



**Molecular Systematics of the Protozoan
Parasite *Giardia intestinalis*: Identification of
Cryptic Species**

Paul T. Monis
B.Sc. (Hons.)
Department of Microbiology and Immunology
The University of Adelaide

TABLE OF CONTENTS

	Page
Declaration	i
Acknowledgments	ii
ABSTRACT	iii
1 INTRODUCTION	1
1.1 Background	1
1.2 Parasitology	3
1.3 Parasites: why study them?	5
1.4 Systematics: a brief history	7
1.5 Methods used in systematics	9
1.5.1 Traditional and numerical taxonomic methods	9
1.5.2 Cladistics	10
1.5.3 Molecular systematics	11
1.6 Inherent problems in parasite systematics	13
1.7 <i>Giardia</i>: introduction	17
1.8 <i>Giardia</i> taxonomy	18
1.8.1 Trophozoite morphology	18
1.8.2 Cyst morphology	18
1.8.3 Taxonomic position	21
1.9 <i>Giardia</i> biology	25
1.9.1 Life cycle	25
1.9.2 The disease process	26
1.9.3 Host range	27
1.9.4 Prevalence	31
1.9.5 Biochemistry	33
1.9.6 Genomic organisation	34
1.9.7 Genetic characterisation	38
1.10 Objectives and rationale of thesis	41
1.11 Aim	43

2	GENERAL MATERIALS AND METHODS	44
2.1	Materials	44
2.1.1	Enzymes and reagents	44
2.1.2	Solutions	44
2.2	Methods for the isolation & culture of <i>Giardia</i>	45
2.2.1	Purification of cysts from faecal samples	45
2.2.2	Axenic culture of <i>Giardia</i> trophozoites	46
2.2.3	Growth of <i>Giardia</i> by infection of suckling mice	46
2.2.4	Purification of trophozoites	47
2.3	General methods	47
2.3.1	Genomic DNA isolation	47
2.3.2	DNA quantitation	48
2.3.3	Restriction endonuclease digestion	48
2.3.4	Agarose gel electrophoresis of DNA samples	48
2.3.5	Calculation of DNA fragment sizes	49
2.3.6	Purification of DNA using BresaClean	49
2.3.7	DNA sequencing	50
2.4	Computer software	50
2.4.1	DNA sequence editing and analysis	50
2.4.2	Oligonucleotide construction	51
2.4.3	Multiple DNA sequence alignments	51
2.4.4	Phylogenetic inference	51

3	GENETIC ANALYSES USING ALLOZYME	
	ELECTROPHORESIS	53
3.1	Description of technique	53
3.1.1	General principles	53
3.1.2	Interpretation of electrophoretic data	55
3.1.3	Applications of technique	56
3.1.4	Advantages/disadvantages of technique	59
3.2	Previous electrophoretic studies using zymodeme analysis	60
3.3	Previous electrophoretic studies using allozyme analysis	63
3.4	Genetic characterisation of isolates using allozyme eletrophoresis	66
3.4.1	Specific aims	67
3.4.2	Materials and methods	67
3.4.2.1	Source of isolates	67
3.4.2.2	Preparation and electrophoresis of samples	69
3.4.2.3	List of enzyme loci	69
3.4.2.4	Allozyme analysis	70
3.4.2.5	Calculating the Index of Association (a test for linkage disequilibrium)	71
3.4.3	Results and discussion	72
3.4.3.1	Comparison of relationships inferred using different distance measures	72
3.4.3.2	Analysis of isolates using complete data set	82
3.4.3.3	Reliability of the inferred genetic relationships	88
3.4.3.4	Effect of varying the number of loci examined	92
3.4.3.5	Test for linkage disequilibrium	97
3.4.3.6	Comparison of results with previous work	99
3.5	Summary and conclusions	99

4	GENETIC ANALYSES USING NUCLEOTIDE SEQUENCE DATA	103
4.1	Overview of available techniques	103
4.1.1	DNA hybridisation and related techniques	104
4.1.2	DNA fingerprinting techniques	105
4.1.3	Nucleotide sequence analysis	106
4.2	Review of DNA-based characterisation studies of <i>Giardia</i>	107
4.3	General aims	111
4.4	Phylogenetic analysis using 18S rRNA and <i>efl</i> α nucleotide sequences	112
4.4.1	Specific aims	113
4.4.2	Materials and methods	113
4.4.2.1	List of isolates	113
4.4.2.2	18S rRNA primer design	113
4.4.2.3	Optimised 18S rRNA PCR conditions	114
4.4.2.4	Elongation factor 1 α primer design	115
4.4.2.5	Optimised <i>efl</i> α PCR conditions	116
4.4.2.6	DNA sequencing	116
4.4.2.7	Phylogenetic analyses	116
4.4.2.8	Estimation of among-site rate variation	117
4.4.3	Phylogenetic analysis of 18S rRNA nucleotide sequences	118
4.4.3.1	Optimisation of 18S rRNA PCR	118
4.4.3.2	Detection of contaminating host DNA in suckling mouse derived isolates	119
4.4.3.3	Alignment of 18S rRNA nucleotide sequences	119
4.4.3.4	Determination of 18S rRNA sequence parameters	124
4.4.3.5	Distance-based phylogenetic analysis of 18S rRNA nucleotide sequences	128
4.4.3.6	Maximum likelihood analysis of 18S rRNA nucleotide sequences	131

4.4.3.7	Cladistic analysis of 18S rRNA nucleotide sequences	131
4.4.4	Phylogenetic analysis of <i>efl</i> α nucleotide sequences	134
4.4.4.1	PCR amplification of <i>efl</i> α	136
4.4.4.2	Alignment of <i>efl</i> α nucleotide sequences	136
4.4.4.3	Determination of <i>efl</i> α sequence parameters	140
4.4.4.4	Distance-based phylogenetic analysis of <i>efl</i> α nucleotide sequences	147
4.4.4.5	Maximum likelihood phylogenetic analysis of <i>efl</i> α nucleotide sequences	150
4.4.4.6	Cladistic analysis of <i>efl</i> α nucleotide sequences	150
4.4.5	Summary and conclusions	157
4.5	Phylogenetic analysis of the <i>gdh</i> and <i>tpi</i> nucleotide sequences	160
4.5.1	Background	160
4.5.2	Aims	161
4.5.3	Materials and methods	161
4.5.3.1	List of isolates	161
4.5.3.2	Design of primers for the <i>gdh</i> locus	163
4.5.3.3	<i>Gdh</i> PCR conditions	163
4.5.3.4	Design of <i>tpi</i> primers	166
4.5.3.5	<i>Tpi</i> PCR conditions	166
4.5.4	Phylogenetic analysis using <i>gdh</i> nucleotide sequences	168
4.5.4.1	Optimisation of PCR conditions	168
4.5.4.2	PCR controls	169
4.5.4.3	Detection of samples comprising multiple genotypes using DNA sequencing	169
4.5.4.4	DNA sequencing of <i>gdh</i> PCR products	170
4.5.4.5	Alignment of <i>gdh</i> nucleotide sequences	170
4.5.4.6	RFLP analysis using the 1.17kb <i>gdh</i> PCR product	172
4.5.4.7	Determination of <i>gdh</i> sequence parameters	172

4.5.4.8	Distance-based phylogenetic analysis of <i>gdh</i> nucleotide sequences	179
4.5.4.9	Maximum likelihood phylogenetic analysis of <i>gdh</i> nucleotide sequences	182
4.5.4.10	Cladistic analysis of <i>gdh</i> nucleotide sequences	184
4.5.5	Phylogenetic analysis of <i>tpi</i> nucleotide sequences	188
4.5.5.1	PCR amplification of <i>tpi</i>	188
4.5.5.2	Alignment of <i>tpi</i> nucleotide sequences	189
4.5.5.3	Examination of <i>tpi</i> sequence parameters	189
4.5.5.4	Distance-based phylogenetic analysis of <i>tpi</i> nucleotide sequences	200
4.5.5.5	Maximum likelihood phylogenetic analysis of <i>tpi</i> nucleotide sequences	200
4.5.5.6	Cladistic analysis of <i>tpi</i> nucleotide sequences	200
4.5.6	Summary and conclusions	204
5	DISCUSSION	209
5.1	Comparison of allozyme and nucleotide sequence analyses	209
5.2	Resolution of allozyme and nucleotide sequence analysis	211
5.3	Evidence for introgression between <i>G. muris</i> and <i>G. ardeae</i>	214
5.4	Molecular epidemiology of <i>G. intestinalis</i>	217
5.5	Molecular systematics of <i>G. intestinalis</i>	219
5.6	Host-preference of <i>Giardia</i>: re-incarnation of a taxonomic marker?	220
5.7	Correlation between growth characteristics and genotype	220
5.8	<i>Giardia</i> ecology	223
5.9	Implications for zoonosis	224

5.10	Comparison with the systematics of other asexual organisms	225
6	CONCLUSIONS	228
7	FUTURE DIRECTIONS	230
	GLOSSARY	231
Appendix A	Prevalence of <i>G. intestinalis</i> .	234
Appendix B	Allele frequencies (%) for operational taxonomic units of <i>Giardia</i> .	235
Appendix C	Aligned <i>Giardia gdh</i> nucleotide sequences	236
Appendix D	Molecular genetic analysis of <i>Giardia intestinalis</i> isolates at the glutamate dehydrogenase locus. (copy of Monis <i>et al.</i> (1996) published journal article)	237
Appendix E	Novel lineages of <i>Giardia intestinalis</i> identified by genetic analysis of organisms isolated from dogs in Australia. (copy of Monis <i>et al.</i> submitted journal article)	238
Appendix F	Genetic Analysis of <i>Giardia</i> from Hoofed Farm Animals Reveals Artiodactyl-Specific and Potentially Zoonotic Genotypes. (copy of Ey <i>et al.</i> submitted journal article)	239
	BIBLIOGRAPHY	240

DEDICATION

To my wife Tracey. It's finally finished (and just before the baby too!!).

DECLARATION

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan or photocopying.

Paul Monis

ACKNOWLEDGEMENTS

I would like to thank my supervisors Assoc. Prof. Graham Mayrhofer, Dr. Peter Ey and Dr. Ross Andrews for their help and support through the course of this project. I wish to acknowledge Graham Mayrhofer for originally proposing this project, and for supplying me with isolates of *Giardia* from his collection and unpublished data regarding them. I would also like to thank Peter Ey for his concentrated efforts in the critical review of this thesis, and for the many helpful suggestions and contributions that he has made during this project. I wish to make a special acknowledgment to Ross Andrews for providing me with his unpublished allozyme data, and for his expert tutelage and assistance regarding parts of the project. Thanks also to Ross for his comments during the drafting of this thesis, our many useful Friday afternoon discussions, and for his continued good faith and trust in me.

Special thanks to my mates Steve Donnellan and Mark Adams for their efforts in reviewing this thesis, and for the many hours that we spent in interesting and useful discussion. Their knowledge and advice in matters regarding evolutionary biology have been invaluable. Thanks also to the rest of the crew at EBU - Jan, Terry, Steve, Malcolm, Carolyn and Sally, and to the other staff and affiliates of the Division of Natural Science, who made my time at the South Australian Museum so enjoyable and productive.

Thanks to Dr. Trevor Petney (University of Heidelberg, Germany) and Dr. Neil Chilton (University of Melbourne) for their review and comments of different parts of this thesis, and Dr. Grant Schiller, Dr. Rhett Swanson and especially Michelle Morris for their efforts in the proofing and preparation of this thesis.

Further thanks to all of the members of the Department of Microbiology and Immunology who have helped me during my time in the department, especially John Mackrill for his efforts in growing the *Giardia* isolates for me, and Jocelyn Darby for her technical assistance.

Finally, I would like to thank my family, Gaby and Andre Monis, for their support during my years at University, and Anne Johnson for helping out during some tough financial times. Last but not least I would like to thank my wife for her love and support during all stages of the preparation of this thesis.

ABSTRACT

The aim of this research was to investigate the phylogeny of isolates of the intestinal protozoan parasite *Giardia intestinalis* using molecular systematic techniques; namely, analysis of allozyme and nucleotide sequence data. Classical taxonomic methods are unable to distinguish isolates of *G. intestinalis* which exhibit substantial heterogeneity in genetic and other biological parameters. The inferred phylogeny would provide a sound foundation from which to further examine the systematics of this species.

The approach adopted combined both contemporary and traditional methods of genetic characterisation and utilised multiple independent gene loci for analyses. In contrast to most other studies, which use axenic cultures of *G. intestinalis*, almost all of the isolates used in this study were propagated by the infection of suckling mice.

Isolates were characterised allozymically and their genetic relationships were inferred using clustering methods. Seven lineages of isolates were identified, five comprising animal-derived *G. intestinalis*, and two comprising human- and animal-derived *G. intestinalis*. The genetic distances separating these lineages were similar to those separating morphologically distinct species of *Giardia*. To complement these data, nucleotide sequences from segments of the genes encoding 18S rRNA, elongation factor 1 α , glutamate dehydrogenase and triose phosphate isomerase were determined for isolates representing the lineages comprising *G. intestinalis* and phylogenetically analysed using distance, parsimony and maximum likelihood methods. Without exception, the results of the separate phylogenetic analyses were consistent with the allozyme-inferred relationships and each other.

The results of these analyses support the proposal that *G. intestinalis* is a species complex, identifying at least seven cryptic species using the criteria of the phylogenetic species concept. Furthermore, the distances separating these lineages are similar to those separating species of other protozoan parasite genera, or to those separating different genera of bacteria. The isolates comprising some of these phylogenetic species appear to exhibit distinct host-preferences, and possess different growth characteristics in either axenic culture or infection of suckling-mice. Although more extensive sampling and characterisation will be required in some instances, it is suggested that these phylogenetic species are worthy of formal taxonomic recognition. This research demonstrates the utility of combining different approaches to assess the systematics of organisms to which the biological species concept cannot be applied.



INTRODUCTION

1.1 Background

The foundations of this thesis lie within the disciplines of parasitology and systematics. A parasite is an organism "... exhibiting a varying but obligatory dependence on another organism, its host, which is detrimental to the survival and/or fecundity of that host ..." (Anderson & May, 1982). As such, parasites are of major medical, veterinary, ecological and evolutionary importance as pathogens. The study of these organisms is therefore of considerable significance. However, it is not possible to study any organism without at least a rudimentary knowledge of its life history and how it is related to other organisms (ie., its systematic relationships). Accurate classification and naming systems (ie., nomenclature) are not only important in biology, but also play important roles in other sciences, such as chemistry. By analogy, the organic chemist uses the structural "morphology" of a molecule to identify it: ie., the number, type and arrangement of atoms which constitute an organic molecule, in much the same way as a taxonomist would use the morphological, biochemical, and genetic characteristics of a plant or animal to identify it.

The philosophy of systematics, which is to produce a global picture of natural relationships by combining information from a number of areas (eg., taxonomy, evolutionary biology, ecology), is central to various disciplines in science. All of the different branches of science are inter-related, and it requires a general knowledge of these for a scientist to be successful. For instance, much of biology requires statistics to test hypotheses, mathematics to provide models of relationships, or the physical sciences to examine/interpret fundamental biological processes. This has long been recognised, and was the theme of C.M. Wenyon's preface to "Protozoology: a manual for medical men, veterinarians and zoologists" (1926), in which he pointed out that, although specialisation in science is inevitable, a general knowledge of all related disciplines is required for a well rounded study. The same is true today, and success in parasitological research will rely upon an holistic approach combining information and expertise from a number of different areas, such as systematics, genetics, ecology, biochemistry and mathematical modelling to create a sum which is greater than the parts (Bush *et al.*, 1995). This multidisciplinary approach is in the spirit of the philosophy of systematics.

If systematics forms the basis of how we investigate the world around us, then phylogeny would have to be the most basic component required for systematics. This has been best illustrated by Dobzhansky's dictum (1973), which stated that "Nothing in biology makes sense except in the light of evolution.". This dictum has been closely supported by others (eg., Hillis & Bull, 1991). In this thesis I will show, using the example of the medically important intestinal protozoan parasite *Giardia intestinalis*, that a phylogenetic framework is an essential requirement for the rational biological investigation of an organism. In the absence of such a framework, a meaningful interpretation of data from comparative studies of biological variation becomes virtually impossible.

Isolates of *G. intestinalis*, which are morphologically indistinguishable, exhibit a wide range of diversity, both behaviourally (eg., pathogenicity, host specificity, host infectivity, growth characteristics) and genetically. A survey of the literature shows that studies of this organism (eg., biochemistry, drug resistance) have been restricted to only a few genotypes which may not be representative of those comprising natural populations. Furthermore, despite evidence that *G. intestinalis* is a species complex (Andrews *et al.*, 1989; Mayrhofer *et al.*, 1995), there have been no attempts to appropriately assess the systematic relationships within this group of organisms. It is the objective of this thesis (as outlined in section 1.10) to address this shortcoming by applying molecular techniques to independently examine the systematics of the *G. intestinalis* species complex. As described in the aims (section 1.11), both allozyme and nucleotide sequence data were used to determine the genetic and evolutionary relationships between isolates of *G. intestinalis*, thereby providing a framework for the future biological investigation of this medically important group of organisms.

It is important to recognise that the problems besetting *Giardia* systematics are not unique, and that the approach adopted herein to resolve this issue can also be applied to other genera of parasites. To provide a general perspective, this introduction is divided into four sections, namely: parasitology, systematics, the application of systematics in parasitology, and the biology of *Giardia*. Each of these topics could easily form complete chapters. However, only those areas and issues of direct relevance will be considered herein. As a final note, there are many terms and concepts used in molecular biology, systematics and parasitology that may be unfamiliar to readers from other disciplines. It is also the case that terms may have different definitions, depending upon the discipline in which they are used (eg., systematics), or that they may have incorrect common usage (eg., homology). In order to avoid any confusion, I have provided a glossary of the terms used within the context of this thesis.

1.2 Parasitology

The term “parasite” is derived from the Greek *parasitos*, meaning “fellow guest”. Its origin was in ancient Greece, where it was used to describe a person who was given a free meal in exchange for witty conversation, usually at the expense of the host. Currently there are many different ideas regarding the definition of a parasite (eg., Price, (1980) lists several definitions and complains about the variety). However, fundamental to all of them is the original concept that the guest (parasite) is getting a free meal to the detriment of the host. Traditionally, the discipline of parasitology has been restricted to the study of parasitic protozoa, helminths and arthropods. In more recent times, other parasitic taxa, such as fungi, bacteria, viruses, and a variety of animals and plants have also been included (Price, 1980).

Parasitic organisms are ubiquitous, comprising over 70% of the animal kingdom (Price, 1980; 75% of insects, Price, 1977). For most free-living organisms, many others exist that are capable of parasitising them. In many cases, the distinction between parasitism and other forms of association, such as commensalism or symbiosis, is often blurred. As discussed by Lapage (1956), these associations are “... categories imposed by the human mind upon living things ...”, and “... because they are alive, they are continually undergoing change ...”. In other words, life is not static, but in a state of constant flux, something of which biologists should never lose sight. As a result, symbiotic relationships may become parasitic, or the converse may occur as circumstances change (Lapage, 1956). It is the constant interplay between host and parasite that makes parasitology so fascinating, particularly for ecologists and evolutionary biologists.

The effects of parasitic infections upon humans and animals have long been recognised. The following is a précis of some of the landmarks in parasitology, which has been reviewed by Hoeppli (1959). Malaria was one of the first parasitic diseases to be recognised. Ancient Chinese physicians were able to differentiate between tertian and quartan forms of malaria based upon the patterns of fever, and Roman and Greek physicians also associated ague with malaria. In the time of Moses, it was recognised that insect and animal pests were vectors for plague, and that the consumption of contaminated meat could cause disease. The Greeks of 350 B.C. described tapeworms, and Hippocrates not only diagnosed echinococcal disease but also described a method for the removal of cysts from infected human tissue.

In 1684, the first book on parasitology was published by Redi (Faust *et al.*, 1976). It was also around this time that Leeuwenhoek (1632 - 1723), using a simple microscope, described both free-living and parasitic protozoans from the intestines of frogs, although the term “protozoa” was not used until 1817 by Goldfuss (Wenyon, 1926). Leeuwenhoek was also the first person to describe a protozoan, now thought to be the parasite *Giardia intestinalis*, from a human source (his own faeces). Following the development of more advanced microscopes in the mid 19th century, Gross was the first to describe the parasitic amoeba *Entamoeba gingivalis*. By the late 19th century, it had been discovered that insects were the vectors for the transmission of trypanosomes and malaria (reviewed by Cole, 1926; Foster, 1965; Faust *et al.*, 1976).

The 20th Century has seen many advances in the field of parasitology, in areas such as parasite identification, and in the diagnosis and treatment of parasitic diseases. Advances in microscopy have allowed the accurate morphological characterisation of many parasitic species, and the combination of microscopy with immuno-labelling methods, eg., immunofluorescence, has proven to be a useful tool for parasite diagnosis in the case of schistosomes (Rijpstra, 1975) and *Giardia* (Lalova, 1976). Ultrastructural characterisation of parasites by electron microscopy has provided new characters for use in taxonomy and identification, and aided the investigation of mechanisms of disease, such as *Hexamita* infection in fish (Ferguson, 1979) and giardiasis (Poley, 1983). Treatment of parasitic diseases has improved slowly with the development of modern medicine. An increased understanding of host and parasite physiologies has allowed a better understanding of the mode of action of chemotherapeutic agents, as well as aiding in the selection or design of new drugs, and improving administration of existing ones (eg., use of metronidazole, Brogden *et al.*, 1978; use of tinidazole, Sawyer *et al.*, 1976; identification of biosynthetic pathways as chemotherapeutic targets, Vitti *et al.*, 1987). Over the last 30 years, molecular techniques, such as allozyme electrophoresis and more recently the polymerase chain reaction (PCR) have played increasingly more important roles in parasite systematics and diagnosis, eg., the genetic characterisation of *Giardia intestinalis* (eg., Andrews *et al.*, 1989; Homan *et al.*, 1992; Ey *et al.*, 1996; Weiss *et al.*, 1992).

1.3 Parasites: why study them?

A primary reason for the study of parasites is that they are of medical and veterinary importance (Table 1, p. 7), having a major impact upon human and animal health. They are also of major economic importance because of the costs associated with the treatment of parasitic infections, the implementation of parasite control strategies, the damage or death they cause in livestock, and the reduction of suitable land for rural development. The magnitude of the losses is quite staggering. Herlich (1978) estimated annual worldwide losses of 10 million cattle and buffaloes and 30 million sheep and goats due to helminth infections alone. The money spent on antihelminthics worldwide was estimated at US\$500 million. Losses due to arthropod infestation or arthropod-borne diseases have been reported to be billions of dollars annually (Fallis, 1980). In a study by Finelle (1974), the cost benefit of controlling trypanosomiasis, which affects an area of 10 million km² in Africa, was estimated to be an increased value in meat production of US\$750 million. Apart from the mortality it causes in cattle, trypanosomiasis also reduces milk production, fertility, and the work rate of draft animals, all of which affect the rural development in these areas (Connor, 1994). The cost to human health and productivity is also high. Over 2 million people die every year from malaria, and substantial proportions of the human population are infected with parasites such as roundworms and hookworms. In Thailand, one third of the population (approximately 6 million people) have liver fluke (*Schistosoma*) infection, which can lead to a variety of pathological conditions such as cholangio-carcinoma, costing US\$19.4 million in direct medical care and resulting in wage losses of approximately US\$65 million (Loaharanu & Sornmani, 1991). Prevention of human infection can also be costly. For example, in the U.S.A., compliance with the Safe Drinking Water Act, which requires the monitoring of water sources for microorganisms such as *Giardia*, *Cryptosporidium* and viruses, as well stricter usage of disinfectants, was estimated to cost US\$1.6 billion (Lykins & Clark, 1994).

In addition to their socio-economic impact, parasites and host-parasite interactions are of particular biological interest. The intimate associations between parasites and their hosts provide useful models for the study of evolutionary and ecological processes (Bush *et al.*, 1995; Price, 1977; Simms, 1996). The observation that many parasites rely upon their hosts to provide an environment in which to survive and reproduce (endoparasites) has led to the proposal that a major mechanism of speciation among parasites is by coevolution with their hosts (Klassen, 1992; Price, 1977). This hypothesis has been investigated by comparison of

independently derived phylogenies of both parasite and host taxa (Hafner & Nadler, 1988; Hafner & Page 1995; Page, 1993, 1994).

Demonstration of cospeciation between hosts and parasites has led to the examination of other aspects of evolutionary biology, eg., the tempo or mode of speciation (Hafner & Nadler, 1990; Hafner *et al.*, 1994; Page, 1990, 1991) and it has also provided models for the study of phenomena such as convergent evolution, or changes in the rate or timing of an ancestral developmental process in a descendent (heterochrony). It is believed that heterochrony may have given rise to some of the genera present in the Order Diplomonadida due to a delay in cell division (see Siddall *et al.*, 1992). The effect of host-parasite coevolution upon host susceptibility to infection has also been examined (Morand *et al.*, 1996).

Parasite systems, by virtue of the host-parasite relationship, also make excellent models for studies of ecology (Minchella & Scott, 1991; Noble, 1960; Schad, 1963, 1965). For obligate parasites, the host defines the boundary within which all ecological processes occur. Furthermore, the host provides a habitat which is capable of responding to the actions of the parasite. The constant interplay between host and parasite allows for the study of the mechanisms involved in selection and adaptation, which affect life history strategies and behaviour (Adamson, 1986; Burns, 1994; Dobson & Hudson, 1992; Ladle *et al.*, 1993; Rzhetskii *et al.*, 1987; Weis & Abrahamson, 1985). More recently, parasites have been used as probes for biodiversity, based largely upon the ancient ecological connections between parasites and their hosts and the assumption of coevolution (Gardner & Campbell, 1992), as biological control agents (Spratt, 1990), and as model systems to study aspects of the immune system, eg., the evolution of different types of immunity and the processes involved in the immune response (Mitchell, 1991; Quintans, 1994).

To be able to utilise parasite systems in the ways discussed, it is imperative that the evolutionary relationships within both parasitic and host taxa be understood. A systematic framework (or a robust phylogeny, at the very least) is essential for the study of fundamental evolutionary or ecological processes, and is also required for the efficient identification or diagnosis and management of parasites and parasitic diseases. As in other areas of comparative biology, systematics plays a central role in parasitology, and it forms the central theme of this thesis. This raises the question: what is systematics?

Table 1. A selection of parasites of medical and/or veterinary importance.

Vectors / intermediate hosts are indicated in parentheses.

Parasite	Disease	Hosts
Protozoans		
<i>Acanthamoeba spp</i>	granulomatous amoebic encephalitis	humans
<i>Babesia spp</i>	babesiosis	cattle, humans (ticks)
<i>Cryptosporidium parvum</i>	cryptosporidiosis	humans, other mammals
<i>Entamoeba histolytica</i>	intestinal amoebiasis	humans
<i>Giardia spp</i>	giardiasis	vertebrates, except fish
<i>Leishmania spp</i>	leishmaniasis	humans (sandflies)
<i>Naegleria fowleri</i>	primary amoebic meningoencephalitis	humans
<i>Plasmodium spp</i>	malaria	humans, mammals (mosquito)
<i>Toxoplasma gondii</i>	toxoplasmosis	felines, humans
<i>Trichomonas vaginalis</i>	vaginal trichomoniasis	humans
<i>Trypanosome spp</i>	sleeping sickness, Chagas' disease	mammals (insects)
Helminths		
<i>Schistosoma spp</i>	lung, liver or blood flukes	human (snails)
<i>Fasciola hepatica</i>	liver flukes	herbivores, humans (snails)
<i>Taenia spp</i>	tapeworm infection	mammals
<i>Echinococcus granulosus</i>	hydatid disease	canines, humans
<i>Ascaris lumbricoides</i>	ascariasis	humans
<i>Strongyloides stercoralis</i>	strongyloidiasis	humans
<i>Toxocara, Toxascaris spp</i>	visceral larva migrans	mammals
<i>Trichinella spiralis</i>	trichinosis	mammals
<i>Trichostrongylus spp</i>	trichostrongylosis	humans, herbivores
Arthropods		
Mosquitoes	vector for many diseases	mammals
<i>Sarcoptes scabiei</i>	scabies, mange	canines
Ticks	vector for many diseases	mammals
Flies	vector for disease, fly strike	mammals

1.4 Systematics: a brief history

The discipline of systematics has a long history. In its most primitive form, systematics was used by the earliest of human societies to allow the successful exploitation of their environment (Whitehead, 1990). Hunter-gatherers required a basic knowledge of the life history of the plants and animals that they relied upon, as well as an understanding of the interactions of different organisms with each other and their habitat. This included the ability to identify and name those animals or plants which were in some way useful or were to be

avoided. The practical knowledge thus gained formed the nucleus for areas of study such as taxonomy and nomenclature, ecology and behaviour.

Although it has become more formalised with time, the essentials of systematics have remained the same. The three main components of systematics are **taxonomy**, which is "... the theory and practice of classifying organisms into taxa within a hierarchy ..." (Whitehead, 1990), **nomenclature**, which is "... a specialised system for ensuring the vital one-to-one relationship between members of a taxon and the label (scientific name) applied to it ..." (Whitehead, 1990), and **biological data**. During the last two hundred years, the theory of evolution has become an integral part of systematics, adding an historical perspective which had been lacking from taxonomy.

From the traditional biological perspective, a distinction is made between systematics and its sub-discipline, taxonomy (Marcus, 1991; Whitehead, 1990). It is interesting to note that the classical definition of systematics is equivalent to the "polyphasic taxonomy" used by bacterial systematists (Vandamme *et al.*, 1996), who use taxonomy synonymously with systematics. The new term (polyphasic) is used to describe a method which combines biological data (phylogenetic, biochemical, etc.) with taxonomy.

Although there are a number of interpretations regarding the meaning of systematics (as indicated in the glossary), for the purposes of this thesis the working definition of Whitehead (1990), which encompasses the holistic nature of the discipline, will be used throughout, namely:

"... the scientific discipline that not only classifies and names diversity of the organic world (through taxonomy), but seeks to understand the processes that produce and sustain that diversity. Its results are classifications which aim to show historically derived relationships between organisms at all taxonomic levels and in so doing, to point to genetic lineages, since these underlie all other forms of relationship. The method is comparative and the materials are all living and extinct organisms and all biological sources of characters that illuminate relationships and their history."

This interpretation is widely accepted by biologists and is in the same spirit as the definition given by Jeffrey (1973), who defined systematics as "... the scientific study of the diversity of living organisms ...", and subdivided it into taxonomy, classification and nomenclature.

1.5 Methods used in systematics

The foundations of systematics are taxonomy, and more recently, evolutionary theory. It is through evolutionary theory that models can be constructed to test hypotheses about evolutionary relationships among organisms or ecological systems. The methods used to determine evolutionary relationships have undergone a revolution since the mid-20th century. There have been three major developments during this time that have had an impact on systematics. The first two, which were the development of numerical taxonomy (phenetics) and of cladistics, saw a revision in the way that traditional taxonomic characters, such as morphological features, were treated. The third was the introduction of molecular systematics. Each advance has involved much controversy, with cladists arguing with pheneticists, and traditional morphologists at odds with molecular systematists (eg., Felsenstein, 1978; Bachmann, 1995).

Currently, there are three separate philosophies concerning the role of taxonomy, each of which reflects the different concepts used to define species (Marcus, 1991). Evolutionary taxonomy, which was built upon the ideas arising from Darwin's *Origin of Species* (1859) and the biological species concept (Mayr, 1969), distinguishes taxa on the amount of evolutionary change which has taken place, regardless of the actual phylogeny. A classic example of this is the relationship between reptiles and birds. Phylogenetically, alligators and birds are each others closest relatives, but because birds have undergone such a large amount of change and are so different to reptiles (warm blood, feathers, etc.), they are not placed as a sister taxon to alligators (ie., within Reptilia), but are placed in a separate order, Aves. Phylogenetic systematics, first used by Hennig (1966), holds that taxonomy should reflect phylogeny. The evolutionary (or phylogenetic) species concept (Cracraft, 1983, 1989) is allied with this idea. In contrast, the phenetic approach used by numerical taxonomists makes use of overall similarity, usually morphological or biochemical, to determine taxonomic hierarchies.

1.5.1 Traditional and numerical taxonomic methods

Prior to the middle of the 20th century, taxonomy was based upon the operational approach of comparative morphology first introduced by Carl Linnaeus in 1759. This was almost invariably done by biologists who specialised in a particular group of organisms. Although this approach resulted in classifications which are still in use today, it had a number of deficiencies, such as a small number of people holding systematic authority for each

taxonomic group, the absence of methods for testing the validity of the proposed classifications, the absence of a universal measure for comparing different taxonomic groups, and a clear philosophy regarding which aspects of evolution should be reflected in a classification had not been developed (Avise, 1994). The deficiencies of traditional systematics led to the development of numerical taxonomy, which is a phenetic approach to taxonomy (Sokal & Sneath, 1963).

The basic premise of numerical taxonomy is that organisms should be grouped according to overall similarity (or its converse, difference). According to Sneath and Sokal (1973), the philosophy of numerical taxonomy is "... to develop methods that are objective, explicit, and repeatable ...". Numerical taxonomy utilises many characters, usually morphological or biochemical markers, and weights each character equally in an attempt to produce a robust classification. The collective similarities or differences between taxa in a comparison are quantitative measures, and are supposed to be estimates of the total similarity or difference between taxa if all possible characters were examined. The method is polythetic, ie., the attributes of the taxonomic units are incorporated simultaneously into the scheme rather than one at a time. The inferred relationships between taxa, or operational taxonomic units (OTU's), are usually depicted as a phenogram, which is a branching diagram that links OTU's by estimates of overall similarity (or difference). Phenograms are usually constructed using the Unweighted Pair Group Method Using Arithmetic Average (UPGMA), which is a clustering technique that assumes equal rates of evolution along all branches (for a general discussion of the technique, see Avise, 1994 or Swofford *et al.*, 1996). As used originally by Sneath and Sokal (1973), phenetics produces classifications which reflect overall similarity, but not necessarily phylogeny. The use of overall similarity has been criticised by cladists, who developed cladistics as an alternative philosophy for phylogenetic reconstruction.

1.5.2 Cladistics

The founder of cladistics, which was also known as phylogenetic systematics, was Willi Hennig. Although Hennig first published his ideas in German in 1950, the impact of his methods were not felt until his original paper was translated into English in 1966 (Avise, 1994). The cladistic philosophy holds that evolutionary relationships cannot be deduced by overall phenetic similarity, but can only be determined by a subset of the characters used by pheneticists, namely those characters that display shared derived (synapomorphic) traits, and therefore indicate common descent between two or more taxa. The underlying principle of

cladistics is that each speciation event will yield two descendant taxa, so that any characters unique to the ancestral species will also be present in the descendants. Operationally, cladists use outgroup analysis to polarise character states into two categories - ancestral or derived. Ancestral (plesiomorphic) or unique-derived (autapomorphic) character states provide no phylogenetic information. As a result, those characters which are invariant (ie., plesiomorphic), or which comprise only plesiomorphic and autapomorphic character states, are excluded from the analysis. This leaves only those characters with character states which are shared by at least two but not all of the taxa under investigation. By virtue of its underlying principles, cladistics emphasises the branch splitting component of a phylogeny (cladogenesis) rather than the branch lengths, which are due to the accumulated change along each lineage (anagenesis).

Due to homoplasy in data sets, which occurs when taxa have similar character states due to parallelisms, reversions or convergent evolution rather than common descent, many characters are required in order to reliably reconstruct a phylogeny. Operationally, the cladistic methods used to analyse large data sets are not dissimilar to those used in numerical taxonomy. Parsimony has been adopted to resolve cases in which different characters conflict. In practice, maximum parsimony is used to identify trees which require the fewest character state transformations to arrive at the present data set (minimal length). The dendrograms derived from cladistic analysis (cladograms) show the order of branch splitting, but not branch lengths, although the number of changes along each branch is sometimes used as an indicator of branch length.

As with phenetics, cladistics was originally developed for use with morphological or biochemical characters. The development of molecular biology has provided new types of data for use in phylogenetic reconstruction (reviewed by Mishler, 1994). This relatively new discipline is known as molecular systematics.

1.5.3 Molecular systematics

The methods used in molecular systematics are similar to those used traditionally, although some new approaches (eg., maximum likelihood analysis) have been developed for particular types of molecular data. The distinguishing feature of molecular systematics, compared to numerical taxonomy or the early cladistics, is the nature of the data used, which is usually based upon genotypic rather than phenotypic characters. There are two types of

molecular data: distance (eg., DNA hybridisation, protein immunology) and character state (eg., nucleotide sequence, allozyme profiles). Character state data can be transformed to distance data, but the reverse is not possible. Not all forms of molecular data are suitable for phylogenetic analysis. Some methods, such as the technique using random amplified polymorphic DNA (RAPD), produce data that may violate the assumptions of character homology or independence (Backeljau, *et al.*, 1995; Schierwater, 1995; Smith *et al.*, 1994). This would introduce systematic errors, ie., errors in the analysis procedure due to the violation of assumptions as opposed to the production of the wrong phylogeny, into any phylogenetic analysis using that particular type of data.

Three approaches can be used for phylogenetic reconstruction:- distance-based, cladistic, and maximum likelihood analysis. Swofford *et al.* (1996) divided these methods into those that use algorithms (distance-based), and those that use optimality criteria (cladistic, maximum likelihood), although there is also an algorithm component in the latter group of methods. Distance-based methods are polythetic, and use specific algorithms (eg., UPGMA, Neighbour Joining) to construct a phylogenetic tree (or phylogram) from a distance matrix. Provided that the data satisfy the assumptions of the algorithm used, and that the distance estimates are accurate, distance methods will reliably recover the true phylogeny (Nei, 1991). Unlike numerical taxonomy, the aim of distance-based approaches used in molecular systematics is to produce a phyletic classification. Because of the historical association of UPGMA with phenetics, analysis of distance data by UPGMA is still occasionally referred to as being phenetic, even though the resulting “phenogram” depicts a phylogeny provided that none of the assumptions of the method are violated.

In contrast, cladistic and maximum likelihood methods use discrete characters and employ various optimality criteria to determine which tree best fits the given data. The major criticisms of distance-based methods have been that the data are presented in a summary form (a relic of the original cladist/pheneticist debate) and that the method does not provide a way to examine or compare sub-optimal trees. Despite these apparent drawbacks, simulation studies have shown that, provided the distance estimates are accurate, distance-based tree building methods such as Neighbour Joining (NJ) or Minimum Evolution (ME) are as efficient or better than parsimony methods at recovering phylogenetic relationships (Nei, 1991).

The earliest molecular studies employed protein electrophoresis to examine allelic variation (allozymes) at enzyme loci (described in more detail in section 3.1.1). Allozyme electrophoresis (AE) dominated systematics from the mid 1960's to the early 1980's (Avisé, 1974, 1983; Buth, 1984), and it is still widely used today (reviewed by Andrews & Chilton 1997). In addition to its use for systematic studies, AE has been utilised for studies of related issues such as species boundaries (eg., Donnellan & Aplin, 1989) and hybridisation (eg., Szymura & Barton, 1986). The next development was the analysis of restriction fragment length polymorphisms (RFLP) of DNA (eg., Avisé *et al.*, 1979). However, from a technical perspective, the development of the polymerase chain reaction (PCR) has had the largest impact. Coupled with the development of automated DNA sequencing, this method has facilitated the rapid accumulation of nucleotide sequence data, which is now probably the most widely used type of data in phylogenetic analyses (eg., Medlin *et al.*, 1996).

The discipline of molecular systematics has seen many controversies, the largest being due to the fundamental operational differences between cladistic and phenetic theory (Bachmann, 1995; Felsenstein, 1978). Much of the debate between the different factions has had little foundation, the simple fact being that neither approach is perfect, and that each has assumptions and limitations which can affect the accuracy of an inferred phylogeny. Another topic of debate has been between morphologists and molecular systematists, relating mainly to the suitability (or perceived superiority) of molecular versus morphological characters. While there have been some researchers who have doggedly advocated one method over another, the approach of most systematists today is to use the different methods and types of data in combination with each other (Hillis, 1987; Medlin *et al.*, 1996; Patton & Avisé 1983).

1.6 Inherent problems in parasite systematics

As noted earlier, parasites and host-parasite systems provide excellent vehicles for studying fundamental biological processes. A prerequisite for this purpose is that the systematics of the taxa of interest must be known. Ironically, the features which make parasites so interesting to study also complicate parasite systematics. These difficulties occur at two taxonomic levels. The first is related to the delineation of species, especially for asexual or parthenogenic organisms, while the second is the more general problem of resolving the evolutionary relationships of parasite species, genera, families, etc. Despite the application of traditional and contemporary approaches, many parasite genera remain poorly

characterised or have ill-defined or resolved taxonomies (eg., Ellis *et al.*, 1995; Lichtenfels *et al.*, 1988; Siddall *et al.*, 1995; Siddall, 1995; Smothers *et al.*, 1994).

A major problem in parasite systematics has been that the traditional approaches using comparative morphology have been of limited utility for parasite taxonomy or identification. In some cases, this has been due to a lack of morphological variation, either because of constraints caused by specialised body-plans (eg., flatworms, Mariaux, 1996) or because species are too closely related. In other cases, the characters have not been phylogenetically informative due to homoplasy (caused by convergent evolution) or because the characters are autapomorphic (due to unique, specialised adaptations). For such organisms other biological parameters have often been used, eg., host range, growth characteristics (eg., carbon or nitrogen assimilation, which is used routinely for the identification of bacterial and fungal species) or serology. More recently, molecular techniques, such as allozyme electrophoresis or nucleotide sequencing, have played a major role in the identification or systematics of parasites (eg., *Naegleria*: Adams *et al.*, 1989; *Leishmania*: Andrews *et al.*, 1988; *Plasmodium*: Escalante & Ayala, 1994; nematodes: Hyman, 1990; mosquitoes: Munstermann, 1995; platyhelminths: Rohde, 1990). However, for many parasites, a dearth of biological data (eg., regarding mode of reproduction, ploidy, host range, life cycle) continues to limit progress.

Of all the parasite assemblages, the systematics of the protozoans is possibly the most problematic. There are five aspects of protozoan biology which have hampered the development of a comprehensive systematics of protozoan parasites (Corliss, 1960):

- 1) The absence of fossil records, which have been used to determine ancestral states and to estimate dates of divergence for other taxa.
- 2) The ubiquity of many protozoan taxa, both geographically and ecologically, has meant that concepts such as dispersal and restricted ecological niches cannot be used reliably as aids to trace evolutionary histories.
- 3) The microscopic size of protozoans has made morphological assessment dependent upon advances in microscopy.
- 4) Subcellular organisation, including ultrastructural characters, are not as informative as the multicellular organ systems or skeletal systems of the metazoans.
- 5) The frequent absence of sexuality has caused difficulties in the application of the biological species concept.

Of all these points, the last is perhaps the most important. Species occupy a pivotal position in systematics, forming the basic units used in comparisons. Similarly, populations form the basic units used in the delineation of species. While mutation and selection act upon individuals within a population, speciation occurs at the level of the population, and it is by the comparison of populations that species are defined (regardless of the species concept adopted). Due to the single celled nature of protozoans, it is not practical or in most cases possible to work with individual organisms. As a result, studies on any aspect of the parasite's biology must be carried out on populations of organisms, or at best, on clones derived from single individuals. The same difficulties also apply to bacteria. However, in most cases these organisms can be readily grown axenically, and the use of solid support media for culturing (eg., agar) simplifies the production of clones. In contrast, many parasitic protozoans are difficult to culture axenically, and cloning procedures are more complicated when liquid growth media have to be used. The discrimination of species is also of practical importance, as species-level differences between closely-related taxa often reflect fundamental biological differences, such as parasite behaviour, ecology, biochemistry, pathogenicity or host range. This is true of many parasites, including arthropods (eg., *Anopheles*: Collins & Paskewitz, 1996), helminths (eg., *Hypodontus*: Chilton *et al.*, 1992) and protozoans (eg., *Eimeria*: Barker *et al.*, 1989; *Perkinsus*: Bushek & Allen, 1996; *Plasmodium*: Escalante & Ayala, 1994).

For those protozoan parasites which are asexual, or for which the mode of reproduction is unknown (as is the case for many), the biological species concept cannot be invoked. Although it is possible to determine the phylogenetic relationships between isolates of parasites using molecular data, and hence to invoke the phylogenetic species concept (Cracraft, 1983), there is currently no universally accepted method available for the determination of species from such data; ie., for distinguishing between intra- and inter-species variation within the phylogeny. Many of the proposed species definitions (Table 2, p. 16) could be applied to asexual organisms. However, these definitions tend to be vague (eg., what constitutes a "diagnosable cluster" as used in the phylogenetic species concept?), and with the exception of the biological species concept, do not specify exact biological criteria for species delineation.

Protozoan systematics has advanced considerably since Corliss' comments in the 1960's, due in part to the development of electron microscopy and the introduction of molecular techniques. However, while molecular systematics has enabled the phylogenetic relationships

Table 2. Various definitions of species concepts (adapted from Avise, 1994).

Biological species concept (Dobzhansky, 1937)

“Species are systems of populations: the gene exchange between these systems is limited or prevented by a reproductive isolating mechanism or perhaps by a combination of several such mechanisms.”

Evolutionary species concept (Simpson, 1951)

“... a lineage (ancestral descendant sequence of populations) evolving separately from others and with its own unitary role and tendencies.”

Phylogenetic species concept (Cracraft, 1983, 1989)

A monophyletic group composed of “... the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent.”

Recognition species concept (Paterson, 1985)

“... the most inclusive population of biparental organisms which share a common fertilisation system.”

Cohesion species concept (Templeton, 1989)

“... the most inclusive population of individuals having the potential for cohesion through intrinsic cohesion mechanisms ...”

cohesion mechanisms:

- genetic exchangeability: factors that limit the spread of new genetic variants through gene flow.
- demographic exchangeability: factors affecting the niche and limits spread of new genetic variants through natural drift or natural selection.

Concordance principles (Avise & Ball, 1990)

A suggested means of recognising species by the evidence of concordant phylogenetic partitions at multiple independent genetic attributes.

of many protozoan families and genera to be elucidated (eg., Hashimoto *et al.*, 1995; Kumazaki *et al.*, 1983), the lack of appropriate models for defining populations and species has hindered the development of the species level systematics of asexual parasitic protozoans. The intestinal parasitic protozoan species complex *Giardia intestinalis* exemplifies the difficulties associated with conducting comparative studies in the absence of a sound systematic framework.

1.7 *Giardia*: introduction

Giardia were probably the first human protozoan parasites to be described. Members of the genus *Giardia* are anaerobic flagellates that infect virtually all vertebrate species, including amphibians, reptiles, birds, and mammals (Kulda & Nohýnková, 1978). They are obligate parasites, with no free living stage known outside of the host. Infections can be asymptomatic or they can cause clinical disease (giardiasis) with various symptoms, most commonly acute or chronic diarrhoea. During the last 25 years, *Giardia* have attracted a great deal of attention, firstly with the recognition that they are human pathogens causing what is now known as giardiasis (by *G. intestinalis*), and secondly because they are believed to belong to one of the earliest lineages branching away from the rest of the eukaryotes.

Although there are still many unanswered questions about the life history of *Giardia*, these organisms have been adopted as models for the study of biological processes in protozoans (eg., speciation, ecology), largely because they are relatively easy to culture, they have been reasonably well characterised compared to other protozoa, and because a basic genetic framework for *G. intestinalis* has been established. However, if they are to be useful as models, the systematics of the genus needs to be well defined. Currently, only traditional taxonomic methods using morphology have been used to describe the species within this genus. *G. intestinalis* (syn. *Giardia lamblia*, *Giardia duodenalis*), which is the only species known to infect humans, has been the primary focus of research within this genus, although *Giardia muris* has been used as an animal model for human giardiasis (eg., Roberts-Thompson *et al.*, 1976). Despite the evidence that *G. intestinalis* is a species complex (Andrews *et al.*, 1989; Mayrhofer *et al.*, 1995), its taxonomy has not yet been revised, although it is recognised that the current taxonomy of *G. intestinalis* is inadequate (Kulda & Nohýnková, 1978; Meloni *et al.*, 1995). Clarification of the taxonomic status of *G. intestinalis* is of practical importance, not only to explain the rich genetic (eg., Andrews *et al.*, 1989; Homan *et al.*, 1992; Meloni *et al.*, 1995) and behavioural (eg., Binz *et al.*, 1992,

Kirkpatrick & Green, 1985) diversity exhibited by isolates which are presumed to belong to the same species, but also because species-level differences in other parasitic protozoa often reflect differences in other important characters, eg., pathogenicity or host-preference (eg., *Leishmania*, *Naegleria*, *Plasmodium*).

