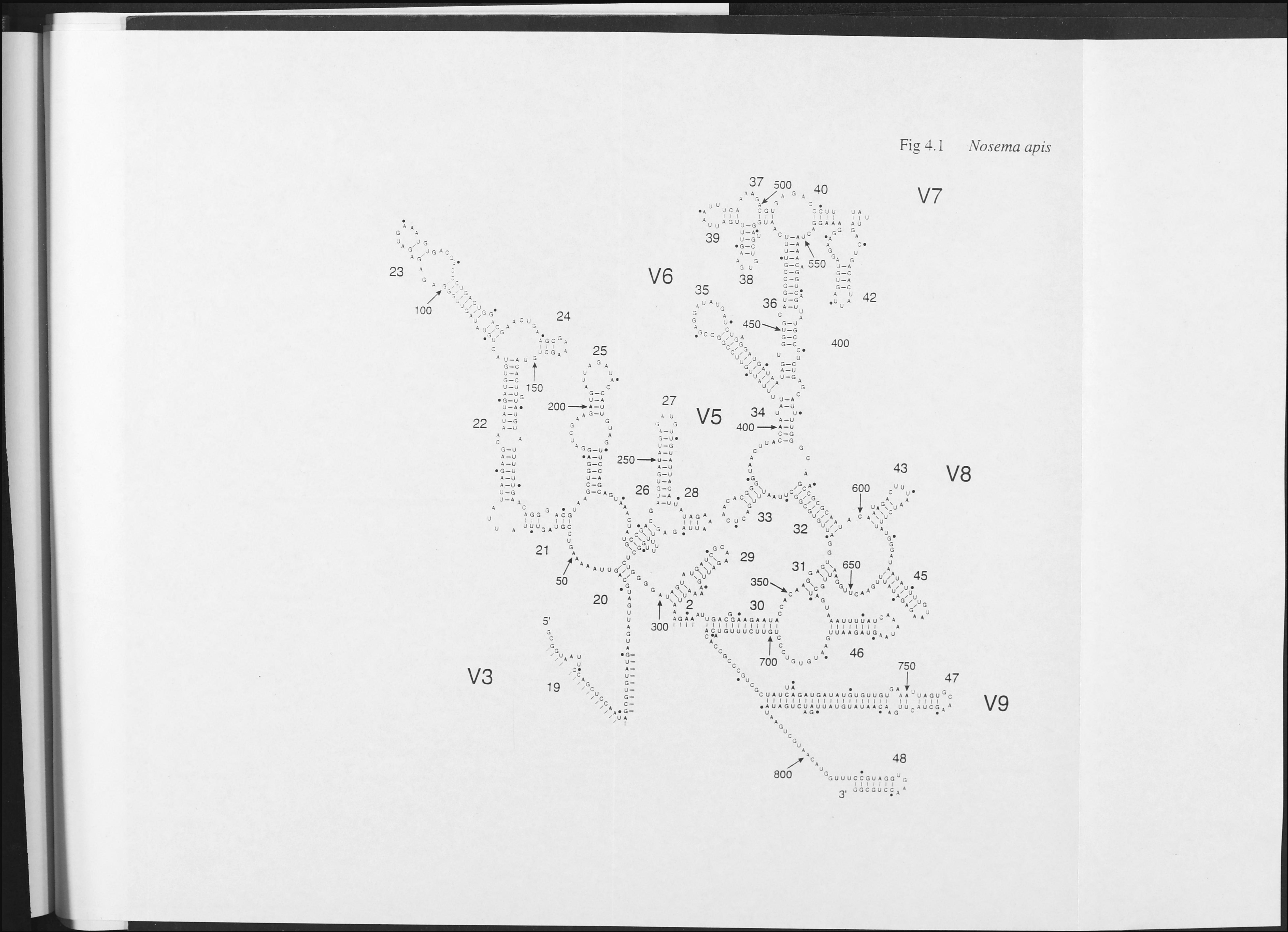
Figures 4.1 and 4.2
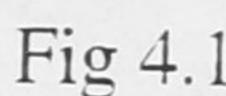
Secondary structure models of 16S rRNA constructed according to Neefs *et al.,* (1991). Structural components are numbered according to Neefs *et al.* (1991). The nucleotides are numbered with position 1 congruent with position 581 of the model of *Plasmodium falciparum* 18S rRNA proposed by McCutchan *et al.* (1988). Every tenth nucleotide is marked with a "." and every fiftieth is marked by an arrow and numbered.

Figure 4.1 *Nosema apis* 16S rRNA.
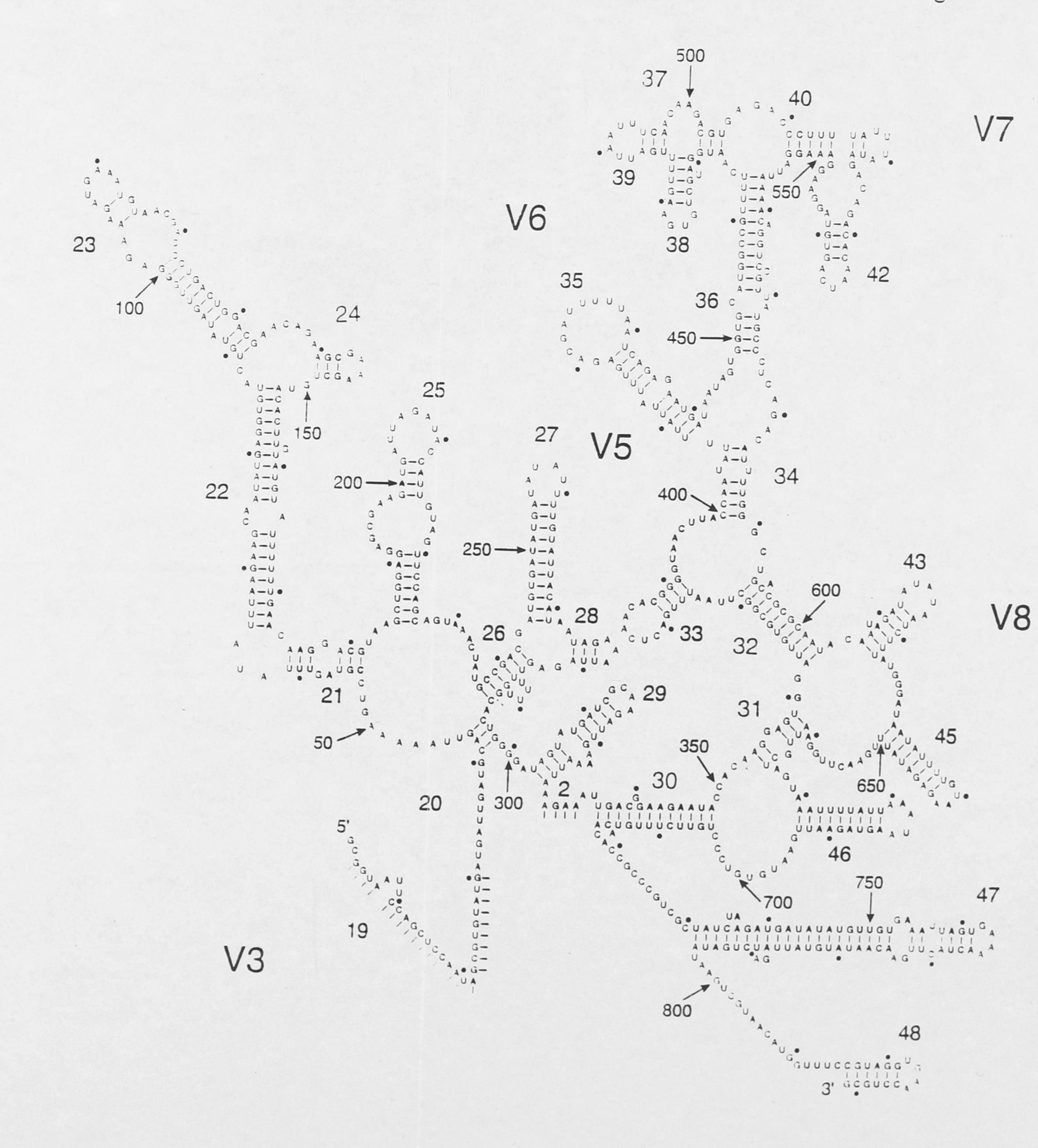
Figure 4.2 "Nosema vespula" 16S rRNA .

Helices have been given a different number if separated by a multi-branched loop (eg. helices 44 and 45) or by a single stranded area that does not form a loop (eg. helix 30). This analysis has opted for maximum Watson-Crick (canonical, A-U, G-C) and non Watson-Crick (non-canonical, G-U) base pairing although it may not always reflect the true secondary structure of rRNA (Zucker and Steigler, 1981).

Fig 4.1   *Nosema apis*

Fig 4.2   "Nosema vespula"

Figures 4.3 to 4.5

Secondary structure models of 18S rRNA constructed according to Neefs *et al.*, (1991) except for helix 35 which was constructed by me. Structural components are numbered according to Neefs *et al.* (1991). The nucleotide at position 1 corresponds to position 581 of the model of *Plasmodium falciparum* 18S rRNA proposed by McCutchan *et al.* (1988). Every tenth nucleotide position is marked by "." and every fiftieth position is marked by an arrow and numbered.
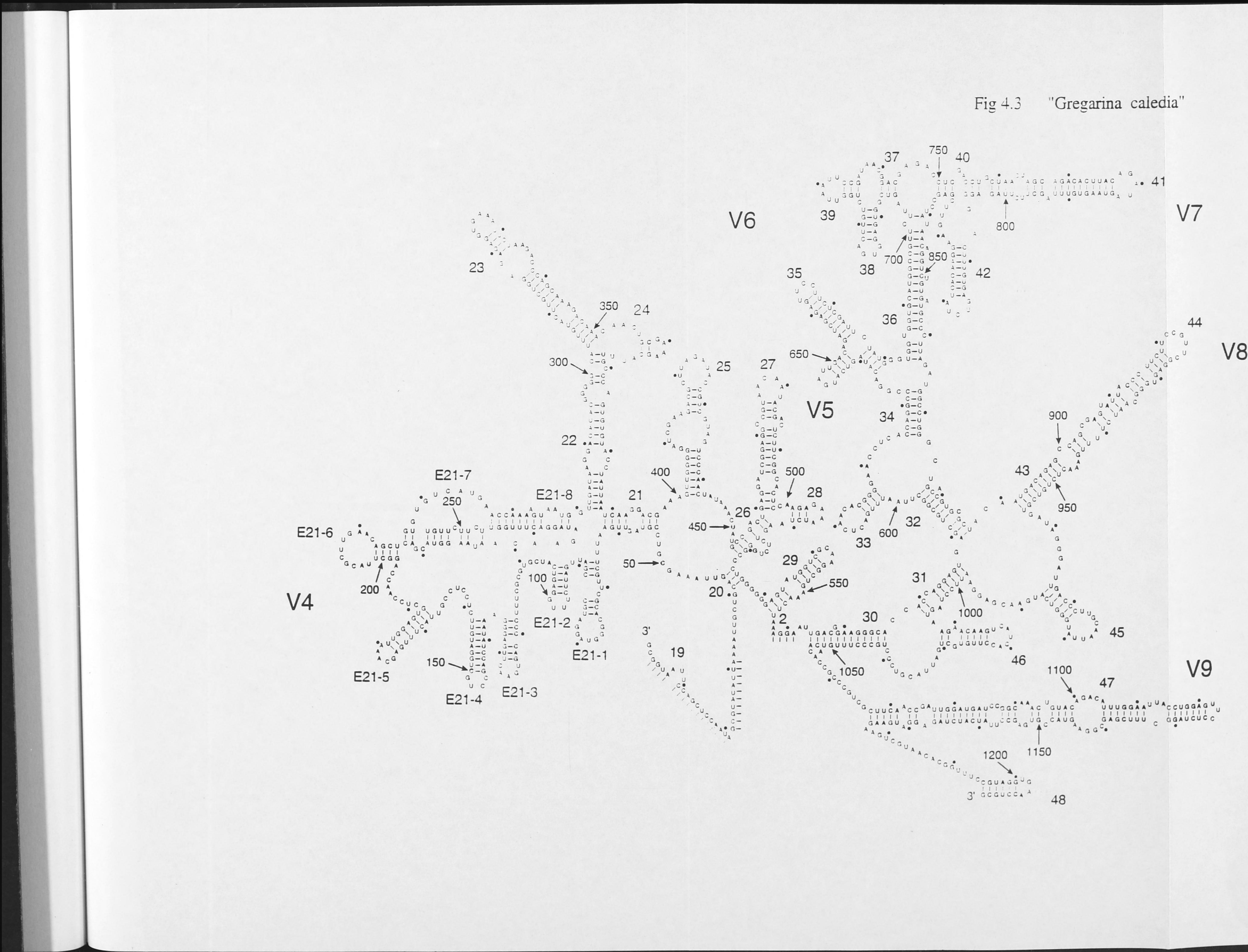
Figure 4.3 "Gregarina caledia" 18S rRNA

Figure 4.4 "Gregarina chortoicetes" 18S rRNA
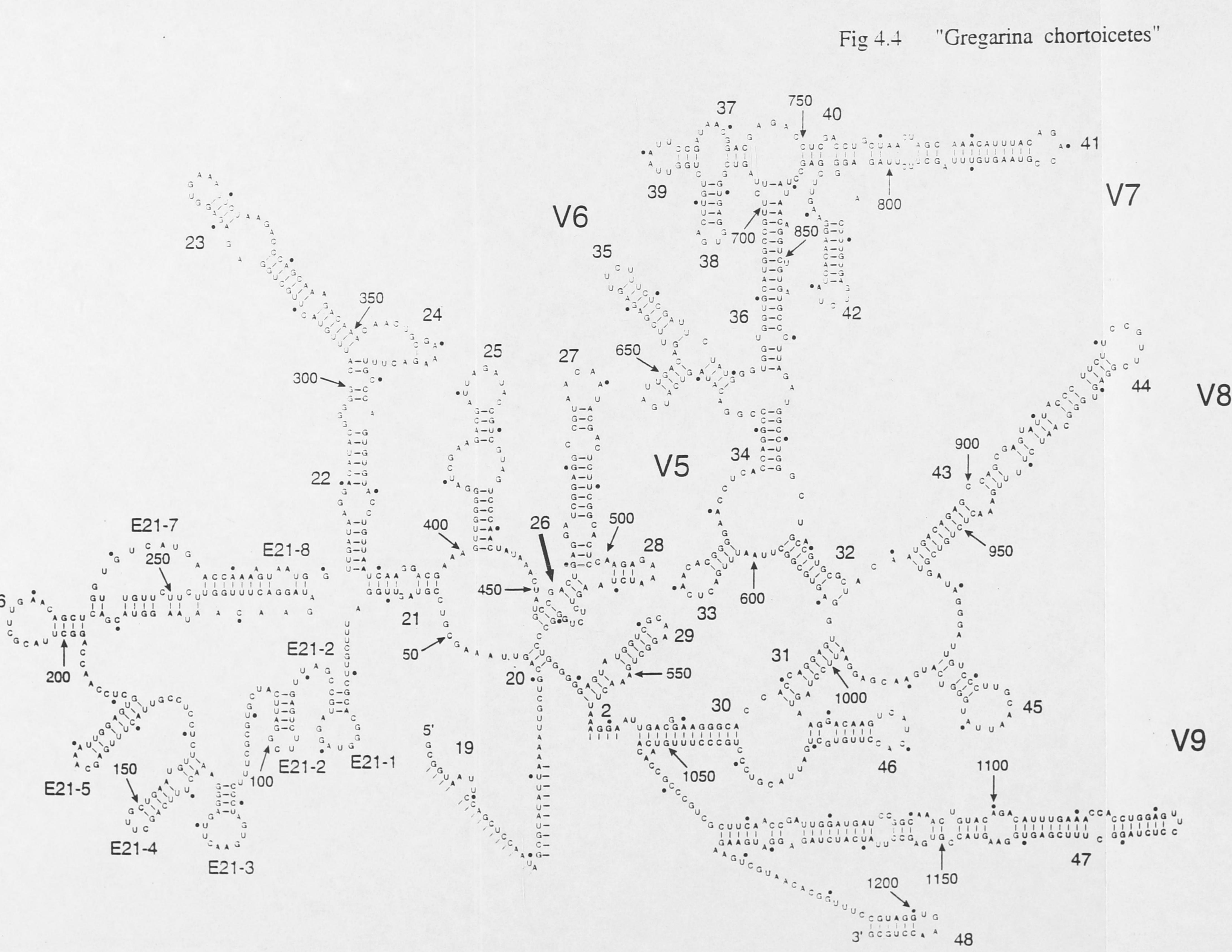
Figure 4.5 "Pseudomonocystis lepidiota" 18S rRNA
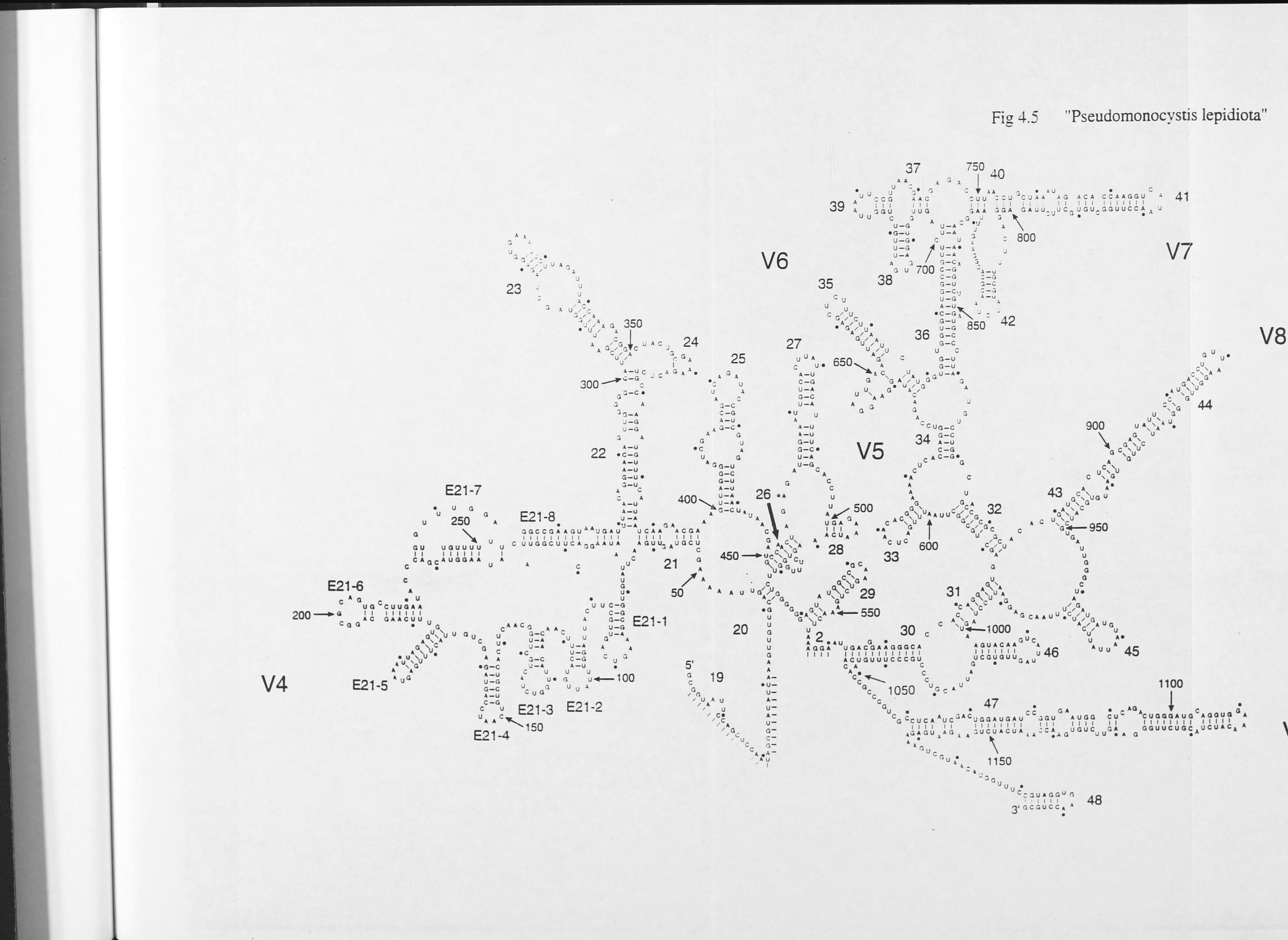
Helices have been given a different number if separated by a multi-branched loop (eg. helices 44 and 45), by a pseudo-knot loop (eg. E21-7 and E21-8), or by a single stranded area that does not form a loop (eg. helix 30). Additional helices specific to the eukaryotic model are numbered Ea-b, where a is the number of the preceding universal helix and b sequentially numbers all helices inserted between universal helices a and a+1 (Neefs *et al.*, 1991). This analysis has opted for maximum Watson-Crick (canonical, A-U, G-C) and non Watson-Crick (non-canonical, G-U) base pairing although it may not always reflect the true secondary structure of rRNA (Zucker and Steigler, 1981).
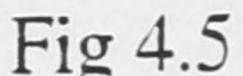
Helix 35 as shown is different from that suggested by Neefs *et al.* (1991) as it has a smaller free energy. This model and the one based on Neefs *et al.*, (1991) are compared in Figure 4.7.

Fig 4.3   "Gregarina caledia"

Fig 4.4 "Gregarina chortoicetes"

Fig 4.5   "Pseudomonocystis lepidiota"

Table 4.1

The empirically derived model of secondary structure of SSU rRNA for Prokaryota (De Rijk *et al.,* 1992) is compared with that for Eukaryota (De Rijk *et al.,* 1992) and with the models deduced in this study, for *Nosema apis*, "Nosema vespula", "Gregarina caledia", "Gregarina chortoicetes" and "Pseudomonocystis lepidiota".

Helices 19, 20, 21, 24, 25, 28, 29, 2, 30, 31, 33, 34, 37, 39, 42, 46 and 48 had very similar secondary structure components and are not shown.

All Eukaryote structures are compared to the generalised structure for Prokaryota

PROKARYOTA generalised secondary strucuture components		MICROSPORA N. apis	"N. vespula"	APICOMPLEXA "G. caledia"	"G. chortoicetes"	"P. lepidiota"	EUKARYOTA generalised secondary structure components
Helix							
V4 p21-1 p21-2	1 hairpin, 1 helix 2 loops 1hairpin, 1 helix partial junction	V4 absent except for partial junction	V4 absent except for partial junction	V4 present E21-1 to E21-8	V4 present E21-1 to E21-8	V4 present E21-1 to E21-8	as in "G. caledia" plus E 21-9 and possible E21-10
22	1 helix, 1loop	pro, plus 1 bulge	pro, plus 1 bulge	pro, plus 1 bulge 1 loop	pro, plus 1 bulge 1 loop	pro, plus 1 bulge 2 loops	pro
23	1 helix, 1 hairpin, 1 loop, part. junc.	pro, plus 1 loop	pro, plus 1 loop	pro, plus 1 loop	pro, plus 1 loop	pro, plus 1 loop	pro
26	1 helix partial junction	pro, plus 1 bulge	pro, plus 1 loop	pro, plus 2 bulges	pro, plus 1 loop	pro, plus 1 loop	pro, plus 1 large bulge
27	1 helix, 1 hairpin partial junction	pro	pro	pro, plus 2 loops	pro, plus 2 loops	pro, plus 1 loop	pro, plus 1 bulge
32	1 helix, 1 bulge partial junction	pro	pro	pro, plus 1 bulge	pro, plus 1 loop	pro, plus 1 bulge	pro
35	2 pronged structure 3 helices, 1 bulge 2 hairpins, 1 junc. (P35-1, P35-2)	1 helix 1 hairpin 1 bulge	1 helix 1 hairpin 1 loop	pro*, but no bulge	pro*, but 1 bulge	pro*, but 1 bulge	1 helix, 1 hairpin, 1 loop, 1 bulge

add footnote: extras = applies only to Helix 41 which may have a three-pronged hairpin terminating the helix

36	1 helix, 1 loop, 1 bulge partial junction	pro, plus 1 bulge	pro, plus 1 bulge	pro, plus 1 loop and 2 bulges	pro,plus 1 loop and 2 bulges	pro, plus 1 loop and 2 bulges	pro, plus 2 bulges
38	1 helix, 1 hairpin	pro, plus 1 bulge	pro, plus 1bulge	pro	pro	pro	pro
40	1 helix 1 partial junction	pro	pro	pro, plus 1 bulge	pro, plus 1 bulge	pro, plus 1 bulge	pro
41	1 helix, 1 hairpin (possible 3 pronged helices each with a hairpin = extras)	pro, but no extras and 1 reduced hairpin	pro, but no extras	pro, but no extras, plus 1 loop 2 bulges	pro, but no extras, plus 1 loop 2 bulges	pro, but no extras, plus 1 loop 3 bulges	pro, but no extras, plus 2 loops
43	1 helix, 1 loop partial junction	pro, plus 1 bulge	pro, plus 1 bulge	pro	pro	pro	pro
44	1 helix, 1 hairpin 1 loop	44 missing	44 missing	pro, plus 2 bulges	pro, plus 2 bulges	pro, plus 4 bulges	pro, plus possible E43-1
45	1 helix, 1 hairpin 1 junction,	pro	pro	pro	pro,plus 1 loop	pro	pro
47	1 helix, 1 hairpin 3 loops	pro, but 2 loops 2 bulges	pro, but 2 loops 2 bulges	pro, plus 4 loops	pro, plus 4 loops	pro, plus 4 loops and 1 bulge	pro

* Secondary structures of helix 35 for the three apicomplexans are shown with structures different from that of De Rijk *et al.* (1992), but with a smaller free energy calculated according to Freier *et al.* (1986).

pro = the structure appears to be the same as that deduced for the Prokaryota;   loop = internal loop;   a partial junction occurs when a helix which is one of several, meets at part of a junction loop

Figure 4.6

Base pairing (*-*) in region 20 of the secondary structure models of SSU rRNA from: (see Figures 4.1 to 4.5)

1. *Drosophila melanogaster* (Neefs *et al.*, 1991).

2. "Gregarina caledia"

3. "Gregarina chortoicetes"

4. "Pseudomonocystis lepidiota"
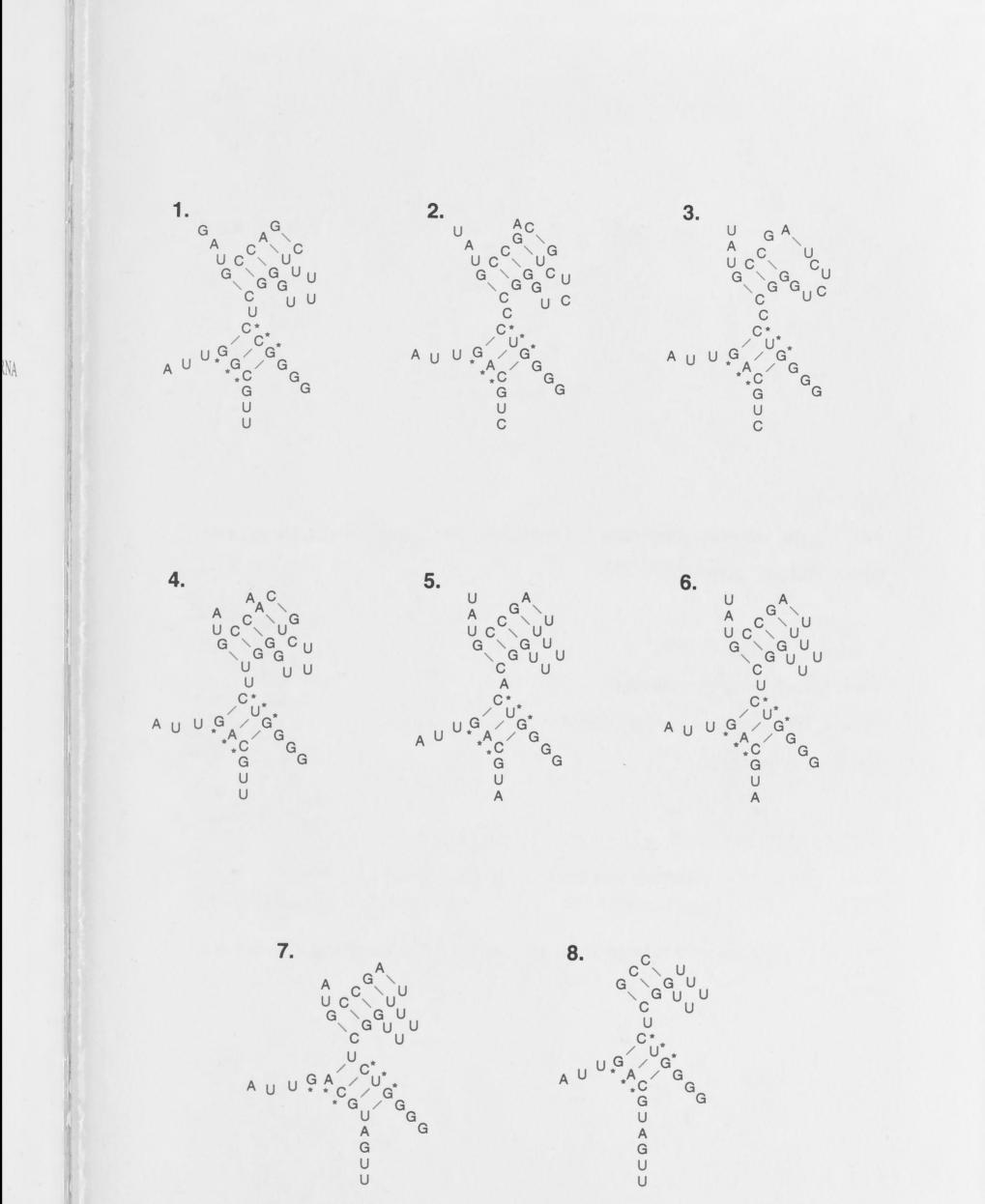
5. "Nosema vespula"

6. *Nosema apis.*

7. *Vairimorpha necatrix* (Neefs *et al.*, 1991).

8. *Vairimorpha necatrix* proposed change determined in this analysis.