1.8 *Giardia* taxonomy

The following description of the cell structure is included primarily to illustrate the characters used for taxonomic purposes. For more detailed descriptions, the reader is directed to the following references for a general review: Kulda & Nohýnková (1978, 1996), Vickerman (1990), and Adam (1991).

1.8.1 Trophozoite morphology

The morphology of the trophozoite, as determined by light microscopy, was originally used by Filice (1952) to define three species of *Giardia*, namely *G. muris*, *G. duodenalis* (now superseded by its precedent, *G. intestinalis*; Kulda & Nohýnková, 1978) and *Giardia agilis*. The salient features used to delineate species are outlined in Figure 1 (p. 19) and Table 3 (p. 20). They will be discussed in more detail in section 1.8.3. In the case of *G. intestinalis*, trophozoites are 10 - 12 μm long, 5 - 7 μm wide, and pyriform in shape. The cytoskeleton consists of two funis (microtubule bands), a median body, a ventral disk, and four pairs of flagella. Other organelles include a pair of oval-shaped nuclei and lysosomal vacuoles. Mitochondria, peroxisomes, smooth endoplasmic reticulum, nucleoli and Golgi apparatus are absent, although Golgi-like bodies have been observed in encysting trophozoites (Lujan *et al.*, 1995; Reiner *et al.*, 1990).

1.8.2 Cyst morphology

Mature *G. intestinalis* cysts are oval shaped and possess a thick, fibrous cyst wall (Figure 2, p. 19). Each cyst contains four nuclei, a double set of flagellar axonemes, and crescent shaped fragments of the ventral disk (Kulda & Nohýnková, 1978; Sheffield & Bjorvat, 1977). Little information appears to be available on the morphology of cysts from *G. agilis* compared to *G. intestinalis* or *G. muris*. According to Thompson *et al.* (1990), two median bodies are also visible. Cysts from *G. muris* and *G. agilis* are similar to those of *G.*

intestinalis, the main distinguishing feature being the shape of the median bodies (Thompson *et al.*, 1990).

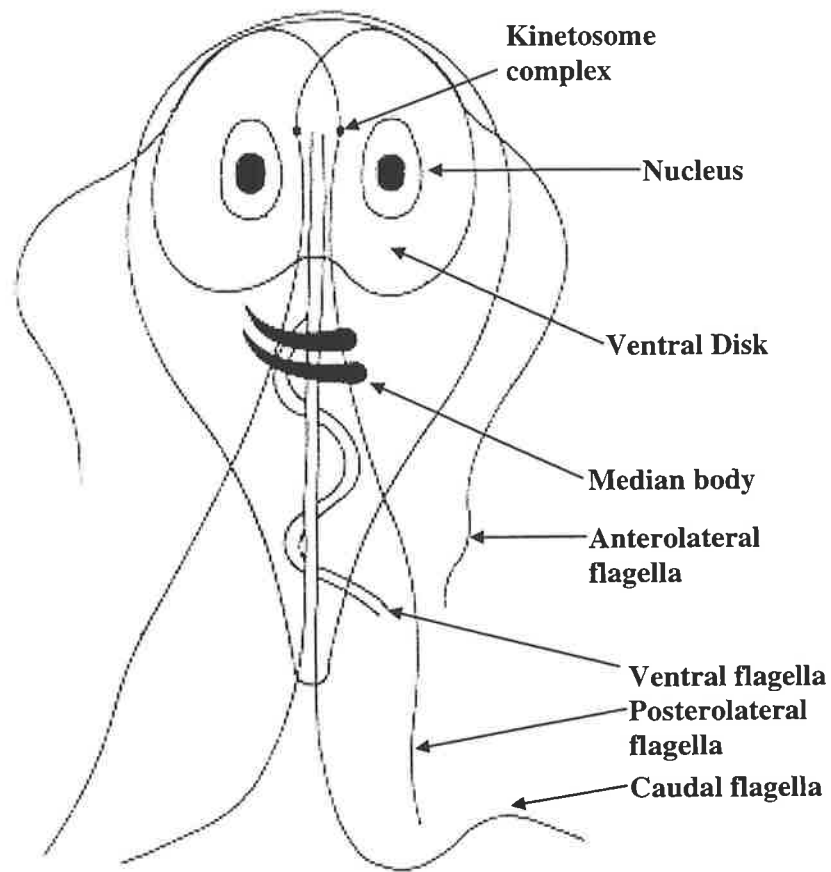


Figure 1. Cartoon of the major morphological features of a *Giardia* trophozoite, adapted from Thompson *et al.* (1990).

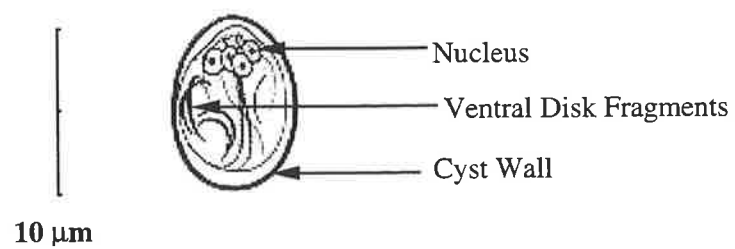


Figure 2. *G. intestinalis* cyst morphology (from Kulda & Nohýnková, 1978).

Table 3. Morphological characters used to identify species of *Giardia* (adapted from Kulda & Nohýnková, 1978).

Feature	<i>G. agilis</i>	<i>G. muris</i>	<i>G. ardeae</i>		<i>G. psicatti</i>	<i>G. intestinalis</i>
Median bodies:						
Shape	long club shaped rods	small rounded	small rounded	claw hammer	claw hammer	claw hammer
Location	anterior third of trophozoite parallel with longitudinal axis	center of body	center of body	parallel with longitudinal axis	transversely across body	transversely across body
Trophozoite shape	long, narrow tear drop	pyriform, short, broad	pyriform, short/medium		pyriform, medium	pyriform, medium
Ventrolateral Disk	1/5 of body length	>1/2 of body length, deeply notched posterior end	>1/2 of body length, deeply notched posterior end		<1/2 body length, incomplete ventrolateral flange (posterior end)	<1/2 body length
Caudal flagellae	paired	single	single		paired	paired
Shape of nuclei	oval	oval	tear-drop		oval	oval

1.8.3 Taxonomic position

Giardia belong to the family Hexamitidae (Kent, 1880), order Diplomonadida (Wenyon, 1926 emend Brugerolle, 1975), Class Trepomonadea, Phylum Metamonada Grassé, 1952 stat. nov. et emend. Cavalier-Smith, 1981 (Cavalier-Smith, 1993). Although perceived as “primitive eukaryotes”, they are in fact highly specialised parasites, adapted to growing in a niche where they have little competition from other organisms. Members of the Hexamitidae are diplozoic flagellated protozoa possessing paired organelles, including two nuclei and kinetids, and lacking a Golgi body and mitochondria. The other members of the Diplomonadida, the enteromonads (family Enteromonadidae; Kulda and Nohýnková, 1978), also lack a Golgi body and mitochondria but are monozoic ie., they possess only one copy of the other organelles. The observation that enteromonads which have failed to complete cell division are morphologically similar to the diplozoic diplomonads has led to the proposal that monozoism represents the ancestral state (Brugerolle, 1975; Brugerolle & Taylor, 1977). Based upon morphological features, behaviour (parasitic versus free living) and life history, Brugerolle (1975) suggested that *Giardia* are the most derived members of the family Hexamitidae (Figure 3, p. 22). This hypothesised phylogeny has been supported by Siddall *et al.* (1992), who used cladistic analysis of 23 ultrastructural characters to demonstrate the monophyly of the Hexamitidae and support the position of the enteromonads. The results of molecular phylogenetic studies, which have not included all members of the Order, are less clear, conflicting with the morphology-based phylogeny as well as with each other (Keeling & Doolittle, 1996; Rozario *et al.*, 1996).

The principle distinguishing character used to delineate *Giardia* from other members of the Hexamitidae is the presence of a ventral adhesive disk, which is a structure supported by a cytoskeleton of microtubules, microfilaments and associated fibrous structures. The ventral disk of *G. intestinalis* has been shown to be composed of tubulin (both α and β subunits) and closely-related proteins called giardins (Crossley & Holberton, 1983; Peattie *et al.*, 1989; Alonso & Peattie, 1992), as well as actin, myosin and tropomyosin (Feely *et al.*, 1982). Surrounding the disk is a marginal groove and ventrolateral flange, although not all isolates of *Giardia* possess a complete flange, eg., *Giardia psittaci* from the budgerigar, *Meliopsittacus undulatus*.

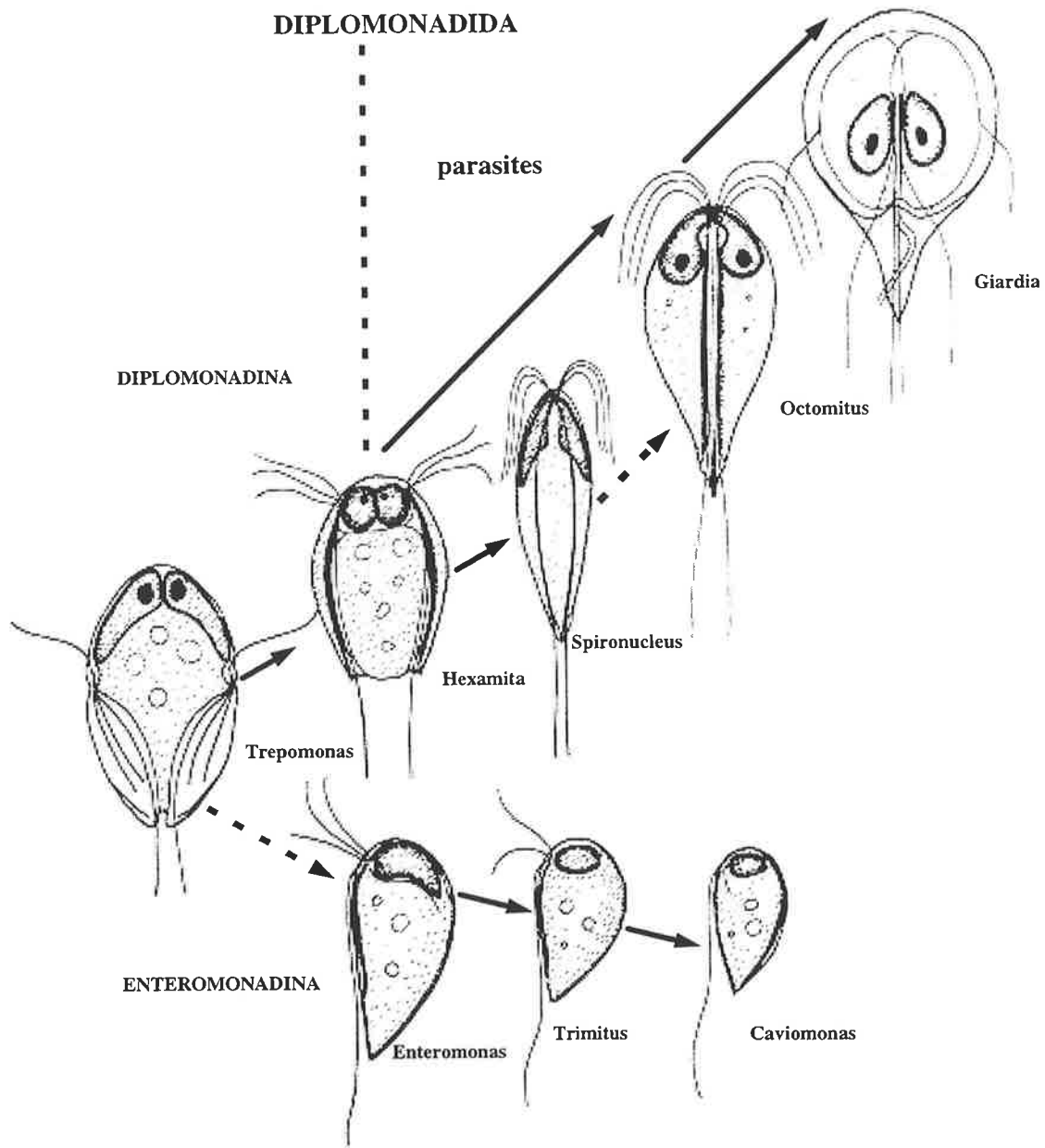


Figure 3. Relationships within the Diplomonadida suggested by Brugerolle (1975). Figure from Vickerman (1990).

Currently, the only method used to define species of *Giardia* has been morphological assessment. Filice (1952) examined all of the criteria used at that time to define *Giardia* species, eg., cell dimensions and presumed host specificity. He concluded that only consistent morphological differences should be used as the basis for defining species. It has since been shown that cross-species transmission can occur, although the success or failure appeared to be isolate and/or dose dependent (discussed in more detail in section 1.9.3).

The taxonomic characters defined by Filice were the shape of the median body and the overall shape of the trophozoite. Using these criteria, *Giardia* was split into three

morphologically distinct species, two of which are pathogenic to mammals (Figure 4). These species have also been called “morphological groups”, in recognition of the shortcomings of the current taxonomy (Kulda & Nohýnková, 1978, 1996). The morphological features of these groups are summarised in Table 3 (p. 20). The first group of organisms, *G. intestinalis*, have pyriform shaped trophozoites which possess a distinctive "claw-hammer" median body and cause disease in a variety of mammals, including humans. The naming of *Giardia* within this morphological group, especially with respect to *Giardia* of human origin, has been controversial. *Giardia lamblia* is extensively used in the United States of America, although Thompson *et al.* (1990) have claimed that *Giardia duodenalis* has priority. Kulda & Nohýnková (1978, 1996) have argued that *G. intestinalis* has priority, and based on the weight of their argument, this will be the species name used in this thesis. Members of the second group, *G. muris*, have rounded median bodies, a rounder trophozoite shape, and primarily infect rodents. Trophozoites of the third group, *G. agilis*, are characterised by long, narrow bodies with relatively short adhesive disks, long, club-shaped median bodies, and have only been isolated from amphibians.

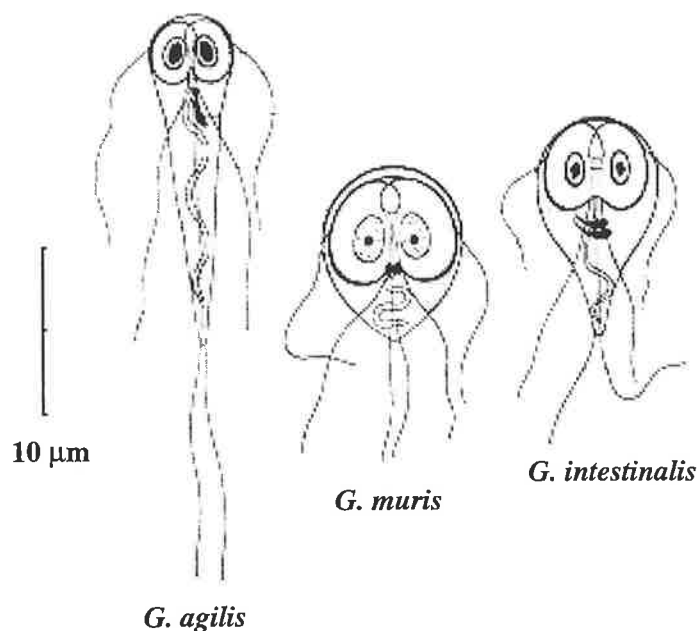


Figure 4. The three morphologically distinct species of *Giardia* originally recognised by Filice (1952). Diagram adapted from Kulda and Nohýnková, 1978.

More recently, scanning electron microscopy (SEM) studies have been used to characterise isolates of *Giardia* from birds. This has resulted in the description of two new species, *G. psittaci* (Erlandsen & Bemrick, 1987) and *Giardia ardeae* (Erlandsen *et al.*, 1990a). Their morphological features are summarised in Table 3 (p. 20).

Giardia psittaci are morphologically similar to *G. intestinalis*, possessing claw-hammer median bodies, pyriform trophozoite shape, etc., but they lack a complete ventrolateral flange and hence a marginal groove, which is a channel formed by the flange. Although *G. psittaci* are currently believed to be host specific, there is evidence to suggest that this may not be the case (Sogayar & Gregorio, 1989). According to Erlandsen & Bemrick (1987), the morphology of the ventrolateral flange is similar for *G. intestinalis*, *G. muris*, and *G. agilis*. However, in comparing *G. muris* and a rat-derived *G. intestinalis* isolate, Sogayar & Gregorio (1989) reported that the distal portion of the marginal groove of the *G. intestinalis* isolate was poorly developed. This description is similar to that used to define *G. psittaci* as a species separate from *G. intestinalis*, and raises the possibility that the *G. intestinalis* isolate described by Sogayar & Gregorio (1989) is related to (or the same as) *G. psittaci*.

Giardia ardeae share morphological features with *G. intestinalis* and *G. muris* (Erlandsen *et al.*, 1990a). The trophozoites of *G. ardeae* have been observed to possess either claw-hammer or rounded median bodies (but not both in a single cell), although the orientation of the claw hammer is different to that observed for *G. intestinalis*. The shape of *G. ardeae* trophozoites is similar to that of *G. intestinalis*. However other features, such as the size of the ventral disk, a notch in the caudal region of the ventral disk, and a single caudal flagellum, are similar to that of *G. muris*. The nuclei of *G. ardeae* are long, slender and tear-dropped in shape, unlike those of either *G. intestinalis* or *G. muris*, which are rounded.

On the basis of cyst morphology, an additional species, *Giardia microti* has been described (Feely, 1988). The cysts of *G. microti* contain two differentiated trophozoites with mature ventral disks, whereas the cysts of *G. intestinalis* contain a single trophozoite which has four nuclei and lacks a ventral disk. Although originally isolated from *Microtus ochrogaster*, *G. microti* has also been recovered from a variety other rodents (eg., Erlandsen & Bemrick, 1988). A seventh potential morphological species, which has distinctive caudal flagellae, has been reported from beavers (Erlandsen & Bemrick, 1988).

1.9 *Giardia* biology

1.9.1 Life cycle

Giardia have a simple bi-phasic life cycle (Figure 5). The cyst is the infectious form, and is found in the environment, either in faeces or in water contaminated by faeces. Cysts are metabolically active, with a respiration rate related directly to the environmental temperature (Paget *et al.*, 1989). They are not completely resistant to dessication, and therefore require moist conditions for survival. Their persistence in the environment is temperature dependent (Bingham *et al.*, 1979). These factors explain why waterborne outbreaks of giardiasis are more common in cooler climates, particularly in Canada and parts of the United States of America and Europe. While it is commonly known that *Giardia* (esp. *G. intestinalis*) cysts frequently contaminate surface waters such as reservoirs and streams, there is less public awareness about the possible contamination of water from rainwater tanks, as reported in the Virgin Islands (Crabtree *et al.*, 1996), or the contamination of marine water (Johnson *et al.*, 1995). Because of the ubiquity of *G. intestinalis* in humans and animals (native and domestic), the possibility that these infections represent a zoonosis has been a topic of debate for several decades.

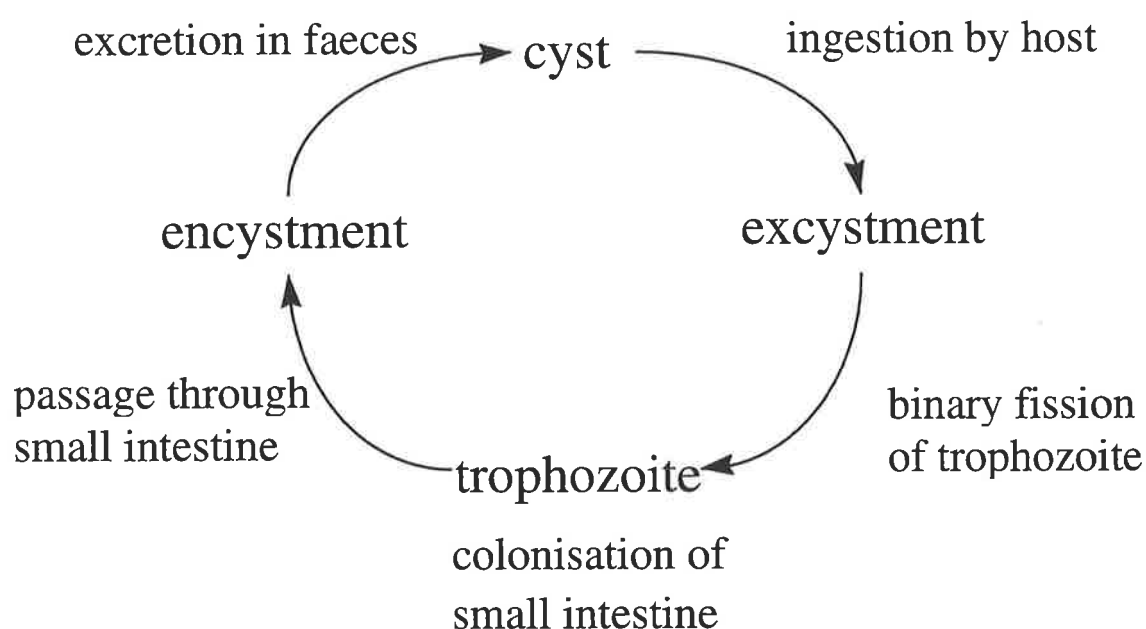


Figure 5. Schematic diagram of the life cycle of *Giardia*.

Infection of the host is initiated by the ingestion of viable cysts, either by drinking contaminated water, or by direct transmission via the faecal-oral route. The actual infectious

dose can be as low as 1 - 15 cysts (Hoff *et al.*, 1985). Excystation occurs following passage through the stomach and entry into the small intestine, yielding vegetative trophozoites. The exact mechanisms responsible for excystment are not known, but gastric factors, such as pH, have been implicated, and excystment can be induced *in vitro* (Feely, 1986; Isaac-Renton *et al.*, 1992). The trophozoites colonise the duodenum and jejunum by adhering to the intestinal villi, and replicate by binary fission, although there have been few reports on the course of cell division. Daughter cells each appear to receive a copy of both nuclei (Cerva & Nohýnková, 1992). Although condensation of chromatin occurs, chromosomes are not visible during mitosis (Cerva & Nohýnková, 1992). There is no experimental evidence to demonstrate the existence of a sexual cycle, although the discovery of homozygous and heterozygous genotypes (Meloni *et al.*, 1995) has been suggested as evidence that it may occur. In suckling mice, encystation occurs in the jejunum (Gillin *et al.*, 1987). The exact signals required to trigger encystment are unknown, but some environmental stimuli (discovered by the study of conditions in the region of gut where encystation occurs) include pH and the composition and concentration of bile salts (Gillin *et al.*, 1987, 1988). Reproduction of these conditions *in vitro* can induce encystation of *G. intestinalis* trophozoites (Schupp *et al.*, 1990). The mechanisms involved in encystment have been examined by using mutant cloned lines of *G. intestinalis* which cannot complete encystment *in vitro* (Reiner *et al.*, 1995). Cysts produced *in vivo* are passed in the faeces into the environment. Some trophozoites may also be passed in diarrhoeic faeces, but they do not survive outside of the host.

1.9.2 The disease process

The most common sources of *G. intestinalis* cysts are either contaminated drinking water or direct contact with infected faeces. Infection ranges from an asymptomatic carrier state to an acute or sometimes chronic disease. The major manifestation of acute infection is a severe, watery and foul smelling diarrhoea. The presence of blood in stools is very rare and can help to differentiate this disease from some of the bacterial enteric diseases. The absence of blood or cellular exudate is consistent with histopathological evidence that *Giardia* are rarely invasive. The mechanism by which the diarrhoea is caused has not been determined. However, use of HeLa and Vero cell monolayers has demonstrated that trophozoite adherence is sufficient to cause cell death *in vitro* (Jyothisri & Baveja, 1988). In chronic infections, malabsorption and anorexia may occur and cause weight loss and failure to thrive in children (eg., reviewed by Faubert, 1996; Thompson *et al.*, 1990).

A number of lines of evidence show that clearance of infection requires a host immune response, in particular the production of secretory IgA antibodies. In immunodeficient humans, especially hypo-gammaglobulinaemic children, the incidence of chronic giardiasis is much higher than that observed among infected immunocompetent patients (Vinayak *et al.*, 1987). The importance of antibody has been demonstrated using mice which are unable to resolve infection after being depleted of B-cells via anti-IgM treatment, (Snider *et al.*, 1985). The involvement of secretory IgA in resolving *Giardia* infections is demonstrated by evidence that bile from immune rats (a good source of secretory IgA) confers protection to recipient animals, whereas bile from normal rats has no effect (Mayrhofer & Waight Sharma, 1988). Experiments with nude mice and nude rats, which become chronically infected, has demonstrated a requirement for thymus-derived (T) cells in protective immunity, presumably to facilitate a humoral response (reviewed by Faubert, 1996). Evidence that T-helper cells (and not cytotoxic T-cells) are involved comes from reconstitution experiments using nude mice, which showed that CD4-positive cells are required for clearance. Cytotoxic/suppressor T-cells and NK cells appear to play no role in host defence against *Giardia* (Heyworth, 1988).

1.9.3 Host range

As a genus, *Giardia* are reported to infect every class of vertebrate, with the exception of fish (Kulda & Nohýnková, 1978). The host range of *G. intestinalis* has not been fully explored, but it is known that isolates with this morphology occur in a wide variety of mammals (Appendix A). The original taxonomy of *Giardia* was based upon presumed host specificity, and a number of species were described (Table 4, p. 28). Cross transmission studies (eg., Erlandsen *et al.*, 1988; Hewlett *et al.*, 1982; Visvesvara *et al.*, 1988) have since shown that this presumption is not strictly correct, and for this reason, only species defined by morphological criteria are currently recognised, although these are often named after their host species (eg., *G. ardeae*, *G. psittaci*).

The ability of morphologically similar *Giardia* to infect both humans and animals which live in close proximity to humans (eg., pets, livestock) has led to the proposal that *G. intestinalis* is zoonotic. There is currently much debate about the validity of this proposal, largely because of the variable results of different cross-infectivity trials (Table 5, pp. 29 - 30). The success or failure of cross-infection appears to be both isolate and host dependent. Some isolates (eg., from muskrat) appear to exhibit host-preference, whereas others (eg., from

beaver) seem capable of infecting different host species, and are therefore potentially zoonotic.

Table 4. Examples of *Giardia* species defined using presumed host-specificity (adapted from Kulda & Nohýnková, 1978).

Host Species	Synonym for <i>G. intestinalis</i> (Lambl, 1859)
Primates	
Man	<i>G. lamblia</i> (Stiles, 1915)
Slender loris	<i>G. wenyoni</i> (Abraham, 1962)
Artiodactyla	
Goat	<i>G. caprae</i> (Nieschulz, 1923)
Ox	<i>G. bovis</i> (Fantham, 1921)
Sheep	<i>G. ovis</i> (Neveu-Lemaire, 1943)
Perissodactyla	
Horse	<i>G. equii</i> (Fantham, 1921)
Lagomorpha	
Rabbit	<i>G. duodenalis</i> (Davaine, 1875)
Rodentia	
Guinea pig	<i>G. caviae</i> (Hegner, 1923)
Norway rat	<i>G. simoni</i> (Lavie, 1924)
Muskrat	<i>G. ondatrae</i> (Travis, 1939)
Squirrel	<i>G. beckeri</i> (Hegner, 1926)
Voies	<i>G. microti</i> (Kofoid & Christiansen, 1915)
Carnivora	
Cat	<i>G. cati</i> (Deschiens, 1925)
Civet	<i>G. dasi</i> (Abraham, 1962)
Dog	<i>G. canis</i> (Hegner, 1922)

Although animals (eg., beavers) have been implicated as sources of waterborne outbreaks of giardiasis (eg., Erlandsen & Bemrick, 1988), the evidence for zoonosis is far from convincing. Human-derived *G. intestinalis* can be infective to animals, but the numbers of cysts used as inocula to establish infections in most experiments were quite high ($>10^5$), and in some instances only young animals were susceptible (eg., Cheng *et al.*, 1993). These inocula are unlikely to resemble the doses to which individuals may be exposed from contaminated water or other sources. For example, utilising the observed levels of *Giardia* cyst excretion from infected humans, and assuming a 1% prevalence rate of *Giardia* infection, Jakubowski (1984) estimated cyst concentrations of 9.6×10^3 cysts/litre in raw sewage. He also estimated that mountain streams contained up to 0.3 - 54 cysts/litre as a result of beaver and muskrat *Giardia* cyst excretion (Jakubowski, 1984). Actual levels have been reported to be 1 - 3 cysts/100 litres in both cistern water supplies in the Virgin Islands (Crabtree *et al.*, 1996) and watersheds of the Olympic Mountains in Washington State, U.S.A (Ongerth *et al.*, 1995). Similar levels have also been reported in watersheds in British Columbia, Canada,

although seasonal variations saw cyst levels as high as 229 cysts/litre (Isaac-Renton *et al.*, 1996). Furthermore, many of the animals implicated in the waterborne outbreaks of giardiasis are commonly infected with *Giardia* which are morphologically distinct from those which infect humans (Erlandsen & Bemrick, 1988). For example, muskrats were found to excrete cysts with the *G. microti* morphology, which have not been reported in humans. Variable infectivity of isolates has also been reported for *G. muris* by Grant and Woo (1978), who found one isolate from a laboratory mouse which appeared to be host specific, and two other isolates from a feral brown rat and a golden hamster which were able to infect other rodents.

Table 5. Cross-infection studies using *G. intestinalis* isolated from different hosts.

Source (no. isolates)	Isolate	Recipient(s)	Inocula (cysts)	Recipient infection (cyst excretion)	Ref.
sewage sediment		s.p.f. dogs	N/A	+	1
human		hamster, rabbit, lab. mice, deer mice, black bear, deer, sheep, cattle	N/A	-	2
human		lab. rat, gerbil, guinea pig, beaver, dog, raccoon	N/A	+	2
human (8)	Portland 1	dog (8)	$10^5 - 10^6$	3/8 [#] +	3
human		dog	10^6 trophozoites	2/3 +	3
beaver, <i>G. muris</i>		gerbil	60000	+	4
human		gerbil	5000	+	4
human	WB	cats (6)	10^6 trophozoites	1/6 + (0/6)*	5
human	WB	gerbil (6)	10^6 trophozoites	4/5 + (6/6)*	5
human		cat (8)	2.25×10^5	2/8 + (0/4)*	5
human		gerbil (6)	5×10^3	6/6 + (6/6)*	5
beaver, dog, human		gerbil	N/A	+	6
human	Guelph82-507	rats (8)	$8 \times 10^2 - 8 \times 10^5$	-	7
human	Guelph82-523	hamster (3)	10^6	-	7
human	Guelph82-524	mice (2)	10^6	-	7
human	Guelph82-525	rat (3)	10^6	-	7
human	10M	hamster (3), rats (4)	10^5	-	7
human	10M	mice (5)	50000	-	7
human	10M	cats (2)	2×10^5	-	7
human	clinical	puppies (4), dogs (3), kittens (3), cats (2)	3×10^5	-	7
human	clinical	suckling rats (10)	5000	-	7
human	WB	kittens (2), hamsters (2)	10^5 trophozoites	-	7
human	WB	mice (5)	8×10^4 trophozoites	-	7
human, dog		jird	N/A	+	8
cat		jird	5×10^3	5/16 +	8
human, dog, cat		AJ mice	N/A	-	8
human	GS/M	human (5)	50000 trophozoites	+	9
human	ISR	human (5)	50000 trophozoites	-	9
human	pooled sample	beaver (6)	<50	-	10
human	pooled sample	beaver (6)	<500	2/6 +	10
human	pooled sample	beaver (3)	<5000	1/3 +	10

Table 5 (cont.).

Donor(s) (no. isolates)	Isolate	Recipient(s)	Dose (cysts)	Recipient infection (cyst excretion)	Ref.
human	pooled sample	beaver (8)	500000	6/8 +	10
human	pooled sample	beaver (3)	550000	2/3 +	10
human	pooled sample	muskrats (8)	30	-	10
human	pooled sample	muskrats (7)	300	-	10
human	pooled sample	muskrats (6)	30000	-	10
human	pooled sample	muskrats (3)	500000	3/3 +	10
muskrat	5 isolates	beavers (9)	125000	-	10
beaver	3 isolates	muskrats (12)	500000	5/8 +, 4 died	10
human (x24)		young gerbils	5000	14/24 +	11
beaver (x4),		young gerbils	5000	4/4 +	11
cats (x2)		young gerbils	5000	2/2 +	11
muskrat (x6)		young gerbils	5000	6/6 +	11
cattle (x2)		young gerbils	5000	1/2 +	11
horse		young gerbils	5000	+	11
dogs (numerous)		young gerbils	5000	-	11
human	1	gerbil (5)	10 ⁴	-	12
human	2	gerbil (12)	10 ⁴	4/12 + (9/12)*	12
human	2	gerbil (12)	50000	- (4/12)*	12
human	3	gerbil (9)	10 ⁵	1/9 +	12
human	4	gerbil (12)	10 ⁵	12/12 +	12
human	4	gerbil (12)	10 ⁵	12/12 + (8/12)*	12
human	4	gerbil (4)	10 ⁵	-	12
human	5	gerbil (8)	10 ⁵	2/8 + (3/8)*	12
human	6	gerbil (2)	10 ⁵	-	12
human	7	gerbil (9)	10 ⁵	-	12
human	7	gerbil (10)	10 ⁴	-	12
human	7	gerbil (10)	250000	-	12
human	8	gerbil (1)	49000	-	12
human	9	gerbil (9)	10 ⁵	6/9 + (2/5)*	12
human	10	gerbil (6)	10 ⁵	-	12
beaver		deer mice (4)	10 ⁵	+ (3/4)*	13
human		deer mice (4)	10 ⁵	-	13

number infected / number inoculated * + for trophozoites at necropsy N/A not available

1 Shaw *et al.*, 1977

2 Davies & Hibler, 1979

3 Hewlett *et al.*, 1982

4 Faubert *et al.*, 1983

5 Kirkpatrick & Green, 1985

6 Wallis & Wallis, 1986

7 Woo & Paterson, 1986

8 Gasser *et al.*, 1987

9 Nash *et al.*, 1987

10 Erlandsen *et al.*, 1988

11 Swabby *et al.*, 1988

12 Visvesvara *et al.*, 1988

13 Cheng *et al.*, 1993

1.9.4 Prevalence

To date, all of the studies on the prevalence of *Giardia* infection have used cyst excretion and cyst morphology for the diagnosis and identification of *G. intestinalis* from faecal samples (Appendix A). In one instance, the morphology of trophozoites has been examined following endoscopy or necropsy (eg., Erlandsen *et al.*, 1990b). The results of numerous prevalence studies are summarised in Table 6 (p. 32). The prevalence of *G. intestinalis* in different hosts varied among studies, and appears to be dependent upon host age and living conditions. A degree of caution must be used when interpreting these studies, as the prevalence can be affected by seasonal fluctuations (Arroyave *et al.*, 1990), and may also be underestimated by the cyclic cyst excretion patterns exhibited by some isolates (Hewlett *et al.*, 1982). Although there have been genetic studies of collections of *G. intestinalis* isolates (eg., Andrews *et al.*, 1989; Homan *et al.*, 1992; Mayrhofer *et al.*, 1995; Meloni *et al.*, 1995; Nash & Mowatt, 1992; Proctor *et al.*, 1989; Upcroft *et al.*, 1995), there have been few studies which have explicitly examined the prevalence of specific genotypes in different host species. Isozyme and karyotypic typing have been used to study the epidemiology of waterborne outbreaks of human giardiasis (Isaac-Renton *et al.*, 1993; Sarafis & Isaac-Renton, 1993). Isolates from a variety of sources were established axenically, and it was found that all of the outbreak-associated isolates were in the same zymodeme. More recently, karyotyping and PCR/Southern hybridisation have been used to assess the prevalence of “Polish” and “Belgian” genotypes in Canadian drinking water (Wallis *et al.*, 1996). Samples were inoculated into gerbils, and the trophozoites from infected animals were then established in axenic culture. Only 41 of 437 samples were capable of infecting gerbils. Of these samples, 12 could be established in axenic culture, and 15 were lost due to contamination. A shortcoming of this study was that only the axenically cultured isolates were subjected to genetic analysis, which was surprising considering the availability of a diagnostic PCR system.

The highest prevalence rates for any animals were in native rodents (eg., muskrats and voles) in North America, where over 90% of trapped animals were found to be infected with *G. intestinalis* (Erlandsen *et al.*, 1990b). For all host species, the prevalence of *G. intestinalis* was highest in the young (eg., 34-53% in puppies versus 6-20% in adult dogs, or 18-28% in calves versus 10-11% in adult cattle; references for these figures are given in Appendix A), although this was not as pronounced in humans, possibly because few of the studies examined both children and adults or counted them separately. Despite the high (8%)

Table 6. Summary of prevalence of *G. intestinalis* in different host species.
(see Appendix A for further details and references)

Location	Host
Africa, Australia, Europe, North America	human 1.2-32%
Asia, Australia, Europe, North, Central & South America	domestic dogs 1.6-53%
Colorado, USA	coyote 5.9% of 34
Australia, Europe, Japan, New Zealand, USA	cats 0.6-17.7%
Canada, Mexico	sheep 7-35.6%
North & Central America, Switzerland	cattle 10.3-27.7%
Mexico	pigs 3%
Granada, Spain	rats 14.7% of 34
USA	beaver 9.2-18%
Canada	wild rat (<i>Peromyscus</i>) 10-98%
USA	muskrat 36.6-100%
Canada	vole (<i>Microtus</i>) 33-98.8 %

prevalence of *Giardia* in children in Ontario day care centres, none of those infected exhibited the classical symptoms of giardiasis (Woo & Paterson, 1986). Health status or living conditions also affect prevalence rates. In the case of stray dogs or dogs kept in kennels, the prevalence rates of *Giardia* are much higher than in pets (eg., 18.6% in a kennel versus 9.8% in pets in Japan, and 7.2% in strays versus 0% in pets in Brisbane, Australia). In humans, areas with poor living conditions tend to have high prevalence rates of *G. intestinalis*, and the

likelihood of reinfection is also high (Farthing *et al.*, 1986). For example, the prevalence rates in Switzerland, where living standards are high, were reported to be 3.6% (Wolff & Eckert, 1979) compared to a prevalence rate of 17.4% in Zimbabwe, where the living standards are lower (Loewenson *et al.*, 1986). In a landmark study, Farthing *et al.* (1986) examined 45 children in rural Guatemala over their first three years of life, and found that 70% of children became infected by the time that they were 1 year of age, and that on average each child had 3.6 *Giardia* infections by the time it reached 3 years of age.

1.9.5 Biochemistry

Studies on *Giardia* metabolism have largely utilised *G. intestinalis*, and have been restricted to a few axenic isolates, predominantly the isolate Portland 1 (eg., Edwards *et al.*, 1993; Lindmark & Jarroll, 1982). These studies have shown that energy production is by fermentation, under both aerobic and anaerobic conditions (Lindmark, 1980). Alanine is one of the major end-products of metabolism (Edwards *et al.*, 1993). The primary energy source is arginine, which is metabolised by the arginine dihydrolase pathway (Schofield *et al.*, 1990a). This metabolic pathway had only previously been found in prokaryotes, and its presence in *G. intestinalis* has been used as evidence to support the proposition that *Giardia* belongs to one of the earliest branching eukaryotic lineages (Knodler *et al.*, 1995). Glucose is also used as an energy source, although it is not absolutely required for growth (Schofield *et al.*, 1990b). Fermentation of glucose results in the production of ethanol, acetate and carbon dioxide (Lindmark, 1980). The enzymes involved in carbohydrate and energy metabolism were found to be in the cytoplasm, and the enzyme activities detected in trophozoite homogenates suggested that energy was produced by substrate level phosphorylation, that glycolysis occurs, and that both the Krebs cycle and cytochrome level oxidative phosphorylation were absent (Lindmark, 1980). Cysts have only 10-20% of the activity exhibited by trophozoites (Paget *et al.*, 1989). Glucose metabolism is via the Embden-Meyerhof-Parnas and pentose phosphate pathways (Jarroll *et al.*, 1981). The enzyme activities and respiration rates of *G. muris* trophozoites and cysts are similar to those of *G. intestinalis* (Paget *et al.*, 1989).

Giardia are obligate parasites, and lack many of the vital *de novo* biosynthetic pathways possessed by free-living eukaryotic cells, such as *de novo* pyrimidine and purine nucleotide synthesis (Lindmark & Jarroll, 1982) and the synthesis of cellular phospholipids and sterols (Jarroll *et al.*, 1981). As a result, trophozoites rely on “salvage” pathways to

utilise metabolites in the intestinal milieu of the host, and take up these substances via specific transport systems, eg., nucleosides (Davey *et al.*, 1992), amino acids (Schofield *et al.*, 1995), and serum lipoproteins, beta-cyclodextrins, and bile salts (Lujan *et al.*, 1996).

There are indications that different isolates of *G. intestinalis* differ in their growth requirements and metabolisms. Andrews *et al.* (1992b) found that different genotypes of *G. intestinalis* comprising mixed infections preferentially grew in either suckling mice or axenic culture. Genetically distinct *G. intestinalis* isolates have been found to have different *in vitro* growth requirements and dynamics (Binz *et al.*, 1992), including differences in the production of acetate, ethanol and alanine during axenic culture (Hall *et al.*, 1992).

1.9.6 Genomic organisation

Despite extensive studies, the organisation of the *G. intestinalis* genome remains poorly understood. The GC nucleotide composition of the *G. intestinalis* genome has been estimated using buoyant density to be 48% (Ortega-Pierres *et al.*, 1990), compared to 42% (Nash *et al.*, 1985) and 46.8% (Boothroyd *et al.*, 1987) determined by Cot analysis. There have also been different estimates of the genome size. Based upon renaturation kinetics, Boothroyd *et al.* (1987) estimated the genome size to be 80 megabases (Mb). However, using densitometric analysis of chromosome bands from pulsed-field gels, Jian-Bing Fan *et al.* (1991) calculated that *G. intestinalis* possess a haploid genome of approximately 12 Mb. The discrepancy was attributed by Jian-Bing Fan *et al.* (1991) to be due to an error in the earlier study. Adam *et al.* (1988) used densitometry to compare *Plasmodium falciparum* and *G. intestinalis* DNA, and estimated the total *G. intestinalis* DNA content of each nucleus to be 40 - 60 Mb. More recently, microdensitometric quantitation of Feulgen stained nuclei has been used to estimate a genome size of 65 - 71 Mb per nucleus (Erlandsen & Rasch, 1994).

Chromosomes are not readily visible during cell division, although 4 - 5 condensed structures have been observed in each nucleus by fluorescent microscopy of 4',6-diamidino-2-phenylindole (DAPI) stained trophozoites (Kabnick & Peattie, 1990) and by confocal microscopic examination of Feulgen stained trophozoites (Erlandsen & Rasch, 1994). Chromosomes have been identified following separation by pulsed-field gel electrophoresis (PFGE) or related techniques (eg., Adam *et al.*, 1988; Upcroft *et al.*, 1989), but their ploidy and distribution between the two nuclei are unknown. Telomeric repeats have been identified

and found to contain ribosomal RNA (rRNA) genes (Adam *et al.*, 1991; Le Blancq *et al.*, 1991), and have been further characterised using the 1.1-Mb chromosome from the isolate WB (Hou *et al.*, 1995). Karyotypic differences have been reported between *G. muris*, *G. ardeae* and *G. intestinalis* (Campbell *et al.*, 1990), as well as between isolates of *G. intestinalis* (eg., Isaac-Renton *et al.*, 1993; Upcroft *et al.*, 1989).

The earliest description of chromosomes using PFGE was by Adam *et al.* (1988), who detected 4-5 major bands and numerous minor bands. The results of Southern analyses suggested that some of the chromosomes were closely related. Using a similar technique, field inversion gel electrophoresis (FIGE), Upcroft *et al.* (1989) identified 5 - 6 distinct chromosomes, and noted that some nuclear material (presumably very large chromosomes) did not enter the gel. De Jonckheere *et al.* (1990) used PFGE to identify 6 - 8 chromosomes ranging in size from 0.8 - 3 Mb. Southern analysis showed that the rRNA genes were located on different chromosomes in different isolates of *G. intestinalis*.

Chromosomal rearrangements, resulting in a change in chromosome size, have been reported for rRNA-containing chromosomes (Le Blancq *et al.*, 1991) and for chromosomes that contain repetitive DNA (Upcroft *et al.*, 1990b). These rearrangements appear to be frequent, occurring in clones and subclones of *G. intestinalis* isolates (Le Blancq *et al.*, 1991; Hou *et al.*, 1995), and have been implicated in the development of drug resistance (Upcroft *et al.*, 1990b). The size changes are apparently due to changes in copy number of the rDNA repeat units (Le Blancq *et al.*, 1991; Adam, 1992). Mapping of a 2-Mb chromosome from the isolate WB and a drug-resistant subclone has suggested that chromosomal duplications occur. In this case, the central region of the duplicated chromosome was retained but one of the termini was deleted (Chen *et al.*, 1994a). Similar observations have been made for a 1.1-Mb chromosome detected in four subclones of the isolate WB, for which a central segment appeared to be conserved (Hou *et al.*, 1995). Based upon these observations, Hou *et al.* (1995) proposed that the *G. intestinalis* genome is tetraploid for the 1.1-Mb chromosome, although their reasoning is not clear. They described four homologous chromosomes, each one from a separate cloned line, but no evidence was presented to show that they all occurred within a single cell. The proposal of tetraploidy therefore appears unjustified.

Determination of the copy number of specific chromosomes by PFGE techniques appears to be problematical since populations of cells, rather than individual trophozoites, must be examined. Although sub-clones have been used in many experiments, the apparent

high frequency of chromosomal rearrangements suggests that sub-clones are unlikely to remain an homogeneous population of cells. This raises the possibility that the major and minor bands observed in some studies merely reflect the karyotypic heterogeneity present in the population of trophozoites sampled. The deduced structure of these *G. intestinalis* chromosomes, which have a central conserved region and variable termini, is similar to that of chromosomes from other protozoa such as *Leishmania* and *Plasmodium* (Lanzer *et al.*, 1995). Despite the apparent rearrangement of chromosomes in *G. intestinalis*, the linkage of some chromosomal markers appears to be conserved between a variety of different *G. intestinalis* isolates (Upcroft *et al.*, 1996).

In contrast to the attention given to the karyotypic analysis of *G. intestinalis*, much less has been given to the determination of the composition or function of the paired nuclei. Using *in situ* techniques, Kabnick and Peattie (1990) studied the incorporation of [³H]-UTP and demonstrated that both nuclei are transcriptionally active, with each nucleus producing an equivalent amount of transcript. Using *in situ* hybridisation, they also demonstrated that both nuclei possess rRNA genes, with evidence of multiple rRNA loci. Examination of DAPI stained cells was used to demonstrate that the nuclei contain equivalent amounts of DNA, and that they replicate synchronously. The number of putative chromosomes observed microscopically by Erlandsen and Rasch (1994) were similar to the numbers determined by the electrophoretic techniques, and to the number of rRNA loci observed by Kabnick and Peattie (1990). On the basis of their results and the published karyotypic data, Kabnick and Peattie (1990) suggested that the nuclei are haploid and functionally equivalent. The demonstration of the presence of Ras-like nuclear protein in both nuclei (Chen *et al.*, 1994b) has further implications for the functional roles of each nucleus, and strengthens the proposition of Kabnick and Peattie (1990). Considering the role of Ras and related proteins in other organisms (eg., regulation of cellular differentiation, cellular development and cell cycle), it would suggest that both nuclei are actively involved in cellular development (Chen *et al.*, 1994b). Currently, the best indication of ploidy has been given by allozyme electrophoresis, which has detected the presence of heterozygous strains, suggesting that trophozoites are functionally diploid (Andrews *et al.*, 1989; Meloni *et al.*, 1995; Andrews pers. comm.). Ultimately, the ploidy and nuclear composition of *G. intestinalis* trophozoites will only be unambiguously determined by *in situ* analysis using single-copy gene loci.

Giardia species, as representatives of the diplomonads, have a number of interesting features about their genomic organisation which distinguish them from other eukaryotes.

Based upon phylogenetic studies of highly conserved genes such as those encoding ribosomal RNA and elongation factors, it has been postulated that *Giardia* belongs to one of the earliest branching eukaryotic lineages (Cavalier-Smith, 1987; Edlind *et al.*, 1996; Hashimoto *et al.*, 1994; Hashimoto *et al.*, 1995; van Keulen *et al.*, 1993). Features possessed by *Giardia* which are consistent with this proposal are their lack mitochondria and the possession of ribosomes closer in size to those seen in prokaryotes. Sedimentation values of 21S and 16S have been determined for the large subunit and small subunit rRNAs respectively, which compares to 25S and 18S found in higher eukaryotes (Montanez *et al.*, 1988), and 23S and 16S found in prokaryotes (Noller, 1984). The sequences of the rRNA genes cloned from *G. intestinalis* (Portland 1 isolate) show that *Giardia* have a compact ribosomal DNA repeat, and the smallest ribosomal RNA genes reported for any eukaryote (van Keulen *et al.*, 1991). The copy number of the rRNA repeats has been estimated to be 63 per genome by densitometric analysis (Boothroyd *et al.*, 1987). Using 5.8S rRNA structural analysis, it has been reported that *G. intestinalis* has 60% similarity to rRNAs from other protozoa and archaeobacteria (Edlind & Chakraborty, 1987). Similar studies have shown that the evolutionary distances between *Giardia* and other nucleated cells, including several protozoans, is greater than expected, providing further evidence for early divergence from the eukaryotic line of descent (Sogin *et al.*, 1988).

Information regarding other aspects of genomic organisation is sparse. The genome contains repetitive elements in the form of minisatellites (Upcroft *et al.*, 1990a), but there have been no reports of microsatellite sequences. A tri-nucleotide repeat has been used by van Belkum *et al.* (1993) for inter-repeat PCR analysis of *G. intestinalis*, but they failed to recognise that this primer was a microsatellite. The assay, therefore did not directly assess microsatellite variation, but presumably amplified the regions between the microsatellite repeats. Since the start of the project described herein in 1992, the number of published *G. intestinalis* gene sequences has been steadily growing, and a number of interesting observations can be made. Despite the characterisation of over 28 protein encoding genes, none have been found to contain introns, and there have been no reports of pseudogenes. All of the genes appear to be present as single copies, although Rozario *et al.* (1996) reported two copies of the glyceraldehyde-3-phosphate dehydrogenase (*gapdh*) gene in *G. intestinalis*. These copies do not appear to be allelic, and only one of them is expressed in trophozoites (Rozario *et al.*, 1996). The *gapdh* gene which is not expressed has less similarity to *gapd* from other organisms compared to the expressed copy, but the signature sequences which characterise the enzyme are present and conserved, indicating that it is unlikely to be a pseudo

gene. Another feature of *G. intestinalis* protein coding sequences is a high GC content at the third codon position (Mowatt *et al.*, 1994).

In addition to their compact rRNA repeats and lack of introns, *Giardia* also possess other features which distinguish them from most other eukaryotes. These include the possession of a single copy of the ubiquitin gene (Krebber *et al.*, 1994) and unusually short leader sequences for gene transcripts (Kirk-Mason *et al.*, 1989; Alonso & Peattie, 1992). However, they possess typical eukaryotic polyadenylation signal sequences (AGTPuAA) downstream of all protein encoding genes (Alonso & Peattie, 1992). Analysis of genes (cytoskeletal proteins, eg., giardins, median body protein; enzymes, eg., glutamate dehydrogenase) have shown that they lack the typical promoter consensus sequences found in other eukaryotes, although conserved regions were identified within the promoter regions of these genes (Yee & Dennis, 1993; Holberton & Marshall, 1995). There have been two reports (Aggarwal *et al.*, 1988; Upcroft *et al.*, 1990b), of genes which contain the sequences typically used by RNA polymerases II and III.

1.9.7 Genetic characterisation

A number of molecular genetic techniques have been employed successfully to characterise *Giardia* isolates. The majority of studies have focused on the differentiation of isolates, rather than the determination of the genetic relationships among isolates. As a result, the genetic heterogeneity among isolates has been well documented, but the origin and biological significance of this heterogeneity is not as clear, especially with respect to the taxonomy of *G. intestinalis*. The following is a brief overview of the different approaches which have been used. These will be discussed in greater detail in the introductory sections of Chapters 3 and 4.

Isozyme and allozyme studies have revealed extensive genetic heterogeneity within the *G. intestinalis* morphological group (Bertram *et al.*, 1983; Meloni *et al.*, 1988; Andrews *et al.*, 1989). As will be discussed in more detail in section 3.1.2, isozyme analysis measures total enzyme activity, including allelic, non-allelic, and processed forms of the enzyme(s). In contrast, allozyme analysis measures allelic differences, giving a more realistic measure of genetic variation for a specific locus. The majority of studies (eg., Baveja *et al.*, 1986; Bertram *et al.*, 1983; Chaudhuri *et al.*, 1994; Homan *et al.*, 1992; Meloni *et al.*, 1988, 1995) have used a variety of methods (eg., starch gel electrophoresis, polyacrylamide gel

electrophoresis) to examine isozyme variation, invariably using the zymodeme approach to analyse the data. With the exception of Stranden *et al.* (1990), who found virtual identity between nine isolates of *G. intestinalis*, all isozyme-based studies have reported significant genetic heterogeneity among isolates of *G. intestinalis*. Numerous zymodemes have been identified, some of which contain isolates from different host species (most recently described by Meloni *et al.*, 1995). Genetic studies using allozyme electrophoresis (AE) have reported that, although genetically heterogeneous, isolates of *G. intestinalis* can be grouped into two major genetic assemblages which comprise discrete genetic groups (Andrews *et al.*, 1989, Mayrhofer *et al.*, 1995). The differences in the results derived by these two approaches (isozyme/zymodeme versus allozyme analysis) may possibly reflect differences in the isolates used, the number of loci scored (26 loci in the allozyme studies compared to 14 but commonly fewer than 10 loci in the isozyme studies) or the intrinsic differences between the two approaches (ie., the zymodeme approach is based upon the phenotype of an isolate, whereas allozyme analysis uses the deduced genotype). The studies utilising allozyme electrophoresis have provided the only quantitative comparisons of *G. intestinalis* isolates using genotypically defined characters.

Early attempts to examine variation at the molecular level utilised random DNA probes to detect restriction fragment length polymorphisms (RFLP). Although most hybridisation patterns were the same (ie., most of the regions probed were not polymorphic), some variation was demonstrated among the 15 isolates of *G. intestinalis* examined (Nash *et al.*, 1985). Nine different banding patterns were observed, with six isolates sharing the same pattern. When whole genomic DNA from two *G. intestinalis* isolates (dog and human origin) were used as probes for the Southern analysis of 8 isolates, little variation was found (Uji *et al.*, 1988). With hindsight, the reasons for the lack of variation in the latter study are very clear. All of the isolates used in these comparisons were grown by axenic culture, a method which preferentially selects for specific genotypes (Andrews *et al.*, 1992b), and allozyme analysis has shown that the isolates used were genetically very similar or identical (Andrews pers. comm.).

Using PFGE and related techniques, karyotypic analyses have been carried out to compare isolates of *G. intestinalis* (both human and non-human derived), *G. muris* and *G. ardeae*. The karyotypes of isolates from the different morphological groups were distinct, while isolates of *G. intestinalis* were similar to each other (Campbell *et al.*, 1990). Southern analyses, using either total genomic DNA or cloned genes from *G. intestinalis* or *G. muris* as

probes, showed no detectable cross hybridisation between the different *Giardia* species, while the *G. intestinalis* gene probe hybridised to different chromosomes in different isolates of *G. intestinalis* (Campbell *et al.*, 1990). Karyotypic differences among isolates of *G. intestinalis* have been used as markers in epidemiological studies (eg., Isaac-Renton *et al.*, 1993; Wallis *et al.*, 1996).

In a study by van Keulen *et al.* (1991), the rRNA genes were cloned from *G. intestinalis*, *G. muris*, and *G. ardeae*, and restriction maps were constructed. The results from this study showed that the three species were distinct, which is significant when considering the morphological similarities between *G. intestinalis* and *G. ardeae*. DNA sequencing of the small subunit rRNA genes has further demonstrated the large divisions which exist between the *Giardia* species, which are similar in magnitude to those separating different genera of other flagellate parasites, eg., *Trypanosoma* and *Crithidia* (van Keulen *et al.*, 1993). The rRNA genes appear to be highly conserved among *G. intestinalis* isolates (van Keulen *et al.*, 1995; Weiss *et al.*, 1992). However, a unique signature sequence was identified, dividing all of the *G. intestinalis* isolates examined into two groups (van Keulen *et al.*, 1995). Other investigators have also identified distinct groups of isolates using Southern analysis with specific gene probes (eg., Homan *et al.*, 1992; Nash & Mowatt, 1992).

Fingerprinting techniques, eg., minisatellites and random amplified polymorphic DNA (RAPD), appear to be useful tools for the differentiation and identification of *G. intestinalis* isolates, but their utility as markers for genetic studies, particularly phylogenetic analysis, is debatable (discussed in sections 1.5.3 & 4.1.1.2). DNA fingerprinting using the phage M13 or human minisatellite sequences as probes has been able to distinguish isolates of *G. intestinalis* that cannot be differentiated by other means (Carnaby *et al.*, 1995; Upcroft *et al.*, 1990a; Upcroft, 1991). RAPD's also appear to have a similar discriminatory power, and the relative similarity of banding patterns between isolates of *G. intestinalis* correlates with isozyme analysis (Morgan *et al.*, 1993; van Belkum *et al.*, 1993).

There is now a growing recognition that isolates of *G. intestinalis* belong to discrete genetic groups. In most instances, these have been defined using qualitative methods, such as the three genetic groups of Nash *et al.* (1985), the "Belgian" and "Polish" groups of Homan *et al.* (1992), and the two demes found by Upcroft *et al.* (1995). The Assemblages A and B of Mayrhofer *et al.* (1995), which include genetic groups I-IV (Andrews *et al.*, 1989), have been quantitatively defined, using distance-based methods to infer the genetic relationships among isolates. It is worth noting for future reference that, with the exception of the allozyme study

of Mayrhofer *et al.* (1995), virtually all of the isolates that have been characterised by other investigators are axenic cultures. Despite evidence that large genetic divisions separate *G. intestinalis* isolates (Mayrhofer *et al.*, 1995; Nash & Mowatt, 1992), the systematic relationships of the genetic groups comprising *G. intestinalis* have not yet been examined.

1.10 Objectives and rationale of thesis

After a review of the literature on *G. intestinalis*, it was clear that this species group requires systematic revision in order to facilitate research on this important and widespread human pathogen. Traditional taxonomic methods have proven incapable of distinguishing between isolates of *G. intestinalis* which exhibited substantial heterogeneity in a range of biological characteristics, such as metabolism, host infectivity and pathogenicity. Despite numerous demonstrations of genetic heterogeneity among isolates of *G. intestinalis*, the phylogenetic relationships of these isolates have not been examined in any depth. The determination of the evolutionary relationships of isolates of *G. intestinalis* would provide a sound foundation from which to examine the mechanisms responsible for the observed biological variation. It was the primary objective of this thesis to provide such a framework through the development and application of appropriate molecular markers which could be used to examine the systematics of *G. intestinalis*.

An integrated approach was adopted to examine the molecular systematics of *G. intestinalis*, combining both allozymic and nucleotide sequence analysis. Allozyme electrophoresis has proven to be a reliable method for the delineation of species in other genera of protozoan parasites, eg., *Leishmania* (Andrews *et al.*, 1988) and *Naegleria* (Adams *et al.*, 1989) and bacteria, eg., *Rhizobium* (Gordon *et al.*, 1995) and *Salmonella* (Selander *et al.*, 1986, 1991), and has been successfully applied to genetically characterise isolates of *Giardia* (Andrews *et al.*, 1989; Mayrhofer *et al.*, 1995). Among the advantages of AE are that it provides genetic data from multiple, independent loci, and that it is a versatile technique which can be applied at different taxonomic levels to examine, for instance, species boundaries, population structure, hybridisation, modes of speciation, paleobiogeography, in addition to its use in systematics. There are two main disadvantages with the application of AE to assess the systematics of *G. intestinalis*. The sample volumes required by this technique often makes the cultivation of isolates unavoidable. In recognition of the effects that culture conditions can have upon the selection of specific genotypes (Andrews *et al.*, 1992b), both axenic and suckling-mouse derived isolates were characterised. Allozyme data

also has a limited ability to resolve phylogenetic relationships when taxa differ at more than 60% of the loci examined (Richardson *et al.*, 1986), which is the case for some isolates of *G. intestinalis* (Andrews *et al.*, 1989; Mayrhofer *et al.*, 1995).

To complement the AE data, PCR systems were developed to allow the amplification and sequencing of segments of conserved structural and housekeeping genes, namely 18S rRNA, elongation factor 1 α (*efl* α), triose phosphate isomerase (*tpi*) and glutamate dehydrogenase (*gdh*). The genetic relationships inferred by AE were used as a basis for the selection of isolates for nucleotide sequencing and phylogenetic analysis. Multiple loci were used for two reasons. First, phylogenies determined using data from a single gene may only represent the evolution of that gene and not the organism, ie., produce a gene tree and not a species tree (Moritz & Hillis, 1996). Second, the genes selected appear to have different rates of evolution, and may therefore provide complementary data for different parts of the phylogeny. The development of these PCR systems were also expected to allow the direct characterisation of isolates without the need for culturing them. Such a detailed phylogenetic analysis is the first of its kind within the field of *Giardia* research.

1.11 Aim

The aim of this thesis was to **examine the molecular systematic relationships of isolates belonging to the morphological species *G. intestinalis*, and to provide the framework around which future comparative biological studies of *G. intestinalis* isolates from humans and other animal host species could be built.** The following specific objectives were chosen to achieve this aim:

- To use allozyme electrophoresis to extend and refine the model of genetic relationships within *G. intestinalis* as proposed by Andrews *et al.* (1989) and Mayrhofer *et al.* (1995).
- To use the polymerase chain reaction and nucleotide sequencing to genetically characterise selected isolates of *G. intestinalis* from human and other animal host species.
- To compare the relative utility of allozyme and nucleotide sequence analysis for the identification and genetic characterisation of *G. intestinalis* isolates.
- To examine the phylogenetic relationships of isolates within the *G. intestinalis* morphological group.
- To construct a systematic framework for the genus *Giardia* using the combined results of the allozyme and nucleotide sequence analyses.
- To assess the applicability of the phylogenetic species concept to organisms within the genus *Giardia*.

This thesis is divided into five chapters, the first of which has been the introduction and aims of this project. The second chapter deals with some of the general materials and methods used. Chapter Three, “Genetic Analyses Using Allozyme Electrophoresis”, describes the application of allozyme electrophoresis in the construction of a genetic model for *G. intestinalis*. In Chapter Four, “Genetic Analyses Using Nucleotide Sequence Data”, the allozyme-based genetic model is tested by phylogenetic analyses of the 18S rRNA, *efl* α , *gdh*, and *tpi* gene sequences. The last chapter discusses the significance of these results with respect to the above aims, and considers future directions for research based on this work.

Chapter 2

GENERAL MATERIALS AND METHODS

2.1 Materials

2.1.1 Enzymes and reagents

Restriction endonucleases were purchased from either Boehringer-Mannheim, New England Biolabs, Pharmacia or Amersham.

Thermostable DNA polymerase (*Taq*) was purchased from Biotech (Perth, W.A.).

Prism Ready Dye Deoxy Sequencing Kit was purchased from Applied Biosystems.

Chemicals were Analar grade. Phenol, sodium dodecyl sulphate (SDS) and sucrose were from BDH Chemicals. Tris was Trizma base from Sigma. Ethylene-diamine-tetra-acetic-acid, disodium salt (EDTA) was from Ajax Chemicals.

Source of oligonucleotides

Oligonucleotides were purchased from either DNA express, Colorado, USA, or Bresatec, Adelaide, Australia, with the exception of oligonucleotides #578, #579, #862, #913, #914, #1051 and #1052, which were synthesised on an ABI 3181A DNA synthesiser in the Department of Microbiology and Immunology, University of Adelaide. Synthesis was typically at the 40 nmol scale, and the oligonucleotides were reverse phase purified by the supplier.

2.1.2 Solutions

Isotonic phosphate

$\text{Na}_2\text{HPO}_4 \cdot \text{H}_2\text{O}$ (3.93 g) and anhydrous Na_2HPO_4 (15.2 g) were dissolved in 1 l of distilled water. The solution's pH was adjusted, if necessary, to 7.2 and it was autoclaved for storage.

PE buffer

20 x PE buffer is 2.66 M Na-phosphate; 20 mM EDTA, at a pH 6.9 (187.5 gm $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$; 207.6 gm Na_2HPO_4 (anhydrous); 7.44 gm Na_2EDTA made up to 1000 ml with distilled water).

Phosphate buffered saline (PBS)

PBS was made by adding 100 ml of isotonic phosphate to 400 ml of sterile physiological saline.

Phosphate buffered saline, modified (PBSm)

The osmolarity of PBS was adjusted to make it more suitable for maintaining the viability of *Giardia* trophozoites. A twenty times concentrated solution was made and diluted as required: KCl (4 g); anhydrous Na_2PO_4 (23 g); anhydrous KH_2PO_4 (4 g) and NaCl (250 g) were dissolved and adjusted to 1 litre with deionised water. The osmolarity of the working solution was 420 mOsm/kg.

TAE buffer

A 40 x stock solution was made using 1.6 M Tris base; 0.8 M Na acetate. $3\text{H}_2\text{O}$; 40 mM $\text{Na}_2\text{EDTA} \cdot 2\text{H}_2\text{O}$ made to a pH of 7.2 with acetic acid.

TE buffer

A 100 x TE stock was made using 1 M Tris-HCl pH 8.0; 100 mM EDTA, which was sterilised by autoclaving. 1 x working stocks were diluted in deionised water and resterilised.

2.2 Methods for the isolation and culture of *Giardia*

Giardia isolated in Adelaide (ie., those bearing the Ad prefix), axenic cultures of trophozoites and trophozoites isolated from suckling mice were provided by Assoc Prof Graham Mayrhofer (Department of Microbiology and Immunology, University of Adelaide). The methods used in the preparation of these materials are described below.

2.2.1 Purification of cysts from faecal samples

Cysts were semi-purified essentially as described by Mayrhofer *et al.* (1995). Faecal samples were emulsified in distilled water and filtered at room temperature through teased cotton wool which had been gently packed into the barrel of a 50 ml syringe. The column was rinsed with distilled water, and the cysts in the collected effluent were pelleted by centrifugation at 1000 rpm for 10 minutes. The cysts were washed 4 times with distilled water. An equal volume of 1 M sucrose was layered underneath the cyst suspension, and the sample centrifuged at 800 rpm for 10 minutes without braking. The interface was collected, and the cysts were washed

twice in 50 mls of distilled water. The cyst pellet was then resuspended in distilled water and cyst numbers determined.

2.2.2 Axenic culture of *Giardia* trophozoites

TYI-S33 medium, originally developed for the culture of *Entamoeba histolytica*, has been modified to support the growth of *G. intestinalis* (Keister, 1983). The composition is: trypticase peptone (BDH), 4 g; Difco bacto yeast extract (10%, autoclaved), 20 ml; glucose, 2 g; L-cysteine.HCl monohydrate (Sigma), 400 mg; NaCl, 400 mg; ox-bile (Oxoid, dessicated), 20 mg; ascorbic acid, 40 mg; anhydrous KH_2PO_4 , 120 mg; anhydrous K_2HPO_4 , 200 mg; ammonium ferric citrate (BDH, 1 g /100 ml), 0.56 ml; Minimal essential medium powder (Eagle; GIPCO), 130 mg; penicillin, 60 mg; streptomycin sulphate, 100 mg; sterile new-born calf serum, 20 ml; sodium bicarbonate (1.4%), 1 ml. This solution was made up to 200 ml with deionised water and sterilised by filtration through a 0.22 μm millipore filter.

Axenic isolates were maintained *in vitro* using modified TYI-S-33 culture medium. Falcon polystyrene tissue culture tubes (13 ml) were used to grow cells at 37°C until confluent. After removing any sediment of unattached or dead cells, subculture was achieved by cooling each tube for ten minutes on ice (Keister, 1983). The cells were detached into the medium using the jet from a Pasteur pipette and 1 to 2 ml aliquots were placed into new sterile tubes. Fresh medium was then used to fill each tube. Care was taken to ensure that cultures were not cross-contaminated by aerosoles in the lamina flow hood by using different bottles of medium for each isolate. Trophozoites were collected by centrifugation of the chilled suspensions at 1000 rpm for 10 minutes at 4°C in a Beckman benchtop centrifuge. The cell pellet was resuspended and washed 3 times in cold PBSm. Prior to the final spin, the cells were counted using a haemocytometer.

2.2.3 Growth of *Giardia* by infection of suckling mice

The following conditions were originally described by Mayrhofer *et al.* (1992). Best results were achieved when suckling mice were dosed with 2 - 3 x 10⁵ cysts/mouse. The minimum dose used was 5 - 6x10⁴ cysts/mouse. The ages of the mice were between 2 - 4 days. Following per-oral inoculation, the suckling mice were left with their mothers and then sacrificed by cervical dislocation 10 days later. The small intestines were then removed, cut into 3 - 4 segments and slit longitudinally. The intestinal segments were then immersed in cold TYI-S-33 medium (3 - 5 intestines in 5 ml) and left on ice for 10 min. The contents were then shaken vigorously. Intestinal debris was allowed to settle and the medium was decanted.

This step was repeated 2 - 3 times to maximise the yield of trophozoites. Trophozoites were pelleted by centrifugation at 4°C (10 min, 1000 rpm), washed twice in cold TYI-S-33 medium and counted. Where necessary, aliquots of 10^5 trophozoites were used to dose further suckling mice.

2.2.4 Purification of trophozoites

Chilled crude suspensions of trophozoites from suckling mice were passed through cotton wool loosely packed in a 5 ml syringe and rinsed through with cold PBSm or TYI-S-33 medium. After counting, at least 10^7 trophozoites were used as samples for allozyme or DNA extraction.

2.3 General methods

2.3.1 Genomic DNA isolation

Two methods were used for the extraction of DNA from trophozoites. The first was generally used for the large-scale purification of DNA from axenically cultured trophozoites, and the second was used as a rapid method for the preparation of DNA for amplification in polymerase chain reactions (PCR).

Method 1

This method was adapted from Campbell *et. al.* (1990). Trophozoites were harvested and washed in PBSm. Cells (5×10^7) were resuspended in 1 ml of cold lysis buffer (0.1 M Tris.HCl, pH 8.0; 0.2 M NaCl; 0.1 M EDTA, pH 8.0), and then mixed with 100 µl of 10% SDS (Sigma) and 20 µl of proteinase K (20 mg/ml). The mixture was mixed gently and incubated overnight at 37°C, then extracted 3 times with TE saturated phenol/chloroform/isoamylalcohol (PCI) (50/48/2, v/v/v), and once with chloroform. The aqueous layer, which was retained in each successive step, was then treated with RNase (10 µl of a 10 mg/ml solution) for 3 hr at 37°C and re-extracted as before. The purified DNA was dialysed overnight at 4°C against 2 l of TE and then precipitated overnight at -20°C by addition of a 1/10 volume of 3 M sodium acetate pH 5.2 and 3 volumes of 100% ethanol. The precipitated DNA was pelleted by centrifugation, rinsed with cold 70% ethanol and then vacuum dried. The dried DNA was redissolved in 100 µl of TE and stored at 4°C.

Method 2

At least 10^7 trophozoites were resuspended in 1 ml of TE to which 100 μ l of 10% SDS and 20 μ l proteinase K (20 mg/ml) were added. The sample was then either incubated overnight at 37°C or for 3 hours at 55°C. Following incubation, the mixture was extracted twice with PCI (as used in Method 1), and then precipitated overnight at -20°C by addition of 1/10 volume of 3M sodium acetate pH 5.2 and 3 volumes of 100% ethanol. It was found that substitution of the sodium acetate with 1/25 volume of 5 M NaCl reduced the co-precipitation of water insoluble solids. Good quality, large molecular weight DNA could also be purified and concentrated using BresaClean (section 2.3.6).

2.3.2 DNA quantitation

The concentration of DNA solutions was determined by measurement of absorption at 260 nm and assuming an A_{260} of 1.0 is equal to 50 mg/ml DNA (Maniatis *et al.*, 1989).

2.3.3 Restriction endonuclease digestion

Samples of DNA were incubated with 2 units of restriction enzyme at 37°C, for 1 - 2 h in a final volume of 20 μ l containing the reaction buffer supplied by the manufacturer. Prior to loading onto a gel, a one tenth volume of tracking dye (15% (w/v) Ficoll, 0.1% (w/v) bromophenol blue, 0.1 mg/ml RNase A) was added.

2.3.4 Agarose gel electrophoresis of DNA samples

Electrophoresis of PCR amplification products or digested DNA was carried out at room temperature on horizontal, 0.8% or 1% (w/v) agarose (Seakem HGT) gels. Gels were run in TBE buffer (67 mM Tris base, 22 mM boric acid, 2 mM EDTA, pH 8.8) at 100 V until the tracker dye had migrated half to two-thirds of the way down the gel. After electrophoresis the gels were immersed in distilled water containing 2 μ g/ml ethidium bromide. DNA bands were visualised by trans-illumination with UV light and photographed.

2.3.5 Calculation of DNA fragment sizes

The size of DNA fragments was calculated by regression analysis from relative mobilities using *EcoRI* digested *Bacillus subtilis* bacteriophage SPP1 DNA. The sizes of the fragments (kilobases, kb) were: 8.51; 7.35; 6.11; 4.84; 3.59; 2.81; 1.95; 1.86; 1.51; 1.39; 1.16; 0.98; 0.72; 0.48; 0.38. (Bresatec catalogue, 1991/1992).

2.3.6 Purification of DNA using BresaClean

The principle of BresaClean and related purification systems is that DNA adheres to the glass beads at certain salt concentrations. Because small fragments (< 100 bp) do not bind efficiently, it is an excellent method for removing unincorporated primers from PCR mixtures. It is also an excellent method for removing salts and other material from DNA samples without resorting to dialysis. With the exception of sterile distilled water and ethanol, the reagents were provided in the kit by the manufacturer, and the composition of the solutions was not specified. The following protocol was modified from the manufacturer's.

The volume of the sample to be cleaned was made up to 100 µl with sterile distilled water. To this was added 300 µl of NaI (Bresa Bind Solution 1) and 5 µl of glass milk (BresaBind). The solution was then mixed by briefly vortexing, and incubated at RT for 20 min, with occasional mixing to keep the glass milk in suspension. Following incubation, the glass milk was pelleted by centrifugation (discarding the supernatant), and washed twice in 400 µl Bresa wash (ie resuspended the pellet in 400 µl BresaWash, spun down, supernatant discarded, then repeated). The pellet was then dried at 60°C. Care was taken not to over-dry the sample, as this can potentially damage the DNA. Once dry, the DNA was eluted from the BresaBind by resuspending the pellet in MQ water (50 µl for PCR mixtures) and incubating (with occasional mixing) at 55 - 60°C for 20 minutes. The glass matrix was pelleted by centrifugation, and the solution was carefully pipetted off. Sample quality was checked by agarose gel electrophoresis.

2.3.7 DNA sequencing

Uncloned PCR amplification products were used as sequencing templates, as this minimises potential sequencing errors due to nucleotide mis-incorporation by *Taq* DNA polymerase, and also allows the detection of PCR contaminants or mixed infections (ie. more than one genotype present in a sample). By using uncloned material, the entire population of amplification products is randomly sampled. This approach is more efficient than using cloned amplification products, which usually necessitates the sequencing of multiple clones to ensure that there is no variation between cloned products (which could be due to polymerase errors or the presence of contaminating DNA). The direct sequencing approach has proven to be highly reliable, provided that the amplified DNA is a single homogeneous product.

Samples were sequenced from both strands using an Applied Biosystems Prism Ready Dye Terminator sequencing kit, following the protocol provided by the manufacturer. All PCR amplification products were purified by BresaClean prior to use as sequencing templates. The sequencing reactions were carried out on a Perkin Elmer GeneAmp PCR System 2400. In the case of the 18S rRNA gene amplification products, it was found that addition of 10% dimethylsulfoxide (final conc.) to the sequencing reaction greatly enhanced the length and quality of the sequences from templates of *G. intestinalis* origin. All analyses were carried out on an Applied Biosystems automated DNA sequencer (model 373A or 377). The resulting raw sequences were edited using the SeqEd package from Applied Biosystems.

2.4 Computer software

The following software packages were used through the course of the present studies. Only a brief description of each package will be given. The theory involved in the various approaches used in phylogenetic inference are reviewed in detail by Swofford *et al.* (1996). In addition, the documentation provided with the majority of the software packages also contains the relevant information regarding the theory and implicit assumptions used, as well as the general operation of the software.

2.4.1 DNA sequence editing and analysis

The raw sequences (chromatogram files) were edited using the SeqEd program (Applied Biosystems), which provides an easy to use interface for the manipulation of sequence

chromatograms, and can be used to align sequences. The edited sequences were analysed using the program DNASIS (version 7.0, Hitachi Software Engineering Company), which was used to conduct sequence similarity searches and identify restriction endonuclease cleavage sites in the DNA sequences.

2.4.2 Oligonucleotide construction

Potential PCR primers were examined for undesirable properties (eg. false priming, formation of primer dimers, hairpin formation) using Primer (version 2.0, Scientific & Educational Software). This program has only limited utility as it cannot analyse primers containing degenerate sites. However, it does provide a useful guide.

2.4.3 Multiple DNA sequence alignments

ClustalW (Thompson *et al.*, 1994) was used to construct multiple sequence alignments. This program computes the pairwise similarity between sequences, and then uses the unweighted pair-group method using arithmetic averages (UPGMA) to construct a dendrogram. The dendrogram is then used as a guide for building the multiple alignment. The default program parameters (eg., for gap penalties, mismatch penalties, etc.) were used.

2.4.4 Phylogenetic inference

A variety of methods were used to infer the phylogenetic relationships among the *Giardia* isolates (section 4.4.2.7). These were implemented using the following computer packages:

- MEGA version 1.01, Kumar *et al.* (1993, 1994).
- METREE version 1.2, available freeware, Rzhetsky and Nei (1992, 1993, 1994).
- PUZZLE version 3, available freeware, Strimmer and von Haeseler (1996).
- PHYLIP version 3.5c, available freeware, Felsenstein, (1989, 1993).
- fastDNAML, available freeware, Gary Olsen, Department of Microbiology, University of Illinois, Hideo Matsuda, Ray Hagstrom & Ross Overbeek, Argonne National Laboratory, Illinois.
- PAUP version 3.1.1, Swofford (1993).
- TREEVIEW version 1.31 (32 bit), Page (1996).

A number of programs (written in BASIC) for the analysis of allozyme data were kindly provided by Mark Adams and Brian James, Evolutionary Biology Unit, South Australian Museum.

Distance analyses of both DNA and allozyme data were conducted using PHYLIP (NEIGHBOUR, GENDIST, DNADIST), while MEGA and METREE were used to analyse DNA and predicted protein sequence data. The default program options were used unless otherwise specified. Where possible, distances were calculated using the pairwise deletion method, which only deletes those sites containing missing/ambiguous data in each individual pairwise comparison. MEGA was used to determine the DNA sequence statistics (eg., nucleotide composition) and to calculate both the uncorrected (p-distances) and corrected (eg., Tamura Nei) distance estimates which were used in the comparisons of the different measures (eg., section 4.4.3.4). Minimum-evolution trees using METREE were generated using the standard search option (ie., "Search for ME tree" option).

Maximum likelihood analyses were conducted using PHYLIP (DNAML), fastDNAML and PUZZLE. The default program options were used unless otherwise specified. The DNAML and fastDNAML analyses used empirical base frequencies calculated from the sequence data, and had the global option in effect. The transition/transversion ratio was originally determined by fastDNAML analysis (eg., section 4.4.3.4), however, PUZZLE can also be used to estimate this parameter. All of the PUZZLE program parameters were automatically calculated by the approximate method using quartet sub-sampling and the Neighbour-Joining tree. The unconstrained Tamura Nei substitution model was used, with either the 2 or 5 category rate heterogeneity model. Of the three programs, PUZZLE would be recommended as it is easier to use, faster, and more flexible.

Cladistic analyses were conducted using PAUP and PHYLIP (DNAPENNY, DNAPARS, PROTPARS). In the case of PAUP, the default options were used (eg. characters were unweighted and treated as unordered), except that ambiguous sites were treated as missing data, and uninformative characters were excluded from the analyses. Analyses using the PHYLIP programs were all set to jumble the sequence input and search for the best tree using ordinary parsimony. The DNAPENNY parameters were set to examine a maximum of 10^6 trees using the more complex branch and bound option (reconsiders order of taxa).

Chapter 3

GENETIC ANALYSES USING ALLOZYME ELECTROPHORESIS

3.1 Description of technique

The technique of allozyme electrophoresis (AE) had its foundation in the mid 1950's, with the first description of starch gel electrophoresis by Smithies in 1955 followed by the report of the histochemical staining of enzymes on gels (Hunter & Markert, 1957). AE provides a direct assessment of the levels of allelic variation at individual enzyme loci, and has facilitated the study of genetic variation in populations, gene flow between populations, hybridisation, and the identification of species boundaries (Murphy *et al.*, 1996; Richardson *et al.*, 1986).

3.1.1 General principles

AE utilises the basic principle that proteins are made up of amino acids (charged or uncharged, depending upon the nature of the side chain), the composition and order of which are genetically encoded. The amino acid sequence, which is the primary structure, determines the secondary and tertiary structure of a protein due to the interactions of the amino acid side chains, such as hydrogen bonding, hydrophobic/hydrophilic interactions, etc., which lead to the formation of β -stranded sheets or α helices. The tertiary and quaternary (if the enzyme is multimeric) structures determine which residues are exposed to the aqueous environment. The net charge of the protein depends on its overall composition and structure, and on the pH of the aqueous environment. Depending upon the support medium used (Table 7, p. 54), AE detects differences between proteins, either due to differences in net charge, or differences in the size and shape of the molecules. Differences in the mobilities of specific enzymes are a reflection of allelic differences at the genetic level.

As with other electrophoretic techniques, samples are loaded onto a support medium eg., cellulose acetate or starch gel, and an electric current is applied. The choice of buffer is important, as it can affect the band resolution (presumably due to the effect of pH on the charge of the protein), and also determines which enzyme staining systems can be used (a practical consideration when gels are to be stained for more than one enzyme). Four different

support media can be used for enzyme electrophoresis: starch, polyacrylamide, cellulose acetate, and agarose. Their properties are summarised in Table 7. Cellulose acetate was chosen for use in this thesis, largely because of its consistency and repeatability (Harris & Hopkinson, 1976), as well as the small sample volumes required. Other advantages are the small volume of stain required, and also ease of handling. Because of its large pore size, cellulose acetate separates only on the basis of charge. A disadvantage of cellulose acetate is that it is subject to electroendosmosis, which is caused by the large pore size, resulting in the increased migration of cationic enzymes, and decreased or reversed migration of anionic enzymes. However, this only becomes a problem if long run times (>5 hours) are used or if an isozyme or zymodeme interpretation is made from the data (Richardson *et al.*, 1986).

Table 7. Properties of the 4 support media used for allozyme electrophoresis
(adapted from Murphy *et al.*, 1996).

Property	Starch	Polyacrylamide	Cellulose acetate	Agarose
Separates by charge	yes	yes	yes	yes
Separates by size	yes	yes	no	no
Simultaneous resolution of cationic/anionic proteins	yes	no	yes	yes
Electroendosmosis ³	yes	no	yes	yes
Running time	4 - 24 h	4 - 6 h	0.3 - 3 h	3 - 4 h
Minimum sample volume	2 µl	2 µl	0.2 µl	1 µl
Volume of stain required	5 - 50 ml	10 - 50 ml	1 - 3 ml	10 - 50 ml
Ease of preparation	no ¹	no ¹	yes ²	no ¹
Ease of handling	yes	no	yes	yes
Ease of staining	no ⁴	no	yes ⁵	no

1 gels must be cast by user:- quality and repeatability may vary.

2 gels are pre-made (under specific conditions), requiring minimal preparation.

3 degree of electroendosmosis is directly proportional to pore size.

4 requires use of agar overlays containing stain.

5 stain is directly applied to cellulose acetate gel sheet.

Following electrophoresis, gels are stained histochemically. This is most straight forward when using cellulose acetate, as the stain can be directly applied to the gel sheets. The staining systems are designed to detect specific enzyme activities, although in some cases there may be cross reactivity between enzymes because of similarities in their substrate specificities. In general, the histochemical stain contains all of the appropriate substrates, cofactors, coenzymes, etc., required for the enzyme to catalyse the reaction, which is in turn

linked to the development of a particular histochemical stain (discussed in detail by Richardson *et al.*, 1986). Some stains are visible under normal light (eg., NBT/PMS, which is used to detect the production of NADH or NADPH), while others produce either a loss (eg., enolase, due to the production of NAD which does not fluoresce) or a gain (eg., esterase, by cleavage of its substrate from a fluorescent compound) in fluorescence and must be viewed with UV light.

3.1.2 Interpretation of electrophoretic data

There are three different approaches to analysing the enzyme banding patterns which are the end result of electrophoresis. These are:

1) Isozyme analysis.

Isozymes are all functionally similar forms of an enzyme, including those resulting from the expression of different alleles at a single locus, expression of enzymes with similar function from different loci, post-translational modifications of a single enzyme, as well as any combination of subunits from different alleles or loci in the case of multimeric enzymes (Murphy *et al.*, 1996; Richardson *et al.*, 1986). Operationally, all bands visualised using a particular histochemical stain are utilised in the analysis, regardless of their origin or genetic basis. This approach does not make a genetic interpretation of the data, and hence avoids any assumptions regarding the possible origins of the bands in question. As a result, raw isozyme data are unsuitable for use in phylogenetic inference, mainly because of uncertainties regarding data independence and character homology of the enzyme bands (Murphy *et al.*, 1996). Isozyme data have most typically been used to assess species boundaries, where fixed differences (which occur when two taxa have no enzyme bands in common at a particular locus, Davis & Nixon, 1992) provide diagnostic markers for species delineation (eg., Good, 1989).

2) Zymodeme analysis.

The zymodeme approach also utilises isozyme banding patterns. However, particular emphasis is placed upon the differences occurring between taxa. Zymodemes are defined by comparison of the isozyme banding patterns of different taxa, and only comprise taxa with identical profiles. Unlike the more conservative fixed difference measure, any difference in

banding pattern between taxa is sufficient to have them placed into a separate zymodeme. This approach is most typically used to differentiate between strains of closely related organisms. A major limitation, especially with respect to microorganisms, is that it cannot detect whether samples contain multiple strains of genetically distinct organisms. Similarity between zymodemes can be examined by phenetic analysis (eg., Meloni *et al.*, 1988). However, the validity of generating any form of genetic distance from such data, which may violate the assumptions of character independence and homology, is debatable (Swofford *et al.*, 1996).

3) Allozyme analysis.

Allozymes are that subset of isozymes caused by allelic variants of the same enzyme at a given locus. This approach requires a genetic interpretation of the electrophoretic results. The presumptive genotype for each locus is determined by analysis of the banding pattern, taking other factors into consideration, such as the quaternary structure of the molecule, the ploidy of the organism, and the time course of staining (which aids in the identification of primary enzyme bands, as distinct from sub-bands due to modifications or degradation of the enzyme). Allozymic interpretation of electrophoretic data is commonly known as allozyme electrophoresis (AE), and of the three approaches discussed is the only one that is extensively used in molecular systematics (Murphy *et al.*, 1996).

3.1.3 Applications of technique

There are a number of applications that allozyme electrophoresis has been used for. These will only be briefly discussed here, but for a more detailed account see Murphy *et al.* (1996). In most cases, AE has been used either as a systematic tool at both the intra and inter-specific level (see Table 8, p. 57), or as a tool to examine evolutionary processes (eg., paleobiogeography, origin of polyploidy, rates of evolution). AE has been applied successfully to a variety of vertebrate and invertebrate taxa, where it has proven to be a versatile tool for both population genetics and species delineation (eg., Riley *et al.*, 1989). In these cases, the biological species concept is most often used, and is invoked when sympatric populations exhibit significant differences in allele frequencies for a given locus (eg., Chilton *et al.*, 1992). In the case of parapatric populations, or where organisms are asexual (eg., bacteria), species delineation is problematic. However, allozyme electrophoresis has been used to assess population structure and species differences in protozoan and bacterial species

(Adams *et al.*, 1989; Andrews *et al.*, 1988; Selander *et al.*, 1986). In some instances (eg., *Salmonella*), allozyme studies have highlighted major deficiencies in current bacterial taxonomic practice (Selander *et al.*, 1991).

Table 8. Applications of AE in systematics.

Intra-species	Inter-species
Population structure	Species boundaries
Inbreeding, outcrossing, dispersal	Phylogenetic relationships
Paternity studies	Hybridisation

In addition to being a comparative tool, AE can also provide basic biological information. This has been exemplified by Tait (1983), who used AE to examine sexual processes in the kinetoplastids. By invoking a genetic basis for the observed banding patterns, it is possible to determine the ploidy and mode of reproduction of an organism. This can be done when multiple banding patterns occur for multimeric enzymes, indicating that either the sample examined comprises a mixture of genotypes, is a heterozygote, or that gene duplication has occurred. These three possibilities can be distinguished from each other on the basis of the observed banding patterns (Figure 6, p. 58). In the case of a sample containing a mixture of two genotypes of either haploid or homozygous diploid organisms, the different polypeptides encoded by either the a or b alleles are expressed in different organisms and are not able to interact to form an ab heterodimer, unlike the case when both alleles are expressed within the same cell of a heterozygous organism. Only two bands representing the a or b homodimers will be observed, the intensities of which are determined by the proportions of each genotype in the sample. Provided that neither the a or b alleles are null alleles (discussed by Richardson *et al.*, 1986), a diploid heterozygote will exhibit a typical 3 banded pattern, with a ratio of 1:2:1 for the a_2 , ab and b_2 dimers respectively. However, this banding pattern can also be observed in haploid organisms which have undergone a gene duplication. These two possibilities can be differentiated by examining the banding patterns exhibited by other multimeric enzymes and by other samples. The detection of apparent homozygotes for both a and b alleles in other samples would be evidence to support a heterozygote diploid, as would the detection of heterozygosity at other loci, both of which are unlikely to occur in a haploid organism (Tait, 1983).

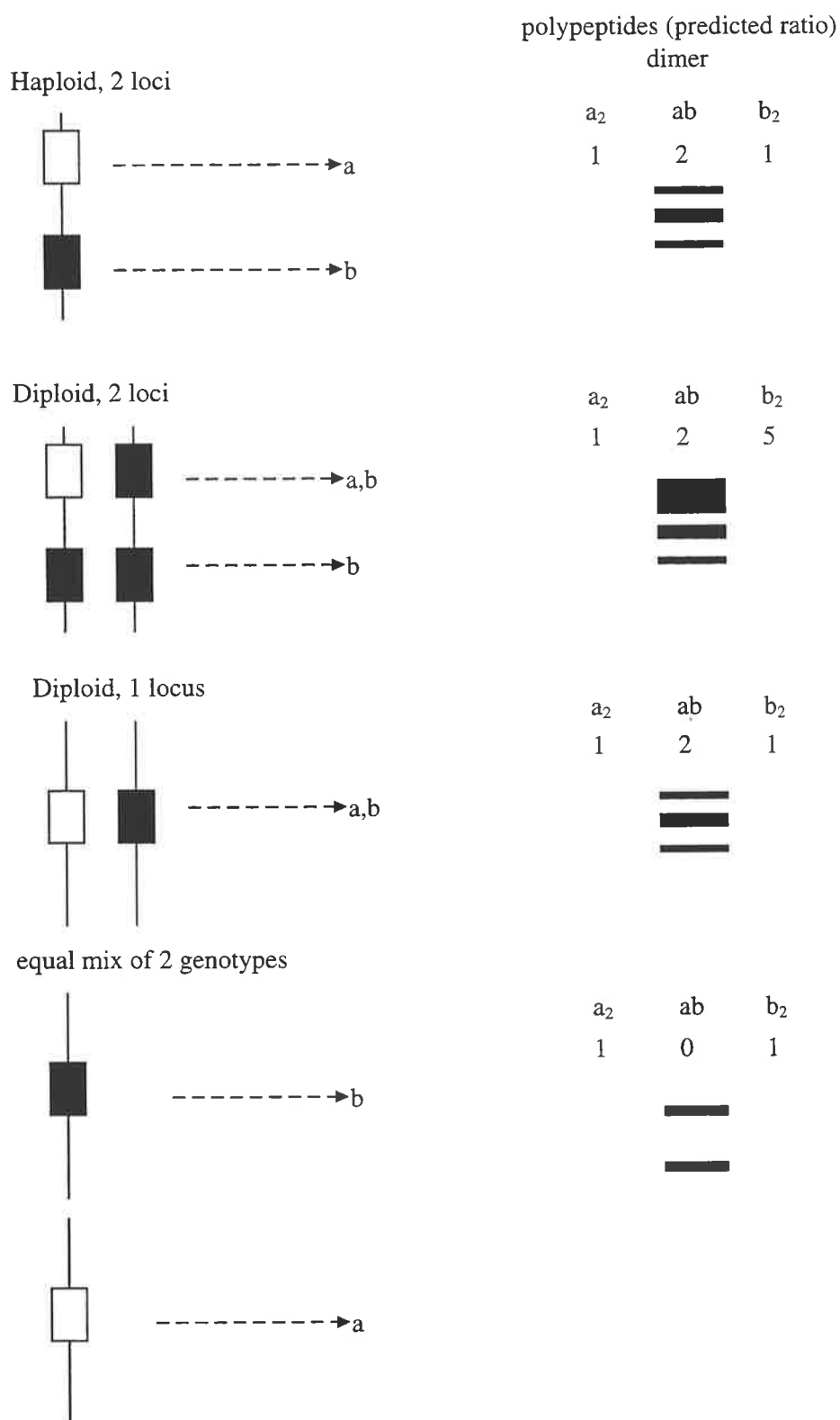


Figure 6. Predicted allozyme banding patterns for a dimeric enzyme according to different genetic models (modified from Tait, 1983).

3.1.4 Advantages/disadvantages of technique

The major advantage of AE over other methods is that it enables genetic data to be obtained from multiple loci (Table 9, p. 60). This effectively samples across the genome, thus minimising the potential pitfalls that may result from the use of a single locus (eg., recovery of gene trees instead of species trees). Another advantage, in terms of project planning and design, is that AE requires a minimal amount of development before it can be applied productively. Many of the biochemical enzyme pathways utilised in AE (eg., the TCA cycle and glycolysis) have been well characterised from a variety of organisms and are almost universal among them. This means that many of the staining systems can be readily adapted from one organism to another. Compared to other methods, AE is highly versatile because it can be used at both the intra- and inter-specific levels, and it is also rapid, allowing the detailed characterisation and analysis of many samples within several days. A further advantage of this technique is that it can detect samples comprising multiple genotypes, and can therefore be used for the detection of strain contamination and the determination of genetic purity.