1.

```
        G
      C A
  G A U C C   U C
    G A U   G G U U
        C   G G   U
        U
        C*  C
      C   * *   G*
A U U G G     G G G
      * *  C
        C
        G
        U
        U
```

2.

```
        A C
    U A U C C   G U C
        A U   G G U G
          C   C C   U
              C
              C*
          C     U
        C   * *   G*
A U U G     G G G
      * *  A
          C
          G
          U
          C
```

3.

```
        G A
    U A U C C   U C U C
        A U   G G U C
          C   C C   G
              C
              C*
          U     U
        C   * *   G*
A U U G     G G G
      * *  A
          C
          G
          U
          C
```

4.

```
        A A C
    A   U C C   G
    U G   G G U C U U
        U U   G G   U
              C
              C*
          U     U
        C   * *   G*
A U U G     G G G
      * *  A
          C
          G
          U
          U
```

5.

```
        G A
    U A U C C   G U U
        U G   G G U U
          C   C A   U
              C
              C*
          U     U
        C   * *   G*
A U U G     G G G
      * *  A
          C
          G
          U
          A
```

6.

```
        G A
    U A U C C   G U U
        U G   G G U U
          C   C U   U
              C
              C*
          U     U
        C   * *   G*
A U U G     G G G
      * *  A
          C
          G
          U
          A
```

7.

```
        A C G
    A   U C C   G U U
    U G   G G U U U
        C   C   U
              C
              C*
          U     U
      C   * *   G*
A U U G A     G G G
        * *  C
            G
            U
            A
            G
            U
            U
```

8.

```
        C C
      G   U C   G U U
          G U U U
              C
              C*
          U     U
        C   * *   G*
A U U G     G G G
      * *  A
          C
          G
          U
          A
          G
          U
          U
```
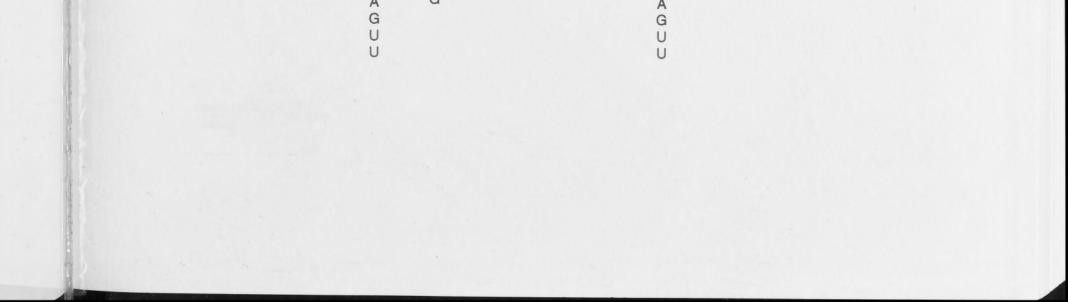
Figure 4.7

Existing and proposed secondary structures, with free energy values, for helix 35 of the SSU 18S rRNA of:
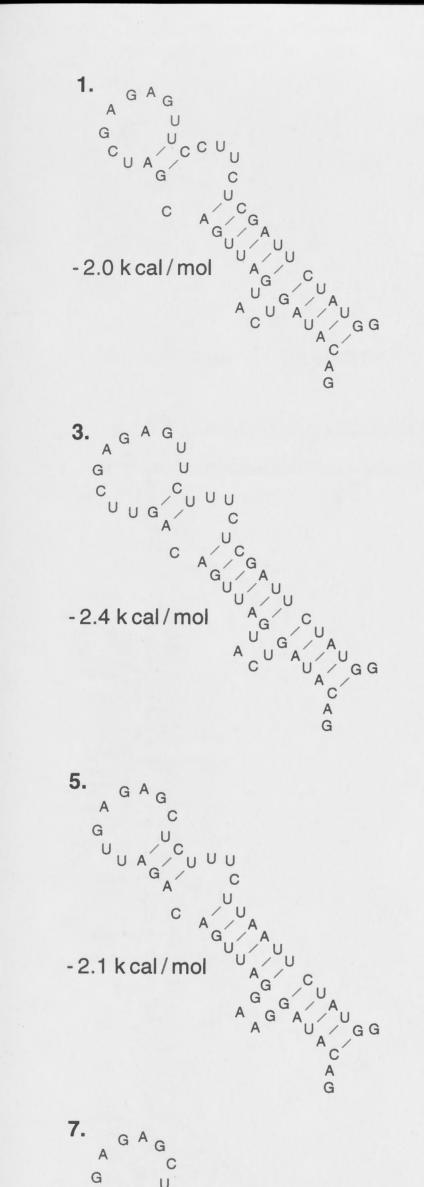
1 and 2 "Gregarina caledia"

3 and 4 "Gregarina chortoicetes"

5 and 6 "Pseudomonocystis lepidiota"

7 and 8 *Euglena gracilis*
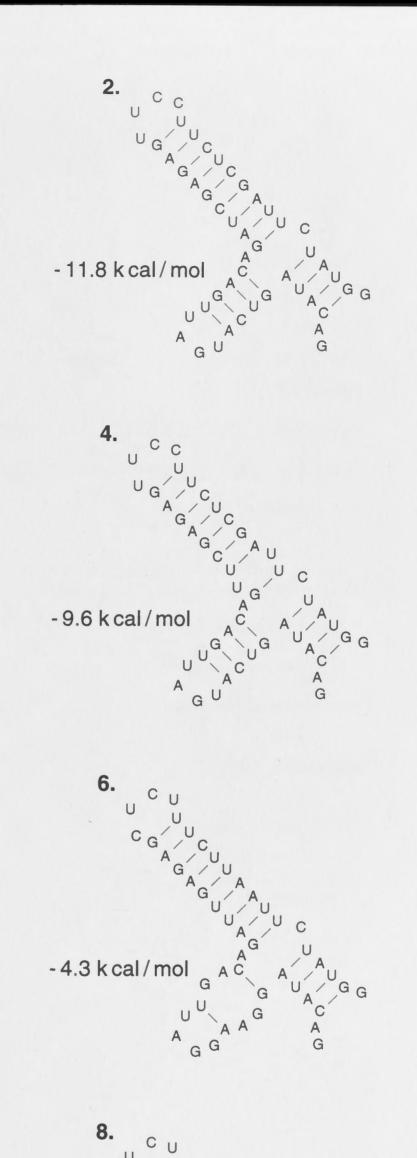
Existing structures based on Neefs *et al.* (1991) (1, 3, 5, and 7)

Proposed structures determined in this analysis (2, 4, 6 and 8)

Free energy values were calculated for 1M NaCl at 37°C according to Freier *et al.* (1986).

1.

−2.0 kcal / mol

2.

−11.8 kcal / mol

3.

−2.4 kcal / mol

4.

−9.6 kcal / mol

5.

−2.1 kcal / mol

6.

−4.3 kcal / mol

7.

−3.9 kcal / mol

8.

−8.7 kcal / mol

Figure 4.8

Part of helix 32 from a SSU rRNA secondary structure model showing Watson-Crick base-pairing (boxed) in microsporans, eubacteria and archaeans, but not in eukaryotes.
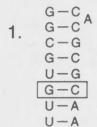
References

* Neefs *et al.*, (1991).

Hartskeel *et al.*, (1993).

+ Sogin *et al.*, (1989).

MICROSPORA   EUBACTERIA   ARCHAEA   EUKARYOTA

MICROSPORA
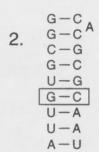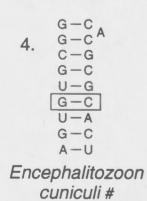
1.
```
G — C A
G — C
C — G
G — C
U — G
[G — C]
U — A
U — A
A — U
```
Nosema apis

2.
```
G — C A
G — C
C — G
G — C
U — G
[G — C]
U — A
U — A
A — U
```
"Nosema vespula"

3.
```
G — C A
G — C
C — G
G — C
U — G
[G — C]
U — A
U — A
A — U
```
Vairimorpha necatrix ∗

4.
```
G — C A
G — C
C — G
G — C
U — G
[G — C]
U — A
G — C
A — U
```
Encephalitozoon cuniculi #

EUBACTERIA

1.
```
G — C A
G — C
U — A
G — C
U — G
[A — U]
C — G
G — C
A — U
```
E. coli ∗

ARCHAEA

1.
```
G — C A
G — C
C — G
G — C
U — G
[C — G]
C — G
G — C
A — U
```
Halobacterium cutirubrum ∗

EUKARYOTA
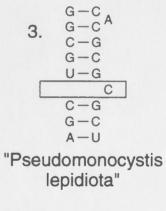
1.
```
G — C A
G — C
C — G
G — C
U — G
[A     C]
U — A
G — C
A — U
```
Euglena gracilis ∗

2.
```
G — C A
G — C
C — G
G — U
U — G
[       C]
C — G
G — C
A — U
```
"Gregarina caledia"

3.
```
G — C A
G — C
C — G
G — C
U — G
[       C]
C — G
G — C
A — U
```
"Pseudomonocystis lepidiota"

4.
```
G — C A
G — C
C — G
G — C
U — G
[C     U]
U — A
G — C
A — U
```
Giardia lamblia +
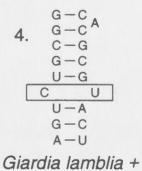
Figure 4.9

Part of Helix 36 from a SSU rRNA secondary structure model showing a similar bulge in microsporans, eubacteria and archaea (boxed) which is reduced to a single base bulge (adenine) in eukaryotes.

References

* Neefs *et al.*, (1991).

Hartskeel *et al.*, (1993).

+ Sogin *et al.*, (1989).

MICROSPORA   EUBACTERIA   ARCHAEA   EUKARYOTA

1.
```
    G—C A
    C—G
    C—G
    G—U
    G—C A
    U—G
    A—U
   ┌─────────┐
   │ C   U   │
   │     A   │
   └─────────┘
    G—U
    U—G
    G—C
    G—C
```
Nosema apis

1.
```
    G—C A
    U—A
    C—G
    G—U
    G—C
    U—A
    A—U
   ┌──────────┐
   │ C   C A  │
   │          │
   └──────────┘
    G—U
    U—G
    C—G
    G—C
```
E. coli *

1.
```
    G—U A
    C—G
    C—G
    G—U
    G—C
    U—A
   ┌──────────┐
   │ A     G  │
   │ C     U  │
   └──────────┘
    G—U
    U—G
    G—C
    G—C
```
Halobacterium cutirubrum *

1.
```
    C—G
    G—C
    C—G A
    C—G
    G—U
    G—C
    U—G U
    A—U
    C—G
    ┌───┐
    G—U│A│
    U—G└──┘
    G—C
    G—U
```
Euglena gracilis *

2.
```
    G—C A
    C—G
    C—G
    G—U
    G—C
    U—G C
    A—U
   ┌──────────┐
   │ C    U   │
   │      A   │
   └──────────┘
    G—U
    U—G
    G—C
    G—C
```
"Nosema vespula"
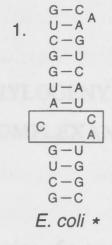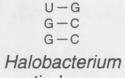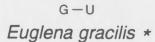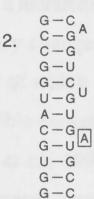
2.
```
    G—C A
    C—G
    C—G
    G—U
    G—C
    U—G U
    A—U
    C—G
    G—U┌──┐
    U—G│A│
    G—C└──┘
    G—C
```
"Gregarina caledia"

3.
```
    G—C A
    C—G
    C—G
    G—U
    G—C C
    U—G
    A—U
   ┌──────────┐
   │ C    U   │
   │      A   │
   └──────────┘
    G—U
    U—G
    G—C
    G—C
```
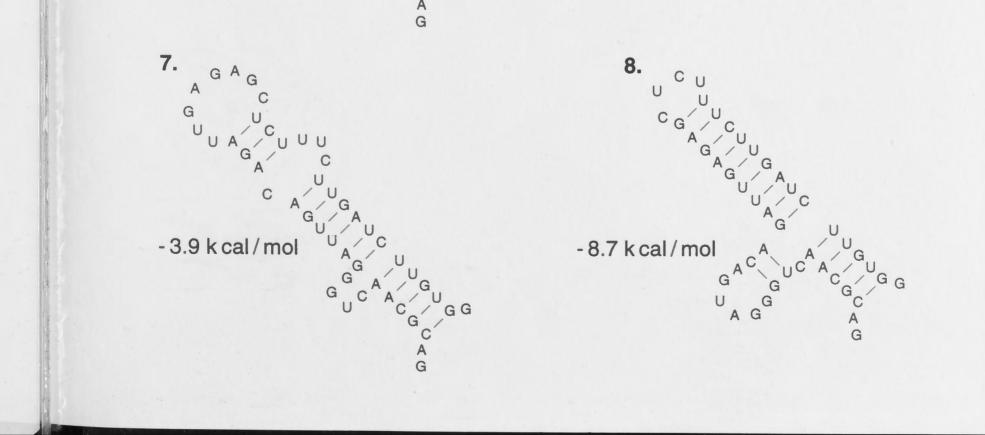*Vairimorpha necatrix* *

3.
```
    G—C A
    C—G
    C—G
    G—U
    G—C U
    U—G
    A—U
    C—G
    G—U┌──┐
    U—G│A│
    G—C└──┘
    G—C
```
"Gregarina chortoicetes"

4.
```
    G—C A
    C—G
    C—G
    G—U
    G—C C
    U—G
    A—U
   ┌──────────┐
   │ C    U   │
   │      A   │
   └──────────┘
    G—U
    U—G
    G—C
    G—C
```
Encephalitozoon cuniculi #

4.
```
    G—C A
    C—G
    C—G
    G—U
    G—C U
    U—G
    A—U
    C—G
    G—U┌──┐
    U—G│A│
    G—C└──┘
    G—C
```
"Pseudomonocystis lepidiota"

5.
```
    G—C A
    C—G
    C—G
    G—U
    G—C
    U—G
    A—U
    C—G
    G—U┌──┐
    U—G│A│
    G—C└──┘
    G—C
```
Giardia lamblia +

CHAPTER 5

PHYLOGENY RECONSTRUCTION
OF THREE APICOMPLEXAN AND TWO MICROSPORAN TAXA

5.1 INTRODUCTION

The genealogical relationships of a group of taxa that have evolved in a hierarchial way can be effectively represented as an evolutionary tree. Evolutionary changes are a function of the rate of evolution and evolutionary distance, neither of which can be observed directly except in the fastest evolving viruses, so the possible evolutionary history is inferred from data using phylogenetic algorithms (Hillis *et al.*, 1992). In earlier phylogenetic analyses of protoctistans using SSU rRNA sequences, trees have been constructed using distance methods (Fitch and Margoliash, 1967; Saitou and Nei, 1987) by Johnson *et al.* (1990) and Goggin and Barker (1993), and by parsimony analysis (Edwards and Cavalli-Sforza, 1963, 1964; Eck and Dayhoff, 1966; Kluge and Farris, 1969) by Barta *et al.* (1991), Ellis *et al.* (1992) and Gagnon *et al.* (1993; Figure 1, Chapter 1). For consistency, the same two methods will be used in this analysis.

The technical aim of tree-building is to discover which species share common ancestors, and in particular which are the most closely related, namely sister species. This study used SSU rRNA (18S, 16S) to establish genealogical relationships, the reasons for which have been stated (Chapter 1, page 5).

Similar character-states can either be homologous or analogous. In trees made from aligned molecular sequences, each aligned character is chosen to be a homologous nucleotide position, and its character-state is the particular nucleotide residue at that position. Gaps which are inserted during the alignment process to produce positional homology may be accepted as a fifth character-state (Needleman and Wunsch, 1970), or may be treated as missing. Homologous character-states are hierarchial and can be either shared-ancestral states (inherited from the common ancestor of the group), or shared-derived states that have been modified in particular lineages. Analogous character-states, or homoplasies, which are non-hierarchial and not inherited from a common ancestor, may arise by convergence, parallelism or

second paragraph, third sentence should read:

"Distance (D) values are functions which express the opposite or converse of similarity values (Equation (6c); Swofford and Olsen, 1990)."

reversal and result in the same character-states being found independently in different lineages (Swofford and Olsen, 1990; Maddison and Maddison, 1992; Stewart, 1993).

In phylogenetic systematics a distinction is made between the studied group or ingroup consisting of the descendants of an assumed single ancestor, and the outgroups consisting of groups closely related to, but not descended from, that ancestor (Maddison and Maddison, 1992). The outgroup is used as a comparison with the ingroup to determine the root of the ingroup (Swofford and Olsen, 1990). The simplest way of choosing the outgroup is to select this from a group of taxa which are closely related to, but not descended from the ancestor of the ingroup (Maddison and Maddison, 1992). Swofford and Olsen (1990) recommend using more than one outgroup taxon to test the assumption of ingroup monophyly.

5.1.1 Distance methods

Distance methods measure pairwise differences between taxa, and from these data, compute trees representing possible evolutionary relationships. This procedure applies to all types of character-state data, but the formulae presented here apply only to aligned sequences. The analysis has three steps.

Firstly, pairwise comparisons between the taxa are made, using some convenient metric, to calculate a distance matrix of percentage similarities or dissimilarities which relates all pairs of taxa. Similarity (S) values may be expressed as percentages ranging from 100 to 0% (Equations (6a) and (6b)). Distance (D) values are a function of the converse of similarity (Equation (6c); Swofford and Olsen, 1990). The most frequently used methods for calculating sequence similarities from aligned sequences use the following formulae:

$$S=M/L \qquad \text{Equation (6a)}$$
$$L=M+U+w_G G \qquad \text{Equation (6b)}$$
$$D=1-S \qquad \text{Equation (6c)}$$

Where S is the similarity, L is the total number of alignment positions or characters, M is the number of identical alignment positions or character-states, U is the number of non-identical alignment positions or character-states , w_G is the weight given to gaps, G is the number of gaps in the alignment and D is the dissimilarity or distance

(Swofford and Olsen, 1990). More subtle measures may focus on particular types of character-state data (e.g. transitions or transversions).

Secondly, the observed distances are transformed to estimated evolutionary distances by statistically correcting for super-imposed nucleotide substitutions (multiple hits) (Swofford and Olsen, 1990), using one of several available correction methods. For example, the Jukes and Cantor (1969) 1-parameter correction method assumes amongst other things, a random substitution among the four nucleotides. The 2-parameter correction method of Kimura (1980) assumes that transitions occur more frequently than transversions. Phylogenetic trees are said to be additive when the sum of the branch lengths separating taxa, namely patristic distances (Felsenstein, 1984), are equal to observed distances. Each branch represents a single taxon and its length represents the expected evolutionary distance between ancestors and their descendants. In practice, although there is an exact relationship between the observed distance data and the inferred tree, these distances may not fit the tree, so that the various correction weightings must be applied to minimise any error. The assumptions of these correction methods mean that gaps cannot be considered to be a fifth nucleotide, and must be eliminated from the analysis. It is possible to infer trees from raw data, but it is claimed that this may cause trees to be lineage-dependent with inconsistent topologies and branch lengths (Golding, 1983; Olsen, 1987).

Thirdly, trees are inferred from estimated evolutionary distances by one or more methods, of which those most frequently used are: neighbor joining (Saitou and Nei, 1987) which was used in this study, maximum likelihood (Felsenstein, 1988), and the more infrequently used distance methods of Sokal and Sneath (1963), Fitch and Margoliash (1967), Li (1981) and Farris (1972). For this analysis, I had considered using maximum likelihood instead of neighbor joining, but could not because of the prohibitive amount of computer time needed for the number of taxa involved. Maximum likelihood is the method of statistical inference applicable to phylogenetic analysis. It involves finding that evolutionary tree which yields the highest probability of evolving to produce the observed data. It does this by applying statistical methods to existing models, the Jukes and Cantor 1-P model (Jukes and Cantor, 1969) or the Kimura 2-P model (Kimura, 1980) of sequence evolution. Its particular strengths are claimed to be that it uses the total sequence including the variable regions which contain information specific to each taxon and which could

thus facilitate their resolution (Gatesy *et al.*, 1993), and that compared to parsimony methods, it is relatively more sensitive to changes in character-states (Felsenstein, 1981), and does not falsely group long branches together (Swofford and Olsen, 1990). The distance methods are united by their use of pairwise distance matrices rather than by their goals or biological assumptions, and may fail if assumptions used to derive the transformation of observed to estimated evolutionary distance fails (Swofford and Olsen, 1990).

5.1.2 Parsimony

The use of parsimony in phylogenetic inference was introduced by Camin and Sokal (Camin and Sokal, 1965). It involved reconstructing changes in a given set of characters on a given tree, counting the smallest number of times that a given event need have happened, and using this as a measure of the adequacy of the evolutionary tree (Felsenstein, 1978). Present day parsimony analyses have developed from the Hennigian school of cladistics (Hennig, 1966). In these and other tree-building analyses, members of the ingroup are arranged as a hierarchial, branching structure of monophyletic clades, which in parsimony analysis are defined by shared-derived characters (Felsenstein, 1988; Swofford and Olsen, 1990; Stewart, 1993). Hennig (1966) defined monophyletic clades as consisting of a common ancestor and all descendants of that ancestor and are distinguished by having at least one putative synapomorphy (shared-derived character state), whereas paraphyletic groups are distinguished by possession of only ancestral (plesiomorphic) character-states with no apparent synapomorphies and polyphyletic groups are distinguished by possession of convergent character-states (Farris, 1974). In principle, if two species share a recent, common ancestor which they do not share with any other taxon in the analysis, they are classified together regardless of their phenetic similarity to other forms (Ridley, 1989). Cladistic analysis discards *a priori*, invariate character-states which do not produce a hierarchial signal and unique character-states (autapomorphies) which exist in only one taxon (Fitch, 1977). Paraphyletic and polyphyletic character-states, which are not visible in sequence data until after they have been used in tree-construction, are discarded *a posteriori* on the basis of being analogous (Ridley, 1989).

Maximum parsimony is a valid criterion (Penny and Hendy, 1985) for the construction of phylogenetic trees and the methods used for its calculation from data

page 41, second paragraph,

page 42, first paragraph,

reference for PAUP should read:

"PAUP (Phylogenetic Analysis Using Parsimony; Swofford (1993)"

sets is one way of inferring trees (Atchley and Fitch, 1991; Hillis *et al.*, 1992). Swofford and Olsen (1990) describe several different algorithms for tree-building by parsimony which employ "exhaustive search" or the faster "branch-and-bound" method (Hendy and Penny, 1989) to identify optimal trees. However because of the large amount of computer time needed for these methods, branch-and-bound methods cannot be used for analysis of more than about 20 taxa. Exhaustive search is impractical when there are more than about 10 taxa. "Heuristic" algorithms, which are faster but do not guarantee to find the optimal tree, usually work well with up to about 50 taxa, and sometimes more, and these are generally used for tree searches on large data sets. All these methods evaluate alternative trees one character at a time, to determine the required number of evolutionary events to fit the data to each tree. Parsimony finds the tree(s) that maximises the number of character-state changes which can be interpreted as single historical events, therefore resulting in the "best" or most parsimonious tree which requires the fewest number of events to fit the data to the tree. However, it is to be noted that Sourdis and Nei (1988) have suggested that the assumption of minimum evolution could be flawed because evolutionary character changes in molecular sequences are not minimal and often involve backward and parallel substitutions. Homoplastic changes which cannot be mapped onto the tree are minimised (Swofford and Olsen, 1990; Maddison and Maddison, 1992).

PAUP (Phylogenetic Analysis Using Parsimony; Maddison and Maddison, 1992) is one of the available computer packages used for tree inference by parsimony. In this study I used version 3.1.1. (Swofford, 1993). PAUP can analyse any type of character data and finds the most parsimonious tree even for relatively large data sets (Stewart, 1993). Gene trees (Nei, 1987) built from molecular sequences will accurately reflect the evolutionary history of those sequences, but only if the molecules being compared are orthologous (divergent through speciation; Swofford and Olsen, 1990) and have not undergone gene conversion (Li and Graur, 1991). However, with any tree-building method, even a completely accurate gene tree does not necessarily match the evolutionary tree of the species from which it is taken, which in distance methods may be caused by the presence of polymorphism which causes the divergence times of genes sampled from different species, to be longer than those from species divergence (Nei, 1987).

page 42, the first sentence should read:

"PAUP (Swofford, 1993) offers several methods (Maddison and Maddison, 1992) which estimate the confidence levels: in the hierarchial structure of the data set (Archie, 1989; Faith and Cranston, 1991), in whole trees and in individual characters."

Estimations of confidence levels are available for the hierarchial structure of the data set (Archie, 1989; Faith and Cranston, 1991) and for whole trees and for individual characters, PAUP offers several methods (Maddison and Maddison, 1992).

A test of confidence for hierarchial structure is the cladistic permutation tail probability, or PTP (Archie, 1989; Faith and Cranston, 1991) tests for the presence of hierarchial structure in the data set. For this test the null model is used in which each character's states are re-assigned randomly to the tree, so that the resulting randomised data set represents random co-variation among the characters. The PTP tests a null hypothesis that the data have no cladistic structure other than that produced by chance. The hypothesis is rejected (say at the 0.05 level) if less than 5 of 100 trees have a length as short or shorter than the original tree (PTP < or = 0.05).

The consistency index, CI (Kluge and Farris, 1969; Farris, 1989) estimates the level of confidence in whole trees by measuring the fit of a character to a tree and determining an appropriate measure of homoplasy in the tree (Goloboff, 1991). The CI is defined as $c = m/s$ where s denotes the amount of change in the character (the number of steps) required parsimoniously by the considered tree and m represents the minimum amount of change that a character may show in a tree. The change may be partitioned into the observation, m and extra steps, h: $h = s-m$. If there is no homoplasy, the CI has a value of 1, and the homoplasy index (HI) has a value of 0 which implies maximum confidence in the data. Higher levels of homoplasy result in a larger number of steps and a lowered CI relative to HI.

Another technique, namely bootstrapping estimates the level of confidence in individual nodes by randomly resampling characters with replacement, to create a series of pseudosamples of the same size as the original data. A phylogeny is constructed from each pseudosample and the frequency of appearance of each particular phylogenetic group on the maximum parsimony tree amongst all the newly estimated trees is called the bootstrap confidence limit (Efron, 1982; Felsenstein, 1985) of that group. Comparative studies by Hillis and Bull (1993) showed that bootstrapping gives an imprecise measure of repeatability, but usually provides a highly conservative estimate of the probability of correctly inferring the corresponding clades. Bootstrapping is used to assess the statistical significance (e.g. at the 95% or 99% levels of statistical significance; even though a level of greater than or equal to 70% usually corresponds to a probability of greater than or equal to

95% (Hillis and Bull, 1993) that each phylogenetic group or node, is supported by the data. Felsenstein (1985) used 50 bootstrap replications and Hedges (1992) propounds that for a 95% level of significance, 2000 replications are needed for statistical accuracy, but because of the prohibitive amount of computer time needed, most authors use 20-100 replications.

A second test of confidence for individual nodes is the *a priori topology-dependent* cladistic permutation tail probability, T-PTP (Faith, 1991) which tests a null hypothesis that a particular hierarchial group in an observed tree could have arisen by chance alone. It does this by constraining the group to be evaluated for monophyly and non-monophyly (paraphyly or polyphyly). T-PTP differs from the PTP (Faith and Cranston, 1991) by being done after the tree(s) is made, using the data matrix of the phylogenetic tree and by preserving the identities of the taxa during the randomisation, thus explaining the relation between the data and a particular topology. From a certain group of taxa, two minimum-length trees made from actual data are constructed by PAUP, in which the hierachial group in question is constrained in one tree for monophyly and in the other for non-monophyly (alternatively, T-PTP can test for non-monophyly by comparing the shortest tree with that constrained for monophyly). The difference-value in the two different tree lengths is calculated by subtracting one from the other. Character-states in the data matrix are randomised amongst the ingroup taxa, for example 100 times, and two minimum-length trees are constructed from each randomisation. The difference-values in the lengths of the differently constrained trees made from the two sets of randomised data are calculated. Minimum tree differences made from the actual data are compared with those from the randomised data. The final proportion of values equal to or greater than the original difference value for actual data trees is referred to as the T-PTP. Faith (1991) suggested that the null hypothesis be rejected at the 95% level of statistical significance, if 95% of the difference values are equal to or less than those for the original constrained trees.

5.1.3 Problems in tree-building methods

The greatest problem in constructing trees from aligned nucleotide sequence data is caused by multiple nucleotide substitutions or multiple hits, at individual positions. For distance methods which employ ancestrally-derived and unique character-states in

addition to shared-derived states, this may result in an under-estimation of the number of evolutionary changes that separate any two sequences, and in parsimony which uses only shared-derived character-states, multiple hits may cause unobserved homoplasies which diminish the phylogenetic signal (Olsen and Woese, 1993).

5.1.4 Comparisons of parsimony and distance methods

It is understood that models of the evolution which determine the performance of methods for phylogenetic inference cannot entirely match the process of evolution. This results in a discrepancy which causes inconsistent results in phylogenetic analyses. The relative performances of tree-making methods are dependent on their robustness to violations of their underlying assumptions (Felsenstein, 1978).

Several simulation analyses (Sourdis and Nei, 1988; Saitou and Imanishi, 1989; Huelsenbeck and Hillis, 1993) and other analyses (Saitou and Nei, 1987) have been done to compare various tree-making methods. Both neighbor joining and parsimony methods are insensitive to differences in the rates of evolution in separate taxa and perform well under conditions of limited amounts of character change and variation in branch-length. Sourdis and Nei (1988) claim that parsimony methods are superior to distance methods when the number of nucleotide substitutions is small and equivalent to them when large data sets (10^3-10^4 nucleotides) are used, in which case neighbor joining and parsimony methods are equivalent, perhaps because parsimony is fairly robust to high proportions of homoplasy in the sequence data (Huelsenbeck, 1991) or alternatively because an increased number of characters does not necessarily mean a proportional increase in homoplasy (Goloboff, 1991). However parsimony methods have been criticised for not using all the available information in the data set, for general inconsistency, and for misleading results when branch lengths are unequal (Felsenstein, 1978). Parsimony methods also tend to erroneously group taxa on long branches together (Hendy and Penny, 1989), a condition which may be diminished by the addition of sequences from additional, appropriate taxa, if they are available (Swofford and Olsen, 1990).

Alternatively, the neighbor joining method performs consistently well when assumptions of distance corrections are met (Huelsenbeck and Hillis, 1993) but because it is only a clustering algorithm for estimating trees under minimum evolution, it only provides a point estimate of the group and does not guarantee an

optimal solution (Nei, 1991). Stewart (1993) argued that overall similarity and true evolutionary relationships may be best considered separate entities, as shown by the confounding of genealogical relationships between closely related species when a shared gene sequence in these evolves at different rates (Shaw *et al.,* 1989; Atchley and Fitch, 1991; Gillespie, 1991). As well, Ridley (1989) has criticised distance methods for using the total data set and thereby increasing homoplasy, while Penny (1982) believed that these methods are weak because information is lost in the transformation of the observed data, as the original data cannot be recovered from the estimated evolutionary distances.

Thus it is apparent that each of these methods has inherent strengths and weaknesses, neither is known to be more robust to violation of its assumptions than the other and neither can be said to infer the absolutely correct tree. Parsimony searches are 'goal-driven': they search for the shortest tree, while the NJ method amongst several others, (e.g. Hennigian 'phylogenetic systematics' and 'clique methods') are 'process-driven': construct a tree by following certain prescribed steps, whether that tree is the shortest or not (Swofford and Olsen, 1990). It would seem that the most sensible way to establish what are most likely to be phylogenetic relationships is to compare trees constructed from the same data set by several different tree-building methods. The greater the number of methods which produce the same inference of relationship, the greater the possibility that the inferred relationship is close to the truth or, at least, is not method-dependent.

5.2 AIM

The aim of the work reported in this chapter was to identify the evolutionary relationships of two microsporans "Nosema vespula" and *Nosema apis* and three apicomplexans namely, "Gregarina caledia", "Gregarina chortoicetes" and "Pseudomonocystis lepidiota", using their aligned SSU rRNA sequences. Phylogenetic trees constructed to confirm the suggested phylogenetic position of the nosemas as a sister group to the eukaryotes (Vossbrinck *et al.,* 1987), necessitated that these included representative sequences from the eubacteria, archea, plant, animal, fungal and protoctistan kingdoms (Table 5.1). The proposal of Levine (1988), that the gregarines form a primitive group of apicomplexans, derived from the dinoflagellates

page 46, 5.3 2a should read:

"Gregarina caledia", "Gregarina chortoicetes" and "Pseudomonocystis lepidiota" with sequences of the apicomplexan *Plasmodium falciparum* for the following reasons. Firstly because the primary malarian sequence was the only aligned apicomplexan sequence available, and secondly because the secondary structure models of the three gregarines (Chapter 4, Figures 4.3 to 4.5) and *Plasmodium falciparum* (De Rijk *et al.,* 1992) all had large insertions in helices 41, 44, and 47 and thus were easier to align with each other than with other protoctistan sequences."

(Chapter 1, page 2), both of which are thought to form a monophyletic group with the ciliophorans (Wolters, 1991), necessitates that phylogenetic trees included representative sequences from all these groups (Table 5.1).

Two tree-building methods were used; the distance matrix method with neighbor joining (NJ; Saitou and Nei, 1987) with the Kimura 2-P correction factor (Kimura, 1980) and maximum parsimony (Phylogenetic Analysis Using Parsimony, PAUP version 3.1.1: Swofford, 1993; Swofford and Olsen, 1990) with the cladistic permutation probability (PTP) test (Archie, 1989; Faith and Cranston, 1991), the *a priori topology-dependent* cladistic permutation tail probability (T-PTP) test (Faith, 1991) and bootstrapping (Felsenstein, 1985). Tree topologies were compared by eye and their patristic distances compared in scatter plots (DIPLOMO: Weiller and Gibbs, 1993, 1995).

5.3. PROCEDURES

5.3.1 Alignment of SSU rRNA sequences.

1. SSU rRNA sequences obtained from the GCG database (Table 5.1) were edited at their 5' and 3' ends to remove approximately 33% of the 5' ends of the sequences that were homologous to those of the three apicomplexans and two microsporans that I had not sequenced.

2. The alignment of SSU rRNA sequences from the three apicomplexans and two microsporans and 25 other taxa obtained from the public databases was done in the following way because of the differential availability of sequences, and to simplify the overall alignment process.

The complete SSU rRNA sequences determined by me were pre-aligned using PILEUP (Devereux *et al.,* 1984) with the following groups of published sequences:

a. "Gregarina caledia", "Gregarina chortoicetes" and "Pseudomonocystis lepidiota" with sequences of the apicomplexan *Plasmodium falciparum* because this sequence was the only aligned apicomplexan available (Neefs *et al.,* 1991) and because the secondary structure models of the three gregarines (Chapter 4, Figures 4.3 to 4.5) and *Plasmodium falciparum* (De Rijk *et al.,* 1992) had large insertions in helices 41, 44 and 47 these were easier to align when they were used together.

page 47, between points 3 and 4 add point 3a:

3a. The conserved primary sequence of *Giardia lamblia* was deduced from the complete 18S rRNA sequence and its secondary structure model (Sogin *et al.*, 1989) and aligned by eye to the other 29 aligned conserved primary sequences.

b. *Nosema apis* and "Nosema vespula" with the sequences of the microsporans *Encephalitozoon cuniculi* and *Vairimorpha necatrix* because these were the only available microsporan primary sequences which had been made into secondary structure models. These models all lacked the V4 region and helices 41 and 48, so the pre-alignment of microsporan sequences in the absence of other taxa, made the alignment easier.

c. The apicomplexan piroplasmids *Babesia bovis, Babesia bigemina, Babesia rodhaini, Theileria annulata* and *Sarcocystis muris* in an aligned form (Ellis *et al.,* 1992), were aligned at a later date with *Plasmodium falciparum* and the three gregarines.

It is noted here that, in each of the three sets of alignments (a, b and c above), the monophyly of each group was assumed. If the pre-alignment was used for tree-making, then the assumed monophyly could potentially result in false, phylogenetic relationships. The pre-alignment was conservative in the sense that there was no *a priori* knowledge of small (up to 5 nucleotides) insertion/deletion events represented as gaps.

3. All pre-aligned sequences (a, b and c above) were manipulated in the sequence editing program LINEUP (Devereux *et al.,* 1984) according to the following aligned, gapped SSU rRNA sequences published by Dams *et al.* (1988): *Trypanosoma brucei, Oxytricha nova, Paramecium tetraurelia, Euglena gracilis, Perkinsus sp. Prorocentrum micans, Plasmodium falciparum, Vairimorpha necatrix, Homo sapiens, Zea mays, Saccharomyces cerevisiae, Halobacterium cutirubrum, Sulpholobus solfataricus, Methanococcus vanniellii, Agrobacterium tumefaciens, Bacillus subtilis* and *Streptomyces coelicolor.* It is to be noted here that the pre-aligned sequences from 2 a,b and c above were subjected to *a priori* knowledge of gaps in this manipulation.

4. The final alignment of the complete SSU rRNA sequences including gaps consisted of 2048 nucleotides which represented approximately 66% of the total 18S rRNA gene from *Plasmodium falciparum* (McCutchan *et al.,* 1988).

5. The conserved regions of the above 30 sequences which were used for tree-making, were selected by visually comparing the aligned primary sequences with secondary structure models of *Plasmodium falciparum, Vairimorpha necatrix* (De Rijk *et al.,* 1992) the three apicomplexans and two microsporans (Figures 4.1 to 4.5, Chapter 4). Conserved nucleotides in the primary sequence alignment (Figure 5.1) were 1-59, 60-255, 256-369, 370-441, 442-524, 525-641 and 643-715 and these corresponded to

page 48, line 8 should read:

"The remainder of characters were excluded from the phylogenetic analysis because they are autapomorphic (taxonomically unique);..."

conserved positions 1-59, 275-456, 520-631, 678-750, 813-890, 957-1069 and 1127-1196 in the secondary structure model (Figure 5.2); position 1 is the same as position 581 in the *Plasmodium falciparum* SSU rRNA sequence given by McCutchan *et al.* (1988). This model of secondary structure (Figure 5.2) made from the primary SSU rRNA sequence from *Pseudomonocystis lepidiota*, represents all other secondary structure models, and shows the placement of conserved nucleotide positions which were used for tree-making. It has been included only as a visual representation of a generalised secondary structure model. The remainder of characters were excluded from the phylogenetic analysis because of they are autapomorphic (taxonomically unique); they could not be aligned with certainty or because not all positions were available for all taxa. Although gaps could well have been included in the parsimony method, they are ignored by the 2-parameter correction algorithm in distance methods and therefore have been omitted from the analyses.

5.3.2 Phylogenetic analyses

For trees calculated by distance methods, tree-making programs were accessed through Australian National Genomic Information Service (ANGIS) (University of Sydney). Distance matrices were constructed using the program in CLUSTAL W (Thompson *et al.,* 1994) with the Kimura (Kimura, 1980) 2-parameter algorithm. Trees were calculated and bootstrapped with 1000 pseudosamples, using the NEIGHBOR program in PHYLIP version 2.0 (Felsenstein, 1993)and drawn with T-DRAW, version 1.4 (J.W.H. Ferguson , personal communication). Trees made by the assumptions of parsimony were calculated by PAUP using the heuristic search algorithm with unweighted characters, 100 random addition sequence starting trees and tree bisection-reconnection branch swapping. Gaps were regarded as 'missing'. Bootstrap resampling with 1000 pseudosamples was used to quantify relative support for branches on trees made by distance and parsimony methods and significant values of 80% to 100% are shown. Bootstrapped trees made by parsimony, were presented separately to show which branches are collapsed to polytomies and thus provide information about instability in the relative phylogenetic positions of taxa.

page 49 heading "Rationale for selection of trees for analysis" should read: "Rationale for selection of trees and outgroups for analysis"

Rationale for selection of trees for analysis

1. Trees were constructed by both methods from sequences representing the "30 multikingdom taxa". These were constructed mainly to establish the positions of the microsporans relative to the prokaryotes and eukaryotes, but they also gave indications of the relationships of the other taxa. In the parsimony analysis the archaeans and eubacteria were defined as the outgroups, as these had been shown before to have not descended from the ancestor of the ingroup (Olsen, 1987, Sogin *et al.,* 1989 and Van Keulen, 1993). In the distance analyses, the default option of rooting the tree midway between the two most distant taxa was used, which assumes that the rate of evolution of these taxa has been the same (Felsenstein, 1984). The ingroup taxa were found by this tree to be monophyletic which supports the parsimony analysis.

2. A "29 taxa" tree was made by parsimony in exactly the same way as the "30 taxa" tree, but without *Giardia lamblia*. This tree acted as a comparison with the "30 taxa" tree, to see the effect, if any, of *Giardia lamblia* in that tree. In another study (Sogin *et al.,* 1989), this zoomastigote was placed more closely to the bacteria than to the eukaryotes, and it is possible that its presence in the "30 taxa" tree may have influenced the inference of the microsporans.

3. Trees were made from eukaryote taxa. Although the "30 taxa" tree inferred the ingroup as being monophyletic, the relationships of the three gregarines within the eukaryote clade was not clarified because of the large evolutionary distances (represented by long branch lengths) between the eukaryotes and the microsporans and the bacteria outgroup. Therefore, to reduce the evolutionary distance, the sequences from the bacteria, microsporans and *Giardia lamblia* were deleted from the data set. Of the remaining eukaryote taxa, the multicellular taxa (*Homo sapiens* and *Zea mays*) the fungus (*Saccharomyces cerevisiae*), the euglenid (*Euglena gracilis*) and the zoomastigote (*Trypanosoma brucei*) were specified as outgroups in the tree made by parsimony, as these were shown in the "30 taxa" tree made by parsimony, to have not descended from ancestors of the ingroup.

Two sets of trees were constructed by neighbor joining and parsimony methods. The first set, a "15 protoctist taxa" tree, with *Trypanosoma brucei* and *Euglena gracilis* as the outgroups (results are not shown), both inferred the apicomplexans as polphyletic with the other protoctists suggesting that the evolutionary distance between the outgroup and the ingroup was still too large. The second set of "19 eukaryote taxa"

trees were made to further decrease the evolutionary distance between the basal ingroup and the basal outgroup, by including *Homo sapiens, Zea mays* and *Saccharomyces cerevisiae* with *Trypanosoma brucei* and *Euglena gracilis* in the outgroup. A third tree made only by parsimony, of "17 eukaryote taxa" was made which also reduced the evolutionary distance, by deleting the more evolutionary distant *Trypanosoma brucei* and *Euglena gracilis* from the outgroup. This also acted as a comparison with the second tree to see if these two taxa affected the inference of the ingroup. Only the "30 taxa" and "19 taxa" selected for the parsimony analysis were investigated for hierarchial structure with the PTP test. The "29, 15 and 17 taxa" data sets are sub-groups of the "30 and 19 taxa" data sets and therefore were assumed to have significant structure.

4. Trees made with fewer taxa which could resolve the position of gregarines, were considered. However, the reported monophyly based on morphological evidence and rRNA sequences (Lynn and Small, 1981; Kumazaki *et al.,* 1983; Johnson *et al.,* 1987 Barta *et al.,* 1991; Gajadhar *et al.,* 1991 and Wolters, 1991) which is thought to exist between the ciliophorans, the dinoflagellates and the apicomplexans, and the very small evolutionary distance between these and *Acanthamoeba castellanii,* left no taxa other than those already used, to act as outgroups.

5.4 RESULTS and DISCUSSION

5.4.1 The "30 taxa" trees

PTP tests (page 42) of the data set showed that there is significant structure to the 423 informative sites (PTP = 0.01 from 100 randomised and 1 actual tree). Figures 5.3 a,b and c are the trees made by NJ and PAUP programs with bootstrapping which represent the relationships of the "30 taxa". The relatively high CI indices and the number of statistically significant bootstrap values indicate that the information in these trees is significant. It can also be seen that the topologies of the neighbour joining and parsimony trees are closely similar, and this was confirmed by plotting the cladistic distances of these trees against one another (Figure 5.4).

Both tree-building methods grouped the data set into three clades. These were: an outgroup clade formed by the bacteria and *Giardia lamblia*, an ingroup clade formed of the eukaryotes and a third, smaller clade composed of the microsporans

which lies between the two larger clades. Both methods show 100% bootstrap support for the bacterial and microsporan clades, but infer the eukaryote clade (without the microsporans) as a poorly resolved polyphyly of multi-cellular and protoctist taxa. These groupings support earlier findings by Olsen (1987), Sogin *et al.* (1989), Knoll (1992) and Van Keulen (1993).

a. Phylogeny of the microsporans

Trees made by both methods (Figure 5.3a and b) showed that, except for the zoomastigote *Giardia lamblia*, the microsporans formed a monophyletic clade which appears to be the sister group to the other eukaryotes. This confirms the findings of Vossbrinck *et al.,* (1987) and satisfies one of the aims of my work. However, there are two reasons why this interpretation must be treated with caution. Firstly, *Giardia lamblia* which is inferred as a sister to the microsporans and *Trypanosoma brucei* which is inferred as derived from the microsporans, are both zoomastigotes (Table 5.1). In this data set, it is not possible to say which of these is in the correct evolutionary position, therefore, this taxon could be the sister group to the other eukaryotes. Secondly, in both trees made by NJ and PAUP, it could be construed that the evolutionary distances which separate the microsporan lineage from the *Trypanosoma brucei* and *Euglena gracilis* lineages, are too small to distinguish which of these taxa is the real sister group to the eukaryotes.

To investigate this resolution, it was decided to test for specific hypotheses of monophyly for the "eukaryotes minus the microsporans" and the "microsporans" alone in two *a priori* T-PTP tests (page 43) of the 30 taxa with the 6 bacteria as outgroups and 423 informative characters. If each of these groups was shown to be significantly monophyletic, then it could be said that the microsporans are the sister group to the other eukaryotes.

Initially, the "eukaryotes minus *Giardia lamblia* and the microsporans" were constrained for monophyly and non-monophyly. A heuristic search resulted in shortest trees of 1745 and 1748 steps respectively, with a difference of +3 steps (Table 5.2). *Giardia lamblia* was excluded from the constraint because it had not been inferred in the eukaryote clade by either tree-making method. However, to justify this exclusion, its phylogenetic relationship to the eukaryotes was tested in two subsequent heuristic searches of the same 30 taxa and outgroups, in which it was constrained firstly to

appear with the other eukaryotes, and then with the other zoomastigote, *Trypanosoma brucei* (Table 5.2). Results showed that compared to the 1745 steps needed for the "eukaryotes minus the microsporans and *Giardia lamblia*", it took 10 extra steps (1755 minus 1745) to force *Giardia lamblia* into the eukaryote clade, and 37 extra steps (1782 minus 1745) to force *Giardia lamblia* into a clade with *Trypanosoma brucei*. The accompanying phylogenetic trees (not shown) inferred that all branches, except those to the microsporans and bacteria were collapsed to polytomies. This justified its deletion from the constrained eukaryote clade. Of the 100 randomised data sets plus one from the actual data set, 100/101 had less than the 3 step difference. Thus parsimony analysis with *a priori* T-PTP testing supported the monophyly of the constrained eukaryote clade (T-PTP = 0.01, 3 steps, p=<0.01) and the null hypothesis was rejected. Thus the "eukaryote clade minus the microsporans and *Giardia lamblia*" can be said to be monophyletic in this data set.

In the second *a priori* T-PTP test the "microsporans" were constrained for monophyly and non-monophyly. A heuristic search resulted in shortest trees of 1745 and 1783 steps respectively, with a difference of +38 steps (Table 5.2). Of the 100 randomised data sets plus one from the actual data set, 100/101 had less than the 38 step difference. Thus parsimony analysis with *a priori* T-PTP testing supported the monophyly of the constrained "microsporan" clade (T-PTP = 0.01, 38 steps, p=<0.01). The significant support for monophyly in both constrained clades indicate that the microsporans are the sister group to the other eukaryotes minus *Giardia lamblia*. This finding confirms the findings of Vossbrinck *et al.* (1987).

Within the microsporan clade, *Encephalitozoon cuniculi* forms a sister group to the other three microsporans and is separated from them by a large evolutionary distance. This supports its taxonomic separation from the others at the class level (Class Haplophasea; haplosis by nuclear dissociation, haploid nuclei throughout the life cycle: Sprague *et al.*, 1992) and its phylogenetic position based on LSU and SSU rRNA sequences (Vossbrinck *et al.*, 1993; Chapter 1, page 9). Unexpectedly, in this tree, the nosemas formed a paraphyletic group, with *Nosema apis* as sister to a clade comprising "Nosema vespula" and *Vairimorpha necatrix*. This conflicts with their current taxonomies and host range. Although these species belong to the same class (Dihaplophasea; Chapter 1, page 9), the chromosome cycles of the nosemas, which parasitize hymenopterans, differ from that of *Vairimorpha necatrix*, which has a

lepidopteran host. These differences are used to place them in different taxonomic orders (Dissociodihaplophasida and Meiodihaplophasida respectively; Sprague *et al.,* 1992). Of the twelve "29 taxa" most parsimonious trees made by parsimony (length, 1698 steps; CI, 0.530; results not shown), all showed the microsporans in a monophyletic cluster which formed a sister group to the eukaryotes. Six trees inferred the microsporan clade in the same way as the "30 taxa" trees and six trees showed *Vairimorpha necatrix* as the sister group to a clade comprising the two nosemas. This suggested that the presence of *Giardia lamblia* in the "30 taxa" tree had little effect on the inference of the microsporans, but that the relationship of the nosemas with *Vairimorpha necatrix* could be the result of chance.

An *a priori* T-PTP test (page 43) was done to test whether the microsporan relationships inferred in the trees could have arisen by chance alone. Following the procedures suggested by Faith (1991), a matrix was constructed which contained the four microsporan sequences and an outgroup "taxon" made from the nucleic acid sequence of a hypothetical ancestor represented by the node (node 39 marked on Figure 5.3b) from which the microsporans were derived. This node had been reconstituted in the PAUP program during the making of the "30 taxon" tree, from the sequences of closely related taxa (John Trueman, Division of Entomology, CSIRO, Canberra, Australia, personal communication) and was retrieved from the data matrix using the PAUP option for creating ancestral states. This was done using the accelerated and delayed transformation options separately, and then combining the results to give character states which are definite for that node. The reason for the use of this node rather than a known sequence from an outgroup taxon, was that of the observed outgroups (Figure 5.3 a and b); the phylogenetic position of *Giardia lamblia* was uncertain and the bacterial outgroup was too evolutionary distant. This left node 39 as the only possible outgroup which was closely related to, but not descended from the ingroup in question. PTP tests of the data set showed that there is significant structure to the 100 informative sites (PTP = 0.01 from 100 randomised and 1 actual tree). For the *a priori* T-PTP, branch-and-bound searches with 100 replicates calculated minimum-length, unrooted trees in which the two nosemas were constrained to appear not as a monophyletic group and as a monophyletic group. These were of 200 and 203 steps respectively, indicating that only 3 steps separated the two states. Of 100 randomised data sets, 46 had more, 13 had the same and 41 had less than the

three step difference. Thus parsimony analysis with *a priori* testing did not support non-monophyly of the constrained group (T-PTP = 0.59, 3 steps), and therefore the null hypothesis was not rejected. Thus the arrangement of the nosemas in the observed trees made from the "30 taxa" matrix, could be a chance event. Thus these molecular data have not provided sufficient evidence to reject the existing taxonomy based on phenotypic characters (Sprague *et al.*, 1992) or the inter-generic relationships suggested by LSU rRNA sequences (Baker *et al.*, 1994; Chapter 1, page 10).

b. Phylogeny of the bacteria and eukaryotes

The "30 taxa" trees also gave information about the relationships of the bacteria and the eukaryotes. Within the eubacterial clade, the two methods (Figure 5.3 a and b) conflicted in the relationships they inferred for *Streptomyces coelicolor, Bacillus subtilis* and *Agrobacterium tumifaciens* which is evidenced in both methods by non-significant bootstrap values, and in parsimony by the collapsing of their branches (Figure 5.3c). As well, parsimony inferred that the archaeans *Halobacterium cutirubrum* and *Methanococcus vaniellii* formed a sister group to the eubacteria with *Sulfolobus solfataricus* being the sister to the other five bacteria supported by 100% bootstrap. By contrast, the neighbour joining tree placed the archaeans and the eubacteria in separate sister clades, which is a more likely result (Olsen, 1987; Woese, 1987). The placement of the eukaryotic zoomastigote *Giardia lamblia* (Table 5.1) with the bacteria has been found in other phylogenetic analyses of SSU rRNA sequences, and is probably erroneous and results from the unusual biased base composition of its genome (Sogin *et al.*, 1989; Hashimoto *et al.*, 1994).

The eukaryote clade was resolved similarly by both methods. Most taxa did not form clear clusters of species thought to be closely related, however the gregarine clade of "Gregarina caledia", "Gregarina chortoicetes" but not including "Pseudomonocystis lepidiota", the piroplasmid clade of *Babesia bigemina, Babesia bovis, Babesia rodhaini* and *Theileria annulata*, the ciliophorans *Oxytricha nova* and *Paramecium tetraurelia* and the *Trypanosoma brucei / Euglena gracilis* lineage had significant bootstrap support.

The scatter plot comparison of patristic distances (Figure 5.4) showed two noteworthy features, in addition to providing evidence that in general, the parsimony and neighbour joining techniques produced similar results. Firstly there is a clear curvi-

linear relationship between the two sets of patristic distances. This may be caused by differences in the proportions of shared-derived characters and unique characters at different levels of relatedness (parsimony only uses the former), the non-linear 2-parameter correction applied to the distance data and/or the distorting effects of multiple hits. The scatter plot comparison also has some outlying values, which were identified by the DIPLOMO method to be mostly pairwise comparisons involving the *Giardia* sequence which probably reflects its base composition bias.

5.4.2 The "19 taxa" trees
a. Phylogeny of the gregarines

PTP tests of the data set showed that there is significant structure to the 212 informative sites (PTP = 0.01 from 100 randomised and 1 actual tree). Figures 5.5 a, b and c represent the trees calculated by distance and parsimony methods, with and without bootstrap values. The NJ tree (Figure 5.5a) infers that the eukaryotes are a polyphyletic grouping of multicellular and protoctist taxa, whereas in the tree made by parsimony (Figure 5.5b) they are well resolved into a monophyletic clade formed from the apicomplexans, the dinoflagellates (*Prorocentrum micans*) and the ciliophorans. Both methods placed the two gregarines "Gregarina caledia" and "Gregarina chortoicetes" and the piroplasmids as separate clades, and this was supported by significant bootstrap values. The NJ tree also placed the two gregarines as a sister to the pseudogregarine "Pseudomonocystis lepidiota". The single most parsimonious PAUP tree (length, 644 steps; CI, 0.533; Figure 5.5b) placed the two gregarines and the pseudogregarine "Pseudomonocystis lepidiota" as paraphyletic with the haemosporids (*Plasmodium falciparum*), which with the piroplasmid clade, share a common ancestor related to the coccidian, *Sarcocystis muris*. Therefore compared to the other apicomplexans, the gregarine taxon must be regarded as derived. This partially disagrees with the suggestion of Levine (1988), who proposed that the gregarines were a primitive apicomplexan taxon which had arisen from the Perkinsasida and therefore would form a sister or paraphyletic stem group to a derived group of haemotrophic coccidians, piroplasmids and haemosporids. However, parsimony with bootstrapping support (Figure 5.5c) suggests that this notion should be treated with caution. This tree (Figure 5.5c) infers structure only for the piroplasmids, the ciliophoran lineage of *Paramecium tetraurelia* and *Oxytricha nova*,the two

page 56, 2: delete the fourth sentence beginning : "However because this result was unexpected...." to the end of the paragraph "... could be a monophyletic group".

gregarines, the *Trypanosoma brucei/Euglena gracilis* and *Homo sapiens* lineages. The other eukaryote branches to *Plasmodium falciparum, Prorocentrum micans, Sarcocystis muris, Perkinsus sp.* and "Pseudomonocystis lepidiota" are collapsed to polytomies which suggests that the phylogenetic positions of these taxa are unstable because of the small evolutionary distances between these lineages. The "17 taxa" tree made by parsimony (page 50) with the higher taxa forming the outgroup, had an identical topology to that of the "19 taxa" tree.

Two *a priori* T-PTP tests (page 43) were done to test whether the monophyly of the two gregarines and the pseudogregarine with the haemotrophs could be the result of chance. These were as follows:

1. Using the data set from the "19 taxa" tree (Figure 5.5b) with the same 5 outgroup taxa (*Euglena gracilis, Trypanosoma brucei, Homo sapiens, Saccharomyces cerevisiae* and *Zea mays*) and 212 informative characters, the 6 haemotrophs, *Babesia bovis, Babesia bigemina, Babesia rodhaini, Theileria annulata, Plasmodium falciparum* and *Sarcocystis muris* were constrained for non-monophyly and monophyly (Table 5.3). Parsimony analysis with *a priori* T-PTP testing did not support the non-monophyly of the constrained group (T-PTP = 0.19, -5 steps) and was not significant. Therefore the null hypothesis was not rejected. Thus, the non-monophyletic arrangement of the haemotrophs in the observed tree made from "19 taxa" could be attributed to chance.

2. Using the same data set, informative characters and outgroups as in 1. (above), the gregarines "Gregarina caledia", "Gregarina chortoicetes" and the pseudogregarine "Pseudomonocystis lepidiota" were constrained with the haemotroph *Plasmodium falciparum* (Table 5.3). Parsimony analysis with *a priori* T-PTP testing supported the monophyly of the constrained group (T-PTP = 0.01, +3 steps, p =< 0.01) and the null hypothesis was rejected. That is, the constrained group was significant in the observed minimum length tree. However because this result was unexpected, the correct form of the test should have been the more computionally demanding *a posteriori* T-PTP test. Non-rejection of the null hypothesis on the *a priori* form of the test would be sufficient to determine that the null hypothesis would not be rejected on the *a posteriori* form of the test. However, rejection of the null hypothesis on the *a priori* test does not necessarily indicate that the null hypothesis would be rejected on the *a posteriori* test (John Trueman, personal communication). The *a posteriori* test was not done, so the

page 57, second paragraph beginning: "The apparent monophyletic arrangement..." should be a sentence which follows on from the previous paragraph on page 56 which ends with the sentence: "That is, the constrained group was significant in the observed minimum length tree.

page 57, third paragraph:

"These differences.." should read "Differences between them are supported by the fact that "Gregarina caledia"...."

outcome of these two tests (1. and 2. above) is based only on the result of test 1; there is a possibility that the haemotrophs could be a monophyletic group.

The apparent, monophyletic arrangement of the gregarines and *Plasmodium falciparum* on the actual tree made by parsimony (Figure 5.5b), may be caused by the grouping together of the long branches. This "19 taxa" tree made by parsimony, shows that the gregarines and *Plasmodium falciparum* are on very long branches (ranging from 2 to 56 character-state changes), compared to the branches leading to the other apicomplexan taxa (range: 1 to 17 character-state changes). However, this feature is not method dependent as a similar grouping together of long branches but without the inference of monophyly, is also evident at the base of the tree made by the neighbour joining method (Figure 5.5a). Further support for the uncertainty of this group is shown in both methods by bootstrap values which are only significant for the "Gregarina caledia"/"Gregarina chortoicetes" clade, and by the topology of the bootstrapped tree (Figure 5.5c) in which the branches to *Pseudomonocystis lepidiota* and *Plasmodium falciparum* are collapsed to unstable polytomies.

The arrangement of the three gregarines as a distinct group within the other apicomplexan taxa, lends support for their tentative classification as belonging to the same taxonomic class (Gregarina; Chapter 1, page 4) based on similar morphological data (Denis Anderson and Ray Akhurst, personal communication). These differences are supported by the fact that "Gregarina caledia" and "Gregarina chortoicetes" both inhabit the guts of acridid orthopterans from the same geographic area, whereas their sister group, the pseudogregarine, "Pseudomonocystis lepidiota" lives in the body cavity of a coleopteran.

b. Phylogeny of the apicomplexans

The "19 taxa" tree made by parsimony (Figure 5.5b) also inferred the monophyly of the apicomplexans with the dinoflagellates (*Prorocentrum micans*) and ciliophorans as sister groups, which lends support for similar findings by other workers (Lynn and Small, 1981; Kumazaki *et al.,* 1983; Johnson *et al.,* 1987 Barta *et al.,* 1991; Gajadhar *et al.,* 1991 and Wolters, 1991). *Perkinsus sp.* is the closest apicomplexan to the dinoflagellates, a finding which partially supports the theory of Levine (1988; Chapter 1, page 2).

c. Phylogeny of the piroplasmids

The two methods cluster the babesias as a paraphyletic taxon with *Babesia bigemina* as their sister group (Figures 5.5 a and b). This is unexpected, as this organism shares a similar vector and host with *Babesia bovis* (Vivier and Desportes, 1989). In the neighbor joining tree, *Babesia bovis* is inferred as the sister to a clade formed from *Theileria annulata* and *Babesia rodhaini* which was also inferred by distance methods in the "30 taxon" tree (Figure 5.3a) and found by Ellis *et al.,* (1992) but in trees made by parsimony and not by the neighbour joining method. *Theileria annulata* and *Babesia rodhaini* share similar merozoites (the parasite stage in the red blood cells of its host; Chapter 1, page 4), and it has been suggested (Melhorn and Schein, 1984) that *Babesia rodhaini* would be more correctly placed in the genus *Theileria*. These positions are reversed in the tree made by parsimony, which infers *Theileria annulata* as a sister taxon to a clade of *Babesia bovis* and *Babesia rodhaini*. The differences inferred by the two methods can be explained by the combination of small evolutionary distances which separate the taxa and the fact that the neighbour joining method uses a greater number of characters than the parsimony method. For example, in the tree made by parsimony in which it is possible to see the exact number of steps, 26 character-state changes separate *Babesia rodhaini* and *Theileria annulata*, and 25 separate *Babesia rodhaini* and *Babesia bovis*. Assuming that the same number of steps separated these taxa in the neighbor joining tree using only the shared derived characters, a further similarity of two ancestral and/or unique character states in *Babesia rodhaini* and *Theileria annulata* would reduce the evolutionary distance between them and change their inference to that shown by the neighbour joining tree.

5.5 CONCLUSION

The resolution of the relative evolutionary positions of the microsporans *Nosema apis* and "Nosema vespula", and the apicomplexans "Gregarina caledia", "Gregarina chortoicetes" and "Pseudomonocystis lepidiota" was only partly successful. The presence of the nosemas in the same clade as *Vairimorpha necatrix* which formed a sister group to the other eukaryotes, confirmed the findings of Vossbrinck *et al.,* (1987). However, the molecular data did not confirm or deny the relative positions of the two nosemas within the microsporan clade, nor those of the three gregarines within

the apicomplexan clade, with what has been respectively found by Sprague *et al.* (1992) and Baker *et al.,* (1994) or suggested by Levine (1988). However it did show that the three gregarines were closer to each other than they were to any other taxon, and that "Gregarina caledia" and "Gregarina chortoicetes" were more closely related to each other than they were to "Pseudomonocystis lepidiota".