The main disadvantage of AE is that it is subjective, requiring a fair amount of experience (in practical terms) to allow a reliable interpretation of the data. The practical limitations of the technique (Table 9, p. 60) are the large sample volume required compared to PCR-based techniques (only a problem when using microorganisms that are difficult to culture), the quality of the sample (as enzymes are labile), and the comparative nature of the technique, which necessitates the use of carryover controls if the data from separate studies are to be combined.

AE can have limited utility in phylogenetic inference, especially when the degree of divergence exceeds 60 - 70% fixed genetic differences (a fixed genetic difference occurs when a pair of samples fail to have an allele in common at a particular locus; Richardson *et al.*, 1986). At these levels, the probability of convergent/parallel evolution is greater, which means that the true genetic distances cannot be accurately determined. Unlike the case with nucleotide sequences, there are no models of allozyme evolution which can correct for multiple substitutions. The cladistic approach is also difficult to employ under these conditions because of the increased chance of homoplasy in the data, as well as the increased number of uninformative characters (due to autapomorphies) which usually occur as

the divergence between taxa increases. The appropriate treatment of heterozygous individuals in phylogenetic inference is also problematic (Murphy *et al.*, 1996; Richardson *et al.*, 1986).

Table 9. Advantages and disadvantages of allozyme electrophoresis.

Advantages	Disadvantages
• Examines multiple independent loci. • Randomly samples throughout genome. • Mutations are neutral (ie., absence of selection). • Time & cost effective. • Can rapidly characterise many samples. • Detects a broad range of genetic variation. • Enzyme pathways conserved among most eukaryotes and prokaryotes. • Well established use in vertebrate and bacterial genetics. • Can detect mixed samples/contamination.	• Subjective. • Comparative; requires carry-over controls. • Is conservative (underestimates differences). • Large sample required (compared to PCR). • Homoplasy of characters common as genetic distance increases, therefore has limited use for phylogenetics.

3.2 Previous electrophoretic studies using zymodeme analysis

There have been numerous electrophoretic studies of *G. intestinalis* but with a few exceptions, the majority have been of limited utility, either using too few characters, small numbers of samples, or both. In some cases the methods used for data analysis are also open to criticism. In light of the demonstration that culturing methods can select for specific genotypes (Andrews *et al.*, 1992b; Thompson *et al.*, 1996), the interpretation of the results of those studies which have used only a single method for establishing isolates (especially those dealing with environmental samples which may comprise mixed genotypes) is questionable. The following is a brief review of the major isozyme studies of *G. intestinalis*.

The majority of isozyme studies published to 1995 are summarised in Table 10 (p. 61). All of them used the zymodeme approach to interpret the data. The majority involved examination of axenic isolates using 10 or fewer enzyme staining systems. Bertram *et al.* (1983) found that human and animal isolates of *G. intestinalis* gave the same banding patterns, and they questioned the validity of using host-origin as a taxonomic marker for *Giardia* species. Most of the subsequent reports revealed a large degree of homogeneity

among isolates (eg., Proctor *et al.*, 1989; Strandén *et al.*, 1990), and in the cases where isolates could be distinguished, no correlation was detected between the course of disease and the zymodeme of the *Giardia* isolate (Proctor *et al.*, 1989; Homan *et al.*, 1992). Using 3 enzyme staining systems, Homan *et al.* (1992) identified 13 zymodemes among 47 isolates of *G. intestinalis*. Of the 13 zymodemes, 2 comprised the bulk of the isolates (named the “Polish” and “Belgian” groups), while the rest of the zymodemes each comprised only 1 or 2 isolates. With the exception of the reports of Meloni *et al.* (1988, 1989, 1992, 1995), none of the isozyme studies have attempted to examine the genetic relationships of the different zymodemes.

Table 10. Published reports characterising *G. intestinalis* isolates by zymodeme analysis.

No. of enzymes examined	No. of samples	Sample origin	Reference
6	5	axenic	Bertram <i>et al.</i> , 1983
4	4	axenic	Baveja <i>et al.</i> , 1986
10	30	axenic, 1 suckling mouse	Meloni <i>et al.</i> , 1988
10	19	axenic	Cedillo Rivera <i>et al.</i> , 1989
10	49	axenic, 1 suckling mouse, 1 cyst	Meloni <i>et al.</i> , 1989
9	32	axenic	Proctor <i>et al.</i> , 1989
14	9	axenic	Strandén <i>et al.</i> , 1990
6	3	axenic	Chaudhuri <i>et al.</i> , 1991
13	38	axenic	Meloni <i>et al.</i> , 1992
3	47	axenic	Homan <i>et al.</i> , 1992
13	97	axenic, 1 suckling mouse, 1 cyst	Meloni <i>et al.</i> , 1995

The most detailed isozyme studies have been by Meloni *et al.* (1988, 1989, 1992, 1995), where 10 - 13 enzyme systems were employed to study 97 isolates in total. Their results confirmed the findings of Bertram *et al.* (1983), with human and animal derived *G. intestinalis* possessing the same, or similar, banding patterns. In addition, a large amount of heterogeneity was detected among isolates of *G. intestinalis* derived from humans. Of the 47 zymodemes identified among the 97 *G. intestinalis* isolates, 33 comprised only single isolates, and 3 zymodemes (M1, M3 & M4) comprised more than 3 isolates. No pattern of relationships, other than groups of zymodemes separated by 36% “fixed enzyme differences”, were described (Meloni *et al.*, 1995). However, Meloni *et al.* (1992) did report a correlation between geographic locality (Western Australia) and a cluster of zymodemes (M8, 7, 12, 9, 11, & 10).

Perhaps the most interesting finding through this series of reports was the presence of apparently heterozygous isolates, along with the corresponding homozygous isolates (Meloni *et al.*, 1995). In addition, many of the cloned isolates which exhibited multiple banding patterns did so for more than one enzyme. If these banding patterns are a true reflection of the underlying genotypes of the isolates, then they would seem to indicate that *G. intestinalis* is functionally diploid, and that recombination or sexual reproduction must have occurred at some stage to produce the apparent heterozygotes (Meloni *et al.*, 1995; Tait, 1983). The degree of heterozygosity observed among isolates of *G. intestinalis* was not as extensive as that observed among species of the asexual protozoan genus *Naegleria*, in which heterozygosity was detected in 60 of 62 isolates examined (Adams *et al.*, 1989).

Comparison of the published enzyme profiles from Meloni *et al.* (1988) and Meloni *et al.* (1992) seem to indicate that zymodemes M4, M5 and M6 have had slight changes in their profiles for the enzymes EST and 6PGD, but the possible reasons for this were not discussed. The placement of some isolates within zymodemes was also sensitive to the number of enzyme systems employed, eg., the cat derived isolate BAC3 was placed within the zymodeme M4 based upon analysis of 10 enzyme systems. However, it was moved to a separate zymodeme (M35) after analysis of 13 enzyme systems. This change is particularly significant considering M4 also contains human-derived *G. intestinalis*, and that the detection of isolates with the same zymodeme infecting different host species has been used as evidence to support zoonosis (eg., Meloni *et al.*, 1995).

Although in many ways comprehensive, the studies by Meloni *et al.* (1988, 1989, 1992, 1995) are flawed by their attempts to interpret their zymodeme data genetically. The zymodeme approach does not make a genotypic interpretation of the data, therefore the application of genotypic distance measures to data analysed in such a fashion would be inappropriate (eg., Meloni *et al.*, 1995). Furthermore, earlier analyses (Meloni *et al.*, 1988; Meloni *et al.*, 1992) used shared presence/absence of enzyme bands as a character rather than a character state for a given enzyme system, raising questions as to independence of the characters used. This was partly addressed by Meloni *et al.* (1995), who proposed the use of the "fixed enzyme difference" measure, which was defined as the failure of 2 zymodemes to share a pattern for a particular enzyme system. Contrary to the claim by Meloni *et al.* (1995), their distance measure is not identical to the fixed genetic difference measure of Andrews *et al.* (1989). As depicted in Figure 7 (p. 63), a fixed genetic difference occurs when a pair of samples do not have an allele in common, which is quite distinct from failing to share a

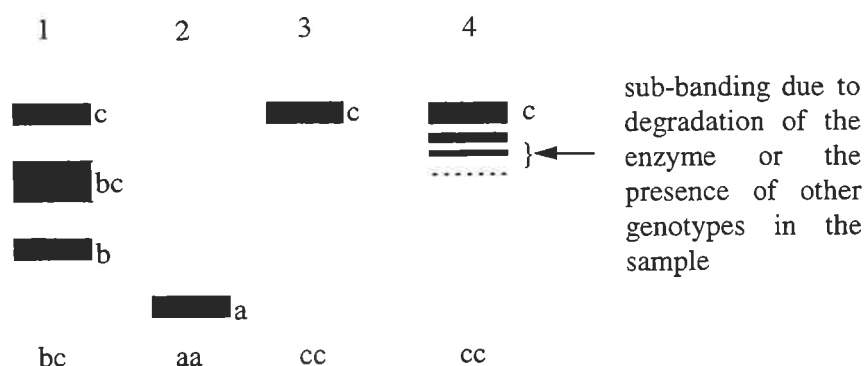


Figure 7. Hypothetical staining patterns of four isolates for a dimeric enzyme encoded at a single locus. The allozyme interpreted genotypes, which would exclude the sub-banding from analysis, are indicated at the bottom of each track. The fixed difference measure of Andrews *et al.* (1989) would indicate a fixed difference between samples 1 & 2, 2 & 3, and 2 & 4 but none between 1 & 3, 1 & 4 and 3 & 4 (they all share allele c). In comparison, the measure used by Meloni *et al.* (1995) would find a “fixed enzyme difference” between all of the samples including 3 and 4 (which have the same genotype) because none of the samples have identical patterns.

banding pattern for a particular enzyme. In effect, the measure of Meloni *et al.* (1995) will estimate a larger distance compared to the measure of Andrews *et al.* (1989) when the same data are analysed. Furthermore, the measure employed by Meloni *et al.* (1995) may still violate the assumptions of character homology and independence, which are inherent assumptions in all forms of genetic analysis (Swofford *et al.*, 1996). The different distance measures employed (eg., Meloni *et al.*, 1988 versus Meloni *et al.*, 1995) did not appreciably affect the inferred relationships or composition of the zymodemes.

3.3 Previous electrophoretic studies using allozyme analysis

There have been two main studies which have used AE to genetically characterise isolates of *G. intestinalis*, the first by Andrews *et al.* (1989) and the second by Mayrhofer *et al.* (1995). These studies utilised 26 and 27 enzyme loci respectively, twice the number of enzyme systems examined by Meloni *et al.* (1995). AE has also been used as a tool to detect samples comprising mixtures of genotypes which exhibit genotype-specific selection in different culture conditions (Andrews *et al.*, 1992b). The original report of Andrews *et al.* (1989) identified four discrete groups of isolates which exhibited fixed differences at 23% - 73% of the loci examined. The levels of difference separating these genetic groups (I - IV) were sufficient to suggest that they might represent cryptic (ie., morphologically identical)

species and that *G. intestinalis* may be a species complex. Genetic group I comprised 20 stocks (isolates) and 47 cloned lines which were monomorphic, while group II, which comprised 7 stocks and 1 cloned line, was polymorphic at 3 loci. The other two genetic groups were each represented by only single isolates. The group III isolate, BAH12 (uncloned), had a multiple banded phenotype at 8 of the 26 loci examined, and the banding patterns were consistent with those expected for a diploid heterozygote. This has also been observed by Meloni *et al.* (1988, 1989, 1992, 1995) for the same isolate.

The study of Mayrhofer *et al.* (1995) consolidated the results of Andrews *et al.* (1989). Additional *G. intestinalis* isolates that had been propagated by infection of suckling mice, as well as an isolate of *G. muris*, were characterised. The findings (Figure 8, p. 65) were fully consistent with the earlier study (Andrews *et al.*, 1989). All human-derived isolates were found to belong to either one of two genetic assemblages. The first, Assemblage A, comprised groups I, II and other related genotypes, and the second, Assemblage B, comprised groups III, IV and numerous other related genotypes. These assemblages were separated by fixed allelic differences at over 60% of loci, while the cat isolate of *G. intestinalis* (Ad-23) exhibited approximately 75% fixed differences compared to the human-derived isolates. *G. muris* exhibited fixed allelic differences at 84% of the loci examined when compared with the *G. intestinalis* isolates.

It is interesting to note that, with the exception of two isolates, none of the new isolates characterised by Mayrhofer *et al.* (1995) were identical either to each other or the representatives of groups I - IV (Figure 8, p. 65). Surprisingly, none of the new isolates were of the group I genotype. Seven isolates clustered with group II, but diverged by 20% - 30% fixed differences. All of the 22 remaining human-derived isolates were within Assemblage B, and exhibited fixed differences at 0 - 30% of the loci examined. Only two of the isolates were identical (Ad-38 & Ad-32), a marked contrast to the homogeneity previously observed within group I by Andrews *et al.* (1989). Similar observations have been made by Meloni *et al.* (1995), who found that the majority of zymodemes were represented by only single isolates, but that three zymodemes comprised 37 of the 97 isolates examined. These 3 zymodemes, M1, M3, and M4, appear to be equivalent to genetic groups I and II based upon the characterisation of the same isolates by Andrews *et al.* (1989) and Meloni *et al.* (1995). These results suggest that these groupings may represent recently dispersed, widespread clones (Mayrhofer *et al.*, 1995).

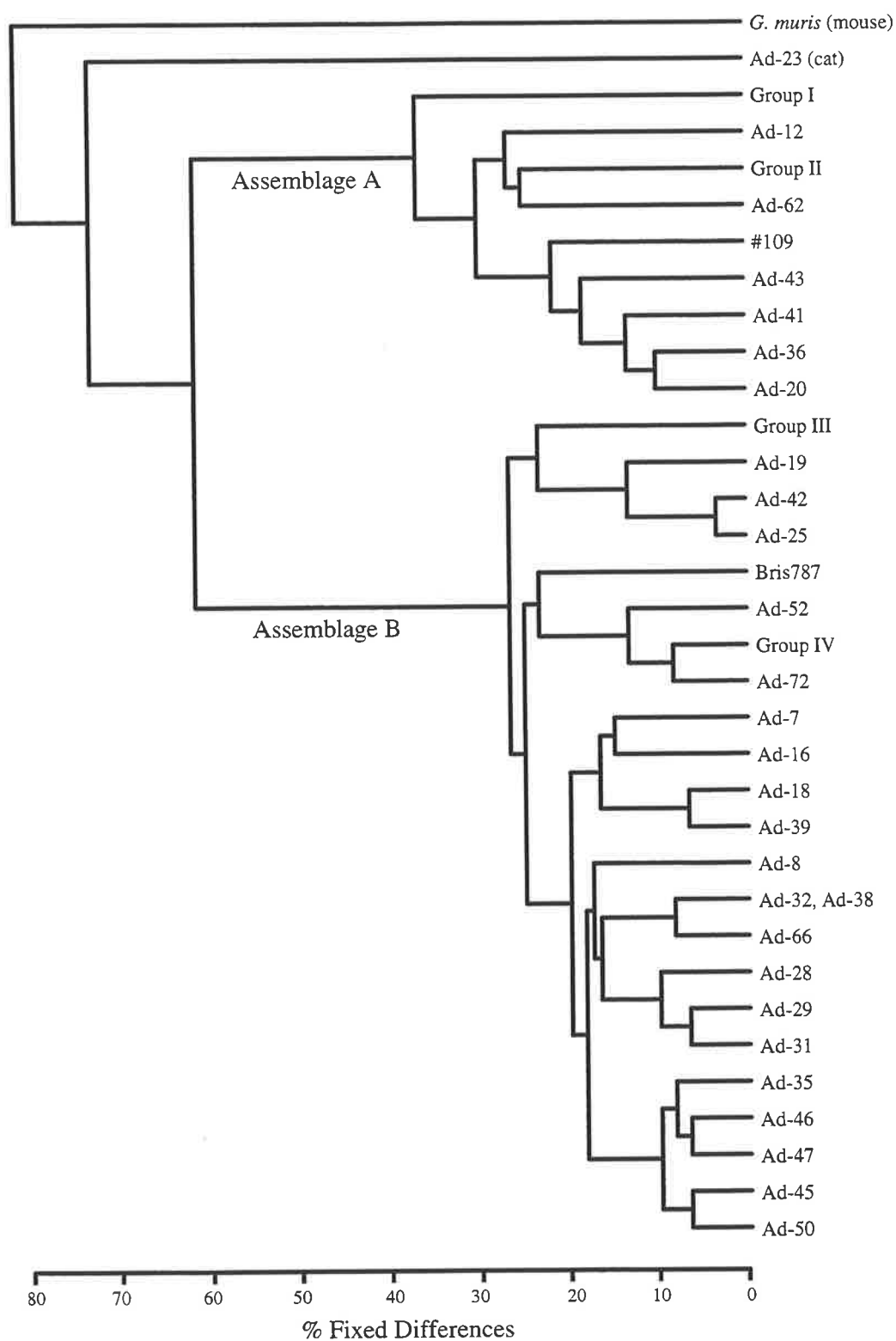


Figure 8. Dendrogram of genetic relationships among *G. intestinalis* isolates determined by allozyme analysis using 27 enzyme loci (from Mayrhofer *et al.*, 1995). Isolates were of human origin except where indicated otherwise.

3.4 Genetic characterisation of isolates using allozyme electrophoresis

The two published AE studies discussed in section 3.3 both concentrated upon isolates of *G. intestinalis* from humans. Animal derived *G. intestinalis* have also been established by growth in suckling mice, but with the exception of a cat derived isolate (Ad-23), their position within the current model of genetic relationships is unknown. Preliminary work conducted at to the commencement of this thesis (R. Andrews, pers. comm. 1993) suggested that some animal derived isolates formed genetic groups external to those originally identified by Andrews *et al.* (1989), but a comprehensive genetic analysis examining all of these isolates was still required. The aim of this section of work was to conduct a comprehensive genetic analysis of isolates of *G. intestinalis* derived from different host species using the technique of AE. Representatives of Assemblages A and B were chosen to cover the spectrum of variation currently discerned within each assemblage. Isolates of the morphologically distinct species *G. muris* and *G. ardeae* were used as outgroups. The genetic relationships inferred by this analysis were then used as a basis for further investigations using nucleotide sequence analyses.

Although allozyme electrophoresis has a demonstrated utility for comparing isolates of *G. intestinalis* (Andrews *et al.*, 1989; Mayrhofer *et al.*, 1995), its utility comparing morphologically defined species has not been tested, apart from a comparison of *G. intestinalis* isolates with a single *G. muris* isolate (Mayrhofer *et al.*, 1995). Comparison of published rRNA sequences from *G. intestinalis*, *G. muris*, and *G. ardeae* (van Keulen *et al.*, 1993) suggests that *G. ardeae* is more similar to *G. intestinalis* than *G. muris*. The evolutionary relationships of these species have not been determined using morphological characters. Although *G. ardeae* trophozoites have a similar shape to *G. intestinalis*, the shape of the nuclei is different (Erlandsen *et al.*, 1990a). The presence of either rounded or claw-hammer median bodies within trophozoites of *G. ardeae* provides conflicting phylogenetic information, and questions the reliability of the median body morphology as a diagnostic marker (Erlandsen *et al.*, 1990a). *G. ardeae* had not been characterised by AE previously, and it was of interest to determine how it was related to the other morphological species, especially in view of the characters that it shared with both *G. intestinalis* and *G. muris*.

3.4.1 Specific aims

- To characterise new isolates.
- To integrate these results with previous allozyme studies.
- To compare the new isolates with previously characterised isolates.
- To compare the three morphologically distinct species *G. intestinalis*, *G. muris*, and *G. ardeae*.
- To determine the levels of divergence which separate the different species.
- To determine the maximum resolution of AE.
- To identify isolates for further molecular characterisation.

Hypotheses to test:

- *G. ardeae* is genetically intermediate between *G. intestinalis* and *G. muris*.
- *G. intestinalis* has a clonal population structure.

Null Hypothesis:

- All of the new *G. intestinalis* isolates are genetically the same as the previously characterised isolates.

3.4.2 Materials and methods

3.4.2.1 Source of isolates

The isolates of *Giardia* examined in this study are listed in Table 11 (p. 68). The isolate WOOF was provided by Dr. Judy Isaac-Renton (Department of Pathology, University of British Columbia), the isolate P-15 was provided by Prof. Jaroslav Kulda (Department of Parasitology, Charles University), and the *G. ardeae* isolate was provided by Prof. Stan Erlandsen (Department of Cell Biology and Neuroanatomy, University of Minnesota Medical School). The isolates bearing the prefix Ad were from a collection established by Assoc. Prof. Graham Mayrhofer (Department of Microbiology and Immunology, University of Adelaide). The majority of the animal-derived faecal samples used to establish isolates were provided by Mr. Michael O'Callaghan, Dr. Peter O'Donohue and Dr. Ian Carmichael (VETLAB, Primary Industries SA), and were all from clinical presentations. The establishment of these isolates in culture by propagation in suckling-mice, as well as any subsequent culturing, was performed in Assoc. Prof. Mayrhofer's laboratory.

Table 11. List of isolates characterised by allozyme electrophoresis.

Species & Isolate	Host	Location	Culture	Analysed herein	Characterised previously by:
<i>G. intestinalis</i>					
Ad-1, Ad-2	human	Adelaide, SA	axenic	yes	Andrews <i>et al.</i> (1989)
Ad-12	human	Whyalla, SA	s/m		Mayrhofer <i>et al.</i> (1995)
Ad-62	human	Alice Springs, NT	s/m		Mayrhofer <i>et al.</i> (1995)
BAH12	human	Wyndham, WA	axenic	yes	Andrews <i>et al.</i> (1989)
Ad-19, Ad-28	human	Adelaide, SA	s/m	yes	Mayrhofer <i>et al.</i> (1995)
BRIS/694	human	Brisbane, Qld	axenic		Andrews <i>et al.</i> (1989)
Ad-45	human	Adelaide, SA	s/m	yes	Mayrhofer <i>et al.</i> (1995)
Ad-52	human	Australia	s/m	yes	Mayrhofer <i>et al.</i> (1995)
Ad-22	human	Mt. Gambier, SA	s/m		Mayrhofer <i>et al.</i> (1995)
Ad-156	marmoset	Adelaide, SA	s/m	yes	R. Andrews, unpub. results
Ad-158	siamang	Adelaide, SA	s/m	yes	R. Andrews, unpub. results
Ad-23	cat	Adelaide, SA	s/m	yes	Mayrhofer <i>et al.</i> (1995)
Ad-142, Ad-154				yes	R. Andrews, unpub. results
Ad-131	cat	Mt. Gambier, SA	s/m	yes	R. Andrews, unpub. results
BAC1	cat	Shenton Park, WA	axenic	yes	Meloni <i>et al.</i> (1988)
BAC2, BAC3	cat	Murdoch, WA	axenic	yes	Meloni <i>et al.</i> (1988)
WOOF	dog	Canada	axenic	yes	Baruch <i>et al.</i> , 1996
D3	dog	Calgary, Canada	s/m	yes	Uji <i>et al.</i> (1988)
Ad-148	dog	Ceduna, SA	s/m		Monis <i>et al.</i> (1997) Appendix E
Ad-140, Ad-141, Ad-138	dog	Port Augusta, SA	s/m		Monis <i>et al.</i> (1997) Appendix E
Ad-147, Ad-145	dog	Alice Springs, NT	s/m		Monis <i>et al.</i> (1997) Appendix E
Ad-143, Ad-137	dog	Port Augusta, SA	s/m	yes	Monis <i>et al.</i> (1997) Appendix E
Ad-139	dog	Alice Springs, NT	s/m	yes	Monis <i>et al.</i> (1997) Appendix E
Ad-136, Ad-146	dog	Alice Springs, NT	clinical	yes	Monis <i>et al.</i> (1997) Appendix E
Ad-116, Ad-112, Ad-109	guinea pig	Adelaide, SA	s/m		R.H. Andrews, unpub. results
Ad-127	sheep	Barmera, SA	s/m	yes	Ey <i>et al.</i> (1997) Appendix F
Ad-152, Ad-153, Ad-134	sheep	Adelaide, SA	s/m	yes	Ey <i>et al.</i> (1997) Appendix F
Ad-144	calf	SA	s/m	yes	Ey <i>et al.</i> (1997) Appendix F
Ad-119, Ad-133	calf	Adelaide, SA	s/m		Ey <i>et al.</i> (1997) Appendix F
Ad-132	pig	Adelaide, SA	s/m		Ey <i>et al.</i> (1997) Appendix F
P-15	pig	Czech Republic	axenic	yes	Koudela <i>et al.</i> (1991)
Ad-149	alpaca	S America?	s/m		R. Andrews, unpub. results
Ad-155, Ad-157	rat	Adelaide, SA	s/m	yes	R. Andrews, unpub. results
Ad-130	<i>N. alexis</i>	Adelaide, SA	s/m		R. Andrews, unpub. results
<i>G. ardeae</i>	blue heron	Minnesota, USA	axenic	yes	Erlandsen <i>et al.</i> (1990b)
<i>G. muris</i>					
Underdown	hamster	Canada	s/m	yes	Roberts-Thomson <i>et al.</i> (1976)
Ad-120	mouse	Adelaide, SA	s/m	yes	Mayrhofer <i>et al.</i> (1995)
Ad-150	mouse	Adelaide, SA	s/m		R. Andrews, unpub. results
Ad-151	mouse	Adelaide, SA	s/m	yes	R. Andrews, unpub. results

key: s/m = propagation by infection of suckling mouse
 SA = South Australia
 NT = Northern Territory

? = uncertain
 WA = Western Australia
 Qld = Queensland

by Mr. John Mackrill (Department of Microbiology and Immunology, Adelaide University). Samples of dog small intestine were collected post mortem (kindly provided by Michael O'Callaghan & Ian Carmichael) and chilled for transport to the laboratory. Trophozoites were purified from these and grown in suckling mice as described in section 2.2.3. A full description of the propagation of the dog-derived isolates is given in Monis *et al.* (1997, Appendix E). Uninfected mouse gut was used as a control to check for host enzymes detected by AE in samples propagated in suckling mice. Axenic isolates were cultured at 37°C in modified TYI-S-33 medium as described in section 2.2.2.

3.4.2.2 Preparation and electrophoresis of samples

Samples were stored at -70°C prior to sonication. Sonicates were prepared from each isolate by adding an equal volume of cold lysing solution (100 µl of β-mercaptoethanol, 10 mg NADP, 100 ml of distilled water; Richardson, *et al.*, 1986) to the thawed isolate, sonicating using short low powered pulses, and then centrifuging at 5000 g for 10 min at 4°C. The resulting supernatants were stored at -20°C prior to use as separate aliquots in microcapillary tubes. To maximise the number of enzyme loci scored for each isolate, allozyme electrophoresis was performed on cellulose acetate ("Cellogel", Chemetron, Milan) as the support medium according to the methods described by Richardson *et al.* (1986). The running conditions were described in detail by Andrews *et al.* (1989) and Andrews *et al.* (1993). Electrophoresis was performed in a cold room at 4°C for 1½ - 1¾ hours at 200 V. Minimal degradation was noted from sonicates which had been stored for at least 3 years at -20°C, although the enzyme activities of these samples were lower compared with earlier analyses using the same materials.

3.4.2.3 List of enzyme loci

A total of 23 enzymes gave sufficient staining intensity and resolution for accurate allozymic (genetic) interpretation. In the case of diaphorase and purine nucleoside phosphorylase, there were 2 presumptive enzyme loci, raising the number of enzyme loci examined to 25. These enzymes, abbreviations and Enzyme Commission (EC) numbers are as follows: acid phosphatase (ACP; EC 3.1.3.2), aldolase (ALD; EC 4.1.2.13), diaphorase (DIA; 1.6.2.2), enolase (ENOL; EC 4.2.1.11), esterase (EST; EC 3.1.1.1), fructose-1,6-diphosphatase (FDP; EC 3.1.3.11), glutamate dehydrogenase (GDH; EC 1.4.1.3), aspartate aminotransferase (syn. glutamate-oxaloacetate transaminase, GOT; EC 2.6.1.1), glucose-6-phosphate dehydrogenase

(G6PD; EC 1.1.1.49), glycerol-3-phosphate dehydrogenase (GPD; EC 1.1.1.8), glucose-phosphate isomerase (GPI; EC 5.3.1.9), alanine aminotransferase (syn. glutamate-pyruvate transaminase, GPT; EC 2.6.1.2), hexokinase (HK; EC 2.7.1.1), malate dehydrogenase (MDH; EC 1.1.1.37), malic enzyme (ME; EC 1.1.1.40), nucleoside diphosphate kinase (NDPK; EC 2.7.4.6), purine nucleoside phosphorylase (NP; EC 2.4.2.1), peptidase (valine-leucine) (PEP-A; EC 3.4.13.11), peptidase (phenylalanine-proline) (PEP-D; EC 3.4.13.9), phosphoglycerate mutase (PGAM; EC 5.4.2.1), 6 phosphogluconate dehydrogenase (6PGD; EC 1.1.1.44), phosphoglycerate kinase (PGK; EC 2.7.2.3), phosphoglucomutase (PGM; EC 5.4.2.2) and triose phosphate isomerase (TPI; EC 5.3.1.1) (Manchenko, 1994).

3.4.2.4 Allozyme analysis

Following histochemical staining, the electrophoretic banding patterns for each isolate were scored alphabetically in order of increasing anodal migration (a, b, c, etc.). A variety of genetic distance measures were used (discussed below). These distance measures were originally proposed for use in population genetics, particularly for the comparison of populations using gene frequency data (Richardson *et al.*, 1986), and they have also been routinely used for the comparison of individual isolates (eg., Adams *et al.*, 1989; Andrews *et al.*, 1988; Selander *et al.*, 1986).

The simplest measure is the proportion of enzyme loci exhibiting fixed genetic (allelic) differences (Richardson *et al.*, 1986), which has been used previously for the genetic characterisation of *Giardia* isolates (eg., Andrews *et al.*, 1989; Mayrhofer *et al.*, 1995), other protozoan parasites (eg., *Leishmania*: Andrews *et al.*, 1988; *Naegleria*: Adams *et al.*, 1989) and other parasite taxa (eg helminths: Andrews & Beveridge, 1990; Chilton *et al.*, 1992 and arthropods: Andrews *et al.*, 1992a). A fixed genetic difference occurs at a locus when the samples being compared fail to share any alleles at that locus. The use of fixed genetic differences as a genetic distance measure between samples, particularly in delineating species and assessing population substructuring, has been shown to correlate with other genetic distance measures such as Nei's D (Richardson *et al.*, 1986). In the case of haploid organisms, or where heterozygosity is absent, this measure is equivalent to the p-distance used in protein or nucleotide sequence comparisons, and may be the most appropriate measure for comparing individual samples.

Nei's genetic distance (Nei, 1972, 1978) was not generally used because it is sensitive to unequal rates of change between loci and lineages (Swofford *et al.*, 1996), which may be the case for the data generated in this study as indicated by the large differences in the levels of polymorphisms among some of the loci examined (eg., 3 alleles for P_{gk} compared to 9 for Me). This genetic distance measure is also affected by intra-taxon heterozygosity (see next paragraph).

Rogers' distance (Rogers, 1972) is another simple measure which is essentially the mean Euclidean distance between allele frequencies at each locus (Swofford *et al.*, 1996). This measure is sensitive to intra-taxon heterozygosity, and will underestimate the distance between taxa when either one is polymorphic, even though they may fail to share any alleles. The chord distance of Cavalli-Sforza & Edwards (1967) is also a Euclidean measure, but it is not sensitive to intra-taxon heterozygosity (Swofford *et al.*, 1996). Chord distances were calculated using the GENDIST program of PHYLIP. Fixed differences, Rogers' distances and allele frequencies were calculated using the BASIC program "pcordrog" (Mark Adams, South Australian Museum). Dendrograms were constructed using either Neighbour Joining or UPGMA as implemented by the NEIGHBOUR program of PHYLIP.

3.4.2.5 Calculating the Index of Association (a test for linkage disequilibrium)

The Index of Association (I_A) was calculated as per the method described by Gordon *et al.* (1995). The equations were originally derived by Brown *et al.* (1980), and refined by Maynard Smith *et al.* (1993). Essentially it is a comparison of the observed and expected variances of the mean number of loci which are different between two taxa. The expected values can be calculated from the allele frequency data. The calculations are as follows:

Consider m loci analysed for n samples. The frequency of the i th allele at the j th locus is p_{ij} . The probability that two samples differ at the j th locus is h_j , which can be calculated by:

$$h_j = 1 - \sum p_{ij}^2$$

K is the mean number of loci different between two sample, calculated using:

$$K = \sum h_j$$

In the absence of linkage disequilibrium, the variance of K , V_E , can be calculated by:

$$V_E = \sum h_j(1 - h_j)$$

The error variance of V_E , $\text{var}(V_E)$, can be calculated by:

$$\text{var}(V_E) = 1/n \left\{ \sum h_j - 7 \sum h_j^2 + 12 \sum h_j^3 - 6 \sum h_j^4 + 2 \left[\sum h_j(1-h_j) \right]^2 \right\}$$

The upper 95% confidence limit (L) for V_E is given by:

$$L = V_E + 2 \left\{ \text{var}(V_E)^{1/2} \right\}$$

The Index of Association is calculated by:

$$I_A = V_O / V_E - 1$$

V_O can be calculated from the observed number of differences for all pairwise comparisons.

The expected value for I_A is zero if there is no linkage disequilibrium. I_A is not significantly different from zero at the 0.05 level of significance if V_O is less than L .

3.4.3 Results and discussion

The isolates examined in this study are listed in Table 11 (p. 68). The observed banding patterns were completely consistent with the earlier studies (Andrews *et al.*, 1989, Mayrhofer *et al.*, 1995 and Andrews unpublished), and no significant degradation was observed for samples or sonicates which had been stored for more than 3 years at -70°C or -20°C . The data from the present study were combined with the data from earlier work (kindly provided by Dr. Ross Andrews) through the use of carry-over controls, and the deduced allelic profiles are presented in Table 12 (pp. 73 - 76).

3.4.3.1 Comparison of relationships inferred using different distance measures

As discussed earlier, Rogers' and Nei's distance measures may be affected by intra-taxon heterozygosity. Although only one isolate (BAH12) appeared to be heterozygous, it would be of interest to see how this would effect the recovered trees. While the chord distance appears to be a more realistic distance measure (Swofford *et al.*, 1996), a practical limitation is that samples with incomplete data must be excluded to enable the calculation of distances by the GENDIST program. To satisfy this requirement, a restricted data set (Table 13, p. 77) was generated for the 23 enzyme loci using isolates which could be scored at all of the loci, and the percentage fixed differences, chord, Rogers' and Nei's distances were calculated. Dendrograms of the inferred relationships were constructed using the Neighbour program of PHYLIP, using both the NJ and UPGMA algorithms, and are presented in Figures 9-12 (pp. 78-81).

Table 12. Allelic profiles of *Giardia* isolates determined by allozyme electrophoresis.

Isolate	Genotype	Enzyme Locus										
		<i>Acp</i>	<i>Ald</i>	<i>Dia-1</i>	<i>Dia-2</i>	<i>Enol</i>	<i>Est</i>	<i>Gdh</i>	<i>Got-1</i>	<i>G6pd</i>	<i>Gpt</i>	<i>Hk</i>
Ad-1	I, A	e	e	c	c	e	d	b	b	c	e	e
D3		e	e	f	c	e	-	b	b	c	e	e
Ad-149		-	g	c	c	e	d	c	b	c	f	e
Ad-109		e	-	-	-	e	-	b	-	c	f	e
Ad-112		e	-	-	-	e	-	b	-	c	f	e
Ad-116		f	g	i	e	-	d	c	c	c	f	e
Ad-127		f	g	b	c	eg	-	b	b	c	ef	de
Ad-134		f	e	c	a	g	b	c	b	b	e	d
Ad-152		f	e	c	a	g	d	c	b	b	e	d
Ad-153		f	e	c	a	g	e	c	b	b	e	d
Ad-119		e	e	b	b	g	d	c	c	-	e	d
Ad-133		f	e	b	a	g	d	c	c	b	e	d
Ad-144		d	e	b	b	g	d	c	c	b	e	d
Ad-132		e	-	a	d	g	d	e	e	c	e	d
P-15		f	e	a	d	g	d	e	e	b	e	d
BAC1		e	g	c	c	e	d	c	b	c	f	e
BAC2		e	h	c	c	e	-	c	b	c	g	e
BAC3		e	h	c	c	e	-	c	b	c	g	e
Ad-2	II, A	e	e	d	c	e	d	b	b	c	f	e
Ad-12		e	e	-	-	e	d	b	b	d	f	e
Ad-62		e	e	-	-	e	d	b	b	c	f	g
Ad-23		f	g	-	-	e	a	a	c	a	e	e
Ad-131		f	g	i	d	e	a	a	c	a	d	e
Ad-142		f	g	h	c	e	a	a	c	a	d	e
Ad-154		f	g	-	-	e	a	a	-	a	e	e

Table 12 (cont). Allelic profiles of *Giardia* isolates determined by allozyme electrophoresis.

Isolate	Genotype	Enzyme Locus											
		<i>Mdh</i>	<i>Me</i>	<i>Ndpk</i>	<i>Np-1</i>	<i>Np-2</i>	<i>Pep-A</i>	<i>Pep-D</i>	<i>Pgam</i>	<i>6Pgd</i>	<i>Pgk</i>	<i>Pgm</i>	<i>Tpi</i>
Ad-1	I, A	a	f	a	a	a	b	d	d	a	d	d	a
D3		a	f	a	a	a	b	d	b	a	d	d	a
Ad-149		a	e	a	a	a	a	d	d	a	d	c	a
Ad-109		a	-	-	-	-	-	-	d	-	-	c	-
Ad-112		a	e	-	-	-	-	-	d	-	-	c	-
Ad-116		a	e	a	a	a	-	d	d	a	d	c	a
Ad-127		a	e	a	a	a	a	d	b	a	d	c	a
Ad-134		a	e	a	b	c	b	d	d	a	d	c	c
Ad-152		b	e	a	b	c	b	d	d	a	d	c	c
Ad-153		b	e	a	b	c	b	d	d	-	-	c	a
Ad-119		a	e	a	b	c	b	-	d	a	-	-	c
Ad-133		a	e	a	b	c	b	d	c	-	-	-	-
Ad-144		a	e	a	b	c	c	d	d	a	-	c	c
Ad-132		a	e	a	b	c	b	d	a	a	-	c	-
P-15		a	e	a	b	c	b	d	d	a	d	-	a
BAC1		a	d	a	a	a	a	d	d	-	d	c	a
BAC2		a	f	a	a	a	b	e	d	a	d	c	a
BAC3		a	f	a	a	a	b	e	d	a	d	c	a
Ad-2	II, A	a	f	a	b	b	b	d	d	a	d	e	a
Ad-12		a	e	a	-	d	b	d	-	a	d	e	a
Ad-62		a	e	a	b	c	b	d	d	a	d	e	a
Ad-23		c	e	c	b	c	b	e	b	a	d	d	d
Ad-131		c	-	b	b	c	b	e	b	a	d	d	c
Ad-142		c	e	b	b	c	b	e	b	a	-	d	c
Ad-154		c	e	b	b	c	b	e	b	a	d	d	c

Table 12 (cont). Allelic profiles of *Giardia* isolates determined by allozyme electrophoresis.

Isolate	Genotype	Enzyme Locus										
		<i>Acp</i>	<i>Ald</i>	<i>Dia-1</i>	<i>Dia-2</i>	<i>Enol</i>	<i>Est</i>	<i>Gdh</i>	<i>Got-1</i>	<i>G6pd</i>	<i>Gpt</i>	<i>Hk</i>
BAH12	III, B	b	df	f	c	d	e	bc	c	f	d	g
Ad-19	B	a	g	f	-	d	d	b	b	f	d	f
BRIS/694	IV, B	b	-	f	b	d	d	b	be	f	d	g
Ad-22	B	b	g	f	b	d	d	b	-	e	d	f
Ad-28	B	b	g	f	b	d	d	b	d	e	d	f
Ad-45	B	b	g	f	b	d	d	b	d	e	d	f
Ad-52	B	b	g	f	b	d	d	b	d	f	d	g
Ad-156		b	g	f	b	d	e	b	c	e	c	g
Ad-158		b	g	-	-	d	e	b	c	f	d	g
WOOF		b	g	f	c	d	f	b	d	e	d	g
Ad-148		c	e	c	c	e	a	c	b	g	b	d
Ad-140		-	a	e	c	b	d	b	a	-	b	c
Ad-147		-	b	e	c	b	d	c	a	f	b	c
Ad-145		-	b	e	c	-	d	b	a	f	b	c
Ad-141		d	b	e	c	b	d	b	a	g	b	c
Ad-138		d	b	e	c	b	d	c	a	g	b	c
Ad-143		d	b	e	c	a	d	c	a	-	b	c
Ad-137		d	b	e	c	b	d	c	a	g	a	c
Ad-139		d	b	f	b	b	d	c	a	g	b	c
Ad-136		d	b	f	b	b	g	c	a	g	b	c
Ad-146		d	b	e	a	b	g	c	a	g	b	c
Ad-155		f	g	a	c	c	f	c	d	e	e	a
Ad-157		f	g	a	c	c	f	c	d	f	e	a
Ad-130		b	-	-	-	d	g	b	d	e	e	e
<i>G.ardeae</i>		b	h	f	d	d	g	c	f	c	g	c
Underdown	<i>G. muris</i>	f	i	g	d	f	i	d	e	c	h	a
Ad-120	<i>G. muris</i>	f	i	g	d	f	i	d	e	c	h	b
Ad-150	<i>G. muris</i>	f	i	g	d	f	h	d	e	c	h	b
Ad-151	<i>G. muris</i>	f	i	g	d	f	h	d	e	c	h	b

Table 12 (cont). Allelic profiles of *Giardia* isolates determined by allozyme electrophoresis.

Enzyme Locus													
Isolate	Genotype	Mdh	Me	Ndpk	Np-1	Np-2	Pep-A	Pep-D	Pgam	6Pgd	Pgk	Pgm	Tpi
BAH12	III, B	b	ac	a	d	d	a	e	bc	cd	b	bc	a
Ad-19	B	b	-	a	-	d	a	d	b	b	b	c	a
BRIS/694	IV, B	b	e	a	c	d	a	e	b	b	d	c	a
Ad-22	B	b	e	a	d	d	b	d	b	b	d	-	a
Ad-28	B	b	c	a	d	d	a	d	b	b	d	c	a
Ad-45	B	b	c	a	d	d	a	e	b	b	d	c	a
Ad-52	B	b	e	a	-	d	-	-	b	b	-	c	a
Ad-156		b	b	a	a	b	a	e	b	b	e	b	c
Ad-158		b	b	a	d	d	a	-	b	b	e	b	a
WOOF		b	b	a	d	d	a	e	b	b	e	-	a
Ad-148		e	g	a	b	b	b	c	b	c	b	a	c
Ad-140		e	g	a	-	-	c	c	b	b	-	a	c
Ad-147		e	i	a	-	-	c	c	b	b	b	a	c
Ad-145		e	i	a	-	-	c	c	b	b	b	a	c
Ad-141		e	f	a	e	e	c	c	b	b	b	a	c
Ad-138		e	c	a	e	e	c	c	b	b	b	a	c
Ad-143		e	c	a	e	e	c	c	b	b	b	a	c
Ad-137		e	f	a	e	e	c	c	b	b	b	a	c
Ad-139		e	i	a	e	e	c	c	b	b	b	a	b
Ad-136		d	h	a	e	e	a	a	b	c	b	a	b
Ad-146		d	h	a	f	f	d	a	b	c	b	a	b
Ad-155		b	f	a	g	g	c	e	b	a	e	b	a
Ad-157		a	f	a	g	g	c	e	b	a	e	b	a
Ad-130		a	-	a	e	d	-	-	a	c	-	b	a
<i>G. ardeae</i>		e	j	-	g	g	c	d	b	-	-	b	b
Underdown	<i>G. muris</i>	c	i	c	h	h	c	c	a	c	d	b	d
Ad-120	<i>G. muris</i>	c	g	c	h	h	c	c	a	c	d	b	d
Ad-150	<i>G. muris</i>	c	g	c	h	h	c	c	a	c	d	b	d
Ad-151	<i>G. muris</i>	c	g	c	h	h	c	c	a	c	d	b	d

Table 13. Restricted allozyme data set used in section 3.4.3.1.

Isolate	Enzyme Locus																						
	<i>Acp</i>	<i>Ald</i>	<i>Dia-1</i>	<i>Dia-2</i>	<i>Enol</i>	<i>Est</i>	<i>Gdh</i>	<i>Got-1</i>	<i>G6pd</i>	<i>Gpt</i>	<i>Hk</i>	<i>Mdh</i>	<i>Me</i>	<i>Ndpk</i>	<i>Np-1</i>	<i>Np-2</i>	<i>Pep-A</i>	<i>Pep-D</i>	<i>Pgam</i>	<i>6Pg</i>	<i>Pgk</i>	<i>Pgm</i>	<i>Tpi</i>
Ad-1	e	d	c	c	e	c	b	b	c	e	e	a	f	a	a	a	b	c	d	a	b	d	a
Ad-2	e	d	d	c	e	c	b	b	c	f	e	a	f	a	b	b	b	c	d	a	b	e	a
BAH12	b	ce	f	c	d	d	bc	c	f	d	g	b	ac	a	d	d	a	d	bc	cd	b	bc	a
Ad-28	b	f	f	b	d	c	b	d	e	d	f	b	c	a	d	d	a	c	b	b	b	c	a
Ad-45	b	f	f	b	d	c	b	d	e	d	f	b	c	a	d	d	a	d	b	b	b	c	a
Ad-156	b	f	f	b	d	d	b	c	e	c	g	b	b	a	a	b	a	d	b	b	c	b	c
Ad-148	c	d	c	c	e	a	c	b	g	b	d	e	g	a	b	b	b	b	b	c	a	a	c
Ad-141	d	b	e	c	b	c	b	a	g	b	c	e	f	a	e	e	c	b	b	b	a	a	c
Ad-138	d	b	e	c	b	c	c	a	g	b	c	e	c	a	e	e	c	b	b	b	a	a	c
Ad-137	d	b	e	c	b	c	c	a	g	a	c	e	f	a	e	e	c	b	b	b	a	a	c
Ad-139	d	b	f	b	b	c	c	a	g	b	c	e	i	a	e	e	c	b	b	b	a	a	b
Ad-136	d	b	f	b	b	f	c	a	g	b	c	d	h	a	e	e	a	a	b	c	a	a	b
Ad-146	d	b	e	a	b	f	c	a	g	b	c	d	h	a	f	f	d	a	b	c	a	a	b
Ad-152	f	d	c	a	g	c	c	b	b	e	d	b	e	a	b	c	b	c	d	a	b	c	c
Ad-134	f	d	c	a	g	b	c	b	b	e	d	a	e	a	b	c	b	c	d	a	b	c	c
P-15	f	d	a	d	g	c	e	e	b	e	d	a	e	a	b	c	b	c	d	a	b	c	a
Ad-155	f	f	a	c	c	e	c	d	e	e	a	b	f	a	g	g	c	d	b	a	c	b	a
Ad-157	f	f	a	c	c	e	c	d	f	e	a	a	f	a	g	g	c	d	b	a	c	b	a
Ad-120	f	h	g	d	f	h	d	e	c	h	b	c	g	c	h	h	c	b	a	c	b	b	d
Underdown	f	h	g	d	f	h	d	e	c	h	a	c	i	c	h	h	c	b	a	c	b	b	d
Ad-150	f	h	g	d	f	g	d	e	c	h	b	c	g	c	h	h	c	b	a	c	b	b	d
Ad-151	f	h	g	d	f	g	d	e	c	h	b	c	g	c	h	h	c	b	a	c	b	b	d

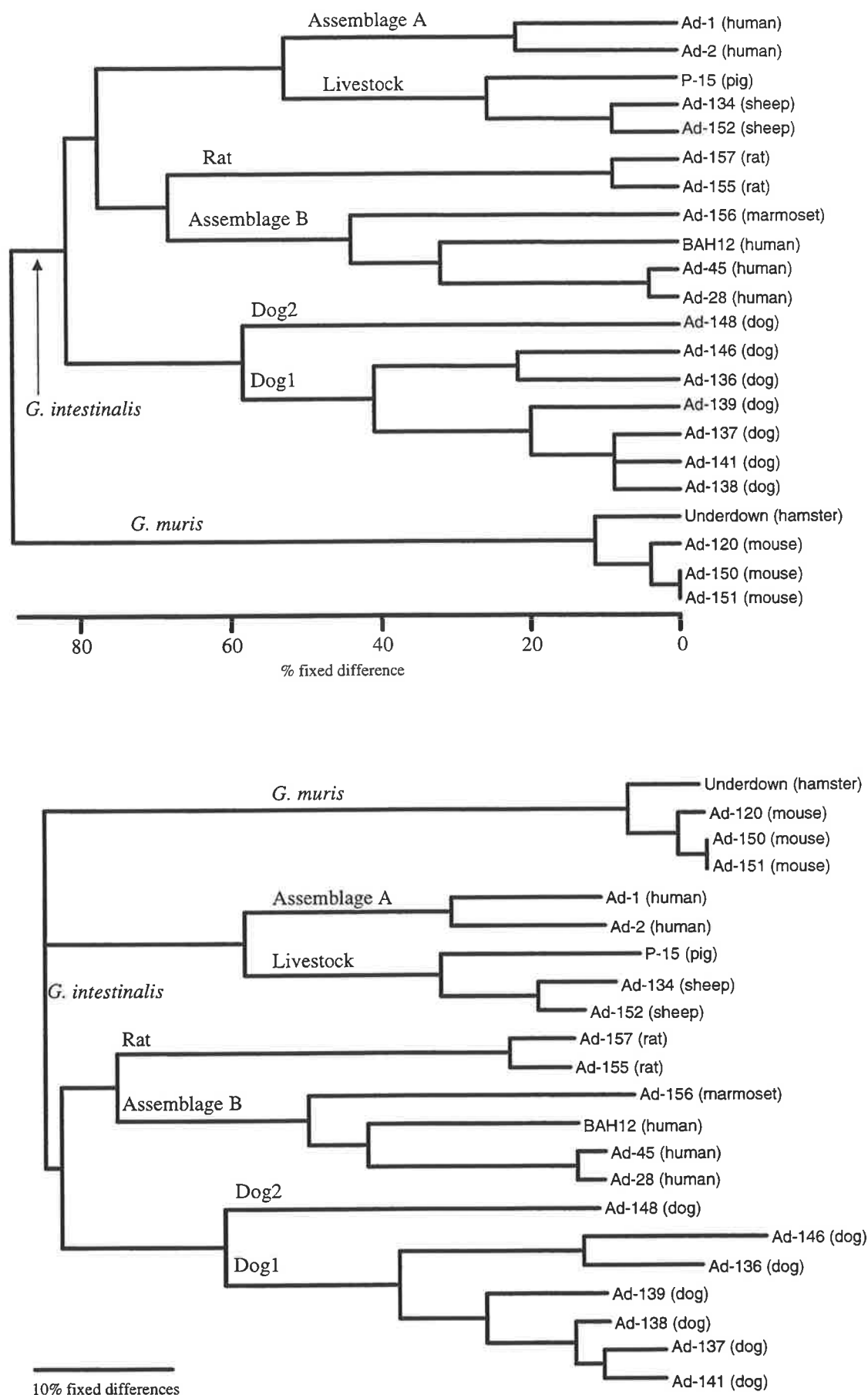


Figure 9. Dendrograms constructed by UPGMA (top) and NJ (bottom) using % fixed differences calculated from the data in Table 13 (p. 77). The host origin of each isolate is in parentheses.