In this analysis, the method of aligning the sequences was conserved, and any assumption of monophyly in the pre-alignment (page 47) was removed when the variable nucleotides were deleted from the aligned data set. Methods used to transform observed distances to estimated evolutionary distances, tree-making methods and methods for estimating confidence in trees and individual nodes were carefully done and are frequently used by other workers. The SSU rRNA data sets were shown to have hierarchial structure, and the CI indices indicated less than 50% homoplasy. However, in spite of this, within the eukaryotes, and particularly within the apicomplexans, the evolutionary distances were relatively small, so that the two tree-construction methods could not accurately infer the relative positions of the microsporan and apicomplexan taxa. The almost identical results of the neighbour joining and PAUP tree-building programs for the microsporans and the gregarines and pseudogregarine indicated that the results were not method-dependent. If any thing, parsimony tended to mislead by giving the impression that it was superior to the distance method in its resolution of the apicomplexans. With deeper analysis by the *a priori* T-PTP test however, this tree was found to have inferred an improbable monophyly of the gregarines with the haemotrophs so that the number of resolved apicomplexans was equivalent in both methods. However, the neighbour joining tree inferred a more realistic inference for the gregarines and pseudogregarine as sister groups to the other apicomplexans but at the same time showed the eukaryotes as polyphyletic which was supported by bootstrap data in both this method and parsimony. Thus when using neighbour joining and parsimony tree-building methods with aligned SSU rRNA sequences it can be seen that the relative positions of the gregarines and nosemas of this thesis can not found. In the future, the availability of SSU rRNA sequence data from other gregarines and microsporans may improve the resolution of these relationships.

Taxon	Access Number	Reference	Acron-nym	Phylum	Class	Order
Halobacterium cutirubrum	K029711	Hui and Dennis (1985)	HACU	Halobacteria		
Methanococcus vannielii	M36507	Jarsch and Boeck (1985)	MEVA	Methanobacteria		
Sulpholobus solfataricus	X03235	Olsen *et al*. (1985)	SUSO	Sulfobacteria		
Bacillus subtilis	K00637	Stewart and Bott (1983)	BASU	Firmicuta(Low G+C)		
Streptomyces coelicolor	Y00411	Bayliss and Bibb (1988)	STCO	Firmicuta(High G+C)		
Agrobacterium tumefaciens	M11223	Yang *et al*. (1985)	AGTU	Proteobacteria		
Vairimorpha necatrix	Y00266	Vossbrinck *et al*. (1987)	VANE	Microspora$	Dihaplophasea	Meiodihaplophasida
"Nosema vespula"	L31842	This analysis	NOVE	Microspora+	Dihaplophasea	Dissociodihaplophasida
Nosema apis	X73894	This analysis	NOAP	Microspora$	Dihaplophasea	Dissociodihaplophasida
Encephalitozoon cuniculi	L07255	Hartskeerl *et al*. (1993)	ENCU	Microspora$	Haplophasea	
Homo sapiens	X03205	McCallum and Maden (1985)	HOSA	Chordata		
Zea mays	K02202	Messing *et al*. (1984)	ZEMA	Magnoliophyta		
Saccharomyces cerevisiae	M27607	Rubtsov *et al*. (1980)	SACE	Ascomycotina		
"Gregarina chortiocetes"	L31841	This analysis	GRCH	Apicomplexia$	Gregarinia	Eugregarinida
"Gregarina caledia"	L31700	This analysis	GRCA	Apicomplexia$	Gregarinia	Eugregarinida
"Pseudomonocystis lepidiota"	L31843	This analysis	PSLE	Apicomplexia+	Gregarinia+	unknown
Perkinsus sp.	L07375	Goggin and Barker (1993)	PERK	Apicomplexia$	Perkinsasida*	Perkinsorida*
Plasmodium falciparum	M19172	McCutchan *et al*. (1988)	PLFA	Apicomplexia$	Haematozoa	Haemosporidia
Babesia rodhaini	M87565	Ellis *et al*. (1992)	BARO	Apicomplexia$	Haematozoa	Piroplasmida
Babesia bovis	M87566	Ellis *et al*. (1992)	BABO	Apicomplexia$	Haematozoa	Piroplasmida
Babesia bigemina	X59604	Reddy *et al*. (1991)	BABI	Apicomplexia$	Haematozoa	Piroplasmida
Theileria annulata	M64243	Gajadhar *et al*. (1991)	THAN	Apicomplexia$	Haematozoa	Piroplasmida
Sarcocystis muris	M34846	Gajadhar *et al*. (1991)	SAMU	Apicomplexia$	Coccidia	Eimeriida
Prorocentrum micans	M14649	Maroteaux *et al*. (1991)	PRMI	Dinoflagellata$		
Oxytricha nova	X03948	Elwood *et al*. (1985)	OXNO	Ciliphora$		
Paramecium tetraurelia	X03772	Sogin and Elwood (1986)	PATE	Cilophora$		
Trypanosoma brucei	M12676	Sogin *et al*. (1986)	TRBR	Zoomastigina$		
Euglena gracilis	M12677	Sogin *et al*. (1986)	EUGR	Euglenida$		
Acanthamoeba castellanii	M13435	Gunderson and Sogin (1986)	ACCA	Rhizopoda$		
Giardia lamblia	M54878	Sogin *et al*. (1989)	GILA	Zoomastigina$		

Table 5.1 Taxa used in phylogenetic analysis References: Neefs *et al*. (1991); $Margulis *et al*. (1989); +Denis Anderson, personal communication; *Levine (1988).

Figure 5.1

An alignment of conserved SSU rRNA sequences from 30 taxa from all kingdoms made with the computer programs PILEUP and LINEUP (Devereux *et al.*, 1984). Position 1 is congruent with position 581 of *Plasmodium falciparum* 18S rRNA (McCutchan *et al.*, 1988). The nucleotides were selected by visually comparing aligned primary sequences with secondary structure models of *Plasmodium falciparum* (De Rijk *et al.*, 1992), the three apicomplexans "Gregarina caledia", "Gregarina chortoicetes" and "Pseudomonocystis lepidiota" and the two microsporans "Nosema vespula" and *Nosema apis*. Gaps are denoted by ".", c = conserved, Frg = fragment

Acronyms are defined as follows:

Gila	*Giardia lamblia*
Hacu	*Halobacterium cutirubrum*
Meva	*Methanococcus vanniellii*
Vane	*Vairimorpha necatrix*
Nove	"Nosema vespula"
Noap	*Nosema apis*
Encu	*Encephalitozoon cuniculi*
Basu	*Bacillus subtilis*
Agtu	*Agrobacterium tumifaciens*
Stco	*Streptomyces coelicolor*
Suso	*Sulpholobus solfataricus*
Hosa	*Homo sapiens*
Zema	*Zea mays*
Sace	*Saccharomyces cerevisiae*
Grch	"Gregarina chortoicetes"
Grca	"Gregarina caledia"
Psle	"Pseudomonocystis lepidiota"
Plfa	*Plasmodium falciparum*
Perk	*Perkinsus sp.*
Prmi	*Prorocentrum micans*
Baro	*Babesia rodhaini*
Than	*Theileria annulata*
Samu	*Sarcocystis muris*
Babo	*Babesia bovis*
Babi	*Babesia bigemina*
Acca	*Acanthamoeba castellanii*
Oxno	*Oxytricha nova*
Pate	*Paramecium tetaurelia*
Trbr	*Trypanosoma brucei*
Eugr	*Euglena gracilis*

```
                1                                                          50
Gila-C.Frg      GCGGUAAUUC CAGCUCGGCG AGCGUCGCGC GGCGCUGCUG CAGUUGAAAC
Hacu-C.Frg      GCGGUAAUAC CGGCAGUCCG AGUGAUGGCC GAUCUUAUUG GGCCUAAAGC
Meva-C.Frg      GCGGUAAUAC CGACGGCCCG AGUGGUAGCC ACUCUUAUUG GGCCUAAAGC
Vane-C.Frg      GCGGUAAUAC UUGUUCCAAG AGUGUGUAUG AUGAUUGAUG CAGUUAAAAA
Nove-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUGUAUG AUGAUUGAUG CAGUUAAAAA
Noap-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUGUAUG AUGAUUGAUG CAGUUAAAAA
Encu-C.Frg      GCGGUAAUAC CUGCUCCAAU AGUGUCUAUG GUGGAUGCUG CAGUUAAAAU
Basu-C.Frg      GCGGUAAUAC GUAGGUGGCA AGCGUUUUCC GGAAUUAUUG GGCGUAAAGG
Agtu-C.Frg      GCGGUAAUAC GAAGGGGGCU AGCGUUGUUC GGAAUUACUG GGCGUAAAGC
Stco-C.Frg      GCGGUAAUAC GUAGGGCGCA AGCGUUGUCC GGAAUUAUUG GGCGUAAAGA
Suso-C.Frg      GCGGUAAUAC CAGCUCCGCG AGUGGUCGGG GUGAUUACUG GGCCUAAAGC
Hosa-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAGUUGCUG CAGUUAAAAA
Zema-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU UAAGUUGUUG CAGUUAAAAA
Sace-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAGUUGUUG CGGUUAAAAA
Grch-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAAUUGCUG CAGUUAAAGC
Grca-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAAUUGCUG CAGUUAAAGC
Psle-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAGUUGUUG CAGUUAAAAA
Plfa-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAAUUGUUG CAGUUAAAAC
Perk-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAGUUGUUG CGGUUAAAAA
Baro-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAACUUGUUG CAGUUAAAAA
Than-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAACUUGUUG CAGUUAAAAA
Samu-C.Frg      GCGGUAAUUC CAGCUCYAAU AGCGUAUAUU AAAGUUGUUG CAGUUAAAAA
Babo-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAAUUGUUG CAGUUAAAAA
Babi-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAACUUGUUG CAGUUAAAAA
Prmi-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAGUUGUUG CGGUUAAAAA
Acca-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAGUUGUUG CAGUUAAAAA
Oxno-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU AAAGUUGUUG CAGUUAAAAA
Pate-C.Frg      GCGGUAAUUC CAGCUCCAAU AGCGUAUACU UAAGUUGUUG CAGUUAAAAA
Trbr-C.Frg      GCGGUAAUUC CAGCUCCAAA AGCGUAUAUU AAUGCUGUUG CUGUUAAAGG
Eugr-C.Frg      GCGGUAAUUC CAGCUCCGAG GGCGUAUACU AACAUUGCUG CUGUUAAAAC

                51                                                        100
Gila-C.Frg      GCCCGUAGUG CCCCG.AGGA GAGC.GGG.. .CGGGGGCAC CGGUACCGGC
Hacu-C.Frg      GUCCGUAGCU UGGGA.CC.. .GGA.AGA.. .CCUGAGG.. .GGUACGUCU
Meva-C.Frg      GUCCGUAGCU UGGGA.CC.. .GGG.AGA.. .GGACAAG.. .GGUACUCCA
Vane-C.Frg      GUCCGUAGU. ..UUU.AAGA AGCA.AUA.. .UGAGGUGUA CUGUAUAGUU
Nove-C.Frg      GUCCGUAGU. ..UUU.AAGA AGCA.AUA.. .UGAGGUGUA CUGUAUAGUU
Noap-C.Frg      GUCCGUAGU. ..AUU.AAGA AGCA.AUA.. .UGAUGUGUA CUGUAUAGUU
Encu-C.Frg      GUCCGUAGU. ..UGU.GUGG CAGA.GGA.. .CGAGGGGCA CUGGAUAGUU
Basu-C.Frg      GCUCGCAGGU UGAGU.G... .AGA.AGA.. .GGAGAGU.. .GGAAUUCCA
Agtu-C.Frg      GCACGUAGGU UGAGU.AU.. .GGA.AGA.. .GGUAAGU.. .GGAAUUCCG
Stco-C.Frg      GCUCGUAGGU AGAGU.UC.. .GGU.AGG.. .GGAGAUC.. .GGAAUUCCU
Suso-C.Frg      GCCUGUAGCU AGGGG.GC.. .GGG.AGA.. .GGCGGGG.. .GGUACUCCC
Hosa-C.Frg      GCUCGUAGUA UGAUU.AAGA GGGACGGC.. .CGGGGGCAU UCGUAUUGCG
Zema-C.Frg      GCUCGUAGUA UGAUU.AAUA GGGACAGU.. .CGGGGGCAU UCGUAUUUCA
Sace-C.Frg      GCUCGUAGUA UGAUU.AAUA GGGACGGU.. .CGGGGGCAU CGGUAUUCAA
Grch-C.Frg      GUCCGUAGUA UGGUU.GAUA AGGACAUA.. .CGGGGGCAU UUGUACUUGC
Grca-C.Frg      GUCCGUAGUA UGGUU.GAUA AGGACAUA.. .CGGGGGCAU UUGUACUUGC
Psle-C.Frg      GCUCGUAGUA UGAUU.AACA GGAACAGU.. .UGGGGGCAU UCGAAUUGG
Plfa-C.Frg      GCUCGUAGUA CGAUU.AAUA GGAGUAGCU. .UGGGGACAU UCGUAUUCAG
Perk-C.Frg      GCUCGUAGUU GGAUU.AAUA GGGACAG.U. .CGGGGGCAU UCGUAUUUAA
Baro-C.Frg      GCUCGUAGUA UGGUU.AAUA GGAACGG.U. .UGGGGGCAU UCGUACUCGA
Than-C.Frg      GCUCGUAGUA UGGUU.AAUA GGAACGG.U. .UGGGGGCAU UCGUAUUUAA
Samu-C.Frg      GCUCGUAGUA UGAUU.AAUA GGGACAG.U. .UGGGGGCAU UCGUAUUUAA
Babo-C.Frg      GCUCGUAGUA UGGUU.AAUA GGAACAG.U. .UGGGGGCAU UCGUAUUUAA
Babi-C.Frg      GCUCGUAGUA UGGUU.AAUA GGAGCAG.U. .UGGGGGCAU UCGUAUUUAA
Prmi-C.Frg      GCUCGUAGUA UGAUU.AAUA GGGAUAGU.. .UGGGGGCAU UCGUAUUUAA
```

```
Acca-C.Frg    GCUCGUAGUA UGAUU.AAUA GGGAUAGU.. .UGGGGGCAU UAAUAUUUAA
Oxno-C.Frg    GCUCGUAGUA UGAUU.AAUA GGGAUAGU.. .UGGGGGCAU UAGUAUUUAA
Pate-C.Frg    GCUCGUAGUA UGAUU.AAUA GGGACAGA.. .UGGGGGCAU UAGUAUUUAA
Trbr-C.Frg    GUUCGUAGUG GAAUC.AAGG AGGGUAGUU. .CGGGGGAGA ACGUACUGGU
Eugr-C.Frg    ACUUGUAGUA UCGUA.AGUG AUGGGACUGU UCGGGGUGAA AGAUACGGGA

              101                                              150
Gila-C.Frg    CGG.GGACGG GUGAAACAGG AU.GAUCCCG CCGAGACCGC CGGCCGCGCA
Hacu-C.Frg    GGG.GUAGGA GUGAAAUCCU GUA.AUCCUG GACGGACCGC CGGUGGCGAA
Meva-C.Frg    GGG.GUAGCG GUGAAAUGUG UUG.AUCCUU GGAGGACCAC CUAUGGCGAA
Vane-C.Frg    GGG.AGAGAG AUGAAAUGUG AC.GACCCUG ACUGGACGAA CAGAAGCGAA
Nove-C.Frg    GGG.AGAAAG AUGAAAUGUA AC.GACCCUG ACUGGACGAA CAGAAGCGAA
Noap-C.Frg    GGG.AGAGAG AUGAAAUGUG AC.GACCCUG ACUGGACGAA CUGAAGCGAA
Encu-C.Frg    GGG.CGAGAG GUGAAAUGCG AA.GACCCUG ACUGGACGAG CGGAAGCGAA
Basu-C.Frg    CGUUGUAGCG GUGAAAUGC. GUAGAGAUGU GGAGGAACAC CAGUGGCGAA
Agtu-C.Frg    AGU.GUAGAG GUGAAAUUC. GUAGAUAUUC GGAGGAACGC CAGUGGCGAA
Stco-C.Frg    GGU.GUAGCG GUGAAAUGC. GCAGAUAUCA GGAGGAACAC CGGUGGCGAA
Suso-C.Frg    GGA.GUAGGG GCGAAAUCCU UAG.AUACCG GGAGGACCAC CAGUGGCGGA
Hosa-C.Frg    CCG.CUAGAG GUGAAAUUCU U.GGACCGGC GCAAGACGGA CCAGAGCGAA
Zema-C.Frg    UAG.UCAGAG GUGAAAUUCU U.GGAUUUAU GAAAGACGAA CAACUGCGAA
Sace-C.Frg    UUG.UC.GAG GUGAAAUUCU U.GGAUUUAU UGAAGACUAA CUACUGCGAA
Grch-C.Frg    UGG.AGAGAG GUGAAAUUCU A.AGACCCAG CAAAGACAAA CAACUGCGAA
Grca-C.Frg    UGG.AGAGAG GUGAAAUUCU A.AGACCCAG CAAAGACAAA CAACUGCGAA
Psle-C.Frg    UAG.CUAGAG GUGAAAUUCU U.AGAUUUAC CAAAGACGGA CUACUGCGAA
Plfa-C.Frg    AUG.UCAGAG GUGAAAUUCU U.AGAUUUUC UGGAGACGAA CAACUGCGAA
Perk-C.Frg    CUG.UCAGAG GUGAAAUUCU U.GGAUUUGU UAAAGACGAA CUACUGCGAA
Baro-C.Frg    CUG.UCAGAG GUGAAA.UCU U.AGAUUUGU CGAUGA.GCA CGACUGCGAA
Than-C.Frg    CUG.UCAGAG GUGAAAUUCU U.AGAUUUGU UAAAGACGAA CCACUGCGAA
Samu-C.Frg    CUG.UCAGAG GUGAAAUUCU U.AGAUUUGU UAAAGACGAA CUACUGCGAA
Babo-C.Frg    CUG.UCAGAG GUGARAUUCU U.AGAUUUGU UAAAGACGAA CUACUGCGAA
Babi-C.Frg    CUG.UCAGAG GUGAAAUUCU U.AGAUUUGU UAAAGACGAA CUACUGCGAA
Prmi-C.Frg    CUGGUCAGAG GUGAAAUUCU C.GGAUUUGU UAAAGACGGA CUACUGCGAA
Acca-C.Frg    UUG.UCAGAG GUGAAAUUCU U.GGAUUUAU GAAAGAUUAA CUUCUGCGAA
Oxno-C.Frg    UUG.UCAGAG GUGAAAUUCU C.GGAUUUGU UAAAGACUAA CUUAUGCGAA
Pate-C.Frg    UUG.UCAGAG GUGAAAUUCU U.GGAUUUAU UAAAGACUAA CUUAUGCGAA
Trbr-C.Frg    GCG.UCAGAG GUGAAAUUCU U.AGACCGCA CCAAGACGAA CUACAGCGAA
Eugr-C.Frg    GCG.CCAGAG GUGAAAUUCU U.AGAUCGCU GCCAGAUCCA CUGCAGCGAA

              151                                              200
Gila-C.Frg    GGCGCCUGCC .AAGACCGCC UCUGUCAAUC AA.GG...GC GA.AGGCCGG
Hacu-C.Frg    AGCG....CC UCAGGAGAAC GGAUCCGA.C AGUGAGGGAC GA.AAGCUAG
Meva-C.Frg    GGCA....CU UGUCUGGAAC GGGUCCGA.C GGUGAGGGAC GA.AAGCCAG
Vane-C.Frg    AGCUGUACAC .UUGUAUGUA UUUUUUGAAC AA.GG...AC GU.AAGCUGG
Nove-C.Frg    AGCUGUACAC .UUGUAUGUA UUUUUUGAAC AA.GG...AC GU.AAGCUGG
Noap-C.Frg    AGCUGUACAC .UUGUAUGUA UUUUUUGAAC AG.GG...AC GU.AAGCUGG
Encu-C.Frg    GGCUGUGCUC .UUGGACUAA UGUUGCGAUG AA.GG...AC GA.AGGCUAG
Basu-C.Frg    GGCG....AC UCUCUGGUCU GUAACUGA.C GCUGAGGAGC GA.AAGCGUG
Agtu-C.Frg    GGCG....GC UUACUGGUCC AUUACUGA.C GCUGAGGUGC GA.AAGCGUG
Stco-C.Frg    GGCG....GA UCUCUGGGCC GAUACUGA.C GCUGAGGAGC GA.AAGCGUG
Suso-C.Frg    AGCG....CC CCGCUAGAAC GCGCCCGA.C GGUGAGAGGC GA.AAGCCGG
Hosa-C.Frg    AGCAUUUGCC .AAGAAUGUU UUCAUUAAUC AA.GA...AC GA.AAGUCGG
Zema-C.Frg    AGCAUUUGCC .AAGGAUGUU UUCAUUAAUC AA.GA...AC GA.AAGUUGG
Sace-C.Frg    AGCAUUGGCC .AAGGACGUU UUCAUUAAUC AA.GA...AC GA.AAGUUAG
Grch-C.Frg    AGACUUUGCC .CAGUGUGUA CCUGUUAAUC AA.GG...AC GA.AAGUUGG
Grca-C.Frg    AGCAUUUGCC .CAGUGUGUA CCUAUUAAUC AA.GG...AC GA.AAGUUGG
Psle-C.Frg    AGACUCUGCC .AAGGAUGUU UUCAUUAAUC AA.GA...AC GA.AAGUUAG
Plfa-C.Frg    AGCAUUUGUC .UAAAUACU UCCAUUAAUC AA.GA...AC GA.AAGUUAA
Perk-C.Frg    AGCAUUUGCC .AAGGAUGUU UUCAUUGAUC AA.GA...AC GA.AAGUUAG
Baro-C.Frg    AGCAUUUGCC .AAGGACGCU UCCAUUAAUC AA.GA...AC GA.AAGUUAG
Than-C.Frg    AGCAUUUGCC .AAGGACGUU UUCAUUAAUC AA.GA...AC GA.AAGUUAG
```

```
Samu-C.Frg   AGCAUUUGCC .AAAGAUGUU UUCAUUAAUC AA.GA...AC GA.AAGUUAG
Babo-C.Frg   AGCAUUUGCC .AAGGAUGUU UUCAUUAAUC AA.GA...AC GA.AAGUUAG
Babi-C.Frg   AGCAUUUGCC .AAGGAUGUU UUCAUUAAUC AA.GA...AC GA.AAGUUAG
Prmi-C.Frg   AGCAUUUGCC .AAGGAUGUU UUCAUUGAUC AA.GA...AC GA.AAGUUAG
Acca-C.Frg   AGCAUCUGCC .AAGGAUGUU UUCAUUAAUC AA.GA...AC GA.AAGUUAG
Oxno-C.Frg   AGCAUUUGCC .AAGGAUGUU UUCAUUAAUC AA.GA...AC GA.AAGUUAG
Pate-C.Frg   AGCAUUUGCC .AAGGAUGUU UUCAUUAAUC AA.GA...AC GA.AAGUUAG
Trbr-C.Frg   GGCAUUCUUC .AAGGAUACC UUCCUCAAUC AA.GA...AC CA.AAGUGUG
Eugr-C.Frg   GGCGUUCUGC .AAGUGCACG UCCGUCGAUC AA.GA...AU GAG.AGUUCG
```

```
             201                                                 250
Gila-C.Frg   GGGCUAGAAG GCGAUCAGAC ACCACCGUAU UCCC..GGCC GUAAACGGUG
Hacu-C.Frg   GGUCUCGAAC CGGAUUAGAU ACCCGGGUAG UCCU..AGCU GUAAACGAUG
Meva-C.Frg   GGGCGCGAAC CGGAUUAGAU ACCCGGGUAG UCCU..GGCC GUAAACUCUG
Vane-C.Frg   AGGAGCGAAG AUGAUUAGAU ACCAUUGUAG UUCC..AGCA GUAAACUAUG
Nove-C.Frg   AGGAGCGAAG AUGAUUAGAU ACCAUUGUAG UUCC..AGCA GUAAACUAUG
Noap-C.Frg   AGGAUCGAAG AUGAUUAGAU ACCAUUGUAG UUCC..AGCA GUAAACUAUG
Encu-C.Frg   AGGAUCGAAA UCGAUUAGAU ACCGUUUUAG UUCU..AGCA GUAAACGAUG
Basu-C.Frg   GGGAGCGAAC AGGAUUAGAU ACCCUGGUAG UCCA..CGCC GUAAACGAUG
Agtu-C.Frg   GGGAGCAAAC AGGAUUAGAU ACCCUGGUAG UCCA..CGCC GUAAACGAUG
Stco-C.Frg   GGGAGCGAAC AGGAUUAGAU ACCCUGGUAG UCCA..CGCC GUAAACGGUG
Suso-C.Frg   GGCAGCAAAC GGGAUUAGAU ACCCCGGUAG UCCC..GGCU GUAAACGAUG
Hosa-C.Frg   AGGUUCGAAG ACGAUCAGAU ACCGUCGUAG UUCC..GACC AUAAACGAUG
Zema-C.Frg   GGGCUCGAAG ACGAUCAGAU ACCGUCCUAG UCUC..AACC AUAAACGAUG
Sace-C.Frg   GGGAUCGAAG AUGAUCUGGU ACCGUCGUAG UCUU..AACC AUAAACUAUG
Grch-C.Frg   GGGAUCGAAG ACGAUUAGAU ACCGUCGUAG UCCC..AACU AUAAACUAUG
Grca-C.Frg   GGGAUCGAAG ACGCUUAGAU ACCGUCGUAG UCCC..AACU AUAAACUAUG
Psle-C.Frg   GGGAUCGAAG ACGAUCAGAU ACCGUCGUAG UCUU..AACU AU.AACGAUG
Plfa-C.Frg   GGGAGUGAAG ACGAUCAGAU ACCGUCGUAA UCUU..AACC AUAAACUAUG
Perk-C.Frg   GGGAUCGAAG ACGAUCAGAU ACCGUCCUAG UCUU..AACC AUAAACUAUG
Baro-C.Frg   GGGAUCGAAG ACGAUCAGAU ACCGUCGUAG UCCU..AACC UUAAACGAUG
Than-C.Frg   GGGAUCGAAG ACGAUC.GAU ACCGUCGUAG UCCU..AACC AUAAACUAUG
Samu-C.Frg   GGGCUCGAAG ACGAUCAGAU ACCGUCGUAG UCUU..AACC AUAAACUAUG
Babo-C.Frg   GGGAUCGAAG ACGAUCAGAU ACCGUCGUAG UCCU..AACC AUAAACUAUG
Babi-C.Frg   GGGAUCGAAG ACGAUCAGAU ACCGUCGUAG UCCU..AACC AUAAACUAUG
Prmi-C.Frg   GGGAUCGAAG ACGAUCAGAU ACCGUCCUAG UCUU..AACC AUAAACCAUG
Acca-C.Frg   GGGAUCGAAG ACGAUCAGAU ACCGUCGUAG UCUU..AACC AUAAACGAUG
Oxno-C.Frg   GGGAUCAAAG ACGAUCAGAU ACCGUCCUAG UCUU..AACC AUAAACUAUG
Pate-C.Frg   GGGAUCAAAG ACGAUCAGAU ACCGUCGUAG UCUU..AACU AUAAACUAUA
Trbr-C.Frg   GGGAUCAAAG AUGAUUAGAG ACCAUUGUAG UCCA..CACU GCAAACCAUG
Eugr-C.Frg   GGGAGCAAAG AUGAUCAGAC ACCGUCGUAG UCCGGCCACU GUAAACGAUG
```

```
             251                                                 300
Gila-C.Frg   CCGCCGCUCU GGGGGGAGUA UGGCCGCAAG GCUGAAACUU GAAGGCAUUG
Hacu-C.Frg   UCCGCCGCCU GGG..AAGUA CGUCUGCAAG GAUGAAACUU AAAGGAAUUG
Meva-C.Frg   CGAACCGCCU GGG..GAGUA CGGUCGCAAG ACUGAAACUU AAAGGAAUUG
Vane-C.Frg   CCGACGCUCU GGGGAUAGUA UGAUCGCAAG AUUGAAAAUU AAAGAAUUG
Nove-C.Frg   CCGACGCACU GGGGAUAGUA UGAUCGCAAG AUUGAAAAUU AAAGAAUUG
Noap-C.Frg   CCGACGCUCU GGGGAUAGUA UGAUCGCAAG AUUGAAAAUU AAAGAAUUG
Encu-C.Frg   CCGACGGAC GGGGAUAGUA UGCUCGCAAG AGUGAAACUU GAAGAGAUUG
Basu-C.Frg   AGUGCCGCCU GGG..GAGUA CGGUCGCAAG ACUGAAACUC AAAGGAAUUG
Agtu-C.Frg   AAUGUCGCCU GGG..GAGUA CGGUCGCAAG AUUAAAACUC AAAGGAAUUG
Stco-C.Frg   GGCACCGCCU GGG..GAGUA CGGCCGCAAG GCUAAAACUC AAAGGAAUUG
Suso-C.Frg   CGGGCCGCCU GGG..GAGUA CGGUCGCAAG ACUGAAACUU AAAGGAAUUG
Hosa-C.Frg   CCGACGUUCC GGGGGGAGUA UGGUUGCAAA GCUGAAACUU AAAGGAAUUG
Zema-C.Frg   CCGACGUUCC GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Sace-C.Frg   CCGACGUUCU GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Grch-C.Frg   CCGACGUUCU GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Grca-C.Frg   CCGACGUUCU GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Psle-C.Frg   CCAACGUUCU GGGGGGAGUA UGGCCGCAAG UCUGAAACUU AAAGGAAUUG
```

```
Plfa-C.Frg    CCGACGUUCU GGGGCGAGUA UUCGCGCAAG CGAGAAAGUU AAAAGAAUUG
Perk-C.Frg    CCGACGUUCC GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Baro-C.Frg    CCGACGUUCU GGGGGGAGUA UGGUCGCAAG UCUGAAACUU AAAGGAAUUG
Than-C.Frg    CCGACGUUCU GGGGGGAGUA UGGUCGCAAG UCUGAAACUU AAAGGAAUUG
Samu-C.Frg    CCGACGUUCU GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Babo-C.Frg    CCGACGUUCU GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Babi-C.Frg    CCGACGUUCU GGGGGGAGUA UGGUCGCAAG TCUGAAACUU AAAGGAAUUG
Prmi-C.Frg    CCGACGUUCC GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Acca-C.Frg    CCGACGUUCC GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Oxno-C.Frg    CCGACGUUCU GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Pate-C.Frg    CCGACGUUCU GGGGGGAGUA UGGUCGCAAG GCUGAAACUU AAAGGAAUUG
Trbr-C.Frg    ACACCUCUCA GGGGGGAGUA CUGUCGCAAG AGUGAAACUU AAAGAAAUUG
Eugr-C.Frg    CCGGCGUUCA GGGGGGAGUA CUGUCGCAAG GCUGAAACUU AAAGGAAUUG

              301                                                  350
Gila-C.Frg    ACGGAGGGGU ACCACCAGAC GUGGAGUCUG CGGCUCAAUC UGACUCAACG
Hacu-C.Frg    GCGGGGGAGC ACUACAACCG GAGGAGCCUG CGGUUUAAUU GGACUCAACG
Meva-C.Frg    GCGGGGGAGC ACCACAACGG GUGGAGCCUG CGGUUUAAUU GGAUUCAACG
Vane-C.Frg    ACGGAAGAAU ACCAGAAGGA GUGGAUUGUG CGGCUUAAUU UGACUCAACG
Nove-C.Frg    ACGGAAGAAU ACCACAAGGA GUGGAUUGUG CGGCUUAAUU UGACUCAACA
Noap-C.Frg    ACGGAAGAAU ACCACAAGGA GUGGAUUGUG CGGCUUAAUU UGACUCAACA
Encu-C.Frg    ACGGAAGGAC ACCACAAGGA GUGGAGUGUG CGGCUUAAUU UGACUCAACG
Basu-C.Frg    ACGGGGGCCC G.CACAAGCG GUGGAGCAUG UGGUUUAAUU CGAAGCAACG
Agtu-C.Frg    ACGGGGGCCC G.CACAAGCG GUGGAGCAUG UGGUUUAAUU CGAAGCAACG
Stco-C.Frg    ACGGGGGCCC G.CACAAGCG GCGGAGCAUG UGGCUUAAUU CGACGCAACG
Suso-C.Frg    GCGGGGGAGC ACCACAAGGG GUGGAACCUG CGGCUCAAUU GGAGUCAACG
Hosa-C.Frg    ACGGAAGGGC ACCACCAGGA GUGGAGCCUG CGGCUUAAUU UGACUCAACA
Zema-C.Frg    ACGGAAGGGC ACCACCAGGC GUGGAGCCUG CGGCUUAAUU UGACUCAACA
Sace-C.Frg    ACGGAAGGGC ACCACUAGGA GUGGAGCCUG CGGCU.AAUU UGACUCAACA
Grch-C.Frg    ACGGAAGGGC ACCACCAGGA GUGGAG..UG CGGCUUAAUU UGACUCAACA
Grca-C.Frg    ACGGAAGGGC ACCACCAGGA GUGAGC..UG CGGCUUAAUU UGACUCAACA
Psle-C.Frg    ACGGAAGGGC ACCACCAGGA GUGAGC..UG CGGCUUAAUU UGACUCAACA
Plfa-C.Frg    ACGGAAGGGC ACCACCAGGC GUGGAGCUUG CGGCUUAAUU UGACUCAACA
Perk-C.Frg    ACGGAAGGGC ACCACCAGGA GUGGANNUGC .GGCUUAAUU UGAUUCAACA
Baro-C.Frg    ACGGAAGGGC ACCACCAGGC GUG..GCAGG CGGCUUAAUU UGACUCAACA
Than-C.Frg    ACGGAAGGGC ACCACCAGGC GUGGAGCCUG CGGCUUAAUU UGACUCAACA
Samu-C.Frg    ACGGAAGGGU ACCACCAGGC GUGGAGCCUG CGGCUUAAUU UGACUCAACA
Babo-C.Frg    ACGGAAGGGC ACCACCAGGC GUGGAG.CUG CGGCUUAAUU UGACUCAACA
Babi-C.Frg    ACGGAAGGGC ACCACCAGGC GUG..GCAGG CGGCUUAAUU UGACUCAACA
Prmi-C.Frg    ACGGAAGGGC ACCACCAGGA GUGGAGCCUG CGGCUUAAUU UGACUCAACA
Acca-C.Frg    ACGGAAGGGC ACCACCAGGA GUGGAGCCUG CGGCUUAAUU UGACUCAACA
Oxno-C.Frg    ACGGAAGGGC ACCACCAGGA GUGGAGCUUG CGGCUCAAUU UGACUCAACA
Pate-C.Frg    ACGGAAGGGC ACCACCAGGA GUGGAGCCUG CGGCUUAAUU UGACUCAACA
Trbr-C.Frg    ACGGAAUGGC ACCACAAGAC GUGGAGCGUG CGGUUUAAUU UGACUCAACA
Eugr-C.Frg    ACGGAAUGGC ACCACAAGGC GUGGAGUAUG CGGCUUAAUU UGACUCAACG

              351                                                  400
Gila-C.Frg    CG.CGCACCU CACCAGGCCC GGUGGUGCAU GGCCGCU.CC CAGCCCGUG.
Hacu-C.Frg    CCGGACAUCU CACCAGCCCA GGAGGUGCAU GGCCGCC.GU CAGCUCGUA.
Meva-C.Frg    CCGGGCAUCU CACCACGAGA GGUGGUGCAU GGCCAUC.GU CAGCUCGUA.
Vane-C.Frg    CGAGGUAACU UACCAAUAUU AGUGGUGCAU GGCCGUU.UU CAAUGGAUG.
Nove-C.Frg    CGGGGUAACU UACCAAUAUU AGUGGUGCAU GGCCGUU.UU CAAUGGAUG.
Noap-C.Frg    CGGGGUAACU UACCAAUAUU AGUGGUGCAU GGCCGUU.UU CAAUGGAUG.
Encu-C.Frg    CGGGGCAACU UACCGGCUCC GGUGGUGCAU GGCCGUU.UU AAAUGGAUG.
Basu-C.Frg    CGAAGAACCU UACUAGGUCA GGUGGUGCAU GGUUGUC.GU CAGCUCGUG.
Agtu-C.Frg    CGCAGAACCU UACCAGCUCA GGUGCUGCAU GGCUGUC.GU CAGCUCGUG.
Stco-C.Frg    CGAAGAACCU UACCAAGGCA GGUGGUGCAU GGCUGUC.GU CAGCUCGUG.
Suso-C.Frg    CCUGGAAUCU UACCGGGGGA GGAGGUGCAU GGCCGUC.GC CAGCUCGUG.
Hosa-C.Frg    CGGGAAACCU CACCCGGCCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Zema-C.Frg    CGGGGAAACU UACCAGGUCU GGUGGUGCAU GGUCGUU.CU UAGUUGGUG.
```

```
Sace-C.Frg  CGGGGAAACU CACCAGGUCU GGUGGUGCAU GGCCGUUUCU CAGUUGGUG.
Grch-C.Frg  CGGGGAACCU CACCAGGCCU GGUGGUGCAU GGCCGUU.CU UAGUCGGUGA
Grca-C.Frg  CGGGGAACCU CACCAGGCCU GGUGGUGCAU GGCCGUU.CU UAGUCGGUGA
Psle-C.Frg  CGGGAAAACU CACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUGG
Plfa-C.Frg  CGGGGAAACU CACUAGUUUU GGUGAUGCAU GGCCGUU.UU UAGUUCGUGA
Perk-C.Frg  CGGGAAAACU CACCAGGUCC GGUGGUGCAU GGCCGUU.CU UAGUUGGUGG
Baro-C.Frg  CGGGGAACCU CACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Than-C.Frg  CGGGGAACCU CACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Samu-C.Frg  CGGGGAAACU CACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Babo-C.Frg  CGGGGAAACU CACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Babi-C.Frg  CGGGAAACCU CACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Prmi-C.Frg  CGGGGAAACU UACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Acca-C.Frg  CGGGGAAACU UACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Oxno-C.Frg  CGGGAAAACU UACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Pate-C.Frg  CGGGGAAACU UACCAGGUCU GGUGGUGCAU GGCCGUU.CU UAGUUGGUG.
Trbr-C.Frg  CGGGGAACUU UACCAGAUCU GGUGGUGCAU GGCCGCU.UU UGGUCGGUG.
Eugr-C.Frg  CGGGGAAUGU UACCAGGUCC GGUGGUGCAU GGCCGCU.CU UGAUUGGUG.

            401                                              450
Gila-C.Frg  GCGCGAGCCG UCUGCUCCAU UGCGACAACG AGCGACCCCC GCGCGAG..C
Hacu-C.Frg  CCGUGAGGCG UCCUGUUAAG UCAGGCAACG AGCGAGACCC GGCCAAA..G
Meva-C.Frg  CCGCGAGGCG UCCUGUUAAG UCAGGUAACG AGCGAGACCC GGCUAAG..C
Vane-C.Frg  CUGUGAAGU. UUUGAUUAAU UUCACCAAGA CGUGAGACCC UAAUCA...G
Nove-C.Frg  CUGUGAAGU. UUUGAUUAAU UUCAACAAGA CGUGAGACCC UAAUCA...G
Noap-C.Frg  CUGUGAAGU. UUUGAUUAAU UUCAA.AAGA CGUGAGACCC UUAUUA...G
Encu-C.Frg  GCGUGAGCU. UUGUCUUAAG UUGCGUAAGA UGUGAGACCC UUUUGA...A
Basu-C.Frg  UCGUGAGAUG UUGGGUUAAG UCCCGCAACG AGCGCAACCC UGACAAA..C
Agtu-C.Frg  UCGUGAGAUG UUGGGUUAAG UCCCGCAACG AGCGCAACCC UGAUAAG..C
Stco-C.Frg  UCGUGAGAUG UUGGGUUAAG UCCCGCAACG AGCGCAACCC U.UCAAC..U
Suso-C.Frg  UUGUGAAAUG UCCGGUUAAG UCCGGCAACG AGCGAGACCC CG.UAAG..C
Hosa-C.Frg  GAGCGAUUUG UCUGGUUAAU UCCGAUAACG AACGAGACUC UGUUCAG..C
Zema-C.Frg  GAGCGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU CGUUUAG..G
Sace-C.Frg  GAGUGAUUUG UCUGCUUAAU UGCGAUAACG AACGAGACCU CUUCAAG..C
Grch-C.Frg  G.GUGACUUG UCUGGUUAAU UCCGAUAACG GACGAGACCU CGUCUA..UC
Grca-C.Frg  G.GUGACUUG UCUGGUUAAU UCCGAUAACG GACGAGACCU CGUCUA..UC
Psle-C.Frg  A.GUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UAUCUAA..C
Plfa-C.Frg  AUAUGAUUUG UCUGGUUAAU UCCGAUAACG AACGAGAUCU UGUCUAA..C
Perk-C.Frg  AG.UGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UGUUUAA..C
Baro-C.Frg  GAGUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UCCUCAAGCU
Than-C.Frg  GAGUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UCUUCAAGCG
Samu-C.Frg  GAGUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UGUCUAA..C
Babo-C.Frg  GAGUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UCCUCAAGCU
Babi-C.Frg  GAGUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UCUUCAAGCG
Prmi-C.Frg  GAGUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UGUCUAA..C
Acca-C.Frg  GAGUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UGCCUAG..C
Oxno-C.Frg  GAGUGAUUUG UCUGGUUAAU UCCGUUAACG AACGAGACCU UACUAAA..C
Pate-C.Frg  GAGUGAUUUG UCUGGUUAAU UCCGAUAACG AACGAGACCU UGUAAUG..C
Trbr-C.Frg  GAGUGAUUUG UUUGGUUGAU UCCGUCAACG GACGAGAUCC AUUUCGC..G
Eugr-C.Frg  GAGUGAUUUG UCUGGUUGAU UCCGAUAACG AGUGAGACAU CACAAAG..G

            451                                              500
Gila-C.Frg  GG.GAGGACG GCGGGG.CGA UAGCAGGUCU GUGAUGC.CC UCAGA.CGCC
Hacu-C.Frg  CG.GAGGAAG GAACGGGCAA CGGUAGGUCA GU.AUGCCCC ...GAAUGGG
Meva-C.Frg  UA.GAGGAAG GAGCGGGCAA CGAUAGGUCC GC.AUGCCCC ...GAAUCUC
Vane-C.Frg  UG.UAGGAAG GAAAGGAUUA AAACAGGUCC GUUAUGC.CC UCAGA.CAUU
Nove-C.Frg  UG.UAGGAAG GAAAGGAUUA AAACAGGUCC GUUAUGC.CC UCAGA.CAUU
Noap-C.Frg  UG.UAGGAAG GAAAGGACUA AAACAGGUCA GUUAUGC.CC UCUGAGCAUU
Encu-C.Frg  GC.AAGGAGG GGAUGGAAAA GAACAGGUCC GUUAUGC.CC UGAGA.UGAG
Basu-C.Frg  CG.GAGGAAG GUGGGGAUGA CGUCAAAUC. AUCAUGCCCC UUA...UGAC
Agtu-C.Frg  CGAGAGGAAG GUGGGGAUGA CGUCAAGUC. CUCAUGGCCC UUA...CGGG
```

```
Stco-C.Frg   CG.GAGGAAG GUGGGGACGA CGUCAAGUC. AUCAUGCCCC UUA...UGUC
Suso-C.Frg   CG.GAGGAAG GAGGGGGCCA CGGCAGGUCA GC.AUGCCCC ...GAAACUC
Hosa-C.Frg   CA.CC.CGAG AUUGAG.CAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUC
Zema-C.Frg   CC.GCG.AAG UUUGAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUU
Sace-C.Frg   CG.AUGGAAG UUUGAGGCAA UAACAGGUCU GUGAUGC.CC UUAGAACGUU
Grch-C.Frg   AC.AAGGAAG UUCGAGCCUA UAACAGGUCU GUGAUGC.CC UUAGA.UGGC
Grca-C.Frg   AC.AAGGAAG UUCGAGCCUA UAACAGGUCU GUGAUGC.CC UUAGA.UGGC
Psle-C.Frg   GC.ACGGAAG UUUAAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUC
Plfa-C.Frg   AC.AAGGAAG UUUAAGGCAA CAACAGGUCU GUGAUGU.CC UUAGA.UGAA
Perk-C.Frg   AC.AAGGAAG CUUGAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUU
Baro-C.Frg   GC.GGUGAGG UUUAAGGCAA UAACAGGUCU GUGAU.C.CC UUAGA.UCUC
Than-C.Frg   UG.GGGGAAG UUUAAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUC
Samu-C.Frg   GC.AAGGAAG UUUGAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUU
Babo-C.Frg   GC.GGUGAGG UUUAAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUC
Babi-C.Frg   UG.GGGGAAG UUUAAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUC
Prmi-C.Frg   GC.AAGGAAG UUUGAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUU
Acca-C.Frg   CA.GCGGAAG UUUGAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUU
Oxno-C.Frg   AC.AAGGAAG UUUGAGGCAA UAACAGGUCU GUGAUGC.CC UUAGA.UGUC
Pate-C.Frg   ...AUGGAAG UUUAAGGCAA UAACAGGUCU GUGAUGC.CC CUAGA.CGUC
Trbr-C.Frg   CAAGGUGAGA UUUUGGGCAA CAGCAGGUCU GUGAUGC.UC CUCAA.UGUU
Eugr-C.Frg   GC.AUGCAUG CUAGAGCCAA CAGCAGGUCU GUGAUGC.UC CCAGA.UGUC

             501                                               550
Gila-C.Frg   CUGGGCCGCA CGCGCGCUAC ACUGGGACCG CGGGCUGGAA ..CGCCCCC.
Hacu-C.Frg   CUGGGCAACA CGCGGGCUAC AAUGGGAUUG AGGGCUGAAA CUCGCCCUC.
Meva-C.Frg   CUGGGCUACA CGCGGGCUAC AAUGGGAUCG UGGGCUGUAA CUCGCCCAC.
Vane-C.Frg   UUGGGCUGCA CGCGCAAUAC AAUAGGAUAA UAUUUUGUAA ..GAGAUAU.
Nove-C.Frg   UUGGGCUGCA CGCGCAAUAC AAUAGGAUAA UAUUUUGUAA ..GAGAUAU.
Noap-C.Frg   UUGGGCAGCA CGCGCAAUAC AAUAGGAUAA UAUUUUGUAA ..GAGAUAU.
Encu-C.Frg   GCGGGUGCAC CGCGCACUAC GAUAGGAUGA AGCUGUGUAA ..GGGGCUU.
Basu-C.Frg   UUGGGCUACA CACGUGCUAC AAUGGGAUCG CAGUCUGCAA CUCGACUGC.
Agtu-C.Frg   CUGGGCUACA CACGUGCUAC AAUGGGAUUG CACUCUGCAA CUCGAGUGC.
Stco-C.Frg   UUGGGCUGCA CACGUGCUAC AAUGGGAUUG GGGUCUGCAA CUCGACCCC.
Suso-C.Frg   CCGGGCCGCA CGCGGGUUAC AAUGGGAUCG AGGGCUGAAA CCCGCCCUC.
Hosa-C.Frg   CGGGGCUGCA CGCGCGCUAC ACUGGGAUCG GGGAUUGCAA UUAUUCCCC.
Zema-C.Frg   CUGGGCCGCA CGCGCGCUAC ACUGGGAUAG AUCAUUGCAA UUGUUGGUC.
Sace-C.Frg   CUGGGCCGCA CGCGCGCUAC ACUGGGAUAG AGCAUUGUAA UUAUUGCUC.
Grch-C.Frg   CUGGGCUGCA CGUGCGCUAC AAUGGGAUUG ACCCUUGCAA UUAUGGGUC.
Grca-C.Frg   CUGGGCUGCA CGUGCGCUAC AAUGGGAUUG ACCCUUGCAA UUAUGGGUC.
Psle-C.Frg   CUGGGCUGCA CGCGCGCUAC ACUGGGCUAG A.UGAUGUAA UUAUUCAUC.
Plfa-C.Frg   CUAGGCUGCA CGCGUGCUAC ACUGGGAUAG AUUAUUGCAA UUAUUAAUC.
Perk-C.Frg   CUGGGCUGCA CACGCGCUAC ACUGGGAUAG ACGAUUGCAA UUAUUCGUC.
Baro-C.Frg   CUGGGCUGCA CGCGCGCUAC ACUGGGAUUG AUCUUUGCAA UUCUAGAUC.
Than-C.Frg   CUGGGCUGCA CGCGCGCUAC ACUGGGAUUG AUUUUUGCAA UUCUAAAUC.
Samu-C.Frg   CUGGGCUGCA CGCGCGCUAC ACUGGGAUAG AUUAUUGCAA UUAUUAAUC.
Babo-C.Frg   CUGGGCUGCA CGCGCGCUAC ACUGGGAUCG AUUAUUGCAA UUGUUAAUC.
Babi-C.Frg   CUGGGCUGCA CGCGCGCUAC ACUGGGAUAG AUUAUUGCAA UUAUUAAUC.
Prmi-C.Frg   CUGGGCUGCA CGCGCGCUAC ACUGGGAUAG AUUAUUGCAA UUAUUAAUC.
Acca-C.Frg   CUGGGCCGCA CGCGCGCUAC ACUGGGAUAG AUCAUUGUAA UUAUUGAUC.
Oxno-C.Frg   CUGGGCCGCA CGCGUGCUAC ACUGGGAUAG AUCUUUGGAA UUAUAGAUC.
Pate-C.Frg   CUGGGCCGCA CGCGCGCUAC ACUGGGAUAG AUCUUUGCAA UUAUAGAUC.
Trbr-C.Frg   CUGGGCGACA CGCGCACUAC AAUGGGACCG AGUAUUGCAA UUAUUGGUCG
Eugr-C.Frg   CUGGGCCGCA CGCGCACUAC AUUGGGAUAG AUGGUUGCAA CUGUCUGCC.

             551                                               600
Gila-C.Frg   GCGCACCAGG AAUGUCUUGU AGGCGCCCGC CCCC.ACCGC GCGCCGGACG
Hacu-C.Frg   AUGAAGCUGG AUUCGGUAGU AAUCGCGUGU CAGC.AGCGC GCGGUGAAUA
Meva-C.Frg   GUGAAGCUGG AAUCCGUAGU AAUCGCAGUU CAUA.AUACU GCGGUGAAUG
Vane-C.Frg   UUGAACUUGG AAUUGCUAGU AAAUUUUAUU AAAU.AAGUA GAAUUGAAUG
Nove-C.Frg   UUGAACUUGG AAUUGCUAGU AAAUUUUAUU AAAU.AAGUA GAAUUGAAUG
```

```
Noap-C.Frg   UUGAACUUGG AAUUGCUAGU AAAUUUUAUC AAAU.AAGUA GAAUUGAAUG
Encu-C.Frg   CUGAACGUGG AAUUCCUAGU AAUAGCGGCU GACG.AAGCU GCUUUGAAUG
Basu-C.Frg   GUGAAGCUGG AAUCGCUAGU AAUCGCGGAU CAGC.AUGCC GCGGUGAAUA
Agtu-C.Frg   AUGAAGUUGG AAUCGCUAGU AAUCGCAGAU CAGC.AUGCU GCGGUGAAUA
Stco-C.Frg   AUGAAGUCGG AGUCGCUAGU AAUCGCAGAU CAGCAUUGCU GCGGUGAAUA
Suso-C.Frg   GUGAACGAGG AAUCCCUAGU AACCGCGGGU CAAC.AACCC GCGGUGAAUA
Hosa-C.Frg   AUGAACGAGG AAUUCCCAGU AAGUGCGGGU CAUA.AGCUU GCGUUGAUUA
Zema-C.Frg   UUCAACGAGG AAUGCCUAGU AAGCGCGAGU CAUC.AGCUC GCGUUGACUA
Sace-C.Frg   UUCAACGAGG AAUUCCUAGU AAGCGCAAGU CAUC.AGCUU GCGUUGAUUA
Grch-C.Frg   AUGAACGAGG AAUUCCUAGU AAGGACAAGU CAUC.ACCUU GUGCUGAUUA
Grca-C.Frg   AUGAACGAGG AAUUCCUAGU AAGAACAAGU CAUC.ACCUU GUGCUGAUUA
Psle-C.Frg   UUUAACGAGG AAUUCCUAGU AAGUACAAGU CAUU.AGCUU GUGCUGAUUA
Plfa-C.Frg   UUGAACGAGG AAUGCCUAGU AAGCAUGAUU CAUC.AGAUU GUGCUGACUA
Perk-C.Frg   UUCAACGAGG AAUUCCUAGU AAAUGCAAGU CAUC.AGCUU GCGUUGAUUA
Baro-C.Frg   AUGAACGAGG AAUGCCUAGU AUGCGCAAGU CAUC.AGCUU GUGCAGAUUA
Than-C.Frg   AUGAACGAGG AAUGCCUAGU AUGCGCAAGU CAUC.AGCUU GUGCAGAUUA
Samu-C.Frg   UUCAACGAGG AAUGCCUAGU AGGCGCAAGU CAGC.AGCUU GCGCCGAUUA
Babo-C.Frg   GUGAACGAGG AAUGCCUAGU AUGCGCAAGU CAUC.AGCUU GUGCAGAUUA
Babi-C.Frg   UUGAACGAGG AAUGCCUAGU AGGCGCGAGU CAUC.AGCUC GUGC.CACUA
Prmi-C.Frg   UUCAACGAGG AAUUCCUAGU AAGCGCGAGU CAUC.AGCUC GUGCUGAUUA
Acca-C.Frg   UUCAACGAGG AAUUCCUAGU AAGCGCGAGU CAUC.AGCUC GCGUUGAUUA
Oxno-C.Frg   UUGAACGAGG AAUUCCUAGU AAGCGCAAGU CAUU.ACCUU GCGCUGAUUA
Pate-C.Frg   UUGAACGAGG AAUUCCUUGU AAGCACAGGU CAUC.AGCCU GUGCUGAAUA
Trbr-C.Frg   CGCAACGAGG AAUGUCUCGU AGGCGCAGCU CAUC.AAACU GUGCCGAUUA
Eugr-C.Frg   UUGAACGUGG AAUGCCUAGU AUGCCUGGGU CAUC.AGCCC AGACCGAUUG

             601                                            650
Gila-C.Frg   CGUCCCUGCC CCUUGUACAC ACCGCCCGUC GCUCCUACCG A.GCGAAGGG
Hacu-C.Frg   CGUCCCUGCU CCUUGCACAC ACCGCCCGUC AAAUCACCCG A.AUGCGCUG
Meva-C.Frg   UGUCCCUGCU CCUUGCACAC ACCGCCCGUC ACACCACCCG A.UGGCUAGG
Vane-C.Frg   UGUCCCUGUU CUUUGUACAC ACCGCCCGUC GCUAUCUAAG A.AAAACUAC
Nove-C.Frg   UGUCCCUGUU CUUUGUACAC ACCGCCCGUC GCUAUCUAAG A.AAAACUAC
Noap-C.Frg   UGUCCCUGUU CUUUGUACAC ACCGCCCGUC GCUAUCUAAG A.CAAGCUAC
Encu-C.Frg   UGUCCCUGUC CUUUGUACAC ACCGCCCGUC GCUAUCUAAG A.CGGAAGGU
Basu-C.Frg   CGUUCCCGGG CCUUGUACAC ACCGCCCGUC ACACCACGAG A.GCCAGCCG
Agtu-C.Frg   CGUUCCCGGG CCUUGUACAC ACCGCCCGUC ACACCAUGGG A.GGCAGCUA
Stco-C.Frg   CGUUCCCGGG CCUUGUACAC ACCGCCCGUC ACGUCACGAA A.GGGAGCUG
Suso-C.Frg   CGUCCCUGCU CCUUGCACAC ACCGCCCGUC GCUCCACCCG AAGUGGGGGA
Hosa-C.Frg   AGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUACUACCG A.GAAGACGG
Zema-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACCG A.GAAGUCCA
Sace-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUAGUACCG A.GAAUUUGG
Grch-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUUCAACCG A.GGAAGUAC
Grca-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUUCAACCG A.GGAAGUAC
Psle-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCCUCAAUCG A.GGAAGUUC
Plfa-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACCG A.GGAAAAAC
Perk-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACCG A.GGAAGUUC
Baro-C.Frg   .GUCCCUGCC CUUUGUACAC ACCG.CCGUC GCUCCUACCC A.CCAAGUUU
Than-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACCG A.GGAAGUUU
Samu-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACCG A.GGAAGUUU
Babo-C.Frg   CGUCCCUGCC CUUUGUACAC ACCG.CCGUC GCUCCUACCC A.CCAAGUUU
Babi-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACCG A.GGAAGUUU
Prmi-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACC. ...GAAGUU.
Acca-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACCG A.CGAAGUCG
Oxno-C.Frg   AGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACCG A.GAAAAUCU
Pate-C.Frg   CGUCCCUGCC CUUUGUACAC ACCGCCCGUC GCUCCUACCG A.GGAAGUUU
Trbr-C.Frg   CGUCCCUGCC AUUUGUACAC ACCGCCCGUC GUUGUUUCCG A.GAAAGUUC
Eugr-C.Frg   UGUCCCUGCC AUUUGUACAC ACCGCCCGUC GUUGCUACCG A.CGAAAUUG

             651                                            700
Gila-C.Frg   CCGCGA.GCC CCCGCGCCUG GAGGAAGGAG AAGUCGUAAC AAGGUAUCCG
```

```
Hacu-C.Frg   GUCAAA.UCU  ..GGCUCCGC  AAGGGGGAUU  AAGUCGUAAC  AAGGUAGCCG
Meva-C.Frg   GUCGAA.CCU  G.GGCUCAGC  GAGGGGGGUG  AAGUCGUAAC  AAGGUAGCCG
Vane-C.Frg   UU.GAA.CAA  U.AUGUAUUA  GAUCUGAUAU  AAGUCGUAAC  AUGGUUGCUG
Nove-C.Frg   UU.GAA.CAA  U.AUGUAUUA  GAUCUGAUAU  AAGUCGUAAC  AUGGUUUCCG
Noap-C.Frg   UU.GAA.CAA  U.AUGUAUUA  GAUCUGAUAU  AAGUCGUAAC  AAGGUUUCCG
Encu-C.Frg   CU.GAG.UCC  U.GAGUGUUA  GAUAAGAUAU  AAGUCGUAAC  AUGGCUGCUG
Basu-C.Frg   CC.GAA.GGU  G.GGACAGAU  GAUUGGGGUG  AAGUCGUAAC  AAGGUAGCCG
Agtu-C.Frg   AC.CAC.GGU  A.GGGUCAGC  GACUGGGGUG  AAGUCGUAAC  AAGGUAGCCG
Stco-C.Frg   UC.GAA.GGU  G.GGACUGGC  GAUUGGGACG  AAGUCGUAAC  AAGGUAGCCG
Suso-C.Frg   UC.GAA.CUC  C.UUUCCCGC  GAGGGGGGAG  AAGUCGUAAC  AAGGUAGCCG
Hosa-C.Frg   UC.GAA.CUU  G.ACUAUCUA  GAGGAAGUAA  AAGUCGUAAC  AAGGUUUCCG
Zema-C.Frg   UU.GAA.CCU  U.AUCAUUUA  GAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Sace-C.Frg   AC.AAA.CUU  G.GUCAUUUA  GAGGAAGUAA  AAGUCGUAAC  AAG.UUUCCG
Grch-C.Frg   CGUGAG.CCU  U.AUCAUCUA  GAGGAUGAAG  AAGUCGUAAC  ACGGUUUCCG
Grca-C.Frg   CGUGAG.CCU  U.AUCAUCUA  GAGGAUGAAG  AAGUCGUAAC  ACGGUUUCCG
Psle-C.Frg   UGUGAA.CCA  A.AUCAUCUG  AAGAAUGAGA  AAGUCGUAAC  AUGGUUUCCG
Plfa-C.Frg   CGUAAA.UCC  U.AUCUUUUA  AAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Perk-C.Frg   UGCAAA.CCU  U.AUCACUUA  GAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Baro-C.Frg   UGUGAA.CCU  U.AUCACUUA  AAGGAAGGAG  AAGUCGUAAC  AAGGA.....
Than-C.Frg   UGUGAA.CCU  U.AUCACUUA  AAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Samu-C.Frg   UGUGAA.CCU  U.AACACUUA  GAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Babo-C.Frg   UGUGAA.CCU  U.AUCACUUA  AAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Babi-C.Frg   UGUGAA.CCU  U.AUCACUUA  AAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Prmi-C.Frg   AGUGAAACCU  U.AUCACUUA  GAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Acca-C.Frg   AUUGAA.CCU  U.ACCAUUUA  GAGGAAGGAG  AAGUCGUAAC  AAGGUCUCCG
Oxno-C.Frg   AGUAAA.CCA  U.AUCACUUA  GAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Pate-C.Frg   CGUAAA.CCU  U.AUCACUUA  GAGGAAGGAG  AAGUCGUAAC  AAGGUUUCCG
Trbr-C.Frg   ACCGAU.AUU  G.CUUCAAUA  GAGGAAGCAA  AAGUCGUAAC  AAGGUAGCUG
Eugr-C.Frg   GACGAU.GUC  C.AGCCACUA  GAGGAAGCAA  AAGUCGUAAC  AAGGUUGCUG


             701          715
Gila-C.Frg   U.AGGUGAAC  CUGCG
Hacu-C.Frg   U.AGGGGAAU  CUGCG
Meva-C.Frg   U.AGGGGAAC  CUGCG
Vane-C.Frg   U.UGGAGAAC  CAUUA
Nove-C.Frg   U.AGGUGAAC  CUGCG
Noap-C.Frg   U.AGGUGAAC  CUGCG
Encu-C.Frg   U.UGGAGAAC  CAUUA
Basu-C.Frg   U.AUCGGAAG  GUGCG
Agtu-C.Frg   U.AGGGGAAC  CUGCG
Stco-C.Frg   U.ACCGGAAG  GUGCG
Suso-C.Frg   U.AGGGGAAC  CUGCG
Hosa-C.Frg   U.AGGUGAAC  CUGCG
Zema-C.Frg   U.AGGUGAAC  CUGCG
Sace-C.Frg   UGAGGUGAAC  CUGCG
Grch-C.Frg   U.AGGUGAAC  CUGCG
Grca-C.Frg   U.AGGUGAAC  CUGCG
Psle-C.Frg   U.AGGUGAAC  CUGCG
Plfa-C.Frg   U.AGGUGAAC  CUGCG
Perk-C.Frg   U.AGGUGAAC  CUGCG
Baro-C.Frg   ..........  .....
Than-C.Frg   U.AGGUGAAC  CUGCG
Samu-C.Frg   U.AGGUGAAC  CUGCA
Babo-C.Frg   U.AGGUGAAC  CUGCA
Babi-C.Frg   U.AGGUGAAC  CUGCG
Prmi-C.Frg   U.AGGUGAAC  CUGCG
Acca-C.Frg   U.AGGUGAAC  CUGCG
Oxno-C.Frg   U.AGGUGAAC  CUGCG
Pate-C.Frg   U.AGGUGAAC  CUGCG
Trbr-C.Frg   U.AGGUGAAC  CUGCA
```