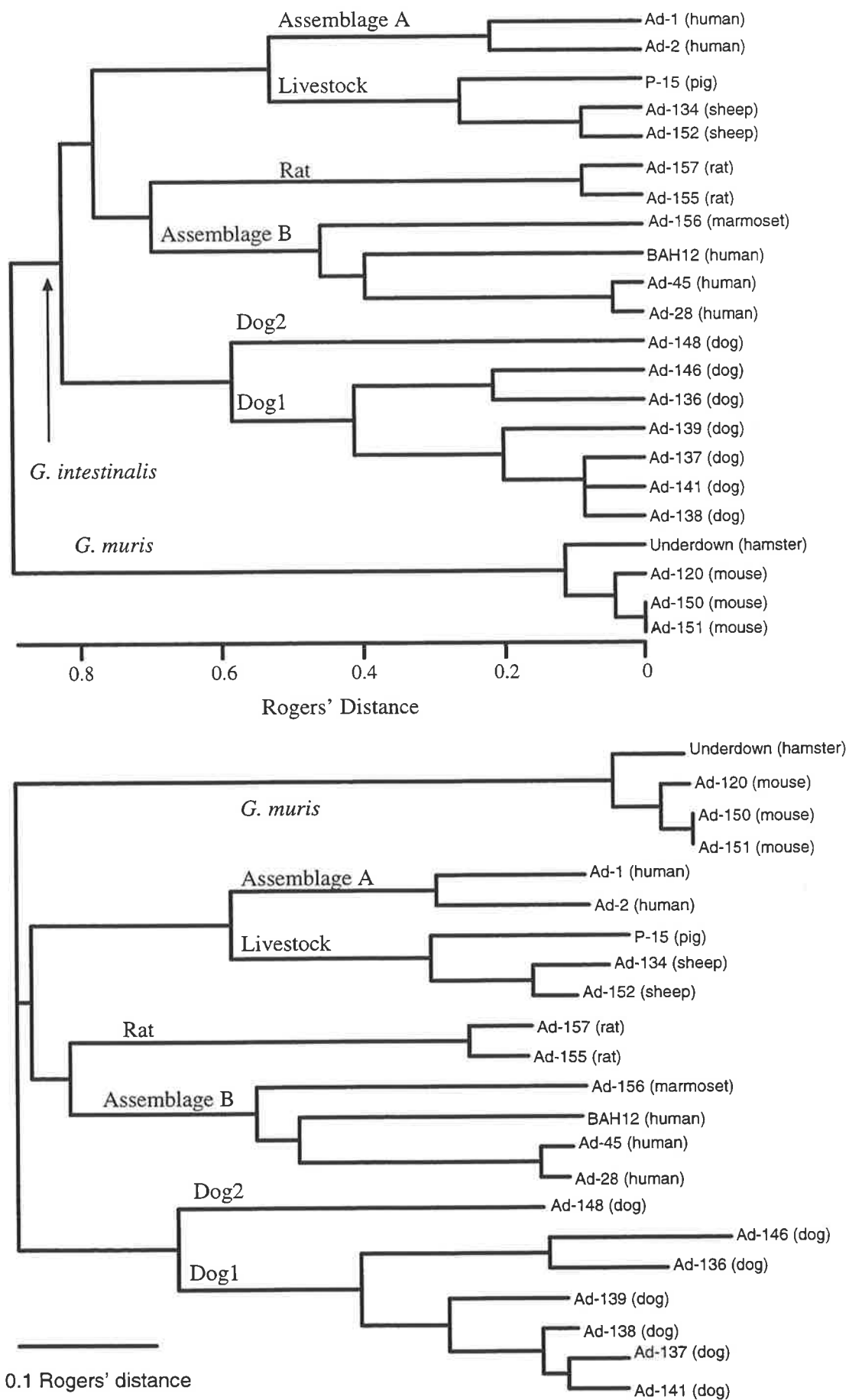


Figure 10. Dendrograms constructed by UPGMA (top) and NJ (bottom) using Rogers' distances calculated from the data in Table 13 (p. 77). The host origin of each isolate is in parentheses.

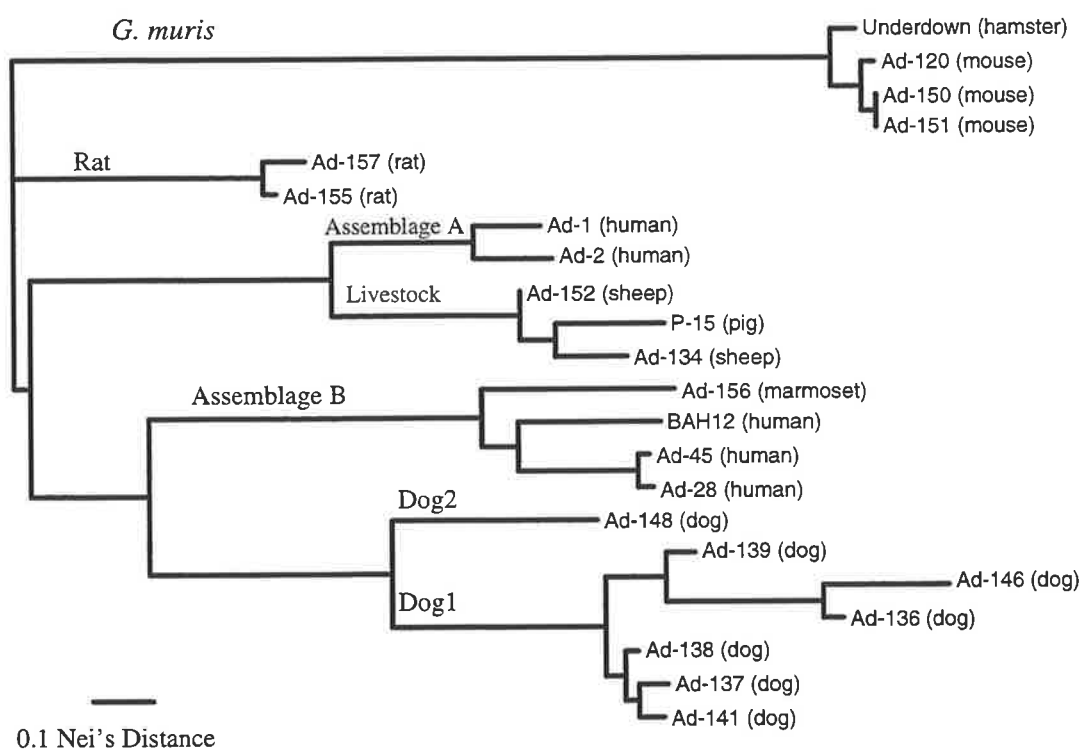
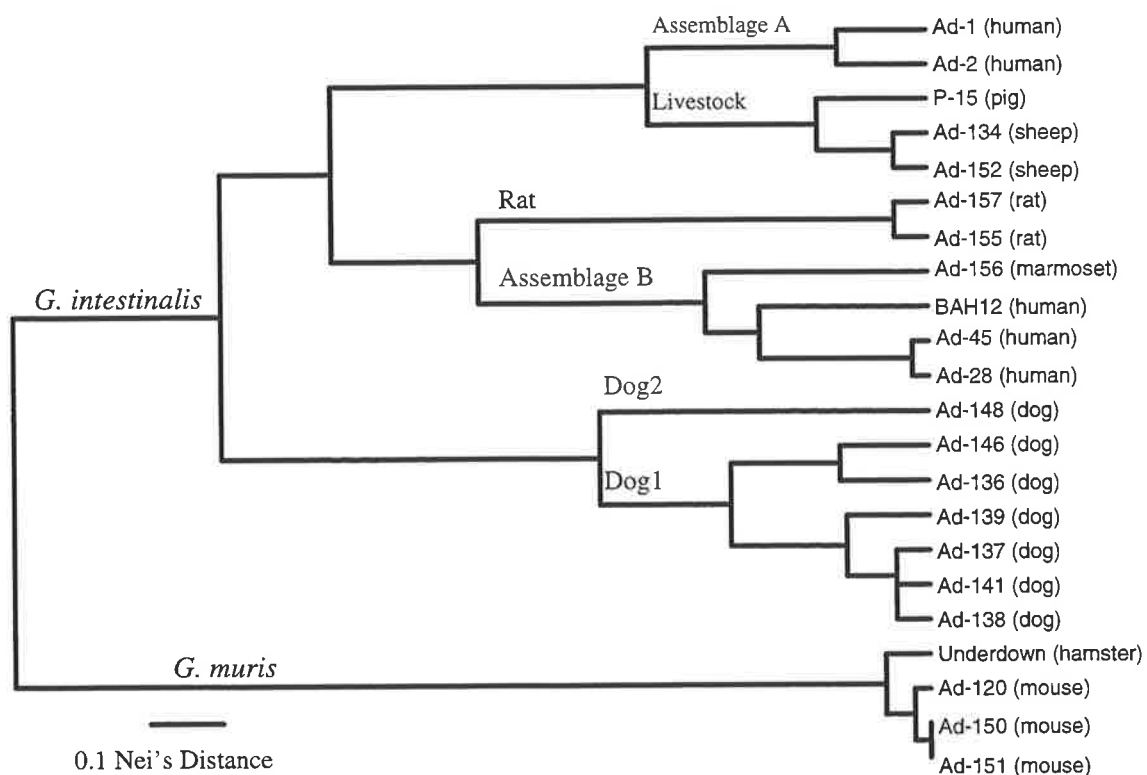


Figure 11. Dendrograms constructed by UPGMA (top) and NJ (bottom) using Nei's genetic distances calculated from the data in Table 13 (p. 77). The host origin of each isolate is in parentheses.

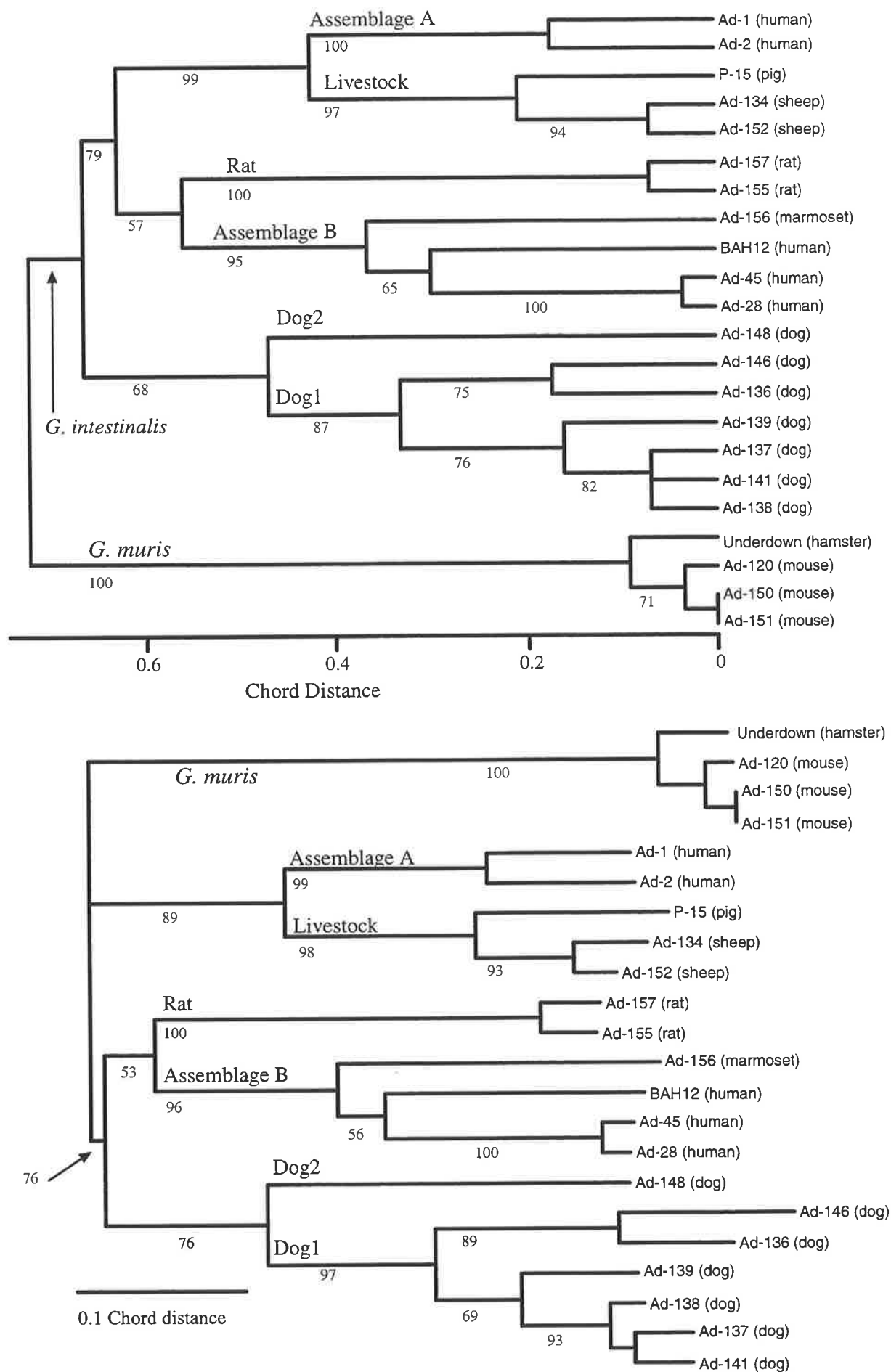


Figure 12. Dendrograms constructed by UPGMA (top) and NJ (bottom) using chord distances calculated from the data in Table 13 (p. 77). The host origin of each isolate is in parentheses. Numbers below branches indicate % bootstrap support from 500 replicates.

As shown in Figures 9 - 12 (p. 78 - 81), the topologies of the UPGMA dendrograms were identical irrespective of the distance measure used, although the branch lengths varied slightly in the different trees. The NJ analyses gave a slightly different result. The trees constructed using either the chord, fixed difference or Rogers' measures were essentially the same, although there was a slight difference between the basal topology of the chord distance NJ trees compared to the other two. The topology of the NJ tree constructed using Nei's D was different to the other trees, as were the relative lengths of some of the branches. This result may be explained by the fact that Nei's D are non-metric, as this can affect the accuracy of additive tree building methods such as NJ when using this distance measure (Swofford *et al.*, 1996). These results indicate that, with the exception of Nei's D, there is little real difference between the different distance measures, at least in the case of the present data set.

3.4.3.2. Analysis of isolates using complete data set

A detailed analysis was carried out using the allelic profiles given in Table 12 (pp. 73 - 76). The sheep-derived isolate Ad-127 was excluded because there was evidence from the banding patterns that it comprised a mixture of genotypes. Isolates which could not be scored at more than 18 loci were also excluded from the analysis. The percentage of fixed genetic differences and Rogers' distances were calculated for the pairwise comparisons of the remaining isolates. Genetic relationships were inferred by both UPGMA and NJ tree building methods, using *G. muris* and *G. ardeae* as outgroups for the NJ analysis. Each analysis was carried out using the NEIGHBOUR program with the "Jumble" option in effect and repeated 20 times to test for any bias caused by the sample input order.

The resulting dendrograms (Figures 13-16, pp. 83 - 86) clearly show the presence of 7 distinct clusters of *G. intestinalis* isolates, two of which correspond to the previously described Assemblages A and B (Mayrhofer *et al.*, 1995). In view of the observed asexual mode of reproduction of *Giardia*, it would seem reasonable to consider these clusters as distinct evolutionary lineages. With the exception of Assemblages A and B, the lineages appear to correlate with host origin. The composition and provisional names of the lineages are given in Table 14 (p. 87).

The dog-derived isolate Ad-148 was distinct from all of the other lineages, and it may represent another major group of isolates. Variation within each of the lineages ranged from 0 - 61% fixed differences (0 - 0.61 Rogers' distance), while that between lineages ranged from 35 - 96% fixed differences (0.35 - 0.95 Rogers' distance). The levels of variation within each

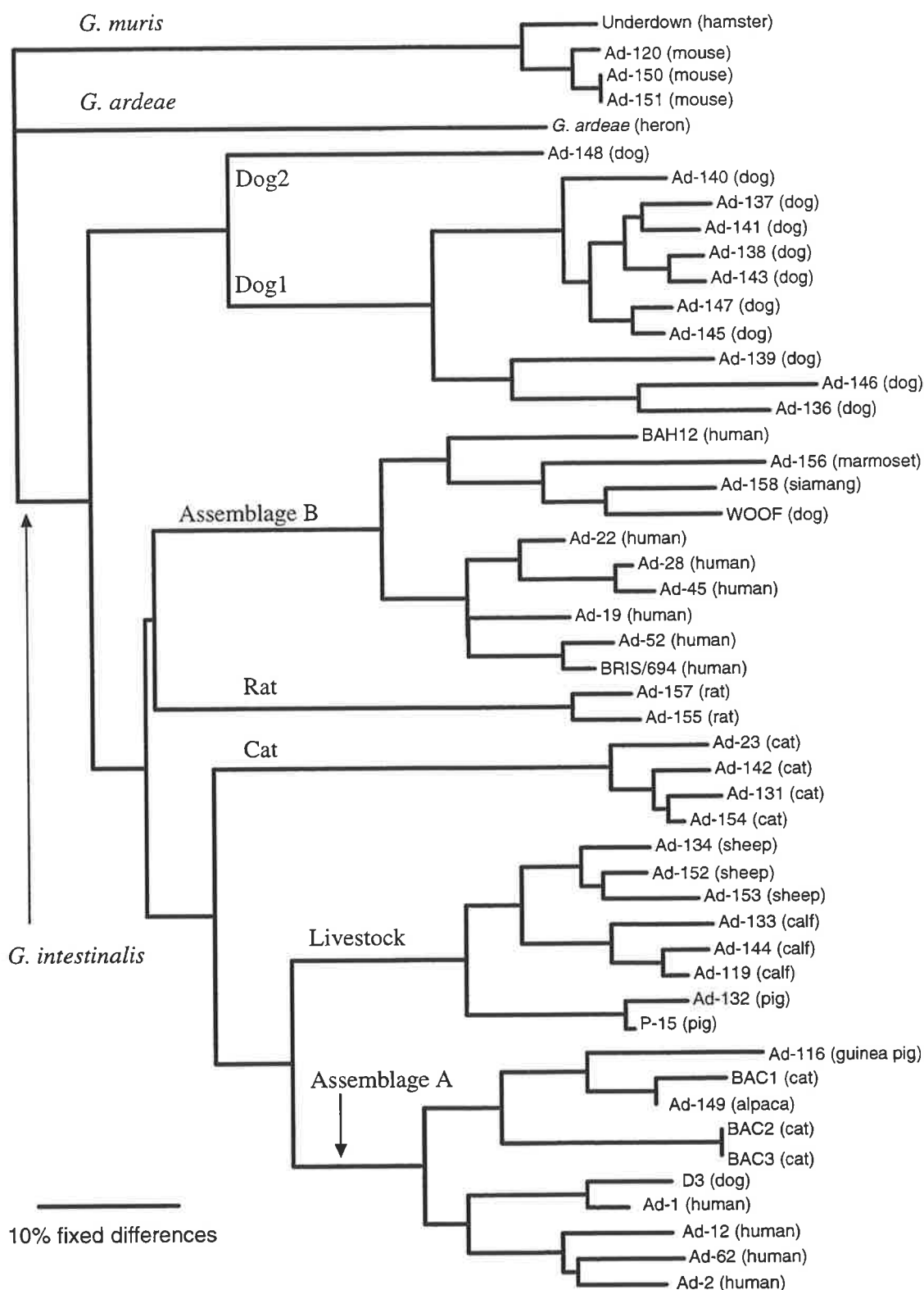


Figure 13. Dendrogram depicting the genetic relationships of isolates of *G. intestinalis* determined by NJ analysis using % fixed differences calculated from the allelic profiles in Table 12 (pp. 73 - 76). The host origin of each isolate is in parentheses.

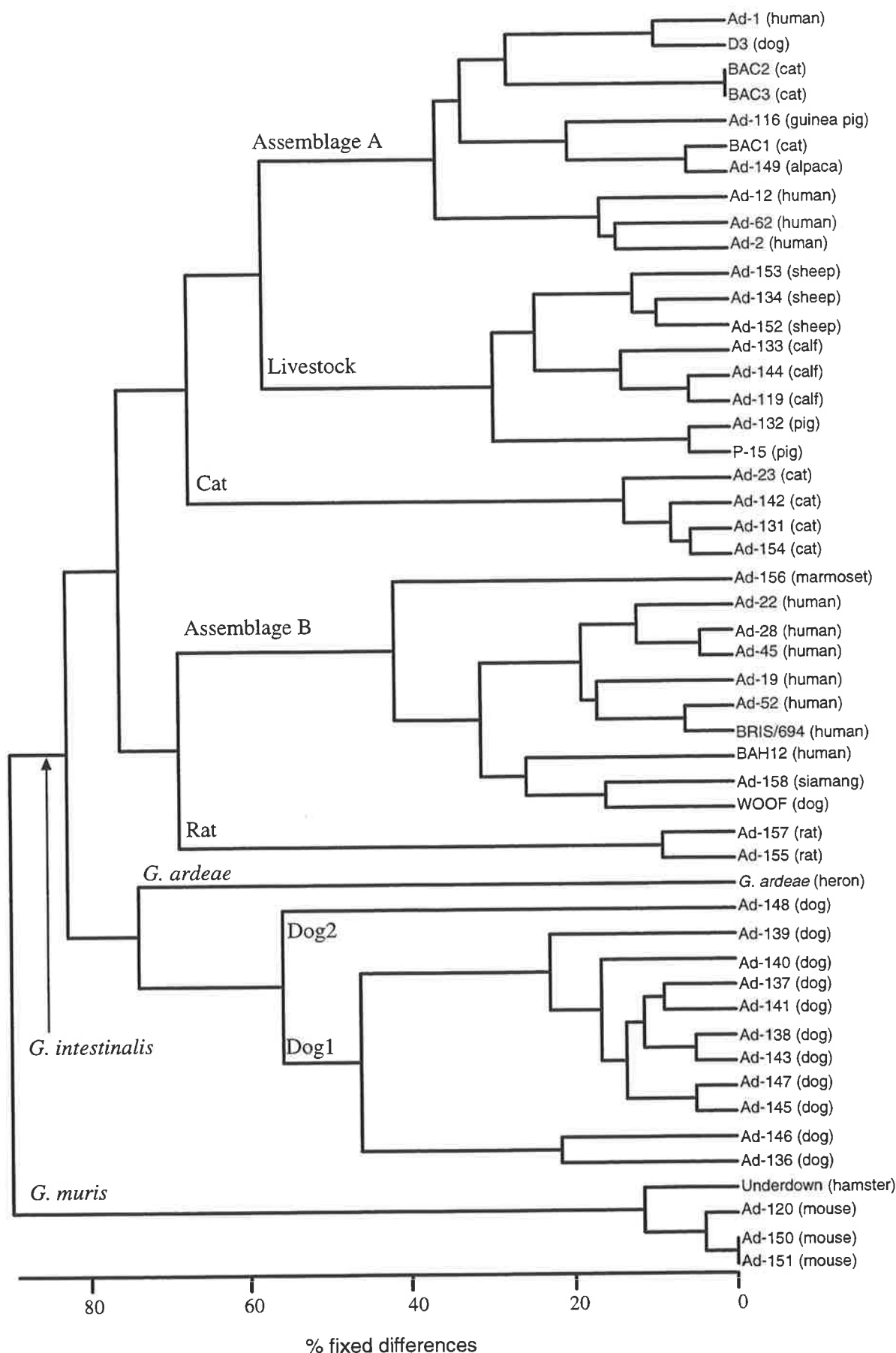


Figure 14. Dendrogram depicting the genetic relationships of isolates of *G. intestinalis* determined by UPGMA analysis using % fixed differences calculated from the allelic profiles in Table 12 (pp. 73 - 76). The host origin of each isolate is in parentheses.

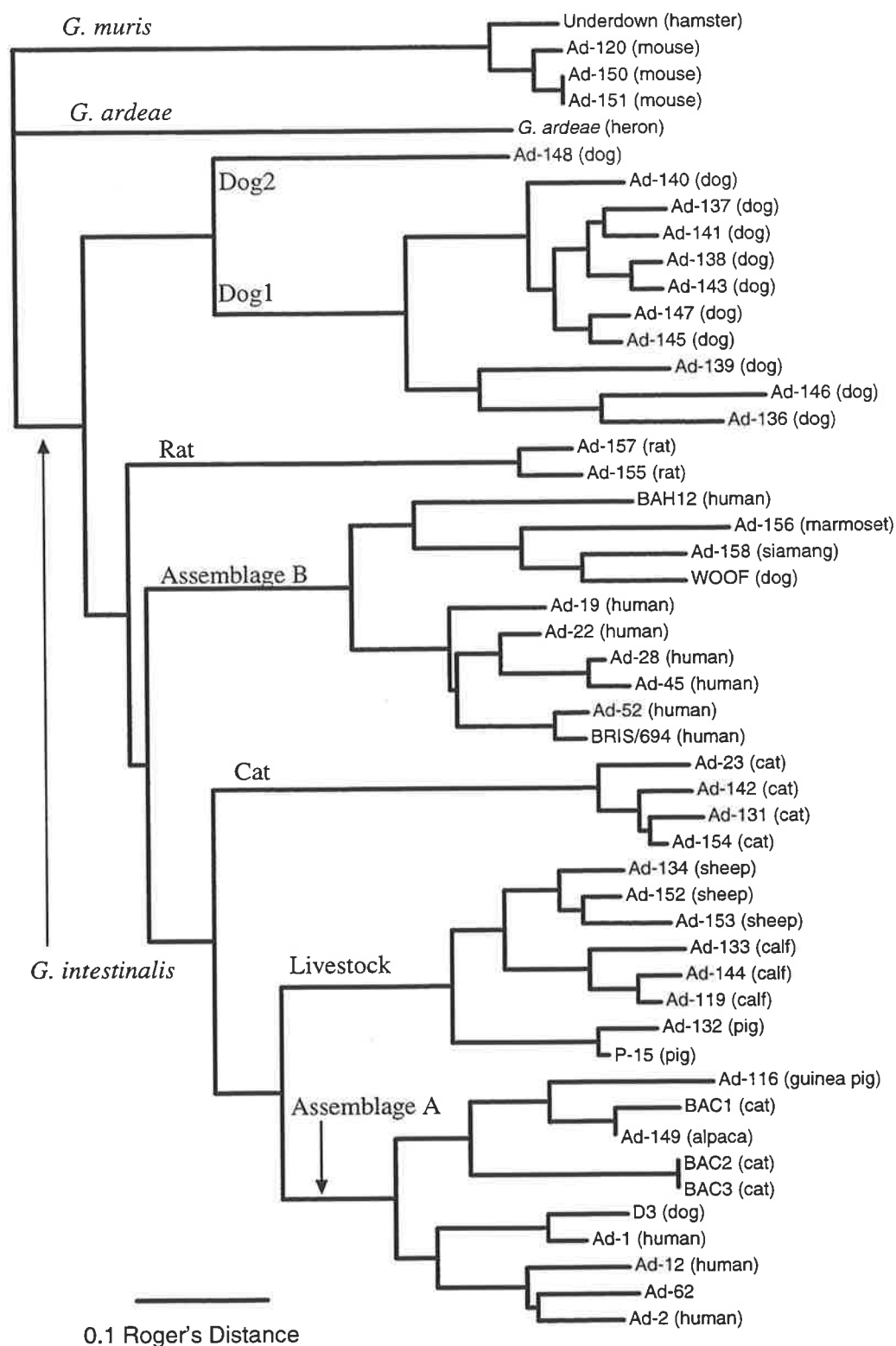


Figure 15. Dendrogram depicting the genetic relationships of isolates of *G. intestinalis* determined by NJ analysis using Roger's distances calculated from the allelic profiles in Table 12 (pp. 73 - 76). The host origin of each isolate is in parentheses.

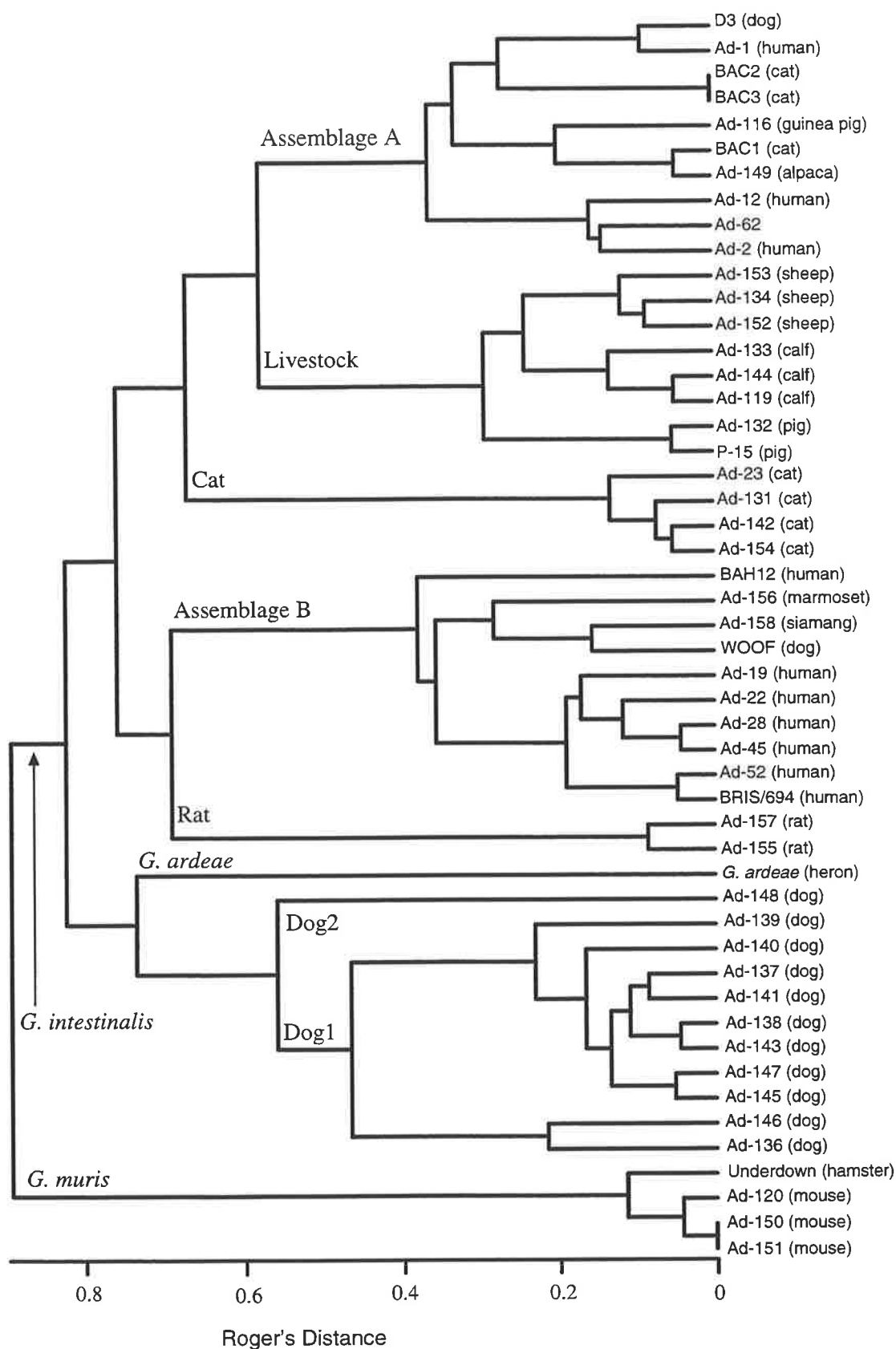


Figure 16. Dendrogram depicting the genetic relationships of isolates of *G. intestinalis* determined by UPGMA analysis using Roger's distances calculated from the allelic profiles in Table 12 (pp. 73 - 76). The host origin of each isolate is in parentheses.

lineage were as great or greater than those which separate the genetic groups described by Andrews *et al.* (1989). The composition of the lineages was consistent in all four dendrograms, as was the grouping of the Cat, Livestock and Assemblage A lineages. All of the isolates of *G. muris* clustered together, and the genetic distances between *G. muris* isolates from different continents were less than those between isolates within the *G. intestinalis* lineages.

Table 14. Major lineages of *G. intestinalis* isolates identified by allozyme analysis.

Lineage	Genetic group	Isolates
Assemblage A	I	Ad-1, D3
	II	Ad-2, Ad-62, Ad-12
		BAC1, BAC2
		BAC3, Ad-116, Ad-149
Livestock		Ad-153, Ad-134, Ad-152
		Ad-133, Ad-144, Ad-119
		P-15, Ad-132
Cat		Ad-23, Ad-131, Ad-142, Ad-154
Assemblage B	III	BAH12
	IV	Ad-19, Ad-22, Ad-28, Ad-45, Ad-52, Bris/694
		Ad-156
		Ad-158, WOOF
Rat		Ad-155, Ad-157
Dog 1		Ad-140, Ad-137, Ad-141, Ad-138, Ad-143, Ad-147, Ad-145
		Ad-139
		Ad-146, Ad-136
Dog 2		Ad-148

UPGMA analyses using both distance measures gave trees with identical topologies with the exception of the intra-lineage relationships within Assemblage B. The most likely explanation for this is that the heterozygosity within isolate BAH12 has affected the Rogers' distance estimates within this lineage, although there was no evidence for this with the reduced data set. The NJ trees were also very similar, although the position of the Rat lineage was slightly different, and there were minor differences in the placement of isolates within Assemblage B, although this was not as pronounced as in the UPGMA dendrograms.

Although the number and composition of the lineages identified by the different tree building methods were consistent, there were differences in the tree topologies, both at the

basal level and within the lineages. Some of these differences are most likely due to the nature of UPGMA, which can distort relationships as a direct result of using group averages. This is exemplified by isolate Ad-139. The NJ tree places it with the isolates Ad-136 and Ad-146, while the UPGMA dendrogram places it with the other group of isolates within the Dog1 lineage. Inspection of the actual distances shows that Ad-139 is no more different to Ad-136 (0.26) than it is to other isolates within Dog1 (0.17 - 0.33), whereas it is very different to Ad-146 (0.43). Since Ad-146 and Ad-136 are only separated by a distance of 0.21, UPGMA would group them first, and calculate a new distance of 0.34 between this cluster and Ad-139. Further distortion also occurs when Ad-139 is grouped with the other isolates. The NJ algorithm is free of this distortion, and will recover the correct tree provided the distance estimates are accurate (Nei, 1991).

3.4.3.3 Reliability of the inferred genetic relationships

As discussed earlier (section 1.5.3), the UPGMA method is predominantly used to determine phenetic relationships. However, it can be used to infer phylogenetic relationships provided certain assumptions are met, such as a constant rate of evolutionary change. If the data satisfy the assumptions inherent in the UPGMA method, an expected outcome would be that outgroup species would form the most distant taxon in the dendrogram. This may provide an indirect means of assessing the accuracy of the proposed phylogeny. The data (eg., Figure 14, p. 84) would appear to partially fulfil this criterion, placing *G. muris* external to the rest of *G. intestinalis*. However, the other outgroup species, *G. ardeae*, was placed within *G. intestinalis*. Because the systematic relationships between *Giardia* species are unknown, and *G. intestinalis* appears to be a species complex, it is not possible to determine directly whether this is either a real result, or whether it reflects a failure of the data and analytical methods to resolve the evolutionary relationships at this level, or an indicator that *G. ardeae* is not an outgroup.

The accuracy of these inferred genetic relationships can also be tested directly by resampling the data, either in the form of “bootstrapping” or “delete-half-jackknifing”. The basis of these methods are described in detail by Swofford *et al.* (1996). In brief, these resampling techniques are used to generate pseudo data sets which are used in place of actual data to estimate sampling variance (Swofford *et al.*, 1996). Multiple pseudo data sets are generated and analysed separately, and a consensus of the resulting trees is determined. The proportion of pseudo data sets that support a particular internal branch can be taken as the

probability that that particular branch is contained in the true phylogeny (Felsenstein & Kishino, 1993). Bootstrap analysis involves the random resampling (with replacement) of the original data set to generate pseudo data sets of the same size as the original (Felsenstein, 1985). Delete-half-jackknifing samples half of the data (without replacement) to generate pseudo data sets which are half the size of the original (Wu, 1986). Both of these methods can be implemented by the SEQBOOT program of PHYLIP. Since this program cannot deal with missing data, a restricted data set for 18 enzyme loci was used (generated by excluding *Dia-1*, *Dia-2*, *Ndpk*, *6Pgd* and *Pgk*, and using the isolates shown in Figure 17, p. 90), in addition to the one used earlier to generate the chord distances (Table 13, p. 77). For each data set, 500 replications were carried out using both methods of resampling. Chord distances were calculated from the replicate samples and analysed using UPGMA. The resulting dendrograms are presented in Figures 12 (p. 81) and 17 (p. 90), along with the percentage bootstrap support values. The results using jack-knifing (not shown) were very similar ($\pm 3\%$ of bootstrap value). In almost all cases there was poor support ($<50\%$) for any of the groupings at the base of the tree. This result supports the observation that AE is unlikely to reliably determine evolutionary relationships when the percentage of fixed genetic differences exceeds 60% (Richardson *et al.*, 1986). The exception to this was the grouping of Assemblage A with the Livestock lineage (89% support for all comparisons). A comparison between the isolates comprising these two lineages revealed that they had fixed differences at between 44 - 77% of loci, whereas the variation between isolates comprising the other lineages generally exceeded 70% fixed differences.

In contrast to the failure to resolve inter-lineage relationships, there was good support for the integrity of the individual lineages themselves, as well as support for some of the subgroupings within lineages. The support for some of the groupings was affected by decreasing the number of loci analysed. In the case of the comparison between the two groups of isolates within the Dog1 lineage, a decrease in the number of loci increased the support for the two groups (Figure 17, p. 90), presumably through the removal of the loci which caused the original conflict. The reverse occurred for the isolates comprising the Livestock lineage, whose branching order became less resolved as the number of loci decreased.

Having identified the main lineages of isolates, each lineage was treated as an operational taxonomic unit (OTU). The advantage of this approach was that all of the

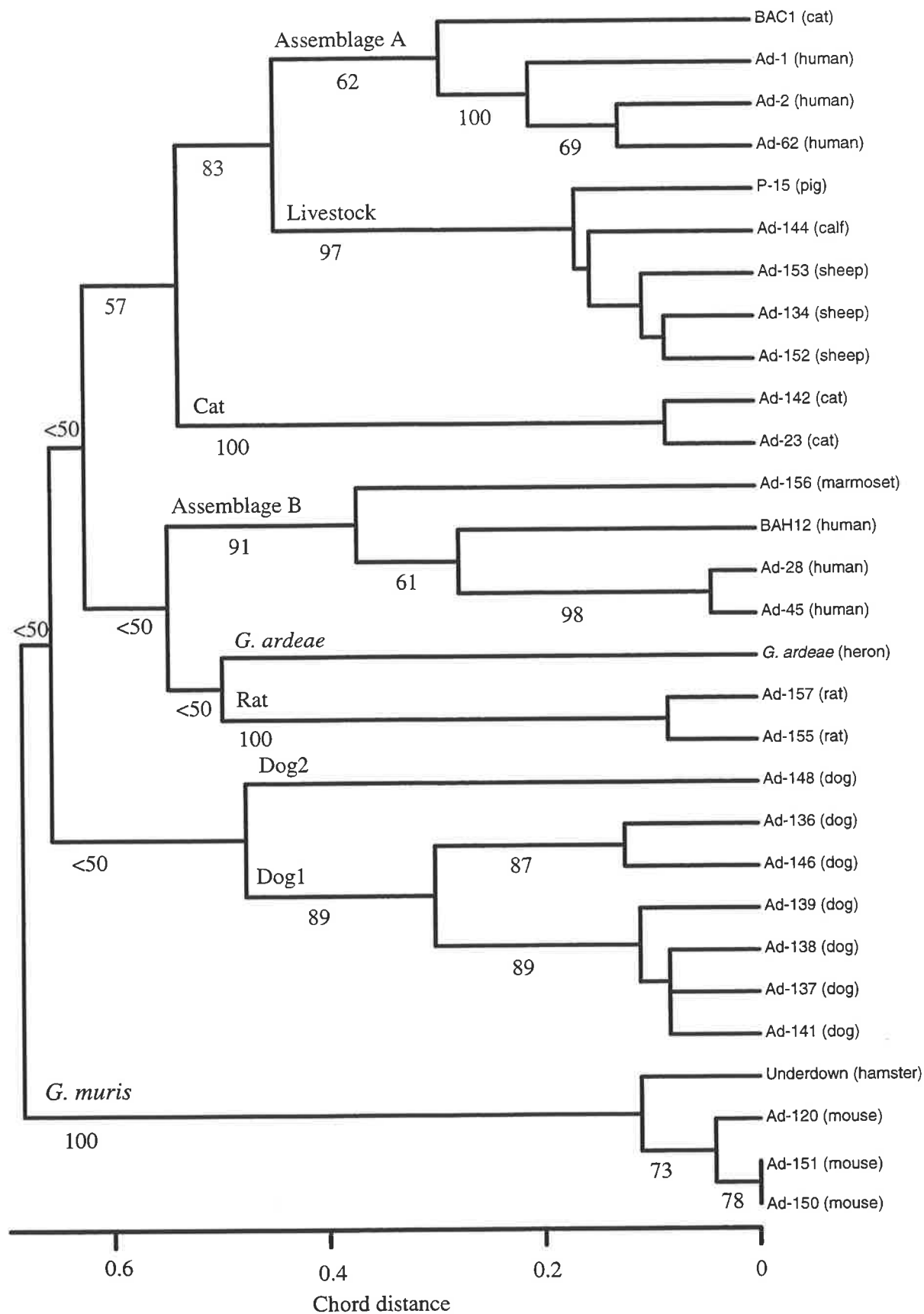


Figure 17. Dendrogram depicting the genetic relationships of isolates of *G. intestinalis* inferred by UPGMA analysis using Chord distances calculated for 18 loci. The host origin of each isolate is in parentheses and % bootstrap support (500 replicates) is indicated under the branches.

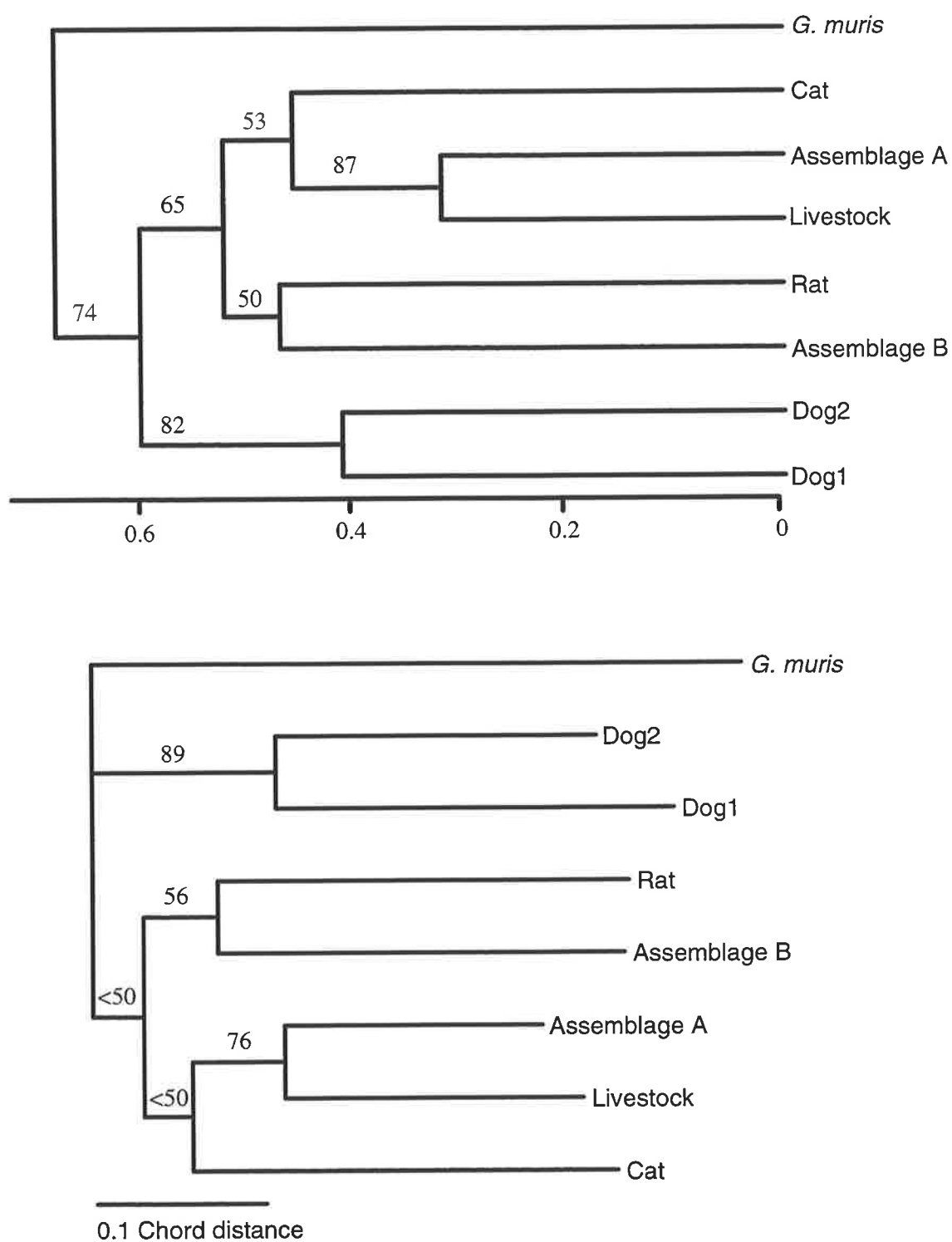


Figure 18. Dendrograms depicting the genetic relationships of the 7 lineages of *G. intestinalis* inferred by UPGMA (top) and NJ (bottom). % bootstrap support (500 replicates) is indicated above the branches.

lineages could be compared using 23 loci, allowing a more appropriate comparison of the lineages by treating them as separate populations. Chord distances were generated using the allele frequencies given in Appendix B. Genetic relationships were inferred using both NJ and UPGMA methods, and the resulting dendrograms are shown in Figure 18 (p. 91). The gene frequency data were also subjected to bootstrap and jack-knife analysis, and the results are indicated on Figure 18 (p. 91). The jack-knife and bootstrap tests both gave similar results ($\pm 3\%$), and the NJ and UPGMA trees both had similar topologies, although these two tree-building methods found different levels of support for different nodes. UPGMA analysis of the replicate data sets found strong support for the grouping of the Dog1 and Dog2 lineages (82%), and the grouping of Assemblage A with the Livestock lineage (87%). NJ analysis of the same data found better support for the grouping of the Dog1 and Dog2 lineages (89%), but lower support for the grouping of Assemblage A with the Livestock lineage (76%).