Eugr-C.Frg   U.AGGUGAAC CUGCG
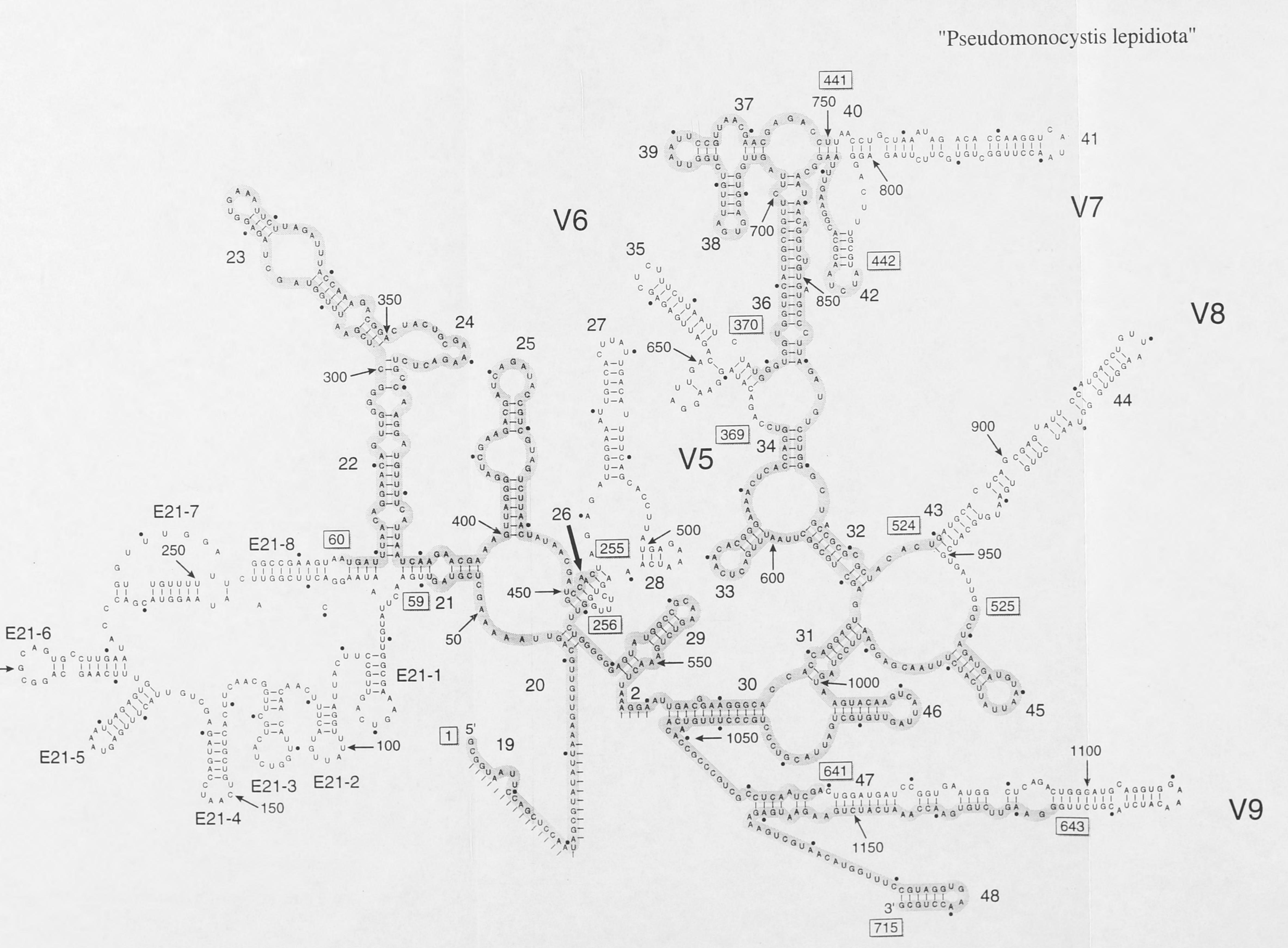
Figure 5.2

A secondary structure model of "Pseudomonocystis lepidiota" indicating the
representative positions of the conserved sequences of 30 taxa selected for the
evolutionary trees (shaded). The boxed numbers at the 5' and 3' ends of each
conserved fragment indicate the positions of each fragment in the aligned
conserved primary sequences (Figure 5.1).

"Pseudomonocystis lepidiota"

legend opposite Figure 5.3b. The second sentence should read:
"Bootstrap proportions are recorded above each branch."

Figure 5.3 Evolutionary trees made from SSU rRNA sequences from 30 taxa representing all kingdoms. Each contains a conserved set of 715 nucleotides.

a. An evolutionary tree made by the neighbor joining method, (Saitou and Nei, 1987), with the Kimura (Kimura, 1980) 2-parameter algorithm. The scale bar corresponds to the evolutionary distance between the nodes of the tree and is represented in the horizontal component of their separation in the figure. Bootstrap proportions from 1000 pseudo-samples are shown in brackets.

b. A 50% Majority-rule consensus of five most parsimonious trees made by PAUP version 3.1.1 (Swofford, 1993) using the heuristic search algorithm with 100 randomised replications. Length, 1839 steps; CI, 0.50. Branch lengths are recorded above each branch. The specified outgroups were the archeans, *Halobacterium cutirubrum, Methanococcus vaniellii,* and *Sulfolobus solfataricus* and the eubacteria *Agrobacterium tumifaciens, Bacillus subtilis,* and *Streptomyces coelicolor.*

T-PTP results are shown beneath the branch. A $*$ indicates significance and "NS" non-significance at the 95% level. PTP testing shows there is significant structure to the 423 informative sites (PTP = 0.01 from 100 randomised trees plus one actual tree).

c. A tree constructed by PAUP using the bootstrap algorithm. Bootstrap proportions from 1000 pseudo-samples are shown in brackets. Length, 1880 steps; CI, 0.499. The specified outgroups were as for b. above.

All T-PTP and PTP tests in this thesis were run for me by John Trueman, Division of Entomology, CSIRO, Canberra.

a.

Agrobacterium tumifaciens
Bacillus subtilis
Streptomyces coelicolor
Halobacterium cutirubrum
Methanococcus vanniellii
Sulpholobus solfataricus
Giardia lamblia
Acanthamoeba castellanii
Zea mays
"Pseudomonocystis lepidiota"
Plasmodium falciparum
Babesia bigemina
Theileria annulata
Babesia rodhaini
Babesia bovis
Sarcocystis muris
Perkinsus sp.
Prorocentrum micans
Oxytricha nova
Paramecium tetraurelia
Saccharomyces cerevisiae
Homo sapiens
"Gregarina caledia"
"Gregarina chortoicetes"
Euglena gracilis
Trypanosoma brucei
Encephalitozoon cuniculi
Nosema apis
"Nosema vespula"
Vairimorpha necatrix

Eubacteria
Archaea
Eukaryota
Microspora

(100) (100) (98) (97) (99) (81) (81) (84) (87) (93) (93) (91) (100) (100) (99)

0.1 scale

b.

Agrobacterium tumifaciens
Bacillus subtilis
Streptomyces coelicolor
Halobacterium cutirubrum
Methanococcus vanniellii
Sulpholobus solfataricus
Acanthamoeba castellanii
Babesia bigemina
Babesia bovis
Babesia rodhaini
Theileria annulata
Plasmodium falciparum
"Pseudomonocystis lepidiota"
Sarcocystis muris
Perkinsus sp.
Prorocentrum micans
Oxytricha nova
Paramecium tetraurelia
Zea mays
Saccharomyces cerevisiae
"Gregarina caledia"
"Gregarina chortoicetes"
Homo sapiens
Euglena gracilis
Trypanosoma brucei
Encephalitozoon cuniculi
Nosema apis
"Nosema vespula"
Vairimorpha necatrix
Giardia lamblia

Eubacteria
Archaea
Eukaryota
Microspora

80 100 100 100 100 100 60 60 60 60 60 100 60 60 100 60 60 60 100 100 100 100 100 100 * Node 39 100 * 100 "NS"

c.

Agrobacterium tumifaciens
Bacillus subtilis
Streptomyces coelicolor
Halobacterium cutirubrum
Methanococcus vanniellii
Sulpholobus solfataricus
Acanthamoeba castellanii
Babesia bigemina
Babesia bovis
Babesia rodhaini
Theileria annulata
"Gregarina caledia"
"Gregarina chortoicetes"
Plasmodium falciparum
"Pseudomonocystis lepidiota"
Sarcocystis muris
Oxytricha nova
Paramecium tetraurelia
Perkinsus sp.
Saccharomyces cerevisiae
Zea mays
Homo sapiens
Euglena gracilis
Trypanosoma brucei
Encephalitozoon cuniculi
Nosema apis
"Nosema vespula"
Vairimorpha necatrix
Giardia lamblia

Eubacteria
Archaea
Eukaryota
Microspora

(100) (100) (100) (100) (100) (100) (88) (100) (88) (88) (88) (88) (100) (100) (100)

Figure 5.3 Evolutionary trees made from SSU rRNA sequences from 30 taxa representing all kingdoms. Each contains a conserved set of 715 nucleotides.
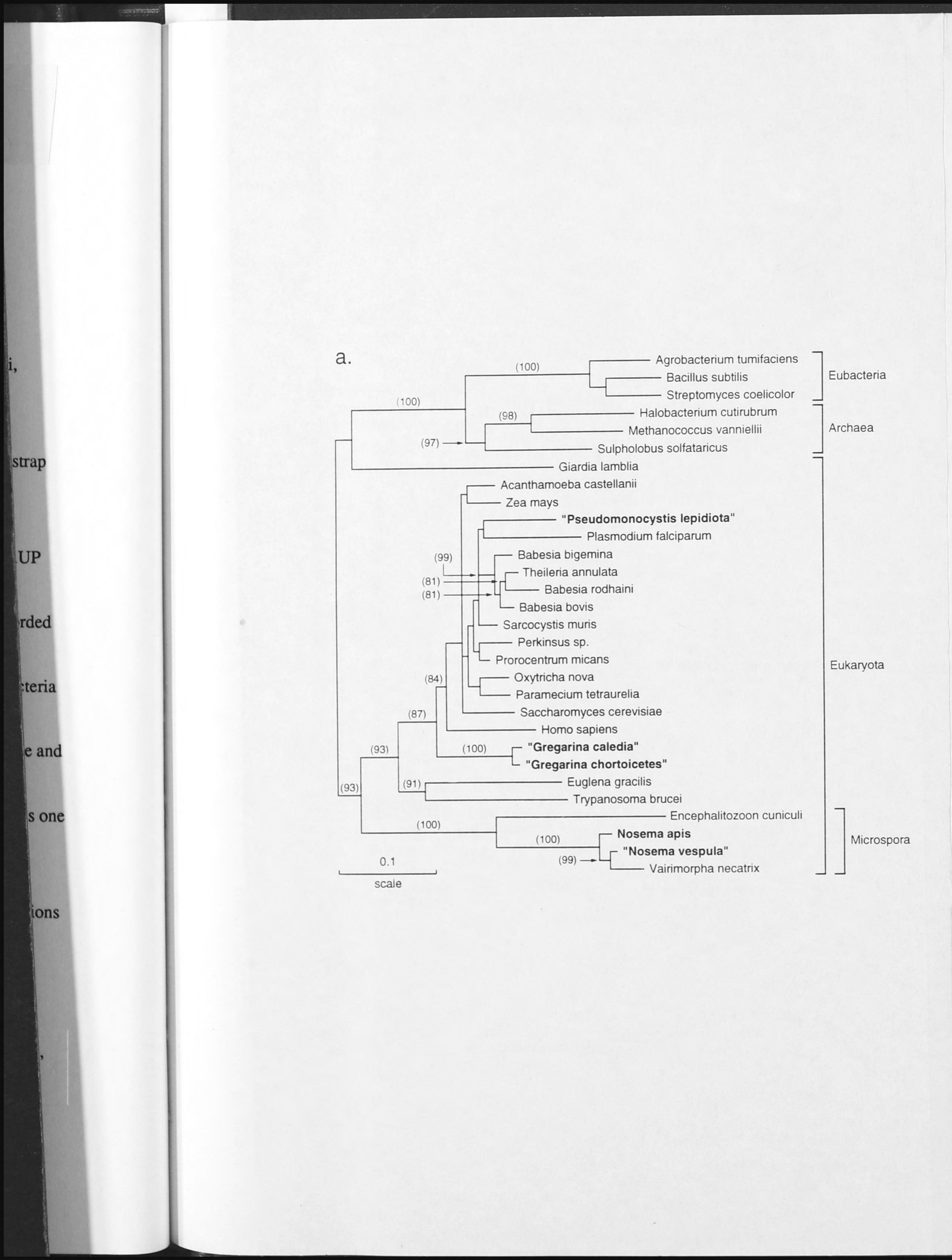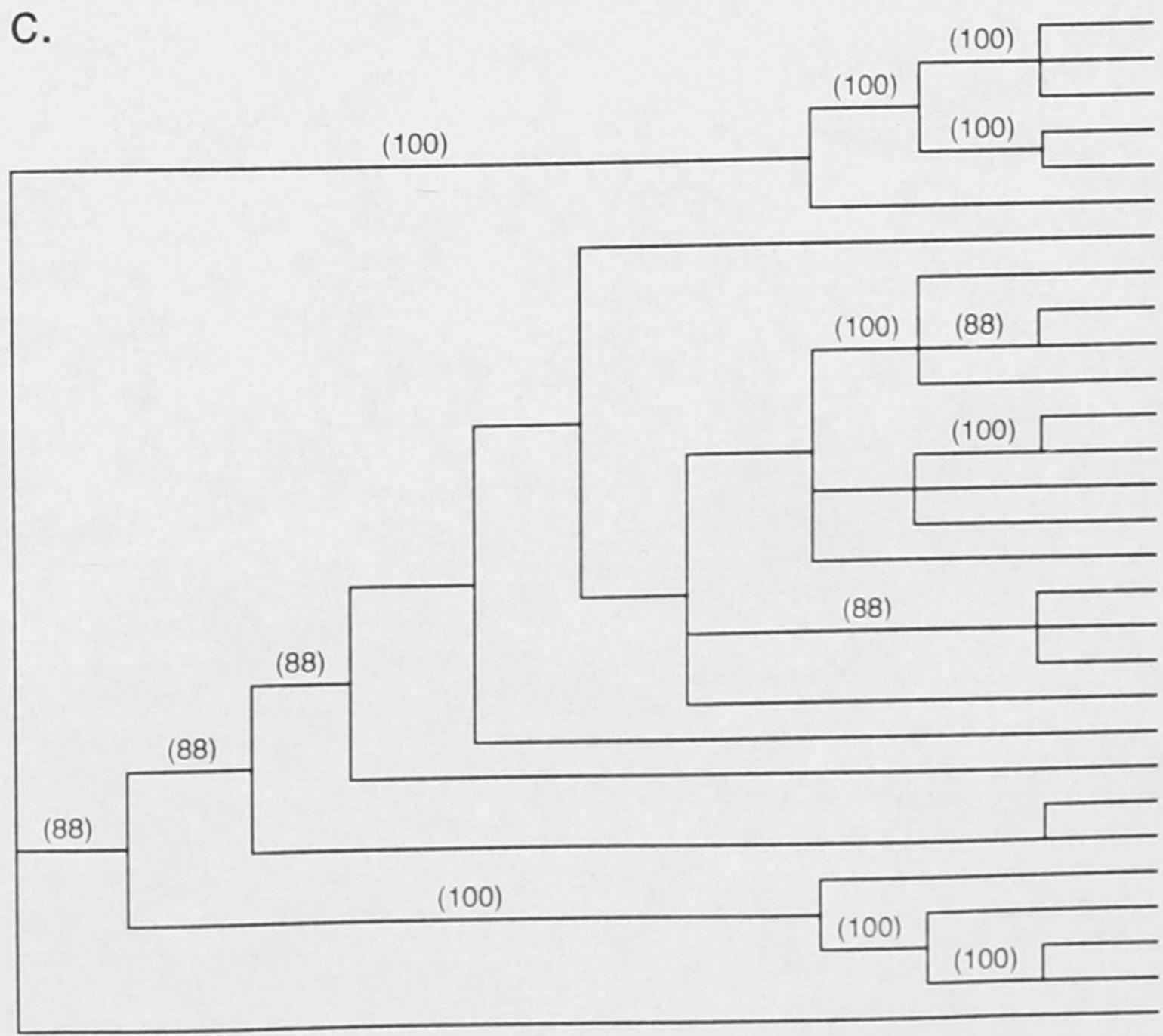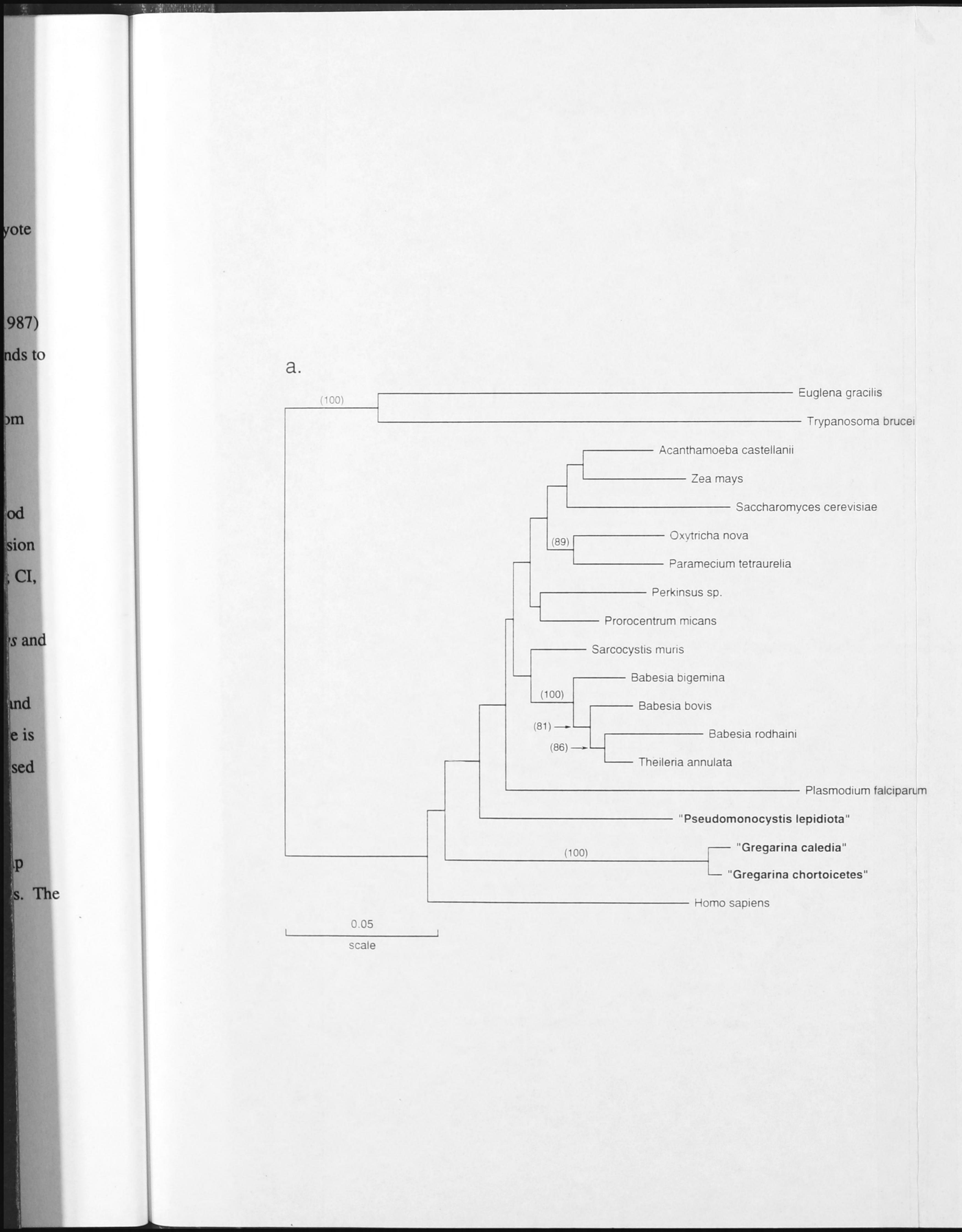
a. An evolutionary tree made by the neighbor joining method, (Saitou and Nei, 1987), with the Kimura (Kimura, 1980) 2-parameter algorithm. The scale bar corresponds to the evolutionary distance between the nodes of the tree and is represented in the horizontal component of their separation in the figure. Bootstrap proportions from 1000 pseudo-samples are shown in brackets.
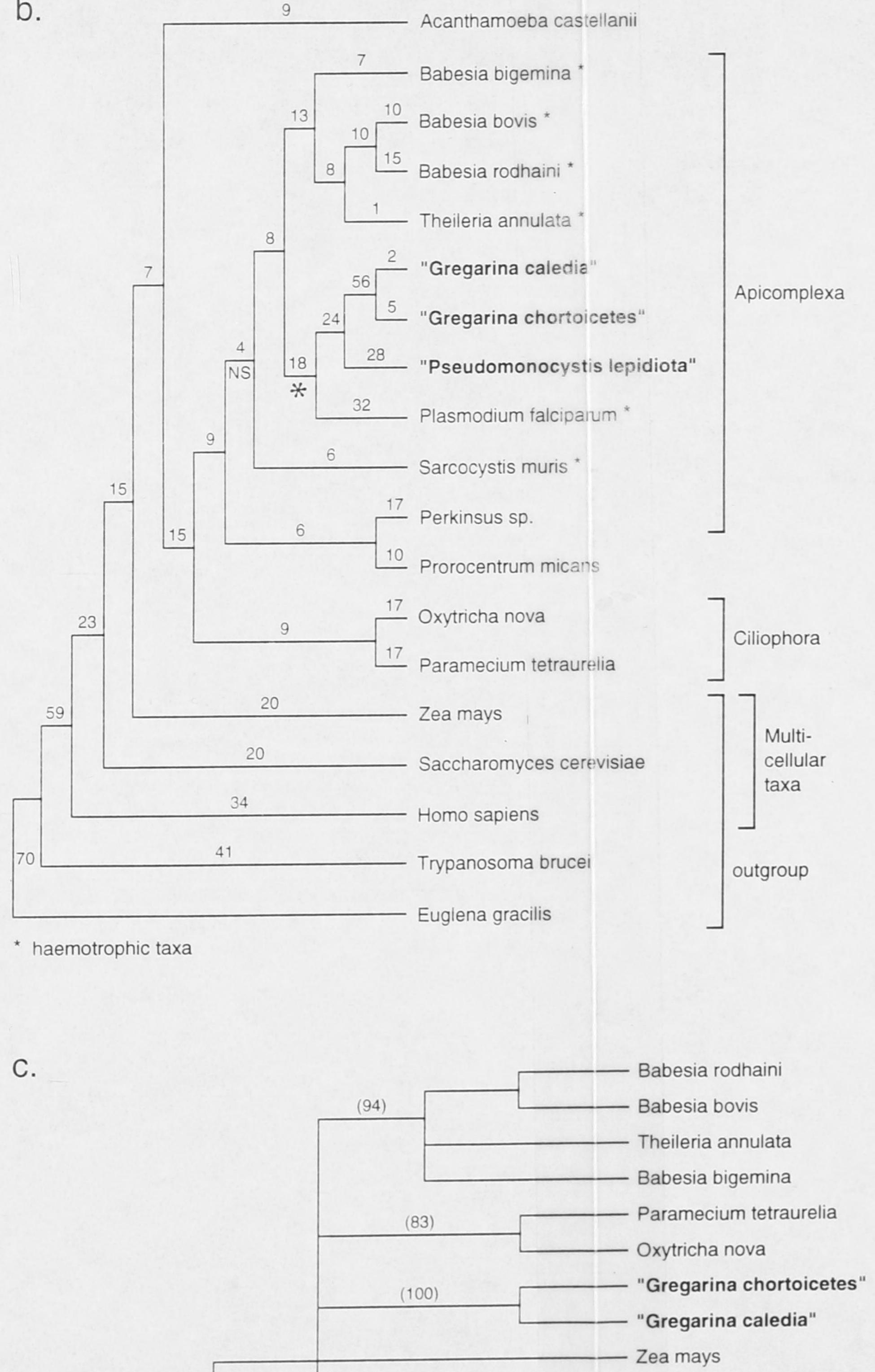
b. A 50% Majority-rule consensus of five most parsimonious trees made by PAUP version 3.1.1 (Swofford, 1993) using the heuristic search algorithm with 100 randomised replications. Length, 1839 steps; CI, 0.50. Branch lengths are recorded above each branch. The specified outgroups were the archeans, *Halobacterium cutirubrum, Methanococcus vaniellii,* and *Sulfolobus solfataricus* and the eubacteria *Agrobacterium tumifaciens, Bacillus subtilis,* and *Streptomyces coelicolor.*

T-PTP results are shown beneath the branch. A $*$ indicates significance and "NS" non-significance at the 95% level. PTP testing shows there is significant structure to the 423 informative sites (PTP = 0.01 from 100 randomised trees plus one actual tree).

c. A tree constructed by PAUP using the bootstrap algorithm. Bootstrap proportions from 1000 pseudo-samples are shown in brackets. Length, 1880 steps; CI, 0.499. The specified outgroups were as for b. above.

All T-PTP and PTP tests in this thesis were run for me by John Trueman, Division of Entomology, CSIRO, Canberra.

a.

(100) Euglena gracilis
Trypanosoma brucei
Acanthamoeba castellanii
Zea mays
Saccharomyces cerevisiae
(89) Oxytricha nova
Paramecium tetraurelia
Perkinsus sp.
Prorocentrum micans
Sarcocystis muris
Babesia bigemina
(100) Babesia bovis
(81)→ Babesia rodhaini
(86)→ Theileria annulata
Plasmodium falciparum
"Pseudomonocystis lepidiota"
(100) **"Gregarina caledia"**
"Gregarina chortoicetes"
Homo sapiens

0.05
scale

b.

9 Acanthamoeba castellanii
7 Babesia bigemina *
13 10 Babesia bovis *
10 15 Babesia rodhaini *
8 1 Theileria annulata *
8 2 "Gregarina caledia"
56 5 "Gregarina chortoicetes"
7 24 28 "Pseudomonocystis lepidiota"
4 18 32 Plasmodium falciparum *
NS * 6 Sarcocystis muris *
9 17 Perkinsus sp.
15 6 10 Prorocentrum micans
15 17 Oxytricha nova
23 9 17 Paramecium tetraurelia
59 20 Zea mays
20 Saccharomyces cerevisiae
34 Homo sapiens
70 41 Trypanosoma brucei
Euglena gracilis

Apicomplexa
Ciliophora
Multi-cellular taxa
outgroup

* haemotrophic taxa

c.

Babesia rodhaini
(94) Babesia bovis
Theileria annulata
Babesia bigemina
(83) Paramecium tetraurelia
Oxytricha nova
(100) **"Gregarina chortoicetes"**
"Gregarina caledia"
Zea mays
Sarcocystis muris
Saccharomyces cerevisiae
"Pseudomonocystis lepidiota"
Prorocentrum micans
Plasmodium falciparum
Perkinsus sp.
Acanthamoeba castellanii
Homo sapiens
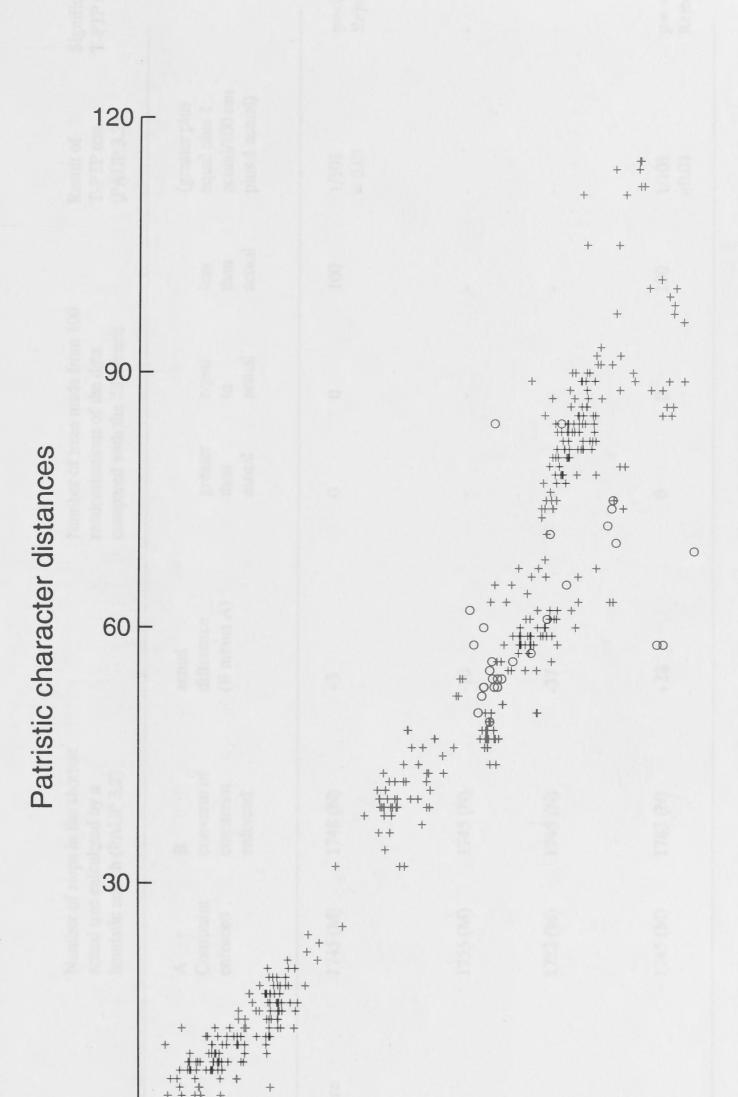Trypanosoma brucei
(100) Euglena gracilis

Figure 5.4

A scatter plot of pairwise comparisons of the patristic distances obtained by two different methods of phylogenetic analysis using sequences of 30 taxa from all kingdoms. The scatter plot was obtained using the DIPLOMO program of Weiller and Gibbs (1993).