3.4.3.4 Effect of varying the number of loci examined

As discussed by Richardson *et al.* (1986), the accuracy of inferred genetic relationships can be affected by the number of loci analysed. This can be seen in some of the above analyses, where a decrease in the number of loci affected some of the groupings of isolates. It was therefore of interest to determine the effect that varying the number of loci analysed would have upon the tree topology. Three sets of data were generated for this purpose, one comprising data from isolates which could be characterised at 27 presumptive loci (Table 15, p. 93), and the other two comprising subsets of data from Table 12 (pp. 73 - 76) for 10 loci (*Est*, *Got-1*, *G6pd*, *Gpt*, *Hk*, *Mdh*, *Me*, *Np-2*, *6Pgd*, *Pgm*) or 13 loci (as for 10 but also including *Acp*, *Gdh*, *Ndpk*). The latter two data sets were similar to those utilised by Meloni *et al.* (1988 & 1992). Percentage fixed differences were used with UPGMA to produce the dendrograms shown in Figures 19-21 (pp. 94-96).

The dendrogram shown in Figure 19 (p. 94) suggests that increasing the number of loci analysed did not affect the overall tree topology. Compared to the UPGMA dendrogram constructed using 23 loci (Figure 14, p. 84), there were some minor changes within Assemblage B (eg., the position of Ad-19 and BAH12) and the position of Ad-139 was also changed within the Dog1 lineage. However, all of the other groupings were consistent.

Table 15. Allelic profiles of *Giardia* isolates determined for 27 enzyme loci.

Isolate	Enzyme Locus																										
	Acp	Ald	Dia-1	Dia-2	Enol	Est	Fdp	Gdh	Got-1	Got-x	Gpd	G6pd	Gpi	Gpt	Hk	Mdh	Me	Ndpk	Np-1	Np-2	Pep-A	Pep-D	Pgam	6Pgd	Pgk	Pgm	Tpi
Ad-1	e	e	c	c	e	d	c	b	b	a	b	c	c	e	e	a	f	a	a	a	b	d	d	a	d	d	a
Ad-2	e	e	d	c	e	d	c	b	b	a	c	c	c	f	e	a	f	a	b	b	b	d	d	a	d	e	a
Ad-12	e	e	-	-	e	d	a	b	b	a	b	d	c	f	e	a	e	a	-	d	b	d	-	a	d	e	a
Ad-62	e	e	-	-	e	d	-	b	b	a	e	c	c	f	g	a	e	a	b	c	b	d	d	a	d	e	a
Ad-19	a	g	f	-	d	d	a	b	b	a	c	f	c	d	f	b	-	a	-	d	a	d	b	b	b	c	a
Ad-22	b	g	f	b	d	d	c	b	-	a	f	e	c	d	f	b	e	a	d	d	b	d	b	b	d	-	a
Ad-28	b	g	f	b	d	d	b	b	d	a	e	e	c	d	f	b	c	a	d	d	a	d	b	b	d	c	a
Ad-45	b	g	f	b	d	d	a	b	d	a	e	e	c	d	f	b	c	a	d	d	a	e	b	b	d	c	a
Ad-52	b	g	f	b	d	d	-	b	d	a	f	f	c	d	g	b	e	a	-	d	-	-	b	b	-	c	a
BAH-12	b	df	f	c	d	e	c	bc	c	a	c	f	d	d	g	b	ac	a	d	d	a	e	bc	cd	b	bc	a
Bris/694	b	-	f	b	d	d	a	b	be	a	f	f	c	d	g	b	e	a	c	d	a	e	b	b	d	c	a
Ad-136	d	b	f	b	b	g	e	c	a	b	c	g	a	b	c	d	h	a	e	e	a	a	b	c	b	a	b
Ad-137	d	b	e	c	b	d	e	c	a	b	a	g	a	a	c	e	f	a	e	e	c	c	b	b	b	a	c
Ad-138	d	b	e	c	b	d	e	c	a	b	a	g	a	b	c	e	c	a	e	e	c	c	b	b	b	a	c
Ad-139	d	b	f	b	b	d	e	c	a	b	c	g	a	b	c	e	i	a	e	e	c	c	b	b	b	a	b
Ad-140	-	a	e	c	b	d	e	b	a	b	b	-	a	b	c	e	g	a	-	-	c	c	b	b	-	a	c
Ad-141	d	b	e	c	b	d	e	b	a	b	a	g	a	b	c	e	f	a	e	e	c	c	b	b	b	a	c
Ad-143	d	b	e	c	a	d	e	c	a	b	-	-	a	b	c	e	c	a	e	e	c	c	b	b	b	a	c
Ad-145	-	b	e	c	-	d	e	b	a	b	a	f	a	b	c	e	i	a	-	-	c	c	b	b	b	a	c
Ad-147	-	b	e	c	b	d	e	c	a	b	a	f	b	b	c	e	i	a	-	-	c	c	b	b	b	a	c
Ad-148	c	e	c	c	e	a	a	c	b	a	b	g	a	b	d	e	g	a	b	b	b	c	b	c	b	a	c
D3	e	e	f	c	e	-	d	b	b	a	b	c	c	e	e	a	f	a	a	a	b	d	b	a	d	d	a
WOOF	b	g	f	c	d	f	b	b	d	a	e	e	d	d	g	b	b	a	d	d	a	e	b	b	e	-	a
Ad-146	d	b	e	a	b	g	e	c	a	a	c	g	a	b	c	d	h	a	f	f	d	a	b	c	b	a	b
<i>G. ardeae</i>	b	h	f	d	d	g	-	c	f	-	-	c	-	g	c	e	j	-	g	g	c	d	b	-	-	b	b
Ad-120	f	i	g	d	f	i	e	d	e	b	-	c	c	h	b	c	g	c	h	h	c	c	a	c	d	b	d
Ad-150	f	i	g	d	f	h	-	d	e	-	-	c	-	h	b	c	g	c	h	h	c	c	a	c	d	b	d
Ad-151	f	i	g	d	f	h	-	d	e	-	-	c	-	h	b	c	g	c	h	h	c	c	a	c	d	b	d
Underdown	f	i	g	d	f	i	-	d	e	-	-	c	-	h	a	c	i	c	h	h	c	c	a	c	d	b	d

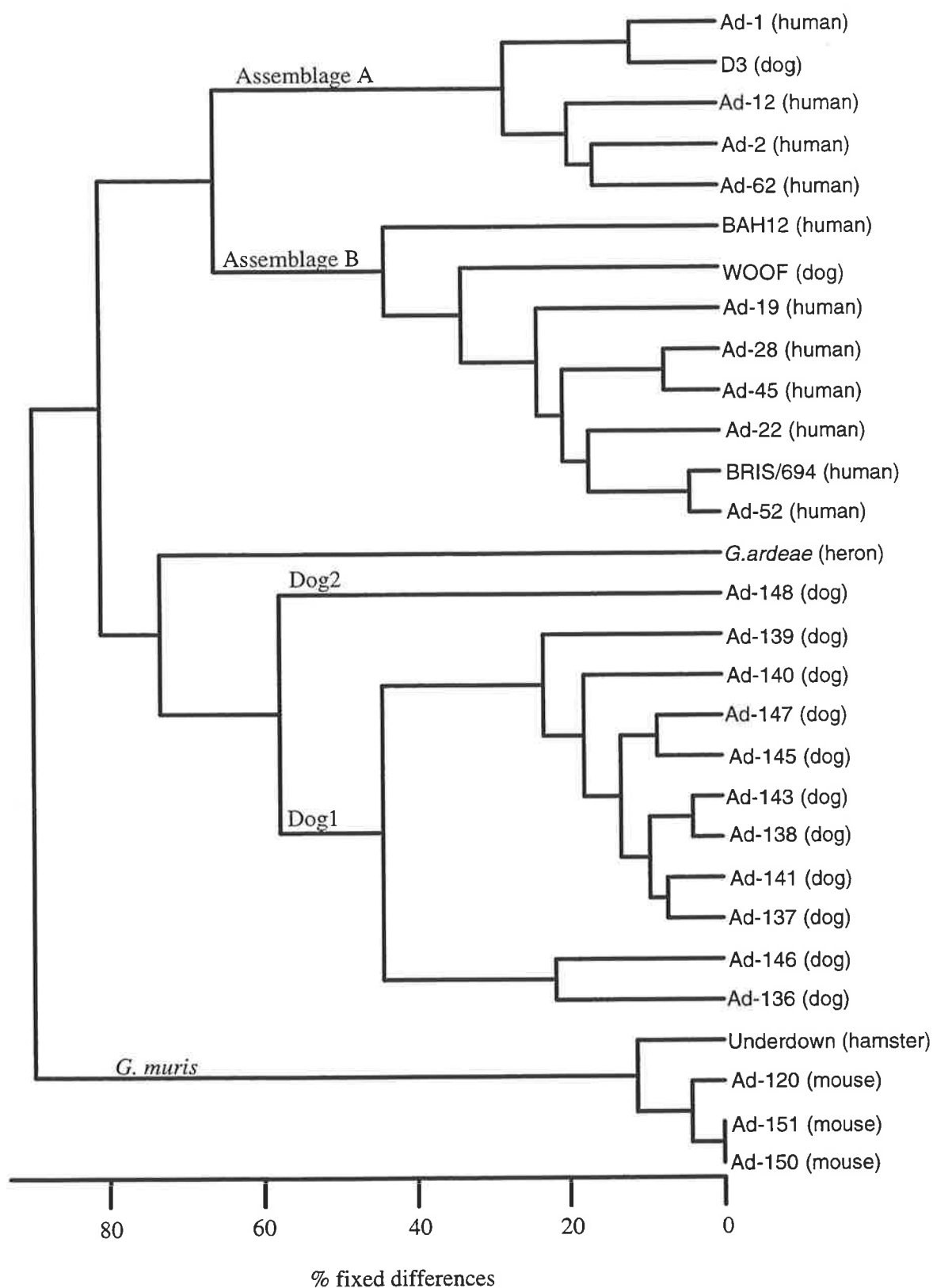


Figure 19. Dendrogram depicting the genetic relationships of isolates of *G. intestinalis* determined by UPGMA analysis using % fixed differences calculated using 27 loci. The host origin of each isolate is in parentheses.

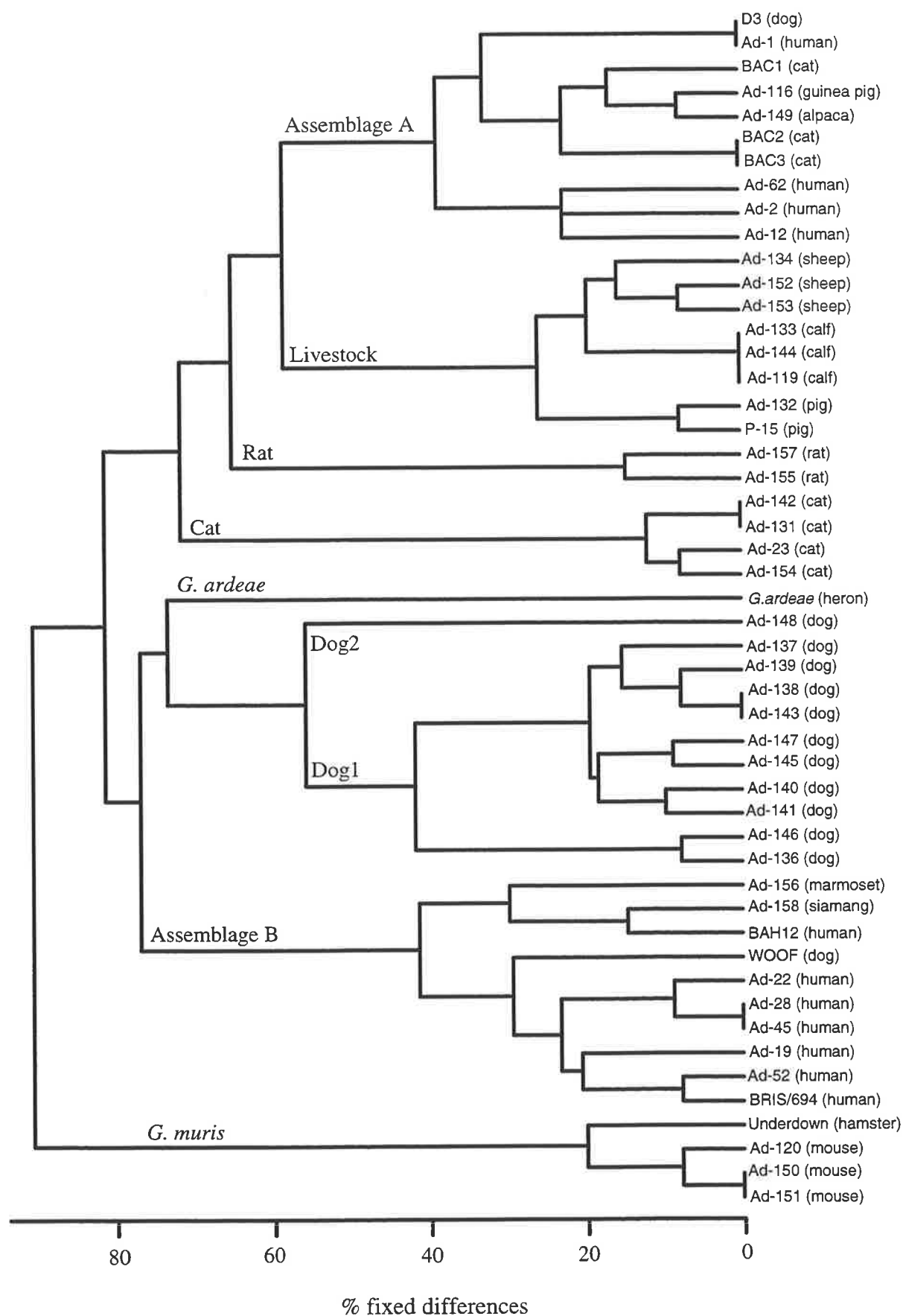


Figure 20. Dendrogram depicting the genetic relationships of isolates of *G. intestinalis* determined by UPGMA analysis using % fixed differences calculated using 13 loci. The host origin of each isolate is in parentheses.

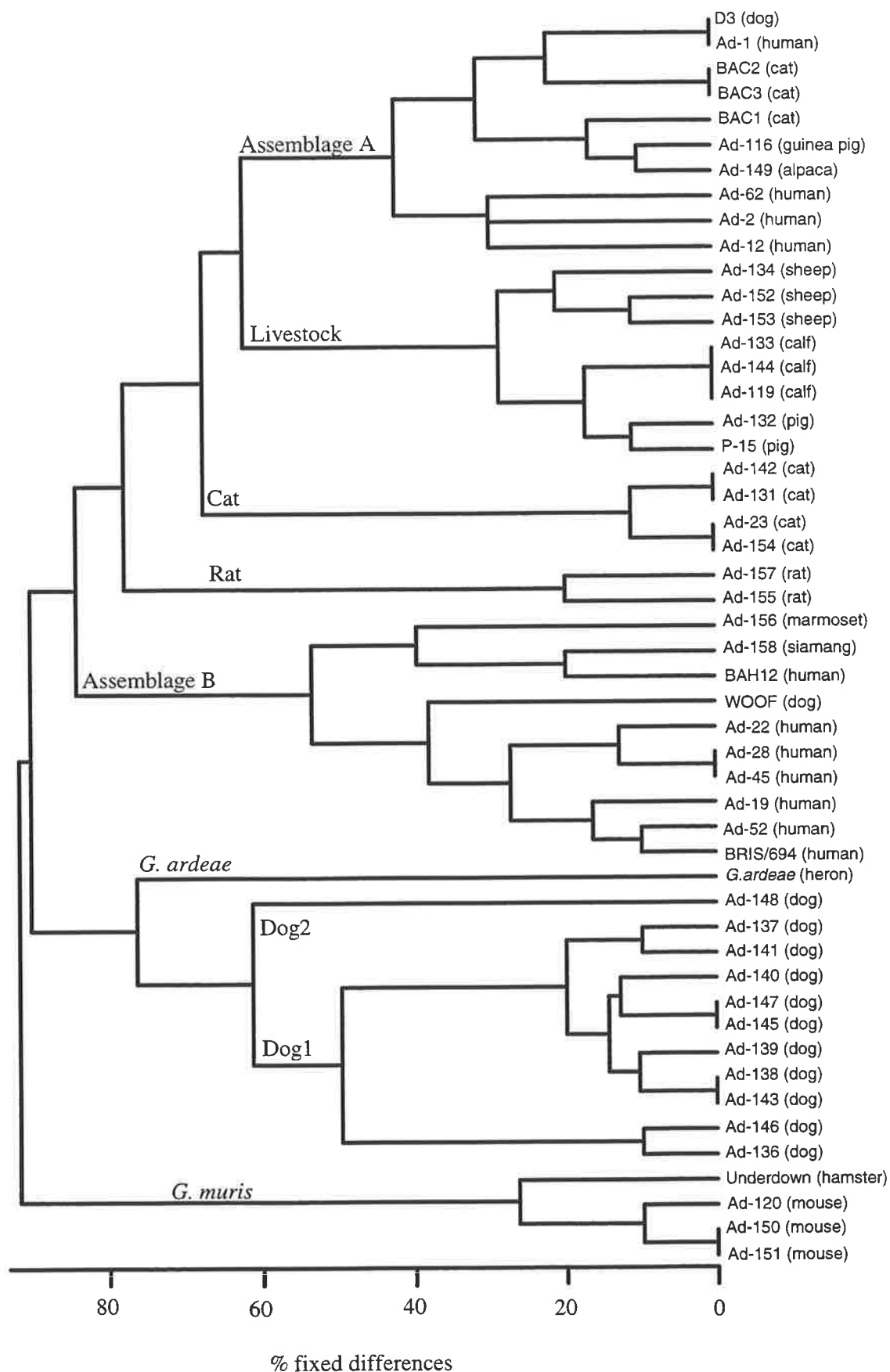


Figure 21. Dendrogram depicting the genetic relationships of isolates of *G. intestinalis* determined by UPGMA analysis using % fixed differences calculated using 10 loci. The host origin of each isolate is in parentheses.

In contrast to the preceding findings, a decrease in the number of loci analysed from 23 (Figure 14, p. 84) to 13 (Figure 20, p. 95) or 10 (Figure 21, p. 96) had a substantial effect upon the levels of variation observed both within and between the different lineages, although the composition of the lineages themselves were not affected. In some cases, the distance between isolates was compressed, eg., Ad-1 versus D3 had 9% fixed differences using 23 loci, but appeared identical using only 10 or 13 loci, whilst in other cases it was extended, eg., the branch lengths separating Underdown from the other *G. muris* isolates. Although the same lineages could be identified readily using these different data sets, the tree topologies tended to differ slightly at both the intra- and inter-lineage level. The levels of difference observed between Assemblages A and B (an average of 81% fixed differences) and between genetic groups I and II (an average of 40% fixed differences) when only 13 loci were analysed were similar to those observed by Meloni *et al.* (1995) who, despite using a zymodeme interpretation, found 87.5% and 46% enzyme differences respectively using the same number of loci (11/13 of which were the same enzymes used between studies). The differences between the results of the study by Meloni *et al.* (1995) and the study herein may be due to the 2 enzyme loci which were unique to each study, or to the different methods used to interpret the data (allozyme versus zymodeme interpretation, section 3.1.2 and Figure 7, p. 63).

3.4.3.5 Test for linkage disequilibrium

The AE data obtained herein were used to test the hypothesis Tibayrenc *et al.* (1991) that *G. intestinalis* has a clonal population structure. This hypothesis can be tested in a number of ways. If *G. intestinalis* is asexual, and recombination between isolates is a rare event, then linkage would be expected to occur between enzyme loci. Linkage disequilibrium can be measured using the χ^2 test (Richardson *et al.*, 1986). However, the sample sizes used in this study were not sufficiently large to allow valid comparisons (expected χ^2 values were less than 5). Another test for linkage disequilibrium is the Index of Association (I_A), which originated from the work of Brown *et al.* (1980) and was further refined by Maynard Smith *et al.* (1993) and Gordon *et al.* (1995). This measure compares the observed and expected variances of the mean of the number of loci at which OTU's (operational taxonomic units, also known as ET's or electrophoretic types) differ. The formula to determine I_A is described in section 3.4.2.5.

An I_A value of zero is the expected outcome if all of the individuals characterised are from a single, large, random-mating population in linkage equilibrium. An I_A significantly

different from zero does not necessarily indicate linkage disequilibrium, as other factors, such as an epidemic population structure, population substructuring, or non-random mating (or recombination) can cause this (Maynard Smith *et al.*, 1993). These competing hypotheses can be tested by recomputing I_A for any sub-groups (which may represent cryptic species or multiple samples of a single clone) identified by the original genetic analysis. This approach has been used to show that the clonal structure of some *Salmonella* species is an artefact caused by an epidemic population structure (reviewed by Maynard Smith *et al.*, 1993).

The Index of Association was computed using the complete data set for the *G. intestinalis* isolates characterised herein, as well as for the individual lineages identified by allozyme analysis (Table 16). The significance of the result was determined by the calculation of L (section 3.4.2.5), which is the upper 95% confidence limit of the expected variance (V_E). An observed variance (V_O) less than L shows that I_A is not significantly different from zero (Gordon *et al.*, 1995).

Table 16. I_A values for *G. intestinalis*.

Lineage	No. of samples	mean K	D	V_E	L	V_O	I_A	Significance
<i>G. intestinalis</i>	45	15.44	0.67	4.61	6.59	25.98	4.63	$P < 0.05$
Assemblage A	7	6.07	0.26	3.30	6.27	6.87	1.08	$P < 0.05$
Assemblage B	10	6.51	0.28	3.64	6.94	8.27	1.27	$P < 0.05$
Livestock	8	4.18	0.18	2.15	4.30	4.76	1.21	$P < 0.05$
Dog1	10	5.43	0.24	3.27	6.23	11.72	2.58	$P < 0.05$

mean K - expected mean number of loci different between any two isolates

D - observed mean genetic distance between any two isolates

The Index of Association was found to be significantly greater than zero for all of the comparisons examined (Table 16). However, the results need to be interpreted with caution. The large I_A value for *G. intestinalis* could be taken as evidence for linkage disequilibrium, or alternatively interpreted as evidence of discrete populations or species. Similarly, the I_A values for the separate lineages could also be affected by the presence of the discrete genetic groups which comprise them. It is interesting to note that V_O for the Assemblage A and Livestock lineages is only slightly larger than L , and may not exceed the upper 99% confidence limit of V_E . Another consideration is the sample sizes which, although adequate if pooled, are not large enough to thoroughly test for linkage disequilibrium within individual lineages, especially if the genetic sub-groups are to be considered separately. Larger sample

sizes would also allow other methods, such as χ^2 contingency analysis, to be used to test for linkage disequilibrium.

3.4.3.6 Comparison of results with previous work

The findings presented herein agree with the existing genetic model (Figure 8, p. 65) with one exception, which was the placement of Ad-19. As shown in Figure 8, Ad-19 had been shown previously to be similar to BAH12 (ie., Group III; Mayrhofer *et al.*, 1995). However in this study it appears to be more closely related to the group IV-like isolates. UPGMA analysis of the data set of Mayrhofer *et al.* (1995, Table 1), represented graphically as a dendrogram in Figure 22 (p. 100), gives the same result as that observed herein. The most likely explanation of this would be that mis-labelling of the published dendrogram has occurred, resulting in the swapping of the Ad-16 and Ad-19 labels.

3.5 Summary and conclusions

Allozyme analysis of isolates of *G. intestinalis* recovered from a variety of host species has allowed the construction of the model of genetic relationships presented in Figure 23 (p. 101). The major lineages identified by this analysis are listed in Table 14 (p. 87). The cohesiveness of the lineages is supported by bootstrap analyses, and their composition is not overtly affected by the number of loci analysed. In addition to the previously described Assemblages A and B, a number of unique lineages can be identified. These lineages were external to Assemblages A and B, and each comprised *G. intestinalis* isolates derived from the same host (or related) species.

None of the new isolates, including those which clustered within Assemblages A and B, were identical to any of the human-derived isolates. Furthermore, with the exception of the dog-derived isolate D3, all of the animal-derived *G. intestinalis* isolates exhibited at least 20% fixed differences compared to the human-derived isolates. Genetic sub-divisions were evident within some of the lineages at levels equivalent to those which were used to originally define groups I-IV. The majority of the sub-groupings within each lineage comprised isolates derived from the same host species (eg., BAC2 & BAC3). The exceptions are the group containing Ad-149 (alpaca), Ad-116 (guinea pig) and BAC1 (cat), and the group containing WOOF (dog) and Ad-158 (siamang).

The AE data presented herein provides additional evidence that *G. intestinalis* is a species complex (Andrews *et al.*, 1989), comprising at least 7 phylogenetic species. The relationships between these species cannot be resolved using AE. There was some support for areas of the tree given in Figure 23 (p. 101), eg., the grouping of Assemblage A with the Livestock lineage. However, statistical analyses in the form of bootstrapping and delete half

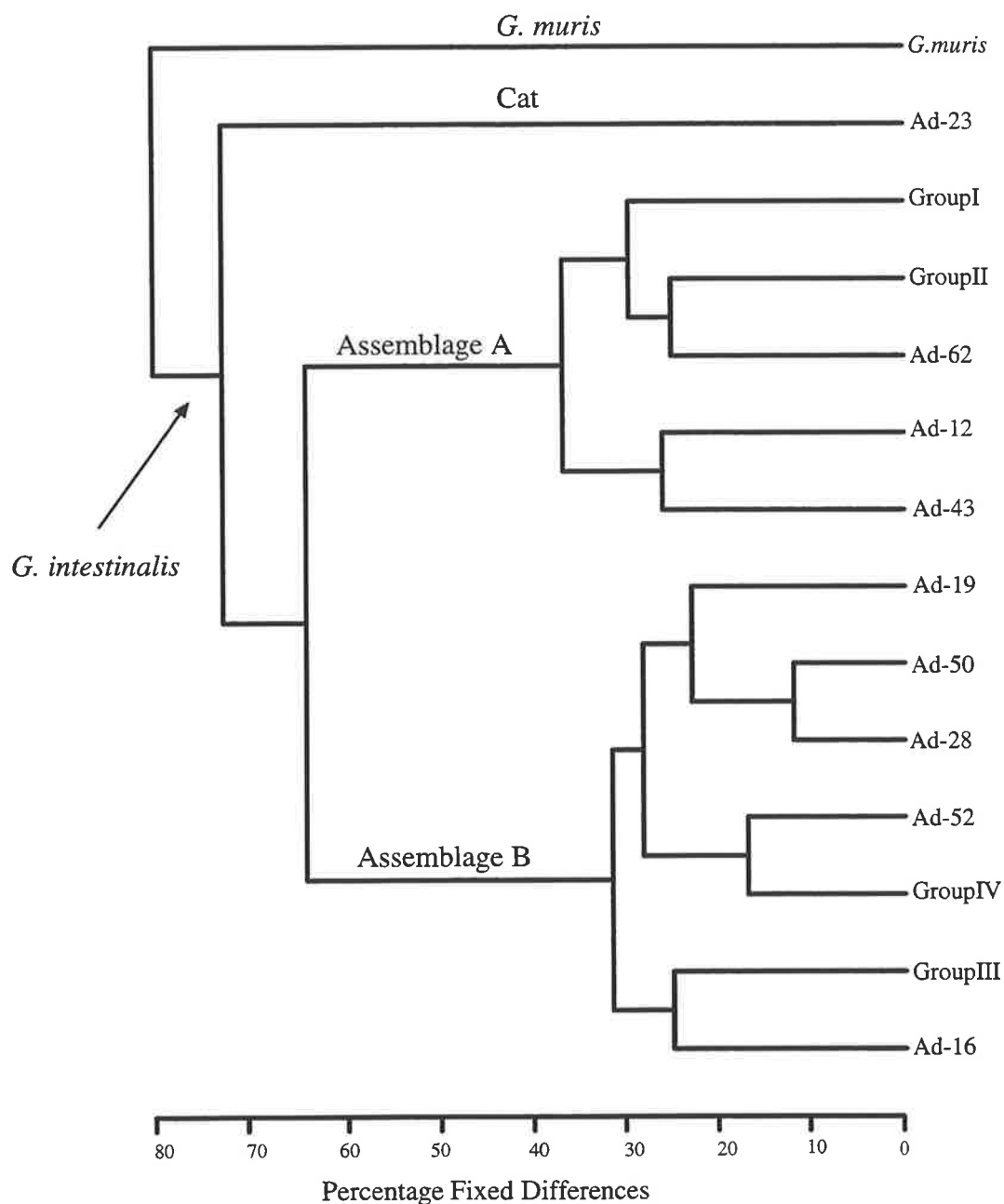


Figure 22. UPGMA dendrogram constructed using the allozyme data from Mayrhofer *et al.* (1995) (Table 1)

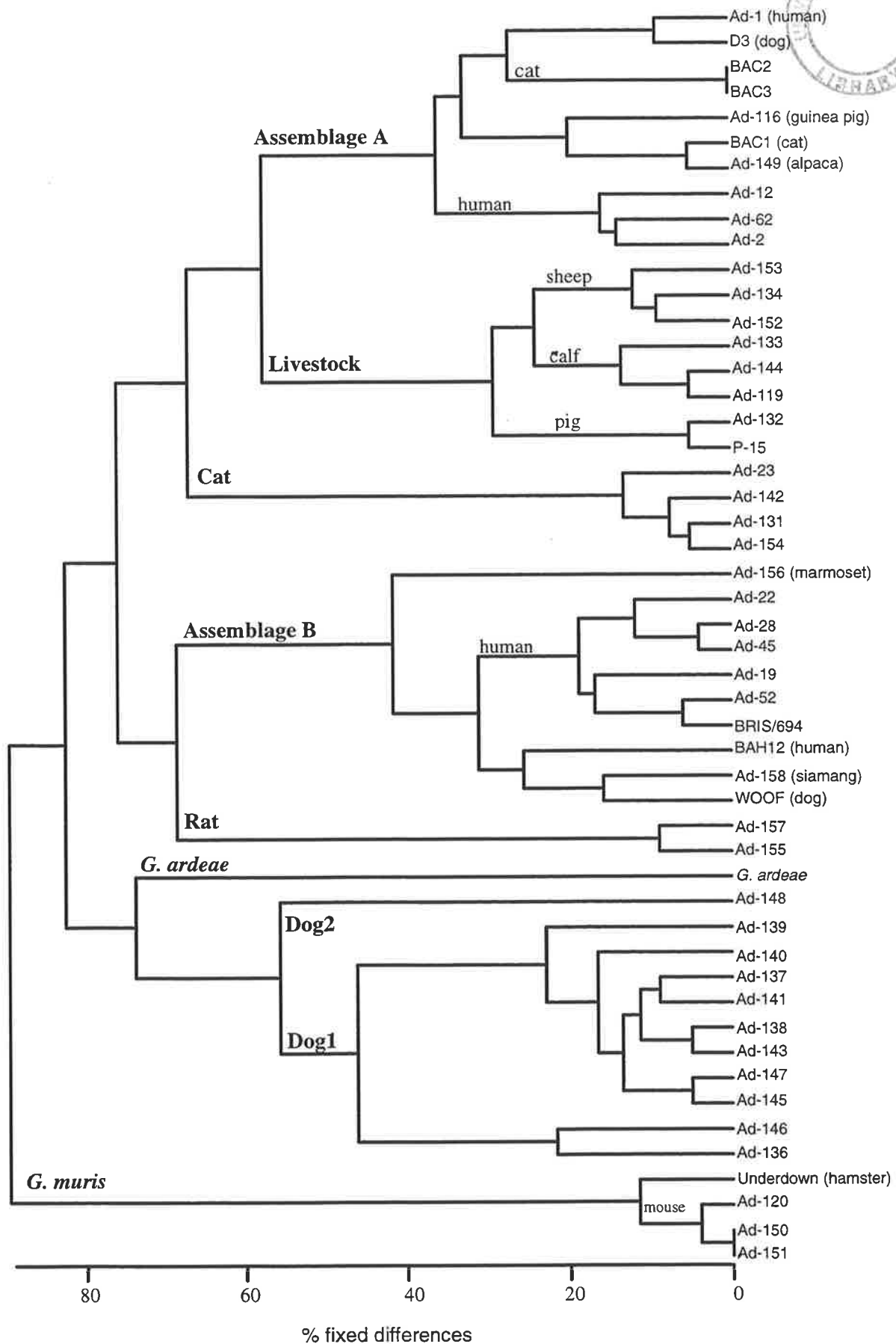


Figure 23. Dendrogram depicting the proposed model of genetic relationships of isolates comprising the *G. intestinalis* species complex (inferred by UPGMA analysis of % fixed differences at 23 loci).

jack-knifing indicated that there was little resolution at the base of the tree, including the relative positions of the outgroups. The relationships between the morphological species of *Giardia* could not be reliably determined.

Examination of the population structure of *G. intestinalis* was inconclusive. There was evidence compatible with the existence of linkage disequilibrium within *G. intestinalis*, which would suggest asexual reproduction with little or no recombination between clones, ie., a clonal population structure. However, apparent linkage disequilibrium will also be observed if distinct species or populations, which can be isolated by a variety of mechanisms (eg., geographic separation, host specificity, incompatibility for sex or recombination), are treated as a single panmictic (ie., random mating) population. The latter possibility is especially important considering the additional evidence provided herein that *G. intestinalis* is a species complex, and that one of the common indicators of clonality, the over-representation of a genotype, was not found in this data set. The determination of the population structure of *G. intestinalis* will require much larger sample sizes of isolates from the separate lineages and genetic groups, which will allow the thorough testing of the two competing hypotheses.

Chapter 4

GENETIC ANALYSES USING NUCLEOTIDE SEQUENCE DATA

The results from the previous chapter were used to propose a model for the genetic relationships of isolates within the *G. intestinalis* species complex (Figure 23, p. 101). Although a number of distinct lineages were identified, the relationships between them was unclear. The magnitude of the genetic distances separating these lineages were such that AE cannot reliably determine their phylogeny. NJ analyses of the AE data were concordant with the UPGMA analyses, however the distance estimates, and hence the observed levels of divergence, may not be accurate. While the AE model could be used directly as a template to examine the systematics of *G. intestinalis*, a more rigorous approach would be to test the genetic relationships using independent techniques, and to determine the phylogenetic relationships of the lineages comprising the *G. intestinalis* species complex. This would provide an evolutionary perspective to the model, and test the assumption (based upon the mode of reproduction) that the clusters of *G. intestinalis* isolates identified by AE are lineages. Aside from its limited utility as a phylogenetic tool, AE also has the disadvantage that it is a comparative technique which requires the use of carry-over controls from previous experiments. The development of DNA-based methods for the characterisation or identification of new isolates would overcome this problem, and provide the means to test the AE-based model of relationships.

4.1 Overview of available techniques

A wide range of DNA-based techniques is available for the genetic characterisation of isolates. These methods can be divided into those that either indirectly or directly assess nucleotide (nt) site variation between samples. The indirect methods include a variety of DNA-DNA hybridisation techniques (which either measure renaturation kinetics or make use of conformational polymorphisms) and fingerprinting techniques, such as Restriction Fragment Length Polymorphisms (RFLP) or Random Amplified Polymorphic DNA (RAPD). The only direct method is nt sequencing, although microsatellites, which differ in the copy number of di-, tri- or tetra-nucleotide repeat sequences, could also be included in this category.

4.1.1 DNA hybridisation and related techniques

There are two different ways in which DNA hybridisation can be utilised for genetic analyses. The earliest described methods used DNA-DNA hybridisation to indirectly compare organisms across the entire genome (discussed in detail by Werman *et al.*, 1996). This approach uses differences in the renaturation kinetics of heteroduplex formation, which are ultimately due to nt sequence differences between the two samples being compared, to estimate a genetic distance. This is usually done using purified single copy nuclear DNA (Werman *et al.*, 1996). The advantage of this technique is that it samples across most of the genome. However it suffers from a similar problem to AE in that it requires direct pairwise comparisons of all samples. This means that all of the previously analysed samples must be compared to each new sample that is added to a data set. Other disadvantages, aside from the objections which could be raised by cladists regarding the use of distance measures, is that the amount of sample required is larger than that needed by other DNA-based techniques or AE, and that it may only have limited utility at specific taxonomic levels ie., it cannot resolve relationships above or below certain levels (eg., within species or between orders).

The second group of techniques either directly utilise DNA-DNA interactions such as heteroduplex analysis or denaturing gradient gel electrophoresis (DGGE), or detect polymorphisms caused by differences in the secondary structure of DNA fragments. All of these approaches, which use discrete segments of DNA (usually generated by PCR), are used to screen for allelic variation (reviewed by Lessa & Applebaum, 1993). The general principle of all of these techniques is that differences in the conformation of DNA, for either double or single stranded molecules, will affect its migration through a gel matrix, and that these differences can be attributed to differences at the nt level. In the case of heteroduplex analysis, the migration of the heteroduplex molecules is affected by the number of mismatches between the hybridising species. DGGE and the related technique of temperature gradient gel electrophoresis (TGGE) make use of the differences in the melting characteristics of homoduplexes, which are sequence specific, to cause different rates of retardation as the duplex migrates through some form of denaturing gradient (Lessa & Applebaum, 1993). Single strand conformational polymorphism (SSCP), as the name suggests, uses conformational differences between single stranded DNA molecules to separate them. In general, these techniques are used as tools for population genetics, although allele frequency data collected for many loci could be used for phylogenetic analysis in the same fashion as AE data.

4.1.2 DNA fingerprinting techniques

A variety of fingerprinting techniques are also available for the characterisation of isolates. In general, these techniques have no true phylogenetic utility, although they do excel as markers for strain delineation and population genetic studies. These methods, of which there are three types, all generate DNA fragment banding patterns. The original “fingerprinting” methods used repetitive DNA elements as probes for Southern analysis of total genomic DNA (described in depth by Dowling *et al.*, 1996). The banding patterns can provide a high level of discrimination, although they can also be difficult to interpret, and because a genetic interpretation is not possible, the data should not be used for phylogenetic analysis. Furthermore, the high mutation rate in these regions (Jeffreys *et al.*, 1988) means that convergence can be a problem and that they may be unsuitable as long-term genetic markers.

RFLP, which are the result of restriction site polymorphisms, are usually easier to interpret because specific genes are the target of the analysis, and the banding patterns are not as complex. This type of data can be generated by using restriction endonucleases to digest genomic DNA, which is then subjected to Southern analysis using a specific probe. Alternatively, PCR products can be digested and the resulting fragments can usually be examined directly by electrophoresis. The latter technique can be a very powerful tool for population genetics studies, where it can be used to generate haplotype or allele frequency data for the study of population substructuring. While RFLP are not suitable as characters for phylogenetic analysis because of the possibility of character non-independence (discussed by Swofford *et al.*, 1996), restriction site data, which can be determined by constructing restriction maps from the RFLP data, is widely accepted, and there are methods for estimating genetic distances for phylogenetic analysis from this type of data (Swofford *et al.*, 1996).

Arbitrarily primed (AP) PCR, also known as RAPD-PCR, uses random primers (usually 10 bp oligonucleotides) to randomly amplify segments of the genome, thus producing a series of DNA fragments which can be separated by electrophoresis. Banding pattern differences between isolates occur through a number of different mechanisms including loss or mutation of primer sites, changes in fragment lengths due to insertions or deletions, or differences in the stoichiometry of the reaction due to differences in the affinities of the competing primer sites. The main advantages of this technique are that it does not require any prior knowledge of the sequences to be amplified and, being PCR-based, can use small

sample quantities. The degree of variation normally detected using this technique makes it useful for population genetics, especially at the level of strain identification in bacteria. Unfortunately, this method also has many disadvantages, including the unreliability of the technique, which is known to be very sensitive to slight changes in the amplification conditions (eg., Muralidharan & Wakeland, 1993). Other disadvantages are that the method cannot be used to identify samples containing multiple genotypes or heterozygotes, and that the data are inappropriate for phylogenetic or even phenetic analysis (Smith *et al.*, 1994).

The major flaw with data produced by RAPD, within the context of phylogenetic inference, is that it often violates the assumption of character independence. As discussed in detail by Backeljau *et al.* (1995), such data are not suitable for cladistic analysis because of issues regarding character independence, the lack of a model of character state transformations, and other undesirable features such as non-codominant inheritance. Furthermore, there is also evidence which shows that RAPD bands are not necessarily homologous, which is an assumption of all methods of analysis (Smith *et al.*, 1994). Despite these theoretical and practical concerns, RAPD's are nonetheless commonly used in phylogenetic analyses. Numerous workers have demonstrated that the results of RAPD analysis often correlate with other forms of analysis (eg., Morgan *et al.*, 1993; Tibayrenc *et al.*, 1993). This is most likely the case because the amount of phylogenetic signal in their data sets far outweighed the noise due to characters which violate the assumptions listed above. Tibayrenc *et al.* (1993) have suggested that the consistency between methods justifies the use of RAPD's in phylogenetic analyses. However, a counter to their reasoning is; "... there is no guarantee that if something is done inappropriately enough times, all will work out in the end (and the amount of systematic error introduced by this shortcut will increase substantially with increasing divergence among the taxa in the analysis) ..." (Swofford *et al.*, 1996). Because of the above concerns, RAPD data should be interpreted with due caution and used appropriately.

4.1.3 Nucleotide sequence analysis

Nucleotide sequences are possibly the most versatile type of molecular data available. They can be analysed phylogenetically by a variety of methods, including parsimony and distance methods, and can also be used to develop markers for other applications, eg., the design of diagnostic PCR systems which can be used in epidemiological studies. More importantly, DNA sequence data can also provide large numbers of characters for analysis, the

data are additive, and new samples can be readily integrated into existing data sets. Through the judicious selection of appropriate genes different levels of evolutionary divergence can be examined, ranging from population genetics using mitochondrial D-loop sequences to the so called “tree of life” using elongation factor sequences.

Technically, nucleotide sequence analysis overcomes many of the limitations of the other methods, eg., it is reliable, objective, and has relatively well understood (or at least examined) processes. Historically, the nt sequencing approach was both too time consuming and expensive for large scale studies, which made some of the indirect methods popular. However, the advent of PCR and automated sequencing has reduced both of these factors, making such an approach feasible. The main disadvantages of a PCR-sequencing approach is that prior information about the region of interest is required for the construction of PCR primers, and that most genes are only suitable for use at specific taxonomic levels; eg., a rapidly evolving sequence would be unsuitable for comparing different families or orders of organisms, and conversely, highly conserved sequences are not suitable for population studies. Like all PCR-based methods, issues relating to sample contamination, and the amplification of orthologous sequences (if present) also need to be taken into account.

4.2 Review of DNA-based characterisation studies of *Giardia*

Isolates of *G. intestinalis* have been characterised by a variety of methods, including RFLP, RAPD-PCR, M13 fingerprinting, and nt sequencing (eg. Morgan *et al.*, 1993; Nash *et al.*, 1985; Upcroft *et al.*, 1990a). None of these studies examined the phylogenetic relationships within *G. intestinalis*, the majority concentrating solely upon the typing of isolates for epidemiological purposes.

One of the earliest DNA-based studies of *G. intestinalis* was by Nash *et al.* (1985), who examined RFLP in whole genome restricted DNA derived from 15 isolates of *G. intestinalis* from humans and animals. Examination of the ethidium bromide stained banding patterns via revealed two major groups, while Southern analyses, using random clones from a *G. intestinalis* genomic library as probes, identified 9 different banding patterns. Seven of the isolates examined were propagated initially by infection of neonatal mice prior to axenization, and three others were trophozoites recovered by *in vitro* excystation. Also of significance was the recognition by the authors that axenic cultures may not be representative of natural populations, something that many of the more contemporary studies do not seem to consider.

Other workers have also used RFLP to characterise isolates of *G. intestinalis*. Ey *et al.* (1992) used Southern analysis to examine RFLP in trophozoite variant-specific surface protein (VSP) genes from 10 axenic isolates, and identified two RFLP-types which corresponded to the genetic groups I and II identified by Andrews *et al.* (1989). The use of PCR and RFLP analysis of conserved segments from a subset of the *G. intestinalis* VSP gene family has been refined to allow identification of isolates belonging to genetic groups I and II, as well as the Livestock lineage identified in section 3.4.3.2 (Ey *et al.*, 1993a, 1993b, 1996, 1997, Appendix F; Monis *et al.*, 1996, Appendix D). Uji *et al.* (1988) examined 8 axenic isolates by Southern analysis using *Hind*III restricted total genomic DNA from 3 of the isolates as probes, but detected no variation. van Keulen *et al.* (1991) reported similar results using RFLP analysis of the rRNA gene repeats, showing that 14 isolates of *G. intestinalis* had identical banding patterns. In contrast, Upcroft & Upcroft (1994) could differentiate genotypes in a mixed infection on the basis of RFLP analysis of rRNA genes.

DNA fingerprinting of *G. intestinalis* isolates has received relatively little attention compared to some of the other approaches. The first report was by Upcroft *et al.* (1990a), who used the bacteriophage M13 as a hybridisation probe to detect minisatellite polymorphisms and found the method to be highly discriminatory. The same technique was used to characterise cloned lines established from a mixed infection (Upcroft & Upcroft, 1994) and was able to discriminate between clones that appeared identical by RFLP analysis of rRNA gene repeats. Butcher *et al.* (1994) also used M13 DNA fingerprinting to demonstrate that a patient with chronic giardiasis had a mixed infection of two different strains of *G. intestinalis*. These strains differed in their sensitivity to metronidazole and mepacrine. The hyper-variability of these minisatellite regions, as demonstrated by Upcroft & Upcroft (1994), raises questions as to their suitability as genetic markers for epidemiological studies, especially when tracing infection patterns over time. This may be a reason why DNA fingerprinting has not been widely adopted for studying *Giardia*.

RAPD analysis has similarly been applied to the study of *G. intestinalis* in only several instances, and has been the only type of DNA-based data used in genetic analyses (Morgan *et al.*, 1993). Depending upon the primers used, RAPD-PCR has detected polymorphisms between different isolates as well as between cloned lines of the same isolate, whereas inter-repeat PCR only resolved differences between isolates (van Belkum *et al.*, 1993). This result, if taken at face value, could indicate that these markers may be evolving too rapidly to be reliable genetic markers for epidemiological studies and strain authentication. An alternative

explanation for the variation between cloned lines of the same isolate could be that the original isolate contained a mix of different genotypes. Mixed (or co-) infections appear to be a relatively common occurrence (Andrews *et al.*, 1992b; Butcher *et al.*, 1994; Mayrhofer *et al.*, 1992; Upcroft & Upcroft, 1994), and cannot reliably be detected by RAPD-PCR (discussed in section 4.1.2). Evidence to support this hypothesis is the observation that, using the same RAPD primer, cloned lines of the isolate Nij-2 were monomorphic and had identical profiles compared to the parental strain, whereas cloned lines of the isolate Nij-5 were highly variable, with only one having a similar profile to the parental strain (van Belkum *et al.*, 1993). If RAPD do undergo a rapid rate of change, then they may prove to be extremely useful for the examination of the evolution of strains and strain population structure, as well as for the study of localised outbreaks or recent epidemics of giardiasis. Genetic analysis of RAPD banding patterns have been shown by Morgan *et al.* (1993) to correlate with the groupings determined by isozyme analysis. However, the long-term stability of the banding patterns of the isolates was not examined.

The amount of DNA sequence information available on genes from *G. intestinalis* has increased substantially over the last 5 years. However, there has been no attempt to investigate the phylogenetic relationships within *G. intestinalis* using sequence data. Instead, attention has been confined to the identification of sequences with diagnostic utility (eg. Weiss *et al.*, 1992), or the evolutionary relationships of *Giardia* with other eukaryotes (eg. Hashimoto *et al.*, 1992; Sogin *et al.*, 1988). Segments of the rRNA repeat, especially the small sub-unit (hereafter referred to as 18S rRNA, as it is equivalent to the 18S rRNA of other eukaryotes despite its size being closer to 16S), have been used to compare isolates of *G. intestinalis*, as well as the morphological species of *Giardia*. van Keulen *et al.* (1993) sequenced the complete 18S rRNA gene for single isolates of *Hexamita sp.*, *G. ardeae*, *G. muris*, and *G. intestinalis*, and compared these to 18S rRNA sequences from other diverse eukaryotes and prokaryotes. Phylogenetic analysis using the NJ algorithm suggested that *G. ardeae* was more closely related to *G. intestinalis* than *G. intestinalis* was to *G. muris*. Rather surprisingly, the Diplomonadida was not found to be a monophyletic group.

A limited number of *G. intestinalis* isolates have also been sequenced for 4 other conserved genes. These encode elongation factors 1 and 2, DNA dependent RNA polymerase and β -tubulin. Analysis of elongation factor and RNA polymerase gene sequences placed *G. intestinalis* as the earliest branch of the eukaryotic lineage (Hashimoto *et al.*, 1992, 1994, 1995; Keeling & Doolittle 1996). Phylogenetic analyses of the deduced amino acid sequences

of elongation factor 1 alpha (EF1 α) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) both supported monophyly of the Diplomonadida (Keeling & Doolittle *et al.*, 1996; Rozario *et al.*, 1996). It has been suggested that the high GC content of the 18S rRNA gene in *Giardia* may bias comparisons with other distantly related taxa (Hashimoto *et al.*, 1992).

A 183-bp segment of the 18S rRNA gene has been used by Weiss *et al.* (1992) to compare 35 axenic *G. intestinalis* isolates. These isolates could be divided into 3 groups based upon two informative nt substitutions. Aside from the 2 informative positions, the rest of the sequences were identical in all isolates. More recently, van Keulen *et al.* (1995) sequenced the entire 18S rRNA gene for 4 isolates of *G. intestinalis*, one of which was an Assemblage A genotype (Portland 1), and another an Assemblage B genotype (AMC-4) (Monis *et al.*, 1996, Appendix D). The maximum sequence divergence (uncorrected) between the isolates was only 3% over the 500 bp used in the comparison. A three nt signature sequence was identified that distinguished Portland 1 from AMC-4, and PCR primers were constructed to this region, allowing the specific amplification of either Assemblage A ($\equiv$ Polish) or Assemblage B ($\equiv$ Belgian) genotypes. Examination of the spacer region between the 18S and 28S rRNA genes indicated that it was highly variable, having less than 61% identity between isolates from the separate assemblages, suggesting that it may have utility for strain identification (van Keulen *et al.*, 1995).

To my knowledge, the only other sequence-based study of genetic heterogeneity within *G. intestinalis* has been by Baruch *et al.* (1996), who used PCR-nucleotide sequencing of the gene encoding triose phosphate isomerase (TPI) to characterise 16 isolates. These isolates were placed within three genetic groups corresponding to the classification of Nash and Mowatt (1992). Less than 1% (presumably uncorrected p-distance) sequence divergence was observed within groups, while groups 1 and 3 were separated by 18% sequence divergence. An RFLP system was also described which could differentiate groups 1 and 3 (Baruch *et al.*, 1996). The finding of identity between human- and animal-derived isolates (eg., WOOF) was of interest, especially in light of the AE results (section 3.4.3.2) which found WOOF to be genetically distinct from all of the Australian human-derived *G. intestinalis* isolates. Three possible explanations for this observation are that *tpi* is too conserved to distinguish between *G. intestinalis* isolates belonging to the same assemblage, that WOOF is truly identical to the North American human-derived isolates which represent genotypes not previously detected in Australia, or that there has been contamination of the axenic WOOF culture.

4.3 General Aims

A review of the literature indicated that a comprehensive phylogenetic analysis of the *G. intestinalis* species complex had not been attempted, and that most attention had been given to the development of diagnostic markers for strain identification. All of the isolates thus far characterised using DNA-based techniques belong to either one of the two genetic groups which are equivalent to Assemblages A and B (van Keulen *et al.*, 1995; Monis *et al.*, 1996, Appendix D). Allozyme analysis (Figure 23, p. 101) has identified discrete lineages of *G. intestinalis* isolates which are external to Assemblages A and B. However, the relationships among these lineages could not be resolved by this technique. Phylogenetic analysis using a different technique will allow an independent test of the AE model of genetic relationships, may increase our understanding of the evolution of these lineages and provide a stronger foundation upon which to build the systematics of the *G. intestinalis* species complex.

The aim of this part of the project was, therefore, to use sequence-based methods to further investigate the phylogenetic relationships of the lineages identified by allozyme analysis (Table 14, p. 87). This aim will be achieved through the sequence analysis of conserved structural and “housekeeping” genes.

Specific aims:

- To identify and assess genes with utility for the genetic characterisation of *G. intestinalis*.
- To use these genes to:
 - test the validity of the allozyme-inferred genetic relationships.
 - determine the phylogenetic relationships of *G. intestinalis* isolates.
 - identify diagnostic probes for the rapid identification of isolates.

4.4 Phylogenetic analysis using 18S rRNA and *efl* α nucleotide sequences

The magnitude of the allozymically-inferred distances separating the lineages comprising *G. intestinalis* (Figure 23, p. 101) is similar to the levels separating the different morphological species of *Giardia*. Statistical analyses of AE allele frequency data (Figures 12 & 17, pp. 81 & 90) indicated that the divergence between these lineages was beyond the phylogenetic resolution of the technique. The results of van Keulen *et al.* (1991) and Weiss *et al.* (1992) suggested that the 18S rRNA gene is highly conserved within *G. intestinalis*. However, the comparisons made by van Keulen *et al.* (1993), which indicated a 16% sequence divergence between *G. intestinalis* (consensus sequence from 5 different isolates) and *G. ardeae* and a 24% sequence divergence between *G. muris* and *G. intestinalis*, would seem to indicate otherwise. These levels of intra-genus divergence are quite high, considering that different genera of ciliates may have less than 1% 18S rRNA sequence divergence between the species comprising them (Schlegel *et al.*, 1991), and that different genera of vertebrates (eg., *Rattus* versus *Mus*) may have less than 5% sequence divergence between them (based on the comparison of sequences available from GenBank, accession numbers X01117 and V00851). Since the divergence between the lineages comprising the *G. intestinalis* species complex is similar to that observed between *G. ardeae* and *G. intestinalis* (from the AE data, Figure 23, p. 101), it was of particular interest to make the same comparisons using the 18S rRNA sequence. This would provide additional information about the relationships within *G. intestinalis*, and also give an indication of the accuracy of the allozyme distance estimates.

To control for the determination of a gene tree rather than species tree, another conserved gene was analysed. The results of Hashimoto *et al.* (1994) suggested that the gene encoding elongation factor 1 alpha (EF1 α) would be a good candidate. It is a highly conserved single copy gene, and more importantly, a *G. intestinalis* DNA sequence was available to aid in the construction of PCR primers. Other potential candidates included the genes encoding elongation factor 2, which is thought to have arisen via an ancient gene duplication event from *efl* α (Hashimoto *et al.*, 1995), ubiquitin, and the tubulins. After consideration of the available sequences these loci were not pursued. Another reason for analysing a second conserved gene was to verify whether *G. ardeae* was the most appropriate outgroup for *G. intestinalis*, as suggested by the results of van Keulen *et al.* (1993).

4.4.1 Specific aims

- To examine the phylogenetic relationships within *G. intestinalis* using 18S rRNA gene sequences.
- To examine the phylogenetic relationships within *G. intestinalis* using EF1 α gene sequences.
- To determine the most appropriate outgroup for *G. intestinalis*.

Hypotheses to test:

- The nucleotide sequence analyses are concordant with the allozyme-inferred genetic relationships of the *G. intestinalis* isolates.
- *G. ardeae* is genetically more similar to *G. intestinalis* than *G. muris* is to *G. intestinalis*.

Null hypotheses:

- The inferred phylogenies from analysis of the 18S rRNA and EF1 α genes are the same.
- There is no variation between *G. intestinalis* isolates in the 18S rRNA gene.

4.4.2 Materials and methods

4.4.2.1 List of isolates

Isolates were chosen to represent the major lineages of *G. intestinalis* identified by AE analysis (Table 14, p. 87), and are listed in Table 17 (p. 114). The rat-derived *G. intestinalis* isolates which comprise the Rat lineage were late inclusions in the AE analysis, and regrettably had to be omitted from the DNA analyses because insufficient samples were available.

4.4.2.2 18S rRNA primer design

The published 18S rRNA sequences of van Keulen *et al.* (1991) were obtained for *G. muris*, *G. ardeae* and *G. intestinalis* from the GenBank database and aligned using ClustalW (Thompson *et al.*, 1994). Conserved regions were identified, and candidate primers were examined for any undesirable characteristics using the computer program Primer (section 2.4). The primer sequences for G18S2 and G18S3 are shown in Figure 24 (p. 114). The source of the oligonucleotides used are briefly described in section 2.1.1.

Table 17. List of isolates.

Species	Lineage	Isolate	Host	Location	Propagation method
<i>G. intestinalis</i>	Assemblage A	Ad-1	human	Adelaide, SA	axenic
	Assemblage A	Ad-2	human	Adelaide, SA	axenic
	Assemblage B	Ad-7 ¹	human	Adelaide, SA	s/m
	Assemblage B	Ad-19	human	Adelaide, SA	s/m
	Assemblage B	Ad-28	human	Adelaide, SA	s/m
	Cat	Ad-23	cat	Adelaide, SA	s/m
	Assemblage B	WOOF	dog	Canada	axenic
	Dog2	Ad-148	dog	Ceduna, SA	s/m
	Dog1	Ad-147	dog	Alice Springs, NT	s/m
	Dog1	Ad-137	dog	Port Augusta, SA	s/m
	Dog1	Ad-136	dog	Alice Springs, SA	s/m
	Livestock	P-15	pig	Czech Republic	s/m
<i>G. ardeae</i>		<i>G. ardeae</i>	blue heron	Minnesota, USA	axenic
<i>G. muris</i>		Underdown	hamster	North America	s/m
		Ad-120	mouse	Adelaide, SA	s/m
		Ad-160	mouse	Adelaide, SA	s/m

key:

s/m = suckling mouse

SA = South Australia

NT = Northern Territory

1. Mayrhofer *et al.*, 1995**Figure 24.** Design of 18S rRNA PCR primers G18S2 & G18S3.

G. intestinalis TCCGGTCGATCCTGCC
G. ardeae TCCGGTTGATCCTGCC
G. muris TCCGGTTGATCCTGCC
G18S2 5' **TCCGGTYGATCCTGCC** 3' →

G. intestinalis AGCAGCCGCGGTAATTCCAG
G. ardeae AGCAGCCGCGGTAATTCCAG
G. muris AGCAGCCGCGGTAATTCCAG
G18S3 3' **TCGTCGGCGCCATTAAGGTC** 5' ←

4.4.2.3 Optimised 18S rRNA PCR conditions

DNA was extracted from isolates of *Giardia* using the methods described in section 2.3.1, and then used as templates in amplification reactions (hereafter the term “isolate”, when used in relation to PCR or sequencing, will refer to the isolate’s DNA). All PCR’s were performed on a Perkin Elmer GeneAmp PCR System 2400 (94°C for 2 min, then 35 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 1 min, with a final 72°C extension for 7 min), using 0.2 ml thin

4.4.2.5 Optimised *efl*α PCR conditions

Amplifications (94°C for 2 min, then 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min, with a final 72°C extension for 7 min) were performed as described in section 4.4.2.3 but using 5% DMSO (final conc.) and 0.8 μM of the glongF and glongR primers.

4.4.2.6 DNA sequencing

DNA sequencing was carried out as described in section 2.4.9.

4.4.2.7 Phylogenetic analyses

As discussed in the introduction, various approaches can be used for phylogenetic inference. These methods have different sets of assumptions and their performance, in terms of recovering the correct tree, can vary depending upon the conditions in which they are used (eg., Kuhner & Felsenstein, 1994; Nei, 1991). Despite the dogged advocacy that some workers may hold for a particular method (eg., Miyamoto & Cracraft, 1991), the approach most commonly used today is to analyse data by a variety of methods. This approach cannot ensure that the correct phylogeny will always be recovered (the data may be insufficient to resolve the relationships of interest, or may violate the assumptions of all of the methods). However, it allows a thorough examination of the data by different methods which in some ways can be complementary. For these reasons, all of the sequence data obtained in these experiments were analysed using parsimony, distance, and maximum likelihood methods. The phylogenetic analysis of more than one gene should allow the detection of data that is biased towards supporting an incorrect tree. This can occur due to phenomena such as the “Felsenstein effect” (also known as “long branch attraction”, discussed by Swofford *et al.*, 1996), which is caused by unequal rates of evolution among lineages. The software packages used for this purpose are described section 2.4.

Another issue needing consideration was how the data from different sources should be treated. Essentially, the question which needs to be addressed is: should different sets of data be combined? There are currently three schools of thought regarding this issue: (1) to always combine data, (2) to analyse the data sets separately and then compare the independent phylogenies, and (3) to conditionally combine the data provided separate analysis of each data set yields a congruent result (reviewed by Huelsenbeck *et al.*, 1996). Each approach has

advantages and disadvantages. Combining data sets can provide more characters for analysis, but rate heterogeneity between sets can affect the accuracy of the inferred relationships. Separate analyses can avoid this problem, although each individual data set will contain fewer characters and potentially have a reduced likelihood of recovering the true phylogeny. The approach adopted for this study was the conditional combination of data sets, which only combines compatible data partitions after individual analysis of each data set (Huelsensbeck *et al.*, 1996). Separate analyses prior to combining the data can also provide more information about evolutionary processes, such as the relative mutation rates of different classes of genes, and it can also allow the detection of specific events such as recombination (eg., horizontal gene transfer) or introgression between species (S. Donnellan, pers. comm.).

4.4.2.8 Estimation of among-site rate variation

A factor which can affect all methods of phylogenetic analysis is among-site rate variation (Yang, 1996). Simulation studies have shown that all phylogenetic inference methods perform poorly when nt sites in a DNA sequence evolve at very different rates, and that parsimony was particularly sensitive to unequal rates among lineages (Kuhner & Felsenstein, 1994). Among-site rate variation is a well recognised characteristic of protein coding genes, where the 1st, 2nd and 3rd codon positions in various codons evolve at different rates (usually in the order of 3rd > 1st > 2nd). It also occurs within segments of genes encoding functional domains, which would be expected to evolve at a slower rate compared to regions not under any functional constraints. Intuitively, only sequences not under any functional constraints would be expected to have equal rates of variation among nt sites.

The determination of among-site rate variation appears to be a rather circular process, initially requiring a phylogeny to determine the number of substitutions (or steps) which have occurred at each site. The distribution of the number of substitutions per site, which approximate a gamma distribution, can be used as an indicator of the amount of among-site rate variation (reviewed by Yang, 1996). A shape parameter, α , can be calculated for the observed distribution. For $\alpha > 1$, the distribution is bell shaped, indicating that most sites exhibit moderate rates of substitution, and only few sites exhibit very high or very low rates of substitution (Yang, 1996). The distribution approaches a binomial distribution (ie., equal rates among sites) as α approaches infinity. When $\alpha \leq 1$, the distribution is L-shaped, indicating that most sites exhibit very low rates of substitution and a few sites have very high rates of substitution. Values of $\alpha < 0.5$ are considered evidence for extreme rate

heterogeneity. Sequences exhibiting extreme among-site rate variation are likely to be saturated for change, and therefore not phylogenetically informative.

The α parameter was estimated from the data presented herein using the computer program PUZZLE (Strimmer & von Haesler, 1996; Strimmer *et al.*, 1997), which uses quartet puzzling and the maximum likelihood algorithm to determine phylogenetic relationships. The program uses an iterative process to estimate the parameters in the analysis, such as α (if the gamma model is used) and the transition/transversion ratio. The Tamura Nei substitution model was used, with the approximate parameter estimation using both quartet sub-sampling and the Neighbour Joining tree options in effect. All sites were counted as variable and 5 categories were used for the gamma model. Inaccurate estimates of α can be made as a consequence of the high similarity which can occur between closely related sequences, which will also result in an L shaped distribution even though there may be no significant among-site rate variation. To test this, the constant sites were removed by counting them as invariant sites (as opposed to sites with very low rates of change), and α was then re-calculated.

4.4.3 Phylogenetic analysis of 18S rRNA nucleotide sequences

4.4.3.1 Optimisation of 18S rRNA PCR

The high GC content of the 18S rRNA gene of *G. intestinalis*, (approximately 76% based upon the published sequence, van Keulen *et al.*, 1993) and the consequent likelihood of secondary structures, required that high annealing temperatures and a co-solvent be used for the amplification of DNA from this gene by PCR. Using an annealing temperature of 55°C allowed the successful amplification of the expected sized fragment from *G. muris* DNA, but not from *G. intestinalis* DNA. Increasing the annealing temperature and inclusion of DMSO (5% final concentration) in the reaction mixture improved the specific amplification of the 480-bp segment from the *G. muris* samples, but it made no improvement for any of the *G. intestinalis* DNA samples tested. Increasing the DMSO concentration to 10% (final) was necessary to amplify the expected sized fragment from the *G. intestinalis* DNA samples (lanes 1 & 4, Figure 26, p. 120).

4.4.3.2 Detection of contaminating host DNA in suckling mouse derived isolates

In preliminary experiments aimed at optimising the PCR conditions for the amplification of the 480-bp 18S rRNA gene segment, it was observed that many of the samples, including *G. muris*, yielded amplification products that were 200-bp larger than the expected sized product. The majority of *G. intestinalis* isolates propagated in suckling-mice amplified weak to moderate amounts of a single 700-bp product, while the *G. muris* isolates amplified both the expected 480-bp product in addition to the 700-bp product. The axenic isolates which were examined did not amplify a 700-bp product. This pattern suggested that the 700-bp product may have been the result of amplification from contaminating host DNA. Examination of the published mouse 18S rRNA gene sequence indicated that the amplified product was of the correct size if the primers were amplifying the mouse 18S rRNA gene. Sequence analysis of the purified 700-bp product confirmed that this was the case (not shown). Although the primers G18SF & G18SR were more similar to the *Giardia* 18S rRNA sequences than to the homologous mouse gene, the mouse 18S rRNA gene was amplified preferentially over the *G. intestinalis* sequences, even at an annealing temperature of 60°C and using 5% DMSO (not shown). This situation was reversed for *G. muris* isolates, which only amplified the *G. muris* 18S rRNA gene under these conditions. An increase in the DMSO concentration to 10% allowed the amplification of the *G. intestinalis* 18S rRNA gene segment, with the mouse-derived product no longer detectable (Figure 26 p. 120). The failure to amplify the *G. intestinalis* 18S rRNA gene segment appears to be due to secondary structure, rather than preferential amplification of the mouse product. Dilution of suckling-mouse derived *G. intestinalis* and *G. muris* DNA samples resulted in the loss/lower yield of the contaminating product in all samples (Figure 27, p. 120), while dilution of the *G. muris* DNA sample did not visibly affect the yield of the *G. muris* 18S rRNA product. The diluted *G. intestinalis* DNA samples failed to amplify any product. This suggests that the mouse DNA was unlikely to be responsible for inhibiting the amplification of *G. intestinalis* DNA.

4.4.3.3 Alignment of 18S rRNA nucleotide sequences

The amplified 18S rRNA fragments were directly sequenced for each isolate, and then aligned using ClustalW. The three published *Giardia* sequences from van Keulen *et al.* (1991) were also included. The *G. psittaci* 18S rRNA sequence was kindly provided by Dr. Harry van Keulen (Department of Biology, Cleveland State University). *Hexamita* was not

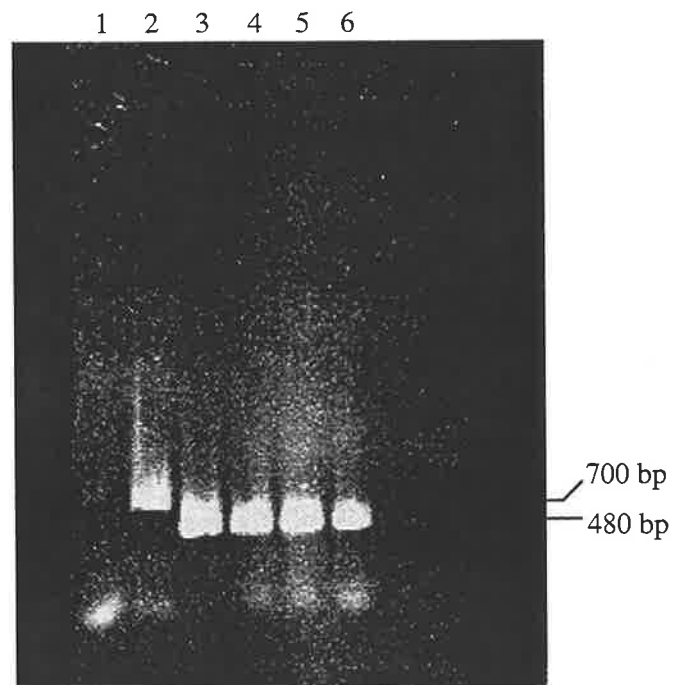


Figure 26. PCR amplification of *Giardia* DNA samples using primers G18S2 & G18S3. Lanes 1 - 3: *G. intestinalis* (Ad-2, axenic), *G. intestinalis* (Ad-7, suckling-mouse derived) & *G. muris* (Ad-120, suckling mouse derived) using 5% DMSO. Lanes 4 - 6: were the same isolates using 10% DMSO.

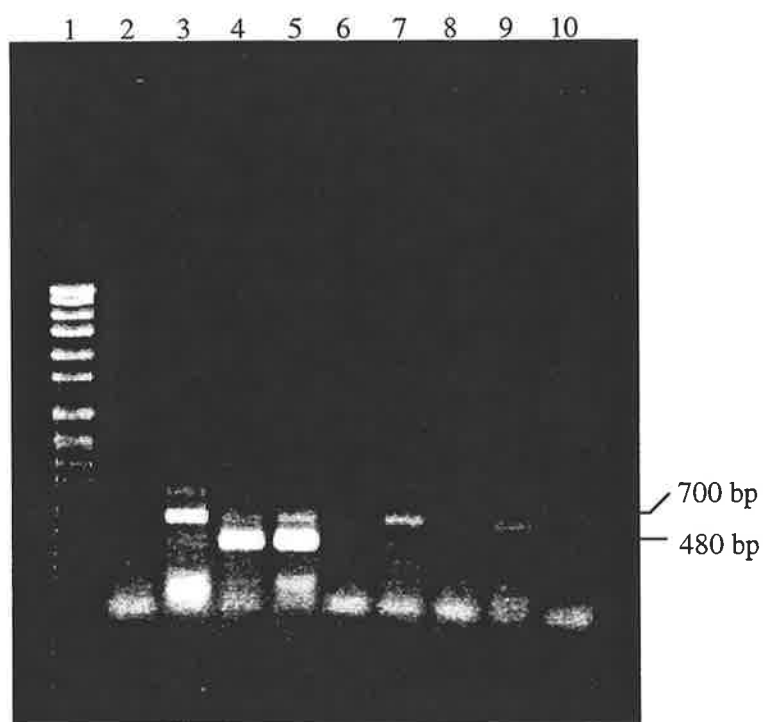


Figure 27. 18S rRNA PCR amplification using primers G18S2 & G18S3 and different amounts of template DNA. Reaction conditions were modified from section 4.4.2.4 to allow amplification of contaminating mouse DNA (55°C annealing temperature, no DMSO added). Lane 1: size marker *Eco*R1 digested bacteriophage SPP-1. Lanes 2 & 3: Ad-156 (Assemblage B) 16 ng or 400 ng. Lanes 4 & 5: Ad-160 (*G. muris*) 12 ng or 300 ng. Lanes 6 & 7: Ad-23 (Cat) 4 ng or 100 ng. Lanes 8 & 9: Ad-7 (Assemblage B) 8 ng or 200 ng. Lane 10: negative PCR control.

included in the alignment because it was far too divergent from the *Giardia* sequences (ca. 37% p-distance, van Keulen *et al.*, 1993) to be a reliable outgroup for *Giardia*. The alignments were edited to remove the primer sequences, and manually adjusted to optimise the position of any postulated indels. Additional published sequences retrieved from GenBank were later included as they were made available (eg., CM & AMC from van Keulen *et al.*, 1995). In instances where isolates were identical (eg., Ad-136, Ad-137, Ad-147), one was chosen as a representative and the rest were excluded. The final alignment used for analysis is shown in Figure 28 (pp. 122-123). Pairwise comparisons of the sequences revealed slightly less similarity (72% - 71%) between *G. muris* and *G. intestinalis* than the 75% reported by van Keulen *et al.* (1993). The comparisons between *G. intestinalis* and *G. ardeae* were similar. This may be a consequence of the shorter sequence length used in this comparison, or slight differences in the alignments used. The 18S rRNA sequence similarity between isolates of *G. intestinalis* was >97%.

Another interesting feature was the difference detected between the published *G. muris* sequence and the three *G. muris* isolates characterised herein. The North American (Underdown) and Australian (Ad-120, Ad-160) isolates had identical 18S rRNA sequences which differed from the published sequence at six positions, three of which were indels (positions 19, 94 & 431 in Figure 28, pp. 122 - 123), and three nt substitutions (transversions at sites 87 & 373 and a transition at 88, pp. 122 - 123). All three indels occurred at repeated bases, and could be due to the different efficiencies of the sequencing methods used, ie., single stranded (Sanger method) compared with cycle sequencing. The indel at position 431 appears unlikely to be just such an artefact because similar differences were also observed between isolates which had been cycle sequenced (eg., P-15 versus Ad-147). A similar difference was also observed between the published Assemblage B ($\equiv$ Polish) sequences and the ones used herein (Ad-7 & Ad-19) at position 351 and involved a GC di-nucleotide repeat of 5 repeats in length. The *G. psittaci*, Ad-7, Ad-19 and WOOF isolates all possessed the full repeat, whereas the rest of the *G. intestinalis* isolates possessed an extra G in the third repeat.

Examination of the 18S rRNA three nt signature sequence identified by van Keulen *et al.* (1995) (Figure 28, p. 122) revealed that it would not be a reliable marker for the "Polish" (Assemblage A) or "Belgian" (Assemblage B) genotypes because the AUC triplet of the "Belgian" isolates is shared with *G. psittaci* and with the *G. intestinalis* isolates from the Dog1 and Dog2 lineages. The GCG triplet of the Polish isolates is also possessed by *G. intestinalis* from the Livestock lineage and *G. ardeae*.

Figure 28. Alignment of *Giardia* 18S rRNA nucleotide sequences. The 3-nt signature sequence of van Keulen *et al.* (1995) is indicated in bold type.

Portland 1 ¹	GGAG CG CGACGCTCTCCCAAGGACG-AAGCCATGCATGCCCGCTCACC CGG--ACGCGGCGGACGGCTCAGGACAACGGTTGCACCCCCCGCGGCGGT	100
Ad-7 ²	GGA AT CCGACGCTCTCCCAAGGACACAAGCCATGCATGCCCGCGCACCCGGG--AGGCGGCGGACGGCTCAGGACAACGGTTGCACCCCCCGCGGCGGT	
CM ¹	GGA AT CCGACGCTCTCCCAAGGACACAAGCCATGCATGCCCGCGCACCCGGG--AGGCGGCGGACGGCTCAGGACAACGGTTGCACCCCCCGCGGCGGT	
AMC-4 ¹	GGA AT CCGACGCTCTCCCAAGGACACAAGCCATGCATGCCCGCGCACCCGGG--AGGCGGCGGACGGCTCAGGACAACGGTTGCACCCCCCGCGGCGGT	
Ad-23	NNNNNNNACGCTCTCCCA--GGACG-AAGCCATGCATGCCCGCTCACC CGG--ACGCGGCGGACGGCTCAGGACAACGGTTGCACCCCCCGCGGCGGT	
P-15	GGAG CG CGACGCTCTCCCAAGGACG-AAGCCATGCATGCCCGCTCACC CGG--ACGCGGCGGACGGCTCAGGACAACGGTTGCACCCCCCGCGGCGGT	
Ad-147 ³	GGA AT CCGACGCTCTCCCAAGGACACAAGCCATGCATGCCCGCACACC CGG--AGGCGGCGGACGGCTCAGGACAACGGTTGCACCCCCCGCGGCGGT	
Ad-148	GGA AT CCGACGCTCTCCCAAGGACACAAGCCATGCATGCCCGCACACC CGG--AAGCGGCGGACGGCTCAGGACAACGGTTGCACCCCCCGCGGCGGT	
<i>G. psittaci</i> ⁴	GGA AT CCGACGCTCTCTCAAGGACGAAGCCATGCATGCCCGCACACC CGG--AAGCGGCGGACGGCTCAGGACGACGGTTGCACCCCCCGCGGCGGT	
<i>G. ardeae</i> ¹	GGAG CG CGACGCTCGCCCCAAGGACACAAGCCATGCACGCCGGCGCGGCACGGCGAAGGGGCGGACGGCTCAGTACACCGTTGCACCTGCGCGGCGGT	
<i>G. muris</i> ¹	GGAG T ACTACGCTACCC--AAGGACAAAAGCCATGCAAGCGGACACGAGGTAT-GAAGTGGCGGACGGCTCGGTACAACGGTACGATCCTGAC-GGGGGT	
<i>G. muris</i> ⁵	GGAG T ACTACGCTACCCCAAGGACAAAAGCCATGCAAGCGGACACGAGGTAT-GAAGTGGCGGACGGCTCGGTACAACGGTACGAGTCTGACCGGGGT	
Portland 1 ¹	CCCTGCTAGCCGGACACCGCTGGCAACCCGGCGCCAAGACGTGCGCGCAAGGGCGGGCGCCCGCG-GGCGAGCAGCGTGACGCAGCGACGGCCCGCCCGG	200
Ad-7 ²	CCCTGCTAGCCGGACACCGCTGGCAACCCGGCGCCAAGACGTGCGCGCAAGGGCGGGCGCCCGCG-GGCGAGCAGCGTGACGCAGCGACGGCCCGCCCGG	
CM ¹	CCCTGCTAGCCGGACACCGCTGGCAACCCGGCGCCAAGACGTGCGCGCAAGGGCGGGCGCCCGCG-GGCGAGCAGCGTGACGCAGCGACGGCCCGCCCGG	
AMC-4 ¹	CCCTGCTAGCCGGACACCGCTGGCAACCCGGCGCCAAGACGTGCGCGCAAGGGCGGGCGCCCGCG-GGCGAGCAGCGTGACGCAGCGACGGCCCGCCCGG	
Ad-23	CCCTGCTAGCCGGACACCGCTGGCAACCCGGCGCCAAGACGTGCGCGCAAGGGCGGGCGCCCGCG-GGCGAGCAGCGTGACGCAGCGACGGCCCGCCCGG	
P-15	CCCTGCTAGCCGGACACCGCTGGCAACCCGGCGCCAAGACGTGCGCGCAAGGGCGGGCGCCCGCG-GGCGAGCAGCGTGACGCAGCGACGGCCCGCCCGG	
Ad-147 ³	CCCTGCTAGCCGGACACCGCTGGCAACCCGGCGCCAAGACGTGCGCGCAAGTGC GGCGCCCGCG-GGCGAGCAGCGTGACGCAGCGACGGCCCGCCCGG	
Ad-148	CCCTGCTAGCCGGACACCGCTGGCAACCCGGCGCCAAGACGTGCGCGCAAGTGC GGCGACGCCCGCG-GGCGAGCAGCGTGACGCAGCGACGGCCCGCCCGG	
<i>G. psittaci</i> ⁴	CCCTGCTAGACGGACAACGCTGGCAACCCGGCGCCAAGACGTGCGCG-AAGTGC GCCCGCCCCCGCGGCGAGCAGCGCAACGCAGCGACGGGCCGCCCGG	
<i>G. ardeae</i> ¹	CGGGGCTAGCCGGACACCGCTGGCAACCCAGCGACAAGACGAGTGCTCAAGAGCGGGGA-----GGGAGCCGGCGCGACGCAGCGAATGGCCGCTGG	
<i>G. muris</i> ¹	GAAGGCTAGACGGATACCGCTGGCAACCCAGCGCCAAGACGAGTGCTCAAGAGCGGGGA-----AGGAAAGCACGCGATGGAGCGAATGCCCGGATGA	
<i>G. muris</i> ⁵	GAAGGCTAGACGGATACCGCTGGCAACCCAGCGCCAAGACGAGTGCTCAAGAGCGGGGA-----AGGAAAGCACGCGATGGAGCGAATGCCCGGATGA	
Portland 1 ¹	GCTTCCGGGGCATCACCCGGTCGGCGCGGTTCGCGGCGCGCCGAGGGCCCCGACGCCTGGCGGAGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGCGAGAC	300
Ad-7 ²	GCTTCCGGGGCATCACCCGGTCGGCGCGGTTCGCGGCGCGCCGAGGGCCCCGACGCCTGGCGGAGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGAGAGAC	
CM ¹	GCTTCCGGGGCATCACCCGGTCGGCGCGGTTCGCGGCGCGCCGAGGGCCCCGACGCCTGGCGGAGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGAGAGAC	
AMC-4 ¹	GCTTCCGGGGCATCACCCGGTCGGCGCGGTTCGCGGCGCGCCGAGGGCCCCGACGCCTGGCGGAGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGAGAGAC	
Ad-23	GCTTCCGGGGCATCACCCGGTCGGCGCGGTTCGCGGCGCGCCGAGGGCCCCGACGCCTGGCGGAGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGCGAGAC	
P-15	GCTTCCGGGGCATCACCCGGTCGGCGCGGTTCGCGGCGCGCCGAGGGCCCCGACGCCTGGCGGAGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGCGAGAC	
Ad-147 ³	GCTTCCGGGGCATCACCCGGTCGGCGCGGTTCGCGGCGCGCCGAGGGCCCCGACGCCTGGCGGAGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGAGAGAC	
Ad-148	GCTTCCGGGGCATCACCCGGTCGGCGCGGTTCGCGGCGCGCCGAGGGCCCCGACGCCTGGCGGAGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGAGAGAC	
<i>G. psittaci</i> ⁴	GCTTCCGCGGCATCACCTGGTCGGCGCGGTTCGCGGCGGGCCGAGGGCGCGACGCCTGGCGGAGAATCAGGGTTCGACTCCGGAGAGCGGGCCTGAGAGAC	
<i>G. ardeae</i> ¹	GCTTCCGCGGCATCACCCCGCGGGCACGGTCGCGGCGTGCCCGGGGTCTACGCCTGGCGGAGGATCAGGGTTCGACTCCGGAGAGCGGGCCTGAGAGAC	
<i>G. muris</i> ¹	GGTTCCGAGGTATTACCTAGTCGGTAGAGTAGTGGTCTACGGAGGGGATGATGCCTGGCGGAGGATCAGGGTTTGACTCCGGAGAACGGGCCTGAGAGAC	
<i>G. muris</i> ⁵	GGTTCCGAGGTATTACCTAGTCGGTAGAGTAGTGGTCTACGGAGGGGATGATGCCTGGCGGAGGATCAGGGTTTGACTCCGGAGAACGGGCCTGAGAGAC	

Figure 28. Alignment of *Giardia* 18S rRNA nucleotide sequences.

Portland 1 ¹	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG-CGCGCGAGGCAGCGACGGGGAGCGCGC-----GAGCGAGGCGGGCC	400
Ad-7 ²	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG--CGCGCGAGGCAGCGACGGGGAGTGCGC-----GAGCGAGGCGGGCC	
CM ¹	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG-CGCGCGAGGCAGCGACGGGGAGTGCGC-----GAGCGAGGCGGGCC	
AMC-4 ¹	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG-CGCGCGAGGCAGCGACGGGGTGTGCGC-----GAGCGAGGCGGGCC	
Ad-23	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG-CGCGCGAGGCAGCGACGGGGAGCGCGC-----GAGCGAGGCGGGCC	
P-15	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG-CGCGCGAGGCAGCGACGGGGAGCGCGC-----GAGCGAGGCGGGCC	
Ad-147 ³	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG-CGCGCGAGGCAGCGACGGGGAGCGCGGA-----GAGCGAGGCGGGCC	
Ad-148	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG-CGCGCGAGGCAGCGACGGGGAGCGCAGA-----GAGCGAGGCGGGCC	
<i>G. psittaci</i> ⁴	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG--CGCGCGAGGCAGCGACGAGGAGTGCCCGGGGCGGAGCGAGGCGGGCG	
<i>G. ardeae</i> ¹	GGCCCCGACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGCGG-GGCGCGAGGCAGCGACGGGGCGTGCCGGGACCCGGCGGGCGTGCGG	
<i>G. muris</i> ¹	GGCCCCGTACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGTGAAGGCGTGAGGCAGCAACGGGGCATCCCATGAAATGGGAAGACTGGGGGG	
<i>G. muris</i> ⁵	GGCCCCGTACATCCAAGGACGGCAGCAGGCGCGGAACCTTGCCCAATGCGTGAAGGCGTGAGGCAGCAACGGGGATCCCATGAAATGGGAAGACTGGGGGG	

Portland 1 ¹	CACAGCCCCCGCCGCGGAGCCGAGGGCAAGGTCTGGTGCC	440
Ad-7 ²	CACAGCCCCCGCCGCGGAGCCGAGGGCAAG-TCTGGNNNN	
CM ¹	CACAGCCCCCGCCGCGGAGCCGAGGGCAAGGTCTGGTGCC	
AMC-4 ¹	CACAGCCCCCGCCGCGGAGCCGAGGGCAAGGTCTGGTGCC	
Ad-23	CACAGCCCCCGCCGCGGAGCCGAGGGCAAG-TCTNNNNNN	
P-15	CACAGCCCCCGCCGCGGAGCCGAGGGCAAGGTCTGGTGCC	
Ad-147 ³	CACAGCCCCCGCCGCGGAGCCGAGGGCAAG-TCTGGTNNN	
Ad-148	CACAGCCCCCGCCGCGGAGCCGAGGGCAAGGTCTGGTGCC	
<i>G. psittaci</i> ⁴	CACAGCCCCCGCCGCGGAGCCGAGGGCAAGGTCTGGTGCC	
<i>G. ardeae</i> ¹	TACAGCCCCACGCGTGAGTCGAGGGAAAGGTCTGGTGCC	
<i>G. muris</i> ¹	TAGATGACCCAGCACAAAGTCGAGGGAAAGGTCTGGTGCC	
<i>G. muris</i> ⁵	TAGATGACCCAGCACAAAGTCGAGGGAAAG-TCTGGTGNN	

1 published sequences (van Keulen *et al.*, 1993; van Keulen *et al.*, 1995).

2 Ad-19, WOOF sequences were identical to Ad-7.

3 Ad-136, Ad-137 sequences were identical to Ad-147.

4 pers. comm. Dr. Harry van Keulen (Department of Biology, Cleveland State University).

5 = Ad-120, Ad-160, Underdown consensus sequence.

4.4.3.4 Determination of 18S rRNA sequence parameters

Prior to phylogenetic analysis, some of the sequence characteristics were examined, such as the nt composition and the extent that multiple nt substitutions had occurred at single sites. The optimal transition/transversion (s/v) ratio also had to be determined as it was required for maximum likelihood (ML) analysis and ML distance estimation. In addition, among-site rate variation in the 18S rRNA sequences was also investigated.

The nt composition of the 18S rRNA sequences of each isolate examined is shown in Table 18. *G. psittaci*, *G. ardeae* and *G. intestinalis* had similar compositions, with GC contents of 75%, 74.5% and 74.6 - 76.6% respectively. The GC content of the *G. muris* 18S rRNA sequence was much lower (60%), largely because of a lower cytosine content. An interesting feature to note was that the adenine content was approximately twice the thymine content in all cases.

The optimal s/v ratio was determined by ML analysis using fastDNAML (section 2.4) on a reduced data set (Table 19, p. 125). Multiple analyses were carried out for different s/v ratios, and the log likelihoods values for the resulting trees were compared (Table 19). The s/v ratio which produced the best (ie., minimum) log likelihood value (in this case, 1) was used as the optimum one.

Table 18. Nucleotide composition (%) of the partial 18S rRNA sequences.

Lineage/Species	Isolate	A	T	C	G	Total
Assemblage A	Portland 1	15.9	7.5	37.1	39.5	428
Assemblage B	Ad-7	16.9	7.4	37.1	38.6	420
Cat	Ad-23	16.0	7.5	37.6	38.8	412
Livestock	P-15	15.7	7.5	37.1	39.7	428
Dog1	Ad-147	17.2	7.8	36.5	38.6	425
Dog2	Ad-148	17.7	7.7	36.5	38.1	430
<i>G. psittaci</i>		17.2	8.0	35.9	38.9	435
<i>G. ardeae</i>		16.7	8.8	34.0	40.5	432
<i>G. muris</i>	Ad-120	26.1	13.8	22.8	37.3	429

Table 19. Log likelihood values for different s/v ratios. Isolates used for analysis were Ad-148, Ad-147, Ad-7, AMC-4, Ad-23, P-15, Portland 1, *G. ardeae*, Ad-120, *G. muris* (published sequence) & *G. psittaci*.

S/V ratio	Log likelihood
0.5	-1406.08043
1	-1402.74657 ← optimal
1.5	-1409.85829
2	-1419.02863
3	-1437.44156

Sequences were examined for multiple nt substitutions at individual sites (commonly referred to as multiple hits, and will be the term used here on) by comparison of corrected (eg., Tamura Nei) and uncorrected (eg., p) distance estimates. Tamura Nei (TN), Kimura 2-parameter (K2p), Jukes Cantor (JC), and p-distances were calculated using MEGA (Kumar *et al.*, 1993), and ML distances (so named because it was originally used in DNAML) were calculated using PHYLIP (Felsenstein, 1993). Corrected distance measures compensate for multiple hits by using a particular model of sequence evolution to calculate a maximum likelihood distance estimate from a pair-wise sequence comparison. In contrast, the p-distance makes no correction for multiple hits, and is the proportion of sites which differ between two sequences. The pair-wise distances estimated using different methods can be directly compared by plotting them against each other. Dot-plots of corrected and uncorrected distances will have a slope of 1 in the absence of multiple hits (ie., the corrected and uncorrected distances for each pair-wise comparison are the same). The occurrence of multiple hits will cause the slope of the curve to deviate from 1, and saturation for change will cause the curve to plateau. The consistency of different corrected distance estimates (which may be affected by the assumed model of sequence evolution) can also be tested by dot-plot comparisons with each other.

Dot-plots of p-distances versus TN distances (Figure 29, p. 126) indicated that multiple hits have occurred at some sites (slope of line $\neq 1$, for distances between *Giardia* species). However, there was no evidence that saturation for change had occurred as the slope did not reach a plateau. For pair-wise comparisons with distances < 0.1 (eg., between isolates comprising different *G. intestinalis* lineages), there was no evidence of multiple hits (slope = 1) from that part of the graph (Figure 29, p. 126). The ML distance estimates were the same as the JC and K2p estimates. However, the TN estimates were slightly larger for distances > 0.2 substitutions/site (Figure 30, p. 126). Distance estimates, using transitions or

transversions only, indicate that multiple hits have occurred in both cases. The TN and K2p distance estimates for transversions only were identical (Figure 31, p. 127), indicating that it was the transition component (Figure 32, p. 127) of the TN estimate which caused it to differ from the other maximum likelihood estimates.

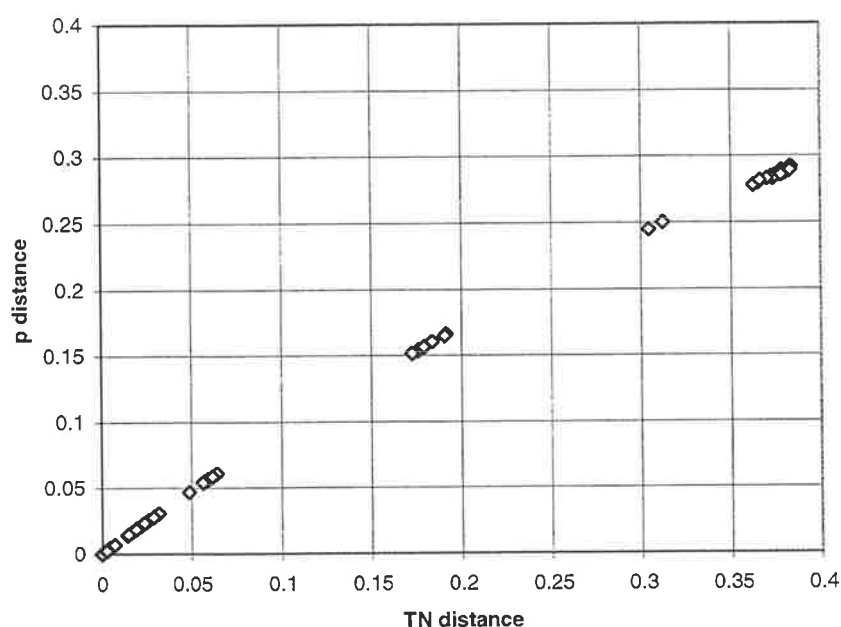


Figure 29. Dot plot of 18S rRNA p-distances versus TN distances calculated for all pair-wise sequence comparisons using all classes of nt substitutions.

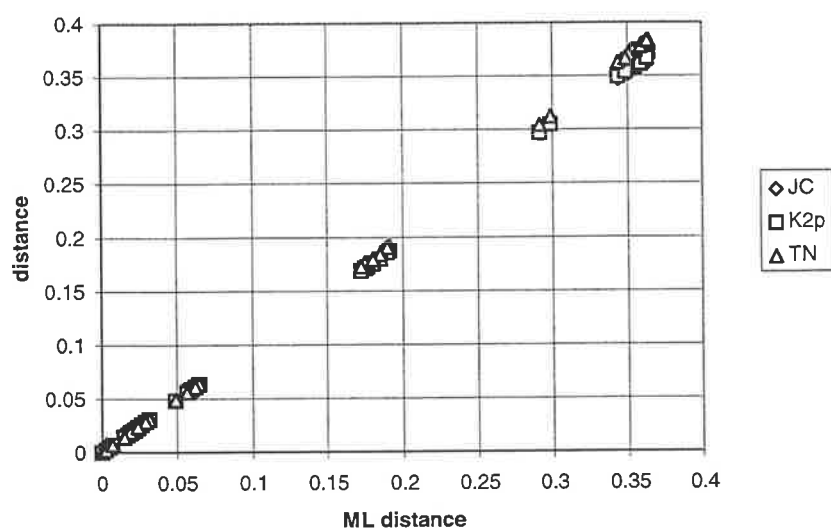


Figure 30. Dot plots of 18S rRNA ML distances versus JC, K2p & TN distances calculated for all pair-wise sequence comparisons using all classes of nt substitutions.