The horizontal axis represents the patristic distances in the tree (Figure 5.3a) calculated by the NJTree program from distances (mutations/site) estimated by the 2-parameter correction methods of Kimura (1980).

The vertical axis represents the patristic distances (number of character changes) in the PAUP tree (Figure 5.3b).

"+" represents pairs of patristic distances from all 30 taxa except *Giardia lamblia*.

"o" represents pairs of patristic distances from pairs of taxa in which one member of the pair is *Giardia lamblia*
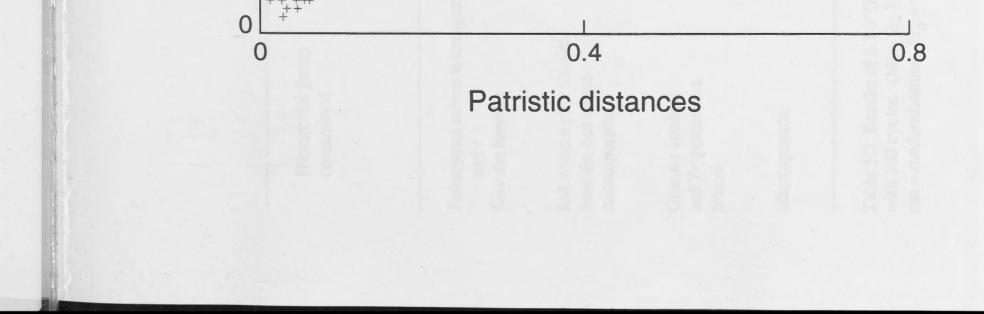
Patristic character distances

Patristic distances

120

90

60

30

0

0

0.4

0.8

Hierarchial group constrained	Number of steps in the shortest actual tree calculated by a heuristic search (PAUP 3.2)			Number of trees made from 100 randomisations of the data compared with the difference			Result of T-PTP test (PAUP 3.2)	Significance of T-PTP test
	A Constraint enforced	B converse of constraint enforced	actual difference (B minus A)	greater than actual	equal to actual	less than actual	(greater plus equal plus 1 actual/100 ran plus 1 actual)	
Eukaryotes minus microsporans and *Giardia lamblia*	1745 (M)	1748 (N)	+3	0	0	100	1/101 = 0.01	$p=<0.01$ Reject H_O
Eukaryotes plus *Giardia lamblia*, but minus microsporans	1755 (M)	1745 (N)	-10	-	-	-	-	-
Giardia lamblia and *Trypanosoma brucei*	1782 (M)	1745 (N)	-37	-	-	-	-	-
Microsporans	1745 (M)	1783 (N)	+38	0	0	100	1/101 =0.01	$p = <0.01$ Reject H_O

Table 5.2 Results of a T-PTP (Faith, 1991) test done on a data matrix made from SSU rRNA sequences from 30 eukaryote taxa with 423 informative characters using heuristic searches with 100 cycles. (M) = the hierarchial group was constrained for monophyly, (N) = the hierarchial group was constrained for non-monophyly, (-) = not done, ran = randomisations, H_O = null hypothesis
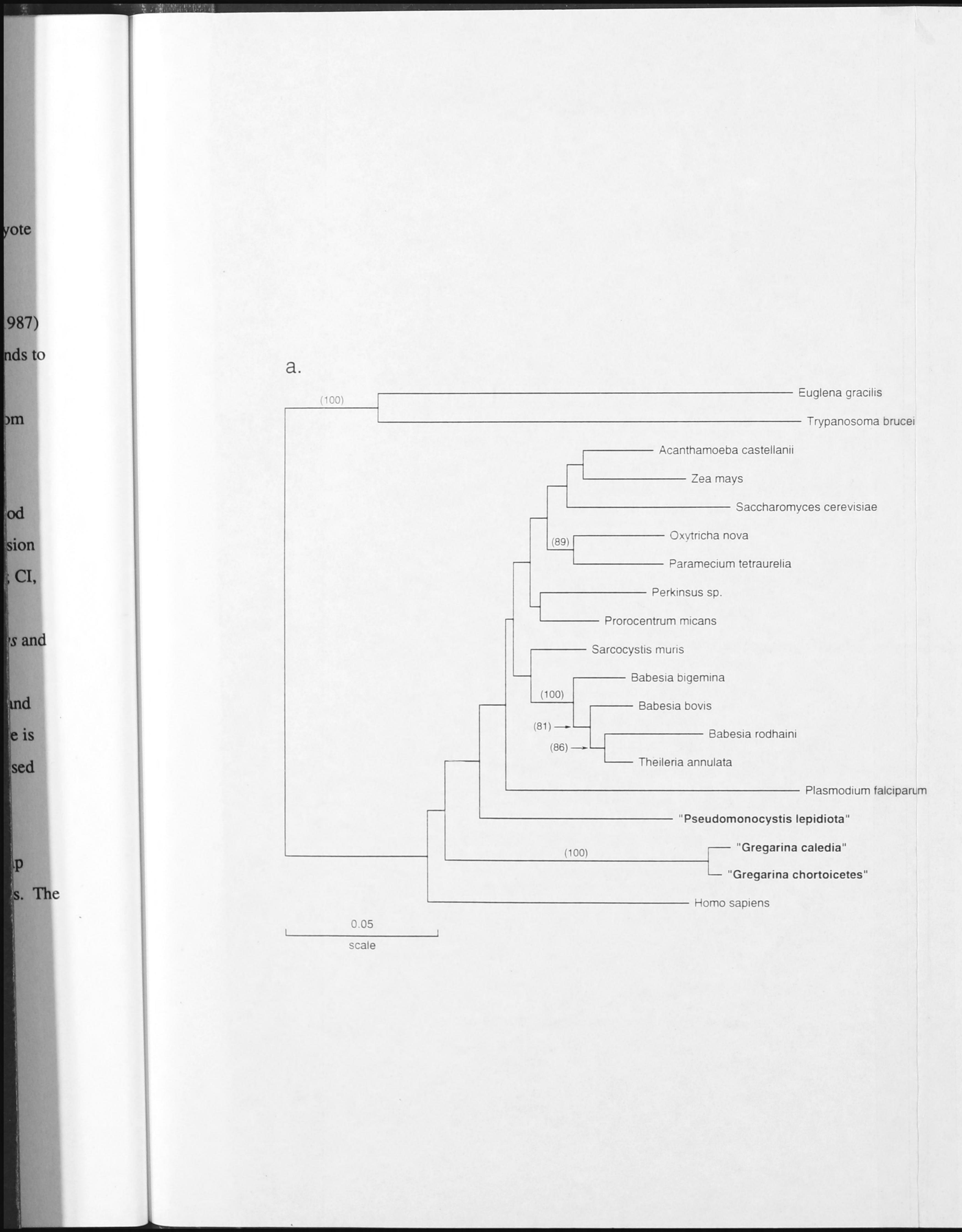
Figure 5.5.   Evolutionary trees made from SSU rRNA sequences from 19 eukaryote taxa, each containing a conserved set of 715 nucleotides.
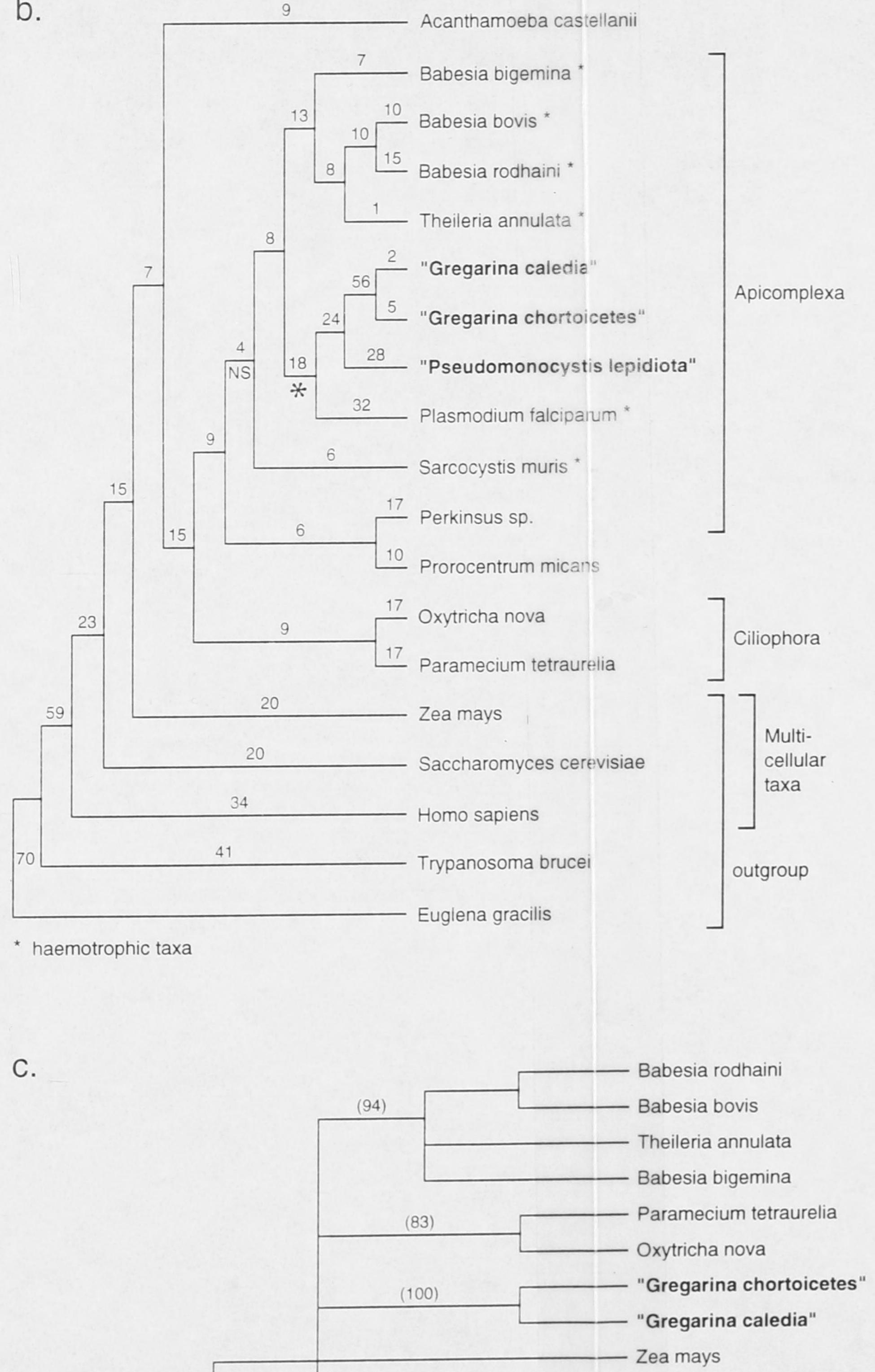
a. An evolutionary tree made by the neighbor joining method (Saitou and Nei, 1987) with the Kimura (Kimura, 1980) 2-parameter algorithm. The scale bar corresponds to the evolutionary distance between the nodes of the tree and is represented in the horizontal component of their separation in the figure. Bootstrap proportions from 1000 pseudo-samples are shown in brackets.
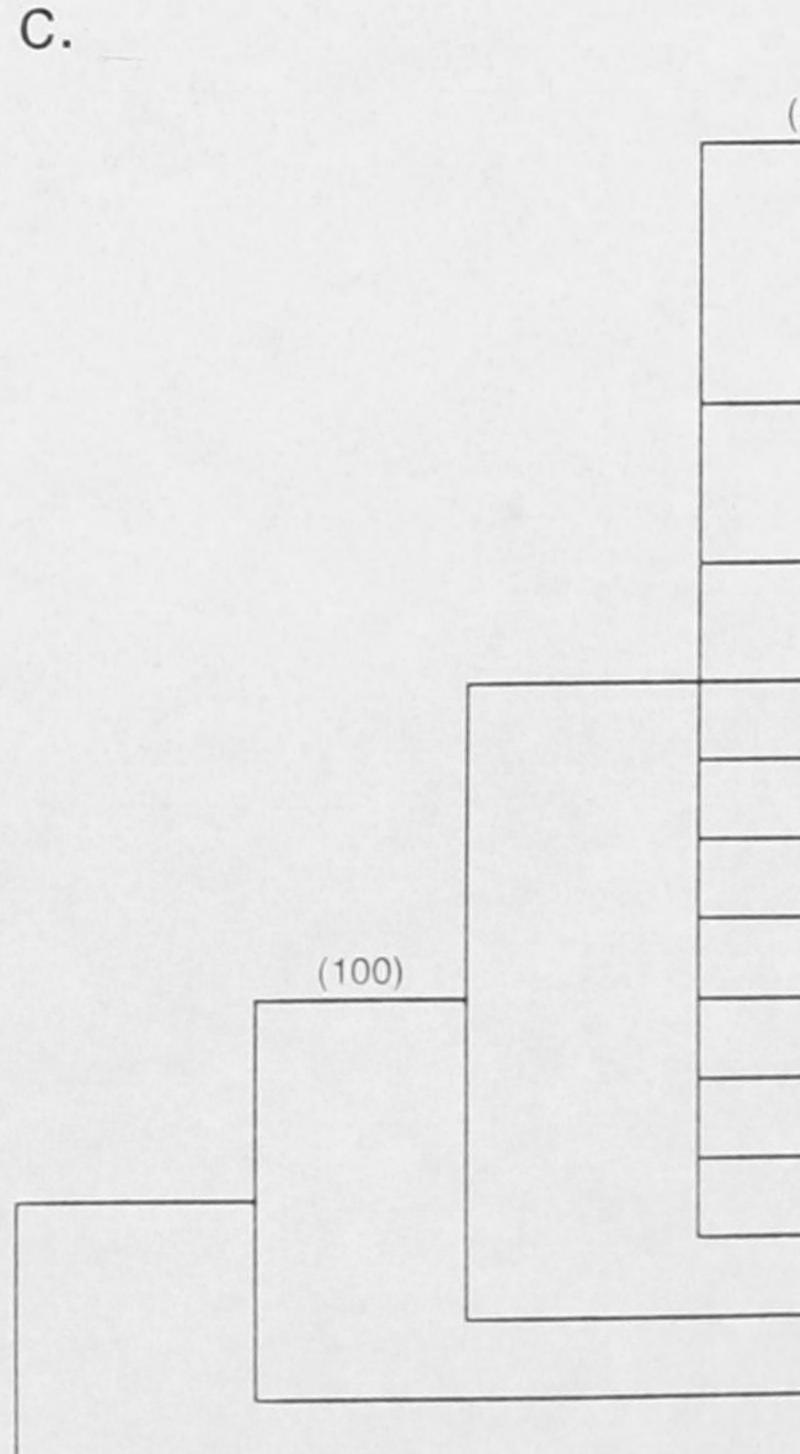
b. A single most parsimonious tree calculated by the maximum parsimony method with the computer program Phylogenetic Analysis Using Parsimony, PAUP version 3.1.1 (Swofford, 1993), using the heuristic search algorithm. Length, 644 steps; CI, 0.533; HI, 0.467. Branch lengths are recorded above each branch. The specified outgroups were *Trypanosoma brucei*, *Euglena gracilis*, *Homo sapiens, Zea mays* and *Saccharomyces cerevisiae*.

T-PTP results are shown beneath the line. A ∗ indicates significance and "NS" non-significance at the 95% level of significance. PTP testing shows there is significant structure to the 212 informative sites (PTP = 0.01 from 100 randomised trees plus one actual tree).

c. A 50% majority-rule consensus tree constructed by PAUP using the bootstrap option. Bootstrap proportions from 1000 pseudo-samples are shown in brackets. The specified outgroups were as for b. above.

a.

Euglena gracilis
Trypanosoma brucei
(100)

Acanthamoeba castellanii
Zea mays
Saccharomyces cerevisiae
Oxytricha nova
(89)
Paramecium tetraurelia
Perkinsus sp.
Prorocentrum micans
Sarcocystis muris
Babesia bigemina
Babesia bovis
(100)
Babesia rodhaini
(81)→
Theileria annulata
(86)→
Plasmodium falciparum
"Pseudomonocystis lepidiota"
"Gregarina caledia"
(100)
"Gregarina chortoicetes"
Homo sapiens

0.05
scale

b.

9
Acanthamoeba castellanii
7
Babesia bigemina *
13   10
Babesia bovis *
10
15
Babesia rodhaini *
8
1
Theileria annulata *
8
2
"Gregarina caledia"
56
5
"Gregarina chortoicetes"
24
18   28
"Pseudomonocystis lepidiota"
4
NS   *
32
Plasmodium falciparum *
6
Sarcocystis muris *
9
17
Perkinsus sp.
15   6
15   10
Prorocentrum micans
23
17
Oxytricha nova
9
17
Paramecium tetraurelia
59
20
Zea mays
20
Saccharomyces cerevisiae
34
Homo sapiens
70   41
Trypanosoma brucei
Euglena gracilis

Apicomplexa

Ciliophora

Multi-
cellular
taxa

outgroup

* haemotrophic taxa

c.

Babesia rodhaini
(94)
Babesia bovis
Theileria annulata
Babesia bigemina
(83)
Paramecium tetraurelia
Oxytricha nova
(100)
"Gregarina chortoicetes"
"Gregarina caledia"
Zea mays
Sarcocystis muris
Saccharomyces cerevisiae
"Pseudomonocystis lepidiota"
(100)
Prorocentrum micans
Plasmodium falciparum
Perkinsus sp.
Acanthamoeba castellanii
Homo sapiens
Trypanosoma brucei
Euglena gracilis

	Number of steps in the shortest actual tree calculated by a heuristic search (PAUP 3.2)			Number of trees made from 100 randomisations of the data compared with the difference			Result of T-PTP test (PAUP 3.2)	Significance T-PTP test
Hierarchial group constrained	A Constraint enforced	B converse of constraint enforced	actual difference (B minus A)	greater than actual	equal to actual	less than actual	(greater plus equal plus 1 actual/100 ran plus 1 actual)	
1. 6 Haemotrophs	649 (M)	644 (N)	-5	9	9	82	19/101 = 0.19	NS H_O not rejected
2. "Gregarina caledia" "Gregarina chortoicetes" "P. lepidiota" *P. falciparum*	644 (M)	647 (N)	+3	0	0	100	1/101 = 0.01	p =<0.01 Reject H_O

Table 5.3 Results of two T-PTP (Faith, 1991) tests done on a data matrix made from SSU rRNA sequences, from 19 eukaryote taxa with 212 informative characters, using heuristic searches with 100 cycles.   (M) = the hierarchial group was constrained for monophyly, (N) = the hierarchial group was constrained for non-monophyly, NS = non-significant, (-) = not done. ran = randomisations, H_O = null hypothesis

CHAPTER 6

GENERAL CONCLUSIONS

The aims of the work reported in this thesis were two fold. Firstly I aimed to determine the morphology and the life cycle of the apicomplexan parasite "Gregarina caledia" found in the grasshopper *Caledia captiva*. Secondly, I attempted to resolve the evolutionary relationships of five protoctists using comparisons of their primary sequences and secondary structure models of their SSU rRNA (18S and 16S).

Apicomplexans belonging to the class Gregarinia have not been evaluated phylogenetically before from their 18S rRNA sequences. However on the basis of morphology and host range, Levine (1988) has conjectured that they are a primitive class of apicomplexans which were derived with the coccidians, from the perkinsasids which inhabit marine polychaetes. These derived forms subsequently infested and became adapted to other marine, freshwater and terrestrial arthropods and were ancestral to the haemotrophic piroplasmids (babesias and theilerians) and haemosporids (plasmodiums) which are parasites of biting insects.

The intermediate phylogenetic position of the microsporans between the prokaryotes and eukaryotes, has been suggested previously using ultra-structure, secondary structure models and phylogenetic trees made from aligned SSU rRNA sequences with taxa from other kingdoms, using the SSU rRNA sequence from *Vairimorpha necatrix* (Vossbrinck *et al.,* 1987). This organism was shown to have a eukaryotic ultra-structure but prokaryotic 70S ribosomes and 16S rRNA that is unique in that it has no V4 region. Thus *Vairimorpha necatrix* is intermediate between the prokaryotes and eukaryotes. Phylogenetic trees have inferred it as a sister group to all eukaryotes, except for *Giardia lamblia* the phylogenetic position of which is obscured because of the biased base composition of its SSU rRNA sequence (Sogin *et al.,* 1989; Hashimoto *et al.,* 1994).

6.1 The morphology and life cycle of the gregarine parasite from *Caledia captiva*

The morphology and life cycle of "Gregarina caledia" which parasitises the grasshopper *Caledia captiva* was investigated to provide an insight into the biology of this organism and to compare this with that of similar organisms. I found this parasite had a monoxenous life cycle with free-living and parasitic stages and early sexual maturity, all of which are typical of most other septate gregarines. Comparisons of size ratios of body parts did not provide any useful phylogenetic information, but the size polymorphism indicated intra-species differences which may be genetically controlled.

6.2 The phylogenetic relationships of the microsporans and the apicomplexans.

My work has confirmed that the secondary structure models of SSU rRNA for the two nosemas, *Nosema apis* and "Nosema vespula" are almost identical and have the same unique secondary structure as that of *Vairimorpha necatrix*, namely that they lack the V4 region and helix 44 and only differ slightly from *Vairimorpha necatrix* in helix 41 and in the most variable helix, helix 35. These models also confirm that in at least two helices (32 and 36), these organisms have components which resemble those of bacteria. Except for the phylogenetic position of the zoomastigote *Giardia lamblia*, which seems to be closer to the bacteria instead of grouping with the other zoomastigote *Trypanosoma brucei* presumably because of the biased base composition of its SSU rRNA sequence, the microsporans, "Nosema vespula", *Nosema apis*, *Vairimorpha necatrix* and *Encephalitozoon cuniculi* formed a monophyletic clade which is a sister group to the other eukaryotes. The taxonomic methods I used were not successful in resolving the relative positions of the nosemas and *Vairimorpha necatrix* .

The secondary structure SSU rRNA models that I produced from the primary sequences of the two gregarines and the pseudogregarine are typically eukaryote, and except for helix 35, which is more stable if drawn as a eubacterial double hairpin, the models do not offer any new phylogenetic insights. The two tree-building models I used gave similar results, but failed to resolve the position of the gregarines, or to conform or deny the proposal of Levine (1988).

The gregarine clade itself is found by both methods to be monophyletic within the apicomplexans, with the two gregarines "Gregarina caledia" and "Gregarina chortoicetes" being more closely related to each other than they are to the

pseudogregarine "Pseudomonocystis lepidiota". This relationship correlates with their positions in their hosts and the morphological differences of their gametocysts; the two gregarines live in the host gut and have gametocysts measuring up to 350μm, and the pseudogregarines inhabit the host body cavity and have gametocyst diameters measuring from 1 to 2mm.

My work has shown that although the relative phylogenetic relationships of microsporans and gregarines with other taxa can be determined from the sequences of their SSU rRNA, the details of intertaxonomic relationships cannot. Other genes which could be informative for resolving phylogenies, are those which evolve more quickly such as those coding for mitochondrial proteins, however these would not be helpful in microsporans which lack these organelles. Other phylogenetically informative genes might include those which code for host-parasite interactions or for specific features such as the apical complex which varies between the apicomplexans. Until more is known about other genes and more apicomplexan and microsporan molecular sequences become available, the phylogenetic relationships of the apicomplexans will remain uncertain.

These uncertain conclusions do however, have an important consequence. For, what I have done is to show with five evidently disparate organisms, that the available computer-based tree-making programs could not un-ambiguously distinguish their relationships using aligned SSU rRNA sequences. If this is so for these organisms, then the implications for deductions of evolutionary relationships for other organisms using similar programs and similar molecular sequences, must be approached with caution.

63
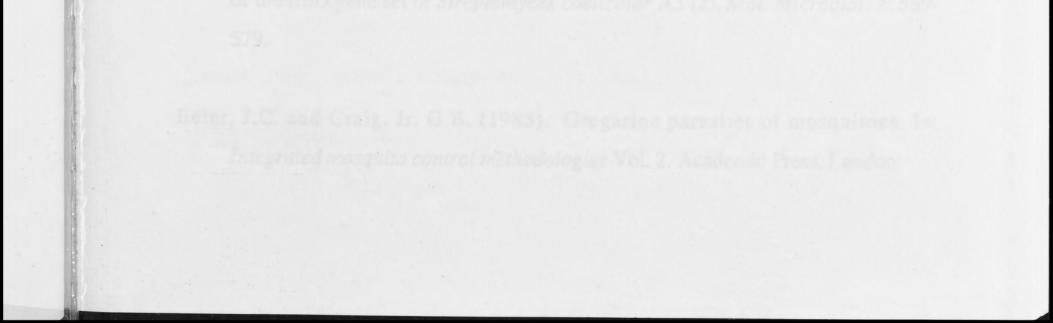
APPENDIX

1. Diet for *Heliothus armigera* (Denis Anderson, personal communication)

86g soy flour, 50g yeast, 60g wheatgerm, 12g agar, 3g nipagin, 3g ascorbic acid, 1g sorbic acid

Yeast was mixed with 200ml water and boiled for 5 minutes. All other ingredients were added except the ascorbic and sorbic acids, and diluted to 1.1 litres. Boiled and simmered for 10 minutes. Cooled to 37°C, before adding sorbic and ascorbic acids. Aliquots poured into paper cups (Lily, Australia) and allowed to set. Covered and stored at 4°C.

Arnheim, N,., Krystal, M., Schmikel, R., Wilson, G., Ryder, O., and Zimmer, E (1980). Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and ape. *Proc. Natl. Acad. Sci. USA* 77: 7323-7327.

REFERENCES

Allegré, C.F. (1948). Contributions to the life history of a gregarine parasite in grasshoppers. *Trans. Am. Microscop. Soc.* 68: 211-226.

Appel, A.G. (1983). Distribution and cyst formation of *Gregarina rhyoparobiae* Watson as a response to host starvation. *J. Parasitol.* 69: (2) 424-425.

Archie, J.W. (1989). A randomisation test for phylogenetic information in systematic data. *Syst. Zool.* 38: 219-252.

Atchley, W.R. and Fitch, W.M. (1991). Gene trees and the origins of inbred strains of mice. *Science* 254: 554-558.

Bailey, L. (1963). *Infectious diseases of the honey bee.* Land Books, London.

Baker, M.D., Vossbrinck, C.R. Maddox, J.V. and Undeen, A.H. (1994). Phylogenetic relationships among *Vairimorpha* and *Nosema* species (Microspora) based on ribosomal RNA sequence data. *J. Invertebr. Pathol.* 64: 100-106.

Barta, J.R., Jenkins, M.C. and Danforth, H.C. (1991). Evolutionary relationships of avian *Eimeria* species among other apicomplexian protozoa: Monophyly of the apicomplexia is supported. *Mol. Biol. Evol.* 8: 345-355.

Bayliss, H.A. and Bibb, M.J. (1988). Transcriptional analysis of the 16S rRNA gene of the rrnD gene set of *Streptomyces coelicolor* A3 (2). *Mol. Microbiol.* 2: 569-579.

Beier, J.C. and Craig, Jr. G.B. (1985). Gregarine parasites of mosquitoes. In: *Integrated mosquito control methodologies* Vol. 2. Academic Press, London.

Chastain, M. and Tinoco, I. Jr. (1991). Structural elements in RNA. In: *Progress in Nucleic Acid Research and Molecular Biology*. 41: 131-177. (W.E. Cohn and K. Moldav eds.) Academic Press, San Diego.

Brimacombe, R. (1980). Secondary structure homology between *Escherichia coli* 16S and *Saccharomyces cerevisiae* 18S Ribosomal RNA. *Biochem. Int.* 1: 162-171.

Britten, E.B. (1979). Coleoptera. In: *The Insects of Australia*. Melbourne University Press.

Bush, S.F. (1928). A study of the gregarines of the grasshoppers of Pieter Maritzburg, Natal. *Ann. Natal Museum* 6: 97-169.

Cali, A. and Owen, R.L. (1988). In: *Laboratory Diagnosis of Infectious Diseases: Principle and Practice, Vol I*. (Ballows, A., Hausler, W.J., Ohashi, M. and Turano, H. eds.), pp. 929-946. Springer-Verlag, New York. USA

Camin, J.H. and Sokal, R.R. (1965). A method for deducing branching sequences in phylogeny. *Evolution* 19: 311-326.

Canning, E.U. (1956). A new eugregarine of locusts, *Gregarine garnhami* n.sp., parasitic in *Schistocerca gregaria* Forsk. *J. Protozool.* 3: 50-62.

Canning, E.U. (1982). An evaluation of protozoal characteristics in relation to biological control of pests. *Parasitology* 84: 119-149.

Canning, E.U. (1989). Phylum Microspora. In: *Handbook of Protoctista* (L. Margulis, J.O. Corliss, M. Melkonian and D.J. Chapman eds.), Jones and Bartlett, Publishers, Boston.

Carleton, H.M. (1980). In: *Carleton's Histological Technique*. (R.A.B. Drury and E.A. Wallington eds.), 5th edition. Oxford University Press.

Cavalier-Smith, T. (1987). Eukaryotes with no mitochondria. *Nature* 326: 332-333.

Cheong, C., Varani, G., and Tinoco, I. Jr. (1990). Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* 346: 680-682.

Dover, G.A. (1982). Molecular drive: a cohesive mode of species evolution. *Nature*. 299: 111-117

Dover, G.A. and Tautz, D. (1986). Conservation and divergence in multigene families: alternatives to selection and drift. *Phil. Trans. R. Soc. Lond* B. 312: 275-289

Clark, C.G. (1987). On the evolution of ribosomal RNA. *J. Mol. Evol.* 25: 343-350.

Clark, C.G., Tague, B.W., Ware, V.C. and Gerbi, S.A. (1984). *Xenopus laevis* 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. *Nucleic Acids Res.* 12: 6197-6220.

Curgy, J., Vavra, J and Vivares, C. (1980). Presence of ribosomal RNA's with prokaryotic properties in microsporidia, eukaryotic organisms. *Biol. Cellulaire.* 38: 49-52.

Dahlberg, A.E. (1989). The functional role of ribosomal RNA in protein synthesis. *Cell* 57: 525-529.

Dams, E., Hendriks, L., Van de Peer, Y., Neefs, J.M., Smits, G., Vandenbempt, I. and De Wachter, R. (1988). Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res.* 16: Supplement: r87-99.

De Rijk, P., Neefs, J.M., Van de Peer, Y. and De Wachter, R. (1992). Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res.* 20: 2075-2089.

Devereux, J. Haeberli, P. and Smithies, O. (1984). A comprehensive set of Sequence Analysis Programs for the VAX. *Nucleic Acids Res.* 12 (1): 387-395.

Edwards, A.W.F. and Cavalli-Sforza, L.L. (1963). The reconstruction of evolution. *Ann. Hum. Genet.* 27: 105. (Also Heredity, 18: 553).

Edwards, A.W.F. and Cavalli-Sforza, L.L. (1964). *Phenetic and phylogenetic classification* (V.H.Heywood and J. McNeill eds.) pp.67-76. Systematics Association Publication N⁰ 6. London.

Eck, R.V. and Dayhoff, M.O. (1966). Inferences from protein sequence comparisons. In: *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Silver Spring, Maryland, USA.

Efron, B (1982). The jack-knife, the bootstrap and other resampling plans. CBNS-NSF Regional Conference Series in Applied Mathematics, Monograph 38. *Society of industrial and applied mathemetics* Philadelphia, USA.

Ellis, M.M. (1913). A descriptive list of the cephaline gregarines of the New World. *Trans. Am. Microscop. Soc.* 32: 259-296.

Ellis, J., Hefford, C., Baverstock, P.R., Dalrymple, B.P. and Johnson, A.M. (1992). Ribosomal DNA sequence comparison of *Babesia* and *Theileria*. *Mol. Biochem. Parasitol.* 54: 87-96.

Elwood, H.J., Olsen, G.J. and Sogin, M.L. (1985). The small-subunit ribosomal RNA gene sequences from the Hypotrichous Ciliates *Oxytricha nova* and *Stylonychia pustulata*. *Mol. Biol. Evol.* 2: 399-410.

Erlandson, M.A., Mukerji, M.K., Ewen, A.B. and Gillot, C. (1985). Comparative pathogenicity of *Nosema acridophagus* Henry and *Nosema cuneatum* Henry (Microsporidia: Nosematidae) for *Melanoplus sanguinipes* (FAB) (Orthoptera: Acrididae). *Canadian Entomol..* 117: 1167-1175.

Faith, D.P. (1991). Cladistic permutation tests for monophyly and nonmonophyly. *Syst. Zool.* 40: 366-375.

Faith, D.P. and Cranston, P.S. (1991). Could a cladogram this short have risen by chance alone? On permutation tests for cladistic structure. *Cladistics* 7: 1-28.

Farris, J.S. (1972). Estimating phylogenetic trees from distance matrices. *Am. Nat.* 106: 645-668.

Farris, J.S. (1974). Formal definitions of paraphyly and polyphyly. *Syst. Zool.* 23: 548-554.

Farris, J.S. (1989). The retention index and the rescaled consistency index. *Cladistics* 5: 417-419.

Felsenstein, J. (1978). Cases in which parsimony and compatability methods will be positively misleading. *Syst. Zool.* 27: 401-410.

Felsenstein, J (1981). Evolutionary trees from DNA sequences. A maximum likelihood approach. *J.Mol.Evol.* 17: 368-376.

Felsenstein, J. (1984). Distance methods for inferring phylogenies: A justification. *Evolution* 38: 16-24.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the Bootstrap. *Evolution* 39: 783-791.

Felsenstein, J. (1988). Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet.* 22: 521-565.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.5 (Computer software package and manual distributed by the author). Dept. Genetics, University of Washington, Seattle, WA. USA.

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R. and Raff, R. (1988). Molecular phylogeny of the animal kingdom. *Science* 239: 748-753.

Fitch, W.M. (1977). On the problem of discovering the most parsimonious tree. *Am. Nat.* 111: 223-257.

Fresco, J.R., Alberts, B.M., Doty, P. (1960). Some molecular details of the secondary structure of ribonucleic acid. *Nature* 188: 98-101.

Fitch, W.M. and Margoliash, E. (1967). Construction of phylogenetic trees. *Science* 155: 279-284.

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986). Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA* 83: 9373-9377.

Gagnon, S., Levesque, R.C., Sogin, M.L. and Gajadhar, A.A. (1993). Short Communication. Molecular cloning, complete sequence of the small subunit ribosomal RNA coding region and phylogeny of *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* 60: 145-148.

Gajadhar, A.A., Marquardt, W.C., Hall, R., Gunderson, J. O., Ariztia-Carmona, E.V. and Sogin, M.L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol. Biochem. Parasitol.* 45: 147-154.

Gatesy, J., De Salle, R. and Wheeler, W. (1993). Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Mol. Phylogenet. and Evol..* 2: 152-157.

Gillespie, J. (1991). *The causes of molecular evolution*. Oxford University Press, NY, USA.

Goggin, C.L. and Barker, S.C. (1993). Phylogenetic position of the genus *Perkinsus* (Protista, Apicomplexa) based on small subunit ribosomal RNA. *Mol. Biochem. Parasitol.* 60: 65-70.

Golding, G.B. (1983). Estimation of DNA and protein sequence divergence. An examination of some assumptions. *Mol. Biol. Evol.* 1: 125.

Hancock, J.M. and Dover, G.A. (1988). Molecular co-evolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. *Mol. Biol. Evol.* 5: 377-391

Goloboff, P.A. (1991). Homoplasy and the choice among cladograms. *Cladistics* 7: 215-232.

Gray, M.W., Sankoff, D. and Cedergren, R.J. (1984). On the evolutionary descent of organisms and organelles: A global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucleic Acids Res.* 12: 5837-5852.

Gunderson, J.H. and Sogin, M.L. (1986). Length variation in eukaryotic rRNAs: Small subunit rRNAs from the protists *Acanthamoeba castellanii* and *Euglena gracilis*. *Gene* 44: 63-70.

Gunderson, J.H., Elwood, H.J., Ingold, A., Kindle, K. and Sogin, M.L. (1987). Phylogenetic relationships between chlorophytes, chrysophytes and oomycetes. *Proc. Natl. Acad. Sci. USA* 84: 5823-5827.

Gutell, R.R., Weiser, B., Woese, C.R. and Noller, H.F. (1985). Comparative anatomy of 16-S-like ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* 32: 155-216.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166: 557-580.

Harry, O.G. (1965). Studies on the early development of the eugregarine *Gregarina garnhami*. *J. Protozool.* 12: (2) 296-305.

Harry, O.G. (1969). A jar for maintaining parasite-free insects and for collecting infected faeces. *Bull. Entomolog. Res.* 58: 833-844.

Harry, O.G. (1970). Studies on gregarines. BSc thesis submitted to the University of Birmingham.

page 71, Hennig, W. (1966) should read:

"Phylogenetic Systematics"

Hillis, D.M. (1987). Molecular versus morphological approaches to systematics. *Annu. Rev. Ecol. Syst.* 18: 23-42

Hillis, D.M. and Dixon, M.T. (1991). Ribosomal DNA: Molecular evolution and phylogenetic inference *The Quarterly Review of Biology* 66: 411-448

Heus, H.A. and Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* 253: 191-193.

Hartskeel, R.A., Schuitema, A.R.J. and deWachter, R. (1993). Secondary structure of the small subunit ribosomal RNA sequence of the microsporidium *Encephalitozoon cuniculi*. *Nucleic Acids Res*. 21 (6): 1489

Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K. and Hasewaga, M. (1994). Protein phylogeny gives a robust determination for early divergences of eukaryotes: Phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol. Biol. Evol*. 11 (1): 65-71.

Hedges, S.B. (1992). The number of replications needed for accurate estimations of the Bootstrap P value in phylogenetic studies. *Mol. Biol. Evol*. 9: 366-369.

Hendy, M.D. and Penny, D. (1989). A framework for the qualitative study of evolutionary trees. *Syst. Zool*. 38: 297-309.

Hennig, W. (1966). *Phylogenetic studies* (translated by D. Dwight Davies and Rainer Zangerl). University of Illinois Press, Urbana, Ill. USA.

Hillis, D.M. and Dixon, M.T. (1989). Vertebrate phylogeny: evidence from 28S ribosomal DNA sequences. In: *The Hierachy of Life* (B. Fernholm, K. Bremer and H. Jornvall eds..) pp. 355-367. Elsevier Science Publishers (Biomedical Division), Amsterdam.

Hillis, D.M. and Bull, J.J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol*. 42 (2): 182-192.

Hillis, D.M., Bull, J.J., White, M.E. Badgett, M.R. and Molineux, I.J. (1992). Experimental phylogenetics: Generation of a known phylogeny. *Science* 255: 589-592.

Holmes, D.S. and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem*. 114: 193-197.

Jaeger, J.A., Turner, D.H. and Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. USA* 86: 7706-7710.

Huelsenbeck, J.P. (1991). Tree length distribution skewness: an indicator of phylogenetic information. *Syst. Zool.* 40: 257-270.

Huelsenbeck, J.P. and Hillis, D.M. (1993). Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* 42 (3): 247-264.

Hui, I. and Dennis, P.P. (1985). Characterisation of the ribosomal RNA gene clusters in *Halobacterium cutirubrum. J. Biol. Chem.* 260: 899-906.

Huysmans, E. and De Wachter, R. (1986). Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res.* 14: (Suppl.) r73-r118.

Ishihara R. and Hayashi, Y. (1968). Some properties of ribosomes from the sporoplasm of *Nosema bombycis. J. Invertebr. Pathol.* 11: 377-385.

Jarsch, M. and Boeck, A. (1985). Sequence of the 16S ribosomal RNA from *Methanococcus vannielii*: Evolutionary implications *Syst. Appl. Microbiol.* 6: 54-59.

Johnson, A.M. and Baverstock, P.R. (1989). Rapid ribosomal RNA sequencing and the phylogenetic analysis of protists. *Parasitol. Today* 5 (4): 102-105.

Johnson, A.M., Murray, P.R., Illana, S., Baverstock, P.J. (1987). Rapid nucleotide sequence analysis of the small subunit ribosomal RNA of *Toxoplasma gondii*: Evolutionary implications for the Apicomplexa. *Mol. Biochem. Parasitol.* 25: 239-246.

Johnson, A.M., Illana, S., Hakendorf, P., Baverstock, P.R. (1988). Phylogenetic relationships of the apicomplexan protist *Sarcocystis* as determined by small subunit ribosomal RNA comparison. *J. Parasitol.* 74: 847-860.

Johnson, A.M., Fielke, R., Lumb, R. and Baverstock P.R. (1990). Phylogenetic relationships of *Cryptosporidium* determined by ribosomal RNA sequence comparison. *Int. J. Parasitol.* 20: 141-147.

Jones, T.C., Yeh, S. and Hirsch, J.G. (1972). The interaction between *Toxoplasma gondii* and mammalian cells. *J. Exp. Med.* 136; 1157-1172.

Jukes, T.H. and Cantor, C.R. (1969). Evolution of protein molecules. In: *Mammalian protein metabolism* (H.N. Munro ed.) pp. 21-32. Academic Press. New York. USA.

Key, K.H.L. (1979). Orthoptera. In: *The Insects of Australia*. Melbourne University Press.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.

Kluge, A.G. and Farris, J.S. (1969). Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18; 1-32.

Knoll, A.H. (1992). The early evolution of eukaryotes: A geological perspective. *Science* 256: 622-625.

Kudo, R.R. (1954). *Protozoology* 4th edition. Publisher C.C. Thomas, U.S.A.

Kumazaki, T., Hori, H. and Osawa, S. (1983). Phylogeny of Protozoa deduced from 5S rRNA sequences. *J. Mol. Evol.* 19: 411-419.

Laird, M. (1959). Gregarines from laboratory colonies of flour beetles, *Tribolium castaneum* Herbst and *T. confusum* Duval, at Montreal. *Can. J. Zool.* 37: 378-381.

Larsen, N. (1992). Higher order interactions in 28S rRNA. *Proc. Natl. Acad. Sci. USA* 89: 5044-5048.

Larsson, R. (1986). Ultrastructure, function and classification of microsporidia. *Progr. Protistol.* 1: 325-390.

Lehninger, A.L. (1982). *Principles of biochemistry*. Worth Publishers Inc. USA.

Levine, N.D. (1970). Taxonomy of the Sporozoa. *J. Parasitol.* 56: (Sect. II) 208-209.

Levine, N.D. (1978). *Perkinsus* gen. n. and other new taxa in the protozoan phylum Apicomplexa. *J. Parasitol.* 64: 549.

Levine, N.D. (1988). *The protozoan phylum Apicomplexia*. CRC Press Inc. Boca Raton, Florida, USA.

Li, W.H. (1981). A simple method for constructing phylogenetic trees from distance matrices. *Proc. Natl. Acad. Sci. USA* 78: 1085-1089.

Li, W.H. and Graur, D. (1991). *Fundamentals of molecular evolution*. Sinauer Associates, Inc., Sunderland, MA., USA.

Lockwood, J.A. and Debrey, L.D. (1990). Direct and indirect effects of a large scale application of *Nosema locustae* (Microsporidia : Nosematidae) on rangeland grasshoppers (Orthoptera : Acrididae). *Entomol. Soc. Am.* 83: 377-383.

Lynn, D.H. and Small, E.B. (1981). Protist kinetids: structural conservatism, kinetid structure and ancestral states. *BioSystems* 14: 377-385.

Maddison, W.P. and Maddison, D.R. (1992). *MacClade: Analysis of Phylogeny and Character evolution* Publishers Sinauer Associates, Inc., Sunderland, MA., USA.

Mandal, R.K. (1984). The organisation and transcription of eukaryotic ribosomal RNA genes. *Prog. in Nucl. Acid Res and Mol. Biol.* 31: 115-125.

Malone, L.A., Broadwell, A. H., McIvor, C.A., Lindridge E.T., and Ninham, J. A. (1993). Ribosomal RNA genes of two microsporidian species, *Nosema apis* and *Vavraia oncoperae* are very variable. *J. Invertebr. Pathol.* 64: 151-152.

Margulis, L. (1974a). Five-kingdom classification and the origin and evolution of cells. *Evol. Biol.* 7: 45-78.

Margulis, L. (1974b). The classification and evolution of prokaryotes and eukaryotes. In: *Handbook of Genetics* (King, R.C. ed.) Vol. 1, pp1-4. New York. Plenum Press.

Margulis, L. (1980). Undulipodia, flagella and cilia. *Biosystems* 12: 105-108.

Margulis, L. (1989). Introduction. In: *Handbook of Protoctista.*(L. Margulis, J.O. Corliss, M. Melkonian, and D.J. Chapman, eds.), Jones and Bartlett Publishers, Boston, USA.

Maroteaux, L., Herzog, M. and Soyer-Gobillard, M.O. (1985). Molecular organization of dinoflagellate ribosomal DNA: Evolutionary implications of the deduced 5.8S rRNA secondary structure. *Biosystems* 18: 307-319.

McCallum, F.S. and Maden, B.E.H. (1985). Human 18S ribosomal RNA sequence inferred from DNA sequence. *Biochem. J.* 232: 725-733.

McCray, Jr. E.M., Fay, R.W. and Schoof, H.F. (1970). The bionomics of *Lankesteria culicis* and *Aedes aegypti*. *J. Invertebr. Pathol.* 16: 42-53.

McCutchan, T.F., De la Cruz, V.G., Lal, A.A., Gunderson, J.H., Elwood, H.J. and Sogin, M.L. (1988). Primary sequences of two small subunit ribosomal RNA genes from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 28: 63-68.

Morgan, E.A. (1982). Ribosomal RNA genes in *Escherichia coli*. In H. Busch and L. Rothblum (eds), *The cell nucleus:* rDNA, Vol.X, Part A, pp.1-29. Academic Press, New York. USA.

Medlin, L., Elwood, H.J., Stickel, S. and Sogin, M.L. (1988). The characterization of enzymatically amplified eukaryote 16S-like rRNA-coding regions. *Gene* 71: 491-499.

Melhorn, H. and Schein, E. (1984). The piroplasms: life cycle and sexual stages. *Adv. Parasitol.* 23: 37-103.

Messing, J. (1983). New M13 vectors for cloning. *Methods Enzymol.* 101: 20-78.

Messing, J., Carlson, J., Hagen, G., Rubenstein, I. and Oleson, A. (1984). Cloning and sequencing of the ribosomal RNA genes in maize: The 17S region. *DNA* 3: 31-40.

Moran, C, (1979). The structure of the hybrid zone in *Caledia captiva*. *Heredity* 42: (1) 13-32.

Moritz, C and Hillis, D.M. (1990). Molecular systematics: context and controversies. In: *Molecular Systematics* (D.M. Hillis and C. Moritz eds.), pp.1-10. Sinauer Associates Inc. Sunderland. MA., USA.

Mullis, K.B. and Faloona, F.A. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction. *Methods Enzymol.* 155: 335-350.

Needleman, S.B. and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48: 443-453

Neefs, J.M., Van de Peer, Y., De Rijk, P., Goris, A. and De Wachter, R. (1991). Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res.* 19: 1987-2015.

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press.

Nussinov, R. and Jacobson, A. (1980). Fast algorithm for predicting the secondary structure of single stranded RNA. *Proc. Natl. Acad. Sci. USA*. 77: 6309-6313.

Pipas, J.M. and McMahon, J.E. (1975). Method for predicting RNA secondary structure. *Proc. Natl. Acad. Sci. USA* 72: 2017-2021.

Nei, M. (1991). Relative efficiences of different tree-making methods for molecular data. Pages 90-128. In *Phylogenetic analysis of DNA sequences* (M.M.Miyamota and J. Cracraft, eds.). Oxford Univ. Press. Oxford, England.

Olsen, G.J. (1987). Earliest phylogenetic branchings: Comparing rRNA-based evolutionary trees inferred with various techniques. *Cold Spring Harbour Symposia on Quantitative Biology* Vol LII Cold Spring Harbour Laboratory.

Olsen, G.J., Pace, N.R., Nuell, M., Kaine, B.P., Gupta, R. and Woese, C.R. (1985). Sequence of the 16S rRNA gene from the thermoacidophilic archaebacterium *Sulfolobus solfataricus* and its evolutionary implications. *J. Mol. Evol.* 22: 301-307.

Olsen, G.J. and Woese, C.R. (1993). Ribosomal RNA: A key to phylogeny. *FASEB J.* 7: 113-123.

Pace, N.R., Smith, D.K., Olsen, G.J. and James, B.D. (1989). Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA - a review. *Gene* 82: 65-75.

Penny, D. (1982). Towards a basis for classification: the incompleteness of distance measures, incompatibility analysis and phenetic classification. *J. Theor. Biol.* 96: 129-142.

Penny, D. and Hendy, M.D. (1985). Testing methods of evolutionary tree construction. *Cladistics* 1: 266-278.

Perkin Elmer Cetus (1990). GeneAmp™ DNA Amplification Reagent Kit with *Amplitaq*™ Recombinant Taq DNA Polymerase. Norwalk, CT., 06859. USA

Perkins, F.O. (1976). Zoospores of the oyster pathogen, *Dermocystidium marinum* I. Fine structure of the conoid and other sporozoan-like organelles. *J. Parasitol.* 62: 959-974.

Nei, M. (1987). Relative efficiencies of different tree-making methods for molecular data. Pages 90-128. In Phylogenetic Analysis of DNA Sequences (M.M.Myamoto and J.Cracraft, eds.). Oxford Univ Press, Oxford, England.

Olsen, G.J. (1987). Earliest phylogenetic branchings: Comparing rRNA-based evolutionary trees inferred with various techniques. Cold Spring Harbor Symposia on Quantitative Biology. Vol LII Cold Spring Harbor Laboratory.

Olsen, G.J., Pace, N.R., Nuell, M., Kaine, B.P., Gupta, R. and Woese, C.R. (1985). Sequence of the 16S rRNA gene from the thermoacidophilic archaebacterium Sulfolobus solfataricus and its evolutionary implications. J. Mol. Evol. 22: 301-307.

Rothschild, L.J. Ragan, M.A. Coleman, A.W. Heywood, P. and Gerbi, S.A. (1986). Are rRNA sequence comparisons the Rosetta Stone of phylogenetics? *Cell* 47: 640

Pace, N.R., Smith, D.K., Olsen, G.J. and James, B.D. (1989). Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA - a review. Gene 82: 65-75.

Penny, D. (1982). Towards a basis for classification: the incompleteness of distance measures, incompatibility analysis and phenetic classification. J. Theor. Biol. 96: 129-142.

Penny, D. and Hendy, M.D. (1985). Testing methods of evolutionary tree construction. Cladistics 1: 266-278.

Perkin Elmer Cetus (1990). GeneAmp[TM] DNA Amplification Reagent Kit with AmpliTaq[TM] Recombinant Taq DNA Polymerase. Norwalk, CT, 06859, USA.

Perkins, F.O. (1976). Zoospores of the oyster pathogen, Dermocystidium marinum. I. Fine structure of the conoid and other sporozoan-like organelles. J. Parasitol.

Pleij, C.W.A., van Belkum, A., Rietreld, K. and Bosch, L. (1986). In: *Structure and dynamics of RNA* (P.H. van Knippenberg, and C.W. Hilbers, eds..) pp. 87-98. Plenum Press.

Reddy, G.R., Chakrabarti, D., Yowell, C.A. and Dame, J.B. (1991). Sequence micro-heterogeneity of the three small subunit ribosomal RNA genes of *Babesia bigemina:* Expression in erythrocyte culture. *Nucleic Acids Res.* 19: 3641-3645.

Reik, E.F. (1979). Hymenoptera. In: *The Insects of Australia.* Melbourne University Press.

Ridley, M. (1989). *Evolution and classification. The reformation of cladism.* Publishers, Langman Scientific and Technical. Hong Kong.

Rubtsov, P.M., Musakhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skryabin, K.G. and Bayev, A.A. (1980). The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from *Saccharomyces cerevisiae. Nucleic Acids Res.* 8: 5779-5794.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. Erlich, H.A. (1988). Primer directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science.* 239: 487.

Saitou, N. and Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4(4): 406-425.

Saitou, N. and Nei, M. (1989). Relative efficiencies of the Fitch-Margoliash, Maximum-Parsimony, Maximum Likelihood, Minimum-Evolution and Neighbour-Joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.* 6: 514-525.

Saitou, N. and Imanishi, T. (1989). Relative efficiencies of the Fitch-Margoliash, Maximum-Parsimony, Maximum-Likelihood, Minimum-Evolution and Neighbour-Joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.* 6: 514-525.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning. A laboratory Manual.* Second edition. Cold Spring Harbour Press. Cold Spring Harbour, New York. USA.

Shaw, D.D., Webb, G.C. and Wilkinson, P. 1976. Population cytogenetics of the genus *Caledia* (Orthoptera: Acridinae) II Variation in the pattern of C-banding. *Chromosoma* 56: 169-190.

Shaw, J.P. Marks, J. Shen, C.C. and Shen, C.J. (1989). Anomalous and selective DNA: Mutations of the old world monkey α-globin genes. *Proc. Natl. Acad. Sci. USA.* 86: 1312-1316.

Sleigh, M.A. (1973). *The biology of the protozoa.* Elsevier, New York. USA.

Sogin, M.L. and Elwood, H.J. (1986). Primary structure of the *Paramecium tetraurelia* small-subunit rRNA coding region: Phylogenetic relationships within the *Ciliophora. J. Mol. Evol.* 23: 53-60.

Sogin, M.L., Elwood, H.J. and Gunderson, J.H. (1986). Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proc. Natl. Acad. Sci. USA* 83: 1383-1387.

Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A. and Peattie, D.A. (1989). Phylogenetic meaning of the kingdom concept: An unusual ribosomal RNA from *Giardia lamblia. Science.* 243: 75-77.

Sokal, R.R. and Sneath, P.H. (1963). *Principles of numerical taxonomy.* W.H. Freeman, San Fransisco.

Studnicka, G.M., Rah, G.M., Cummings, I.W. and Salser, W.A. (1978). Computer method for predicting the secondary structure of single stranded RNA. *Nucleic Acids Research* 5: 3365-3387.

Sourdis, J. and Nei, M. (1988). Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. *Mol. Biol. Evol.* 5: 298-311.

Sprague, V. (1941). Studies on *Gregarina blattarum* with particular reference to the chromosome cycle. *Illinois Biol. Monogr.* 18: 5-57.

Sprague, V., Becnel, J.J. and Hazard, E.I. (1992). Taxonomy of Phylum Microspora. *Crit. Rev. Microbiol.* 18: 285-395.

Staden, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Res..* 10: 4731-4751.

Staden, R. (1987). Computer handling of DNA sequencing projects. In: *Nucleic acid and protein sequence analysis. A practical application.* (M.J. Bishop and C.J. Rawlings eds.). IRL Press.

Stewart, G.C. and Bott, K. (1983). DNA sequence of the tandem ribosomal RNA promoter for *B. subtilis* operon rrnB. *Nucleic Acids Res.* 11: 6289-6300.

Stewart, C-B. (1993). The powers and pitfalls of parsimony. *Nature* 361: 6.3-606.

Swofford, D.L. (1993). *PAUP: Phylogenetic Analysis Using Parsimony, version 3.1.1.* Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.

Swofford, D.L. and Olsen, G.J. (1990). Phylogeny reconstruction. In: *Molecular systematics.* (D.M. Hillis and C. Moritz eds.) Sinauer Associates, Inc. Sunderland, MA., U.S.A.

Thompson, D.W. (1968). *On growth and form*. Cambridge University Press.

Tinoco, I., Jr. Puglisi, J.D. and Wyatt, J.R. (1990). RNA folding. *Nucl. Acids Mol Biol*. 4: 205-226.

Taylor, F.J.R. (1989). Phylum Dinoflagellata. In: *Handbook of Protoctista* (L. Margulis, J.O. Corliss, M. Melkonian and D.J. Chapman eds.), Jones and Bartlett, Publishers, Boston.

Taylor, G.R. (1991). Polymerase chain reaction: basic principles and automation. In *PCR A practical approach* (M.J. McPherson, P. Quirke and G.R.Taylor eds) IRL Press.

Tanada, Y. and Kaya, H.S. (1993). *Insect pathology* Academic Press, Boston. USA.

Tautz, D., Tautz, C., Webb, D. and Dover, G.A. (1987). Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila spp:* Implications for molecular co-evolution in multigene families. *J. Mol. Biol..* 195: 525-542.

Tenter, A.M., Baverstock, P.R. and Johnson, A.M. (1992). Phylogenetic relationships of *Sarcocystis* species from sheep, goats, cattle and mice based on ribosomal RNA sequences. *Int. J. Parasitol.* 22: 503-513.

Theodorides, J. (1984). The phylogeny of the Gregarinia. *Origins of Life*. 13: 339-342.

Thompson, J.D. Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22: 4673-4680.

Titus, D.E. (1991). *Protocols and Applications guide*. Second edition. Promega Corporation, USA.

Trager, W. (1974). Some aspects of intra-cellular parasitism. *Science* 183: 269-273.

Undeen, A.H. and Cockburn, A.F. (1989). The extraction of DNA from Microsporidia spores. *J. Invertebr. Pathol.* 54: 132-133.

Varani, G., Cheong, C. and Tinoco, I. Jr. (1991). Structure of an unusually stable RNA hairpin. *Biochemistry* 30: 3280-3289.

Van Keulen, H., Gutell, R.R., Gates, M.A., Campbell, S.R., Erlandsen, S.L., Jarroll, E.L., Kulda, J., Meyer, E.A. (1993). Unique phylogenetic position of Diplomonadida based on the complete small sub unit ribosomal RNA sequence of *Giardia ardeae, G. muris, G. duodenalis* and *Hexamita sp. FASEB J.* 7: 223-231.

Vivier, E. and Desportes, I. (1989). Phylum Apicomplexa. In: *Handbook of Protoctista* (L. Margulis, J.O. Corliss, M. Melkonian and D.J. Chapman eds.). Published by Jones and Bartlett, Boston. USA

Vossbrinck, C.R. Maddox, J.V. Friedman,S., Debrunner-Vossbrinck, B.A. and Woese, C.R. (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* 326: 411-414.

Vossbrinck, C.R., Baker, M.D., Didier, E.S., Debrunner-Vossbrinck B.A. and Shadduck, J.A. (1993). Ribosomal DNA sequences of *Encephalitozoon hellem* and *Encephalitozoon cuniculi*: Species identification and phylogenetic construction. *J. Euk. Microbiol.* 40(3): 354-362.

Walsh, R.D. and Olsen, J.K. (1976). Observations on the susceptibility of certain culicine mosquito species to infection by *Lankesteria culicis* Ross. *Mosq. News* 36: 154-160.

Weiller, G.F. and Gibbs, A.J. (1993). *DIPLOMO: Distance Plot Monitor version 1.0* A computer program distributed by the Australian National University, Canberra. Australia.

Weiller, G.F. and Gibbs, A.J. (1995). *DIPLOMO*: A new approach to phylogeny. *CABIOS* submitted.

Wiley, E.O. (1981). *Phylogenetics: The theory and practice of phylogenetic systematics* Publisher Wiley, New York, New York. USA.

Wolters, J. (1992). The nature of preferred hairpin structures in 16S-like rRNA variable regions. *Nucleic Acids Research* 20(8): 1843-1850

Williams, A. Jr., and Tinoco, I. Jr. (1986). A dynamic programming algorithm for finding alternative RNA secondary structures. *Nucleic Acids Res*. 14: 299-315.

Williams, J.G.K., Hanafey, M.K., Rafalski, J.A. and Tingey, S.V. (1993). Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol*. 218: 704-740.

Woese, CR. (1973). Evolution of the genetic code. *Naturwissenschaften* 60: 441-446.

Woese, C.R. (1987). Bacterial Evolution. *Microbiol. Rev*. 51: 221-271.

Woese, C.R., Magrum, L.J., Gupta, R., Siegel, R.B., Stahl, D.A. (1980). Secondary structure model for bacterial 16S ribosomal RNA: Phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res*. 8: 2275-2293.

Wolters, J. (1991). The troublesome parasites - molecular and morphological evidence that Apicomplexa belong to the dinoflagellate-ciliate clade. *BioSystems* 25: 75-83.

Wright, J.H and Craighead, E. M. (1922). Infectious motor paralysis in young rabbits. *J. Exp. Med*. 36: 135-140.

Yang, D., Oyaizu, Y., Oyaizy, H., Olsen, G.J. and Woese, C.R. (1985). Mitochondrial origins. *Proc. Natl. Acad. Sci. USA* 82: 4443-4447.

Zhu, X., Wittner, M., Tanowitz, H.B., Cali, A. and Weiss, L.M. (1993). Nucleotide sequence of the small ribosomal RNA of *Encephalitozoon cuniculi Nucleic Acids Res*. 21: 1315.

Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science* 244: 48-52.

Zuker, M. and Steigler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*. 9: 133-149.

Zuker, M., Jaeger, J.A. and Turner, D.H. (1991). A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. *Nucleic Acids Res*. 19: 2707-2714.