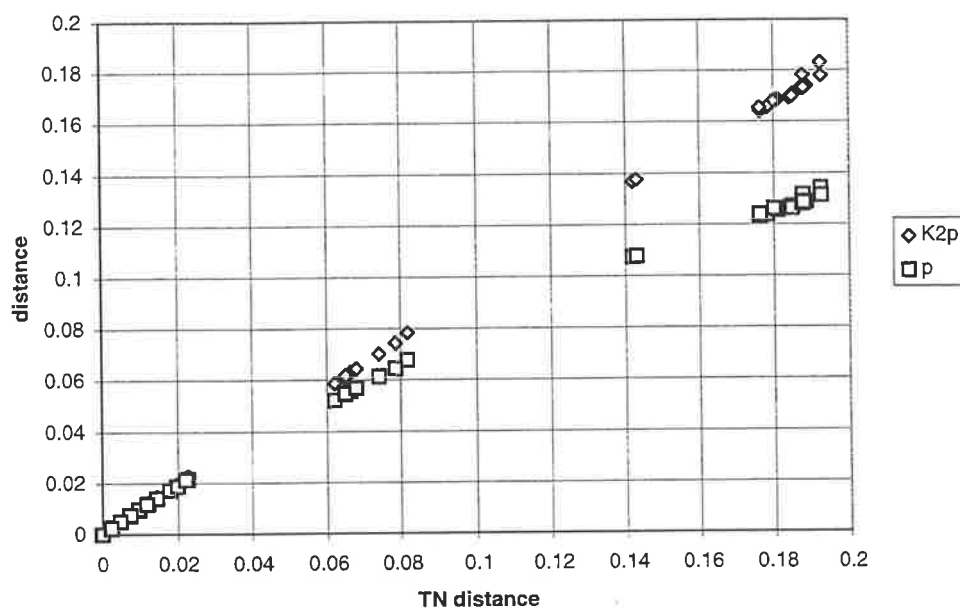


Figure 31. Dot plots of 18S rRNA TN distances versus p and K2p distances calculated for all pair-wise sequence comparisons using transitions only.

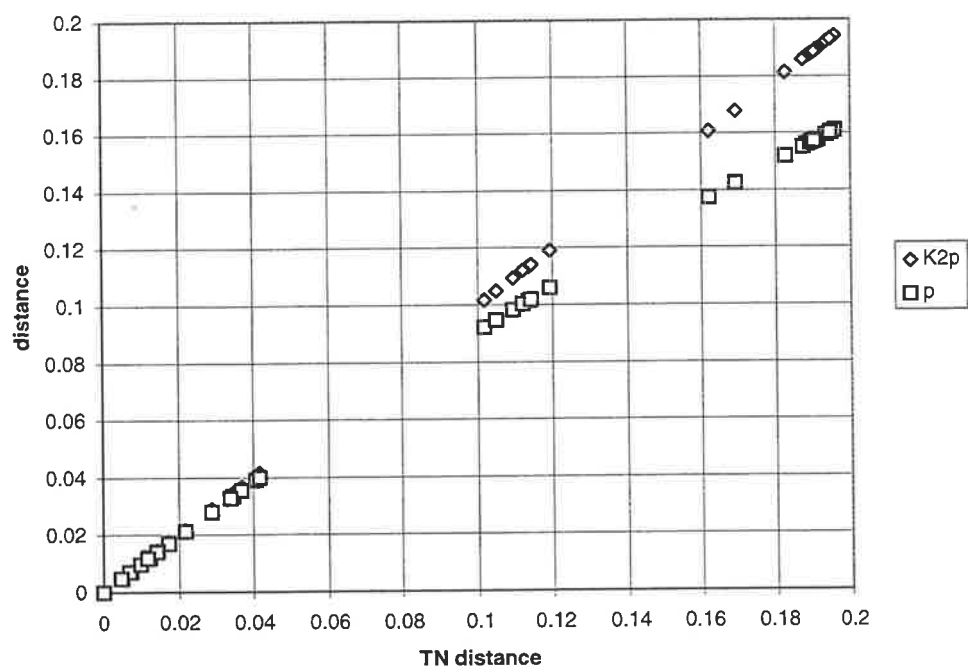


Figure 32. Dot plots of 18S rRNA TN distances versus p and K2p distances calculated for all pair-wise sequence comparisons using transversions only.

Among-site rate variation was examined as described in section 4.4.2.7. The sequences used for analysis were from *G. muris* (Ad-120), Portland 1, Ad-23, P-15, Ad-7, Ad-147, Ad-148, *G. psittaci*, and *G. ardeae*. The estimated value of α was 0.4 when all sites were counted as variable. When constant sites were counted as invariant (frequency = 0.61), the α parameter changed to 1.52, suggesting that there was no significant rate variation among the variant sites. The observation of few or no multiple substitutions for the *G. intestinalis* pair-wise comparisons suggests that the shape of the gamma distribution was due to the relatively recent divergence of the *G. intestinalis* lineages (compared to their divergence from the outgroups), not as a result of extreme rate variation.

Examination of the ML trees inferred with or without correction for among-site rate variation indicated that correction using $\alpha = 0.4$ caused an increase in the branch lengths between the *G. intestinalis* isolates and the outgroups, and a decrease in the resolution within *G. intestinalis* (indicated by decreased support for some branches, see section 4.4.3.6) when compared to the tree obtained using the Poisson model (Figure 33, p. 129). A concern with the analysis was that the nt composition of the *G. muris* 18S rRNA sequence was significantly different from the other sequences, which violates a basic assumption of ML (and therefore affect the estimation of α). However, the effect of this may not be as severe as would be the case if one of the ingroup sequences violated this assumption (Strimmer *et al.*, 1997; PUZZLE manual). To check this, the outgroups were excluded and the data re-analysed. The sequences were highly conserved (92% of sites), which was reflected by an $\alpha = 0.03$. According to Strimmer *et al.* (1997 & PUZZLE manual), high sequence similarity can affect the accuracy of parameter estimation, and values which are close to the parameter's upper or lower bounds may be suspect. Treating the conserved sites as invariable resulted in an $\alpha = 1.14$, which was compatible with the earlier findings and suggested that the high sequence similarity among isolates adversely affected the estimation of α .

4.4.3.5 Distance-based phylogenetic analysis of 18S rRNA nucleotide sequences

Distance analyses were carried out using both minimum evolution (ME) and NJ methods. TN distances were used for NJ analysis in MEGA. K2p distances were used for the ME method (implemented by METREE, Rzhetsky & Nei, 1992, 1993). The resulting phylograms, including percentage bootstrap support, are presented in Figure 34 (p. 130). In general, the tree topologies were the same, irrespective of the distance estimate or method

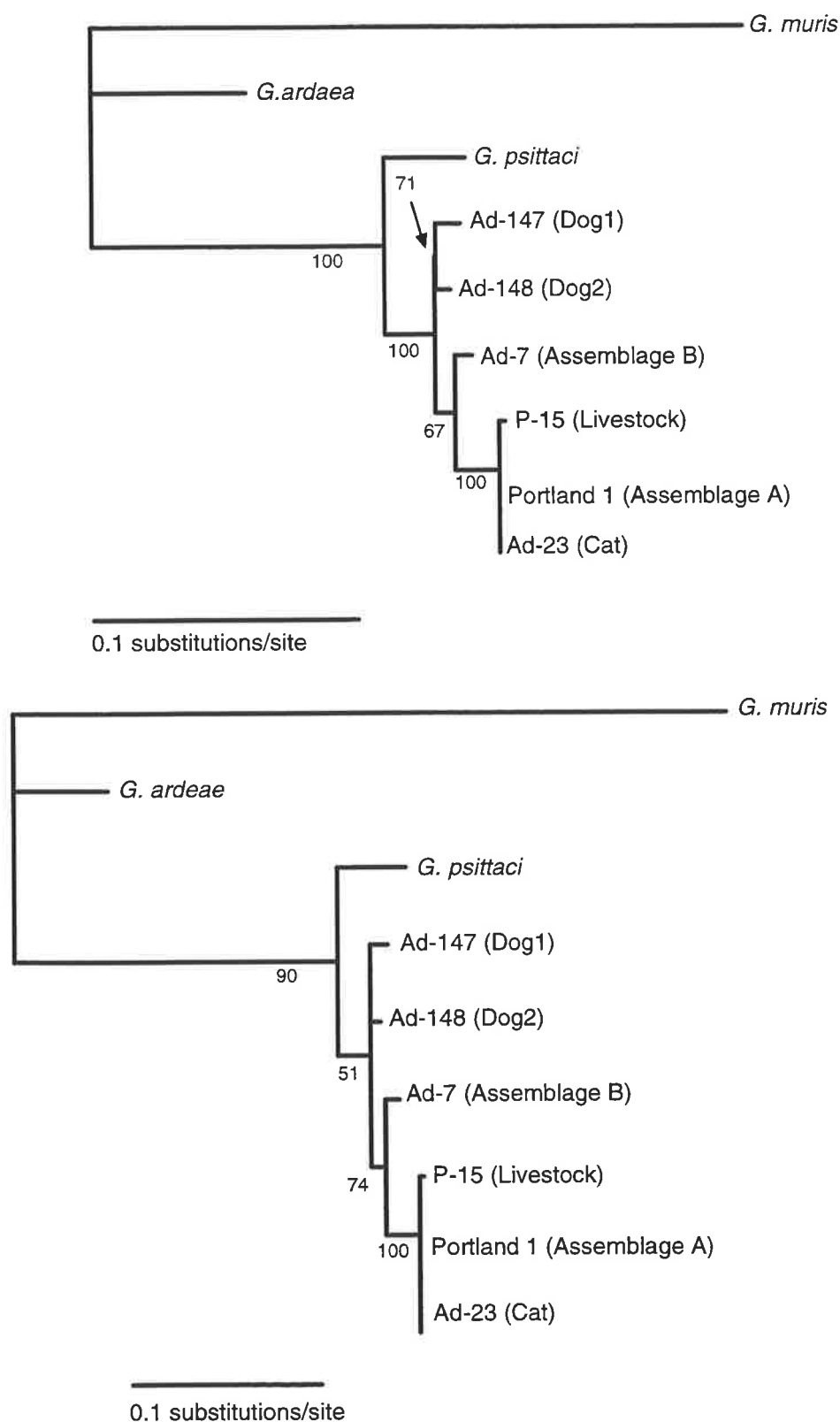


Figure 33. Comparison of 18S rRNA ML trees inferred using either the Poisson (top) or Gamma (bottom) models of site substitution rate distributions. The allozyme defined lineage of an isolate is in parentheses. Percentage support (1,000 quartet puzzles) is indicated under the branches.



A = Assemblage A, B = Assemblage B, published = van Keulen *et al.* (1993).

used. The concordance between the two methods is in agreement with the observations of Rzhetsky and Nei (1992, 1993) that neighbour joining will usually recover the same tree topology as the minimum evolution method unless the number of samples is very large.

The 18S rRNA-inferred phylogenetic relationships of the isolates were the same as those observed in the AE dendrogram (Figure 23, p. 101), confirming the earlier assumption that the AE-inferred clusters represented distinct lineages. Although the branch lengths separating the lineages within *G. intestinalis* were relatively short, some of the relationships could still be resolved. Both ME and NJ bootstrap analyses gave high support for the grouping of the Cat, Livestock and Assemblage A lineages, and also supported the placement of the Livestock and Assemblage A lineages as sister taxa (ie., each others closest relatives). There was also strong support for the monophyly of the Assemblage B isolates. The relationships between the rest of the lineages could not be resolved. The most interesting result was the placement of *G. psittaci*, which appears to be closely related to the *G. intestinalis* species complex. In this case, the 18S rRNA sequence similarities appear to reflect the morphological similarity of these two species.

4.4.3.6 Maximum likelihood analysis of 18S rRNA nucleotide sequences

The results of maximum likelihood analysis of the data are shown in Figure 35 (p. 132). The analysis used an s/v ratio of 1, and assumed a single category of substitution rates for the substitution model. The topology of the ML tree was similar to the trees determined by the distance methods. However, the branching observed for the Livestock, Cat and Assemblage A lineages using the other methods was an unresolved polytomy in the ML analysis. The DNAML tree (not shown) was similar to the tree resulting from ML analysis using quartet puzzling (Figure 35, p. 132). The quartet puzzling support values (equivalent to bootstrap values) supported the monophyly of the *G. intestinalis* species complex, as well as the grouping of the Assemblage A, Livestock and Cat lineages.

4.4.3.7 Cladistic analysis of 18S rRNA nucleotide sequences

The alignment in Figure 28 (pp. 122 - 123) was also used for cladistic analyses, using both the PHYLIP and PAUP packages. Two equally most parsimonious trees, with a length of 66 steps, were found using the exhaustive search option of PAUP (Figure 36, p. 133). Branch and bound analysis, using additional sequences recovered a tree with identical

topology (Figure 36, p. 133). The tree topology was tested by bootstrap analysis using the branch and bound search strategy. There was strong support for the monophyly of *G. intestinalis*, and for the placement of *G. psittaci* as the sister taxon of *G. intestinalis*. However, with the exception of the clade comprising the Assemblage A, Livestock and Cat Lineages, the rest of the *G. intestinalis* lineages formed an unresolved polytomy. The Assemblage B isolates did not form a well supported clade, indicating that although they were virtually identical, not enough

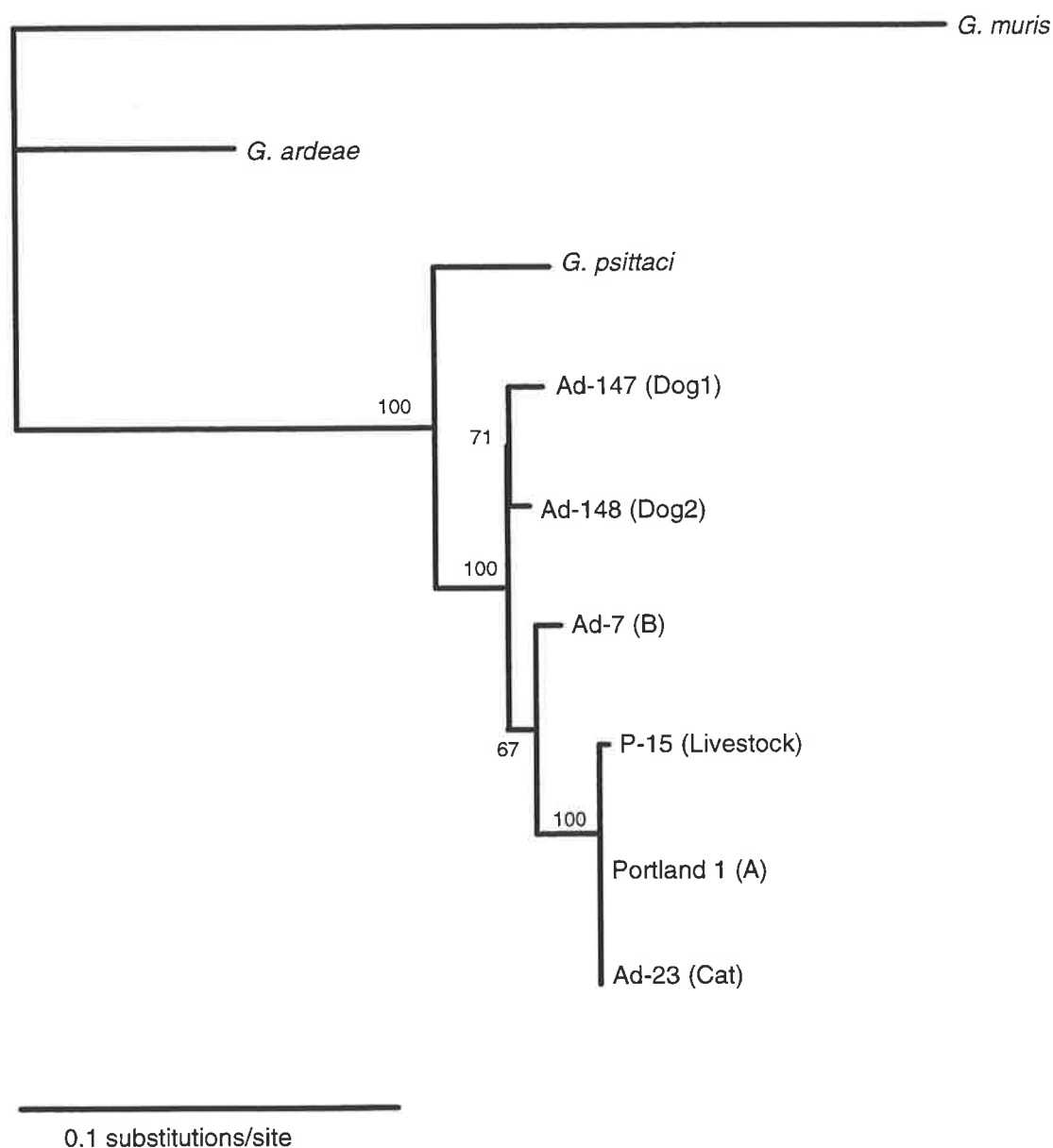


Figure 35. Phylogenetic relationships of *Giardia* isolates inferred from the 18S rRNA sequences using ML analysis and quartet puzzling. Allozyme cluster designations for isolates are in parentheses. Percentage support from 1,000 quartet puzzles tested is indicated at each branch. A = Assemblage A, B = Assemblage B, published = van Keulen *et al.* (1993).

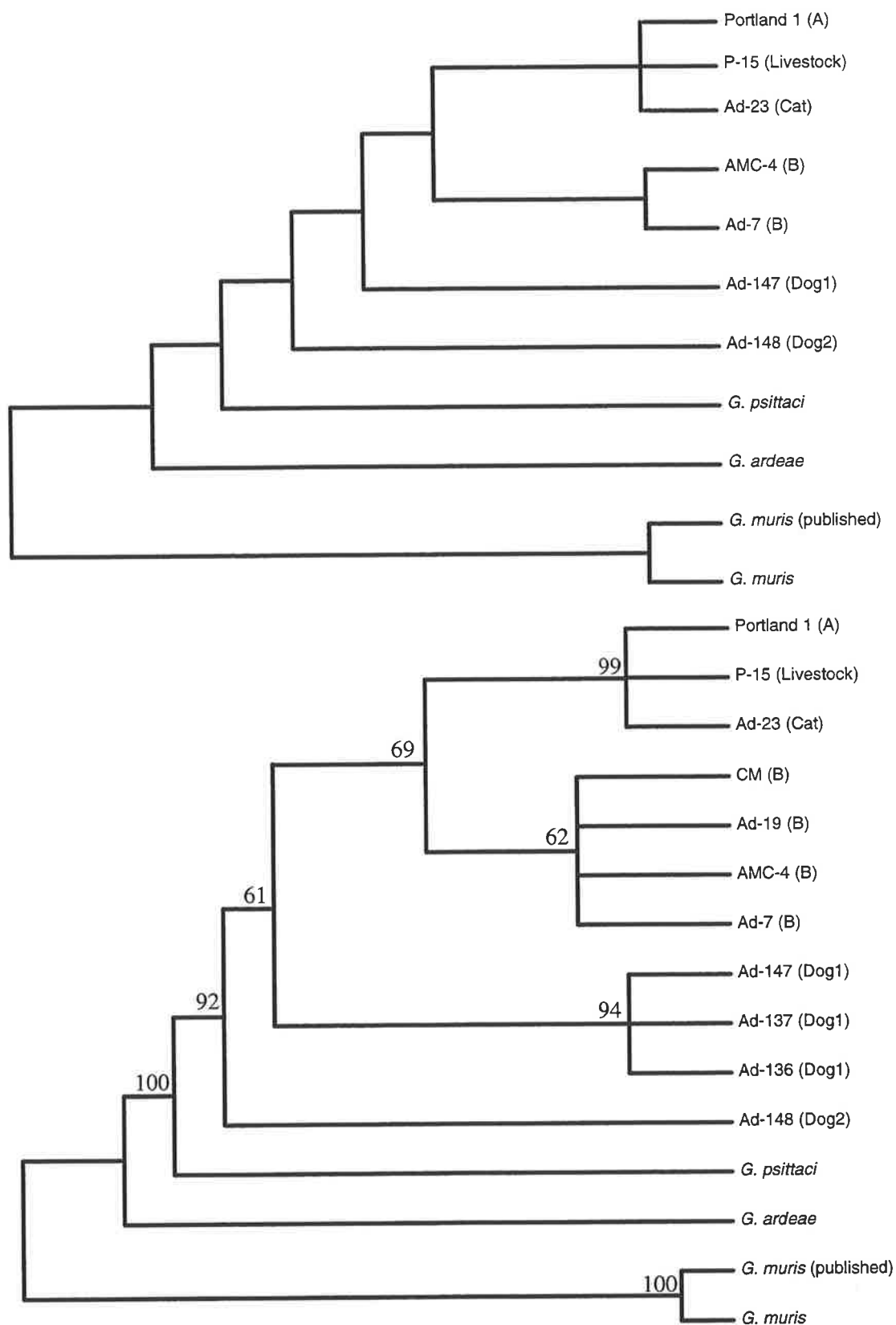


Figure 36. Cladogram inferred by parsimony analysis of *Giardia* 18S rRNA sequences using either the exhaustive search (top) or branch and bound (bottom) algorithms of PAUP. The AE lineage designation of each isolate is indicated in parentheses, and % bootstrap (1,000 replicates) support given above each branch.

A = Assemblage A, B = Assemblage B, published = van Keulen *et al.* (1993).

parsimony informative sites were present. Branch and bound analysis using DNAPENNY found three equally most parsimonious trees with a length of 266 steps. The difference of the tree lengths determined by the two programs is caused by the exclusion of uninformative sites in PAUP, which shortens the number of steps required. The topology was the same as the exhaustive tree from PAUP, and the bootstrap results, determined using the heuristic method with DNAPARS, were also similar.

To determine the amount of phylogenetic signal relative to random noise in the data, the tree-length distributions of the trees examined in the exhaustive search were analysed (Figure 37, p. 135). The presence of strong phylogenetic signal will cause the distribution to be strongly left-skewed, with only one or a few shortest trees well separated from the rest (Hillis, 1991). The skewness can be measured by the g_1 statistic, which is negative for left skewed distributions, 0 if the distribution is symmetrical, and positive if the distribution is right skewed (Hillis, 1991). In the case of 8 taxa, a g_1 value of less than -0.47 indicates that the distribution is unlikely to be due to random noise (99% confidence limit, Hillis, 1991). This critical value increases as more taxa are included. The exhaustive search of PAUP evaluated 2,027,025 trees. The mean tree length was 110.45 ± 7.99 steps, compared to 66 steps for the shortest tree. The g_1 statistic was -2.28, indicating significant skewness and therefore phylogenetic signal. To test the amount of signal within the *G. intestinalis* sequences, the outgroups were removed and the analysis was repeated. The tree distributions are shown in Figure 38 (p. 135). The exhaustive search evaluated 945 trees, which had a mean length of 23.21 ± 3.06 steps. The g_1 was -1.13, which was smaller than the critical value of -0.6 for trees with 7 taxa, and therefore significant.

4.4.4 Phylogenetic analysis of *efl* α nucleotide sequences

The results of the 18S rRNA analysis indicated that this locus was not particularly suitable for the examination of the phylogenetic relationships between genera within the Diplomonadida because it was too variable at this taxonomic level. It also appears to be too conserved to resolve some of the relationships within the *G. intestinalis* species complex. The gene encoding EF1 α was therefore chosen as a candidate to further examine phylogenetic relationships.

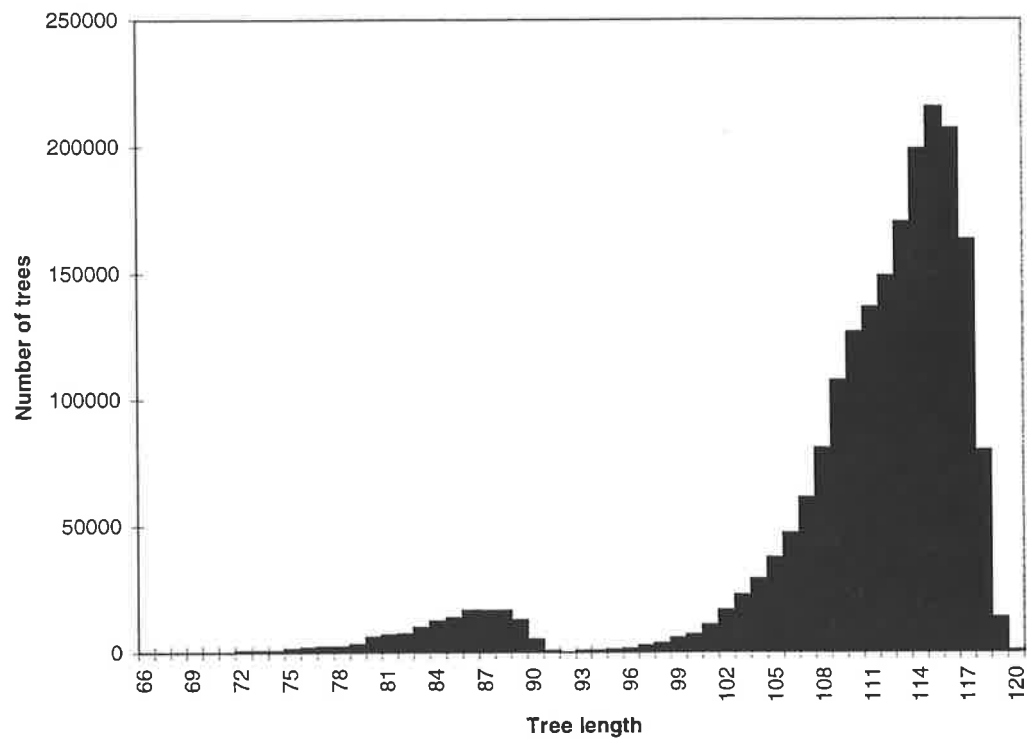


Figure 37. Tree length distributions of trees examined in the 18S rRNA exhaustive search using all taxa.

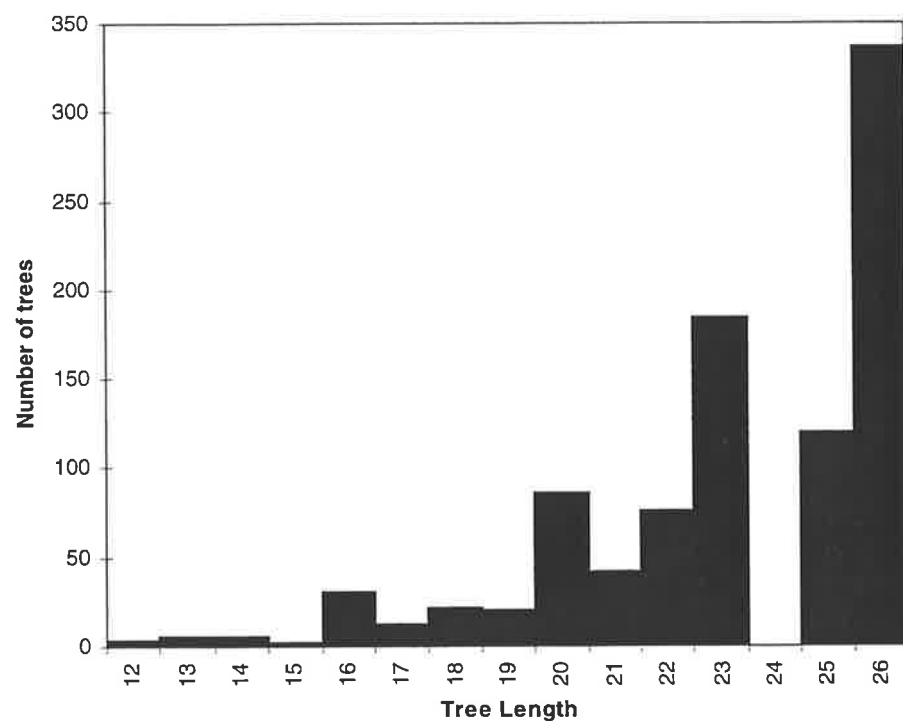


Figure 38. Tree length distributions of trees examined in the 18S rRNA exhaustive search using *G. intestinalis* sequences only.

4.4.4.1 PCR amplification of *efl* α

The primers EF1AF and EF1AR amplified the expected sized fragment (700 bp) under standard PCR conditions from all samples except the Assemblage A representatives. However, closer examination of the Bresa-cleaned products suggested the presence of more than one similarly sized band. This was supported by attempts at sequencing the products using both strands, which failed to produce readable sequences. The possibility of host contamination was ruled out because the expected size of the product from mouse DNA was greater than 3 kb. However, the amplification of contaminating bacterial DNA was a possibility. Amplification of the elongation factor 2 gene (*ef2*) appeared unlikely. Primer EF1AF was < 60% similar to the *G. intestinalis* *ef2* and its best potential binding site was 3' of the EF1AR binding site (which is conserved among *ef1a* and *ef2* sequences), which means that the primers would not be facing each other to allow PCR.

Because of the difficulties associated with EF1AF and EF1AR, new primers were constructed based on the *G. intestinalis* and *T. cruzi* *efl* α sequences (glongF & glongR). These primers proved to be specific, amplifying the expected sized product from the majority of *Giardia* isolates. The inclusion of DMSO (5%) and the use of the glongF and EF1AR primers allowed the successful amplification of *efl* α from representatives of all of the lineages. To ensure that the same segment was being amplified by the different primer combinations, the glongF/R and glongF/EF1AR PCR products amplified from the isolate Ad-7 were sequenced and found to be identical.

4.4.4.2 Alignment of *efl* α nucleotide sequences

The *Giardia* *efl* α sequences determined herein and the published *Hexamita inflata* (GenBank acc. no. U37081), *Spironucleus muris* (GenBank acc. no. U37078) and Diplomonad sp. (GenBank acc. no. U29442) *efl* α sequences were aligned by ClustalW (Figure 39, pp. 137 - 139). The release of *efl* α sequences from other Diplomonads was rather fortuitous, providing suitable outgroups to examine the relationships between *Giardia* species. Following alignment, the position of any indels was checked to ensure that the reading frame was maintained. There were only 2 indels in the alignment, the first a 6 nt deletion in the *S. muris* sequence, and the second a 3 nt insertion in the *S. muris* sequence. The *efl* α sequences were highly conserved, with only 28 variable sites between the *G. intestinalis* sequences.

Figure 39. Alignment of *efl α* nucleotide sequences (AE defined lineage in parentheses, A & B = abbrev. of Assemblages A & B).

Ad-2 (A)	NNC	CTC	GAC	CAG	CTC	AAG	GAC	GAG	CGC	GAA	CGC	GGG	ATC	ACG	ATC	AAC	ATC	GCG	CTC	TGG	AAG	TTC	GAG	ACG	AAG
P-15 (Livestock)	..T	...	...	...	...	...	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23 (Cat)	...	...	...	...	...	...	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-28 (B)	...	...	...	...	...	...	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-136 (Dog1)	...	..G	...	...	...	...	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-148 (Dog2)	..N	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	N..	...	...	...	...	...	...	...	...	...	...	...	...	...
<i>G. ardeae</i>	...	...	...	...	...	...	...	...	...	..G	..T	..T	...	...	...	...	...	..C	...	...	...	...	..C	...	...
<i>G. muris</i>	..N	NN	...	...	T..T	...	...	...	...	..G	..T	..A	...	...	...	...	...	..C	..T	...	...	...	..C	...	...
<i>Spironucleus muris</i>	..T	..G	...	...	..G	...	...	...	...	..G	..T	..T	...	..C	...	...	...	..C	..G	...	...	...	...	...	...
<i>Hexamita inflata</i>	..T	...	..T	..A	...	...	...	..A	..T	...	..T	..T	...	..A	...	...	...	..C	...	...	...	..A	..C	...	...
Diplomonad spp	..T	..T	..T	...	..T	...	..T	...	..T	...	..T	..T	...	..A	...	...	...	..T	...	...	...	...	..C	...	...
Ad-2 (A)	AAG	TAC	ATC	GTC	ACG	ATC	ATC	GAC	GCC	CCG	GGC	CAC	CGC	GAC	TTC	ATC	AAG	AAC	ATG	ATC	ACG	GGG	ACG	TCC	CAG
P-15 (Livestock)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23 (Cat)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-28 (B)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..C	...	...	...
Ad-136 (Dog1)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..C	...	...	...
Ad-148 (Dog2)	...	...	...	...	...	...	...	...	..T	..T	..T	...	...	...	...	...	...	...	...	...	...	..C	...	...	...
<i>G. ardeae</i>	...	...	G..	...	...	...	...	...	...	..C	..G	...	...	...	...	...	...	...	...	...	...	..C	..C	...	...
<i>G. muris</i>	...	...	G..	...	..A	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...	..C	...	..C	...	...
<i>Spironucleus muris</i>	..C	..	...	...	...	...	...	...	...	..A	...	..A..A	...	...	...	...	...	...	..T	..T	..C	G..	...	...	...
<i>Hexamita inflata</i>	...	..T	...	...	..C	...	...	..T	...	..A	..T	...	..A..A	...	...	...	...	...	..T	..T	..C	...	..A	...	...
Diplomonad spp	..A	..T	..C	...	..C	...	...	...	...	..C	...	..T	..T	..T	...	...	...	...	..C	..T	..T	..T	...	...	...
Ad-2 (A)	GCC	GAC	GTC	GCG	ATC	CTC	GTC	GTC	GCG	GCG	GGC	CAG	GGC	GAG	TTC	GAG	GCC	GGG	ATC	TCG	AAG	GAC	GGC	CAG	ACG
P-15 (Livestock)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23 (Cat)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-28 (B)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-136 (Dog1)	..G	...	...	..C	...	...	...	...	..C	...	...	...	...	...	...	...	..G	..C	...	...	...	...	...	...	...
Ad-148 (Dog2)	..A	...	...	..C	...	...	...	...	..C	..A	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
<i>G. ardeae</i>	...	...	...	..C	...	...	...	...	..C	..C	..G	...	..T	...	...	...	...	..C	...	...	...	...	...	...	..C
<i>G. muris</i>	..T	...	...	..C	...	...	..T	...	..C	..C	..G	...	...	..T	...	...	...	..A	...	...	...	...	...	...	..A
<i>Spironucleus muris</i>	..A	...	...	..C	...	..G	..T	A..	..T	..C	...	...	..T	...	...	...	..T	..T	...	..C	...	...	..T	ACA	G..C
<i>Hexamita inflata</i>	..T	...	...	..C	...	...	...	...	..C	..C	..T	..A	..T	..A	...	..A	...	..T	...	..C	GC	..A	..T	..A	..T
Diplomonad spp	..T	..T	...	..C	..T	...	..A	A..T	..T	T..C	..T	...	..T	..A	...	..A	..T	..T	...	..C	...	..G	...	...	..T

75

150

225

Figure 39 (cont.). Alignment of *efl* α nucleotide sequences (AE defined lineage in parentheses, A & B = abbrev. of Assemblages A & B).

Ad-2 (A)	CGC	GAG	CAC	GCG	ACC	CTT	GCG	AAC	ACG	CTC	GGG	ATC	AAG	ACG	ATG	ATC	ATC	TGC	GTC	AAC	AAG	ATG	GAC	GAC	GGC	300
P-15 (Livestock)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Ad-23 (Cat)	...	...	...	...	...	...	...	A	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Ad-28 (B)	...	...	...	...	G	C	...	...	...	...	...	...	...	...	...	...	...	...	A..	...	...	...	...	...	...	
Ad-136 (Dog1)	...	...	...	C	G	C	A	...	...	...	C	...	...	...	...	...	...	...	A..	...	...	...	...	...	...	
Ad-148 (Dog2)	...	...	...	C	...	...	A	...	A	...	C	...	...	...	...	...	...	...	A..	...	...	...	...	...	...	
<i>G. ardeae</i>	...	...	...	C	G	C	...	...	...	...	...	...	...	...	...	...	...	...	T	A.T	...	...	...	...	T	CC.
<i>G. muris</i>	...	...	...	T	...	...	T	...	T	...	C	...	...	...	...	...	T	...	A.T	...	...	...	...	T	CCT	
<i>Spironucleus muris</i>	..T	...	...	T	T	G	C	...	C	...	T	...	GA	C	C	...	TG	GC	A..	...	...	...	...	TC	---	
<i>Hexamita inflata</i>	..T	..A	...	C	T	...	C	...	C	T	T	...	...	C	...	T	G.T	GC	...	...	...	...	...	...	CCA	
Diplomonad spp	..T	...	...	T	T	...	T	C.T	T	T	T	...	...	T	C.C	...	G..	...	T	...	...	...	T	...	CC.	
Ad-2 (A)	CAG	GTC	AAG	TAC	TCG	AAG	GAG	CGC	TAC	GAC	GAG	ATC	AAG	GGC	GAG	ATG	ATG	AAG	CAG	CTC	AAG	AAC	ATC	GGC	T--	375
P-15 (Livestock)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Ad-23 (Cat)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	C	...	...	...	...	...	...	...	...	
Ad-28 (B)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Ad-136 (Dog1)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Ad-148 (Dog2)	...	...	...	...	...	...	C	...	T	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
<i>G. ardeae</i>	...	...	...	...	...	...	...	...	...	...	...	...	...	CG	...	...	...	...	...	...	...	...	...	N	...	
<i>G. muris</i>	...	...	...	T	...	...	T	...	G	...	...	...	ACG	...	...	...	...	...	...	...	...	...	A	...	...	
<i>Spironucleus muris</i>	---	A..	...	T	GAC	C..	A..	T	...	AC	..A	...	T	..AA	...	...	A.	...	TT	...	...	G	...	..AT	...	
<i>Hexamita inflata</i>	..A	...	C	...	T	G.A	..CT	T	...	ACT	..A	...	...	ACT	..A	...	CAA	...	ACT	T..	...	C.A	...	T	...	
Diplomonad spp	...	...	C	T	T	G.A	..CC	T	...	A.G	..A	...	...	..AA	..A	...	CA.	...	A.T	..T	...	C.G	...	T	...	
Ad-2 (A)	-GG	AAG	AAG	GCC	GAG	GAG	TTC	GAC	TAC	ATC	CCG	ACG	TCC	GGC	TGG	ACC	GGG	GAC	AAC	ATT	ATG	GAG	AAG	TCC	GAC	450
P-15 (Livestock)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	C	...	...	...	...	...	
Ad-23 (Cat)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	C	...	...	...	...	...	
Ad-28 (B)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	C	...	...	G	...	...	
Ad-136 (Dog1)	...	...	...	...	...	...	...	...	...	...	...	...	C	...	...	...	...	...	...	C	...	...	...	...	...	
Ad-148 (Dog2)	...	...	...	...	...	...	...	...	...	...	C	...	G	...	...	...	T	...	...	C	...	...	...	...	...	
<i>G. ardeae</i>	...	...	...	G	...	...	...	...	...	C	C	...	...	...	...	G	C	...	...	C	...	...	...	...	...	
<i>G. muris</i>	...	...	...	T	...	...	...	T	...	T	...	C	G	...	...	G	C	...	...	C	...	...	G	...	...	
<i>Spironucleus muris</i>	G.T	...	...	...	...	...	C..	...	...	C	GTT	...	...	...	...	TT	T	...	...	C	...	...	...	...	..G	
<i>Hexamita inflata</i>	..TC	...	C.C	TGG	..A	...	...	...	TT	G..	..A	CTC	...	..T	...	..T	..T	...	...	C	...	..A	GCC	..T	CCA	
Diplomonad spp	..AC	...	...	TGG	..C	..A	..T	..T	..T	...	..C	..C	..T	...	...	..T	..T	...	G	..C	...	...	...	..T	CC.	

Figure 39 (cont.). Alignment of *eflα* nucleotide sequences (AE defined lineage in parentheses, A & B = abbrev. of Assemblages A & B).

Ad-2 (A)	AAG	ATG	CCC	TGG	TAC	GAG	GGC	CCG	TGC	CTG	ATC	GAC	GCG	ATC	GAC	GGG	CTC	AAG	GCC	CCG	AAG	CGC	CCG	ACC	GAC	525
P-15 (Livestock)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..A	...	
Ad-23 (Cat)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Ad-28 (B)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Ad-136 (Dog1)	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..G	
Ad-148 (Dog2)	...	...	..G	...	...	...	...	...	..T	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..A	...	
<i>G. ardeae</i>	...	...	...	...	...	...	..C	...	..C	...	...	...	...	...	...	...	..T	..C	...	...	..C	...	..C	..G	...	
<i>G. muris</i>	...	...	..G	...	...	...	..C	...	..T	..T	...	...	...	...	...	...	...	..A	..T	...	..T	..C	..G	...	...	
<i>Spironucleus muris</i>	..C	...	..A	...	...	ACC	...	AA	...	..C	...	..A	..C	...	..T	..A	..G	...	..C	...	..A	..G	..C	...	...	
<i>Hexamita inflata</i>	...	..CT	..A	...	...	A..	...	AA	...	..C	...	..A	TGC	...	..T	..T	...	...	..T	..A	...	..T	..A	..A	...	
Diplomonad spp	..C	...	..A	...	...	TCC	...	..T	...	..T	..T	..T	..C	...	...	..A	...	...	...	..C	...	A..A	..C	..T	..T	
Ad-2 (A)	AAG	CCC	CTC	CGC	CTC	CCG	ATC	CAG	GAC	GTC	TAC	AAG	ATC	TCG	GGC	GTC	GGG	ACC	GTC	CCC	GCG	GGC	CGC	GTC	GAG	600
P-15 (Livestock)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..C	...	...	..T	...	
Ad-23 (Cat)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..C	...	...	...	...	
Ad-28 (B)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Ad-136 (Dog1)	...	...	...	...	...	...	...	...	...	...	...	...	...	..C	...	...	...	...	...	...	...	...	...	...	...	
Ad-148 (Dog2)	...	...	...	..T	...	...	..A	..T	...	...	...	...	..C	...	...	..A	...	...	..C	...	...	...	...	...	...	
<i>G. ardeae</i>	...	...	..T	...	..C	...	...	...	...	...	...	...	..C	..T	...	...	..G	...	..T	...	..T	..T	..T	...	...	
<i>G. muris</i>	...	..T	..T	...	..T	...	...	...	...	...	...	...	..C	..T	...	...	..G	...	..T	...	..T	..T	..T	...	...	
<i>Spironucleus muris</i>	...	...	..G	..T	...	..A	C..	...	...	..G	...	...	..C	..T	..A	..T	..A	..A	...	..C	...	..T	...	...	...	
<i>Hexamita inflata</i>	...	..A	...	..T	...	..A	...	..A	...	...	...	...	AAC	..T	...	..T	...	..A	..T	..T	..T	..T	...	..A	...	
Diplomonad spp	...	..T	...	..T	...	..C	...	...	...	...	...	...	AAT	..T	...	..T	...	...	...	..T	..T	..T	...	..A	...	
Ad-2 (A)	ACG	GGC	GAG	CTC	GCG	CCC	GGG	ATG	AAG	GTC	GTC	TTC	GCC	CCG	ACG	TCC	CAG	NNN	NNN	NNN						660
P-15 (Livestock)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	GTC	TC	...						
Ad-23 (Cat)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	GTC	TCG	GAG						
Ad-28 (B)	...	...	...	...	...	..C	...	...	...	...	...	...	..G	...	...	...	...	GTC	TCG	GAG						
Ad-136 (Dog1)	...	...	...	...	..C	...	...	...	...	...	...	...	...	..C	...	...	..N	...	...	...						
Ad-148 (Dog2)	..C	..T	...	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	GTC	TCT	GAG						
<i>G. ardeae</i>	..T	...	...	...	..TC	...	..C	...	...	...	...	...	..G	..C	...	..A	...	ACC	TCT	GAG						
<i>G. muris</i>	..C	..T	...	...	ATT	..G	...	...	...	..T	...	..G	..C	..G	..A	...	...	...	...	...						
<i>Spironucleus muris</i>	T..T	..T	..TC	..G	AA	..A	..C	CA	..TC	...	...	...	..C	T..T	GA	..G	..TCC	GGT	GAA							
<i>Hexamita inflata</i>	T..T	..T	..A	...	ATC	..A	..T	...	..T	...	...	...	..A	G..C	GG	..G	..A	AAG	ACT	GAA						
Diplomonad spp	T..C	..T	TT	...	ATC	...	AAC	...	..CC	..T	...	...	..T	..C	T..C	..A	ACC	ACT	GCT	GAA						

Uncorrected sequence comparisons revealed approximately 75 - 80% similarity between the *Giardia* and the other diplomonad sequences, and 88 - 92% similarity among the sequences of the *Giardia* species. There was >94.5% similarity among the lineages comprising the *G. intestinalis* species complex, which was similar to that observed for the 18S rRNA sequence comparisons.

4.4.4.3 Determination of *efl* α sequence parameters

As described in section 4.4.3.4 for the 18S rRNA data, sequence properties such as nt composition and the optimal transition/transversion (s/v) ratio were determined for each data, as well as the extent of among-site rate variation and multiple substitutions. As shown in Table 20, the optimal s/v ratio was found to be 1.

Table 20. Log likelihoods for different s/v ratios.

s/v ratio	log likelihood
0.5	-2874.62271
0.75	-2858.19135
1	-2855.46596 ← optimal
1.5	-2864.99848
2	-2881.22919
3	-2916.64725

The nt compositions are shown in Table 21 (p. 141). The GC content of the outgroup diplomonad sequences was 51 - 56% over all sites, compared to 57.5%, 62%, and 64% for the sequences from *G. muris*, *G. ardeae* and *G. intestinalis* respectively. All of the sequences examined had similar 1st and 2nd codon position GC contents (54% and 38% respectively), but the 3rd codon position GC content varied widely, ranging from 98 - 100% for all of the *G. intestinalis* isolates (except Ad-148 which was 90%), to 93.6% in *G. ardeae*, 80% in *G. muris*, and 61 - 88% in the other diplomonad sequences. This GC bias at the 3rd codon position has been noted before for other *Giardia* protein coding genes (eg., Mowatt *et al.*, 1994), although not (as far as I know) at these extreme levels.

The degree of multiple hits and saturation for change was examined by the comparison of corrected and uncorrected distances. These were calculated using either PHYLIP or MEGA (as for section 4.4.3.4), and then compared graphically (Figure 40, p. 142). The

corrected distances estimated by the different methods were identical, indicating that the assumptions made by the simpler models such as JC (nt composition of 25% for each base and equal rates of substitutions for all sites), were not being violated to an extent which would detectably affect the estimate. Comparisons of these distances with the p-distance (which is uncorrected) showed that multiple hits had occurred, especially for comparisons where the p-distance was greater than 0.15 substitutions per site (eg. between *G. intestinalis* and *G. muris* sequences). Examination of distance estimates, using transitions or transversions only, showed that there had been multiple hits for both classes of substitution (Figures 41 & 42, pp, 142 - 143).

A further breakdown of the data into either 1st and 2nd codon positions or 3rd codon positions showed that multiple hits at the first two codon positions were relatively infrequent (Figure 43, p. 144), but that substitutions at 3rd codon positions were approaching saturation

Table 21. Nucleotide composition (%) of the partial *eflα* sequences (for all sites and each individual codon position).

Lineage/species	Isolate	All positions					1 st codon position				
		A	T	C	G	Total	A	T	C	G	Total
Assemblage A	Ad-2	22.9	13	32	32	646	33.5	11.2	19.1	36.3	215
Livestock	P-15	22.7	13.4	32	32	651	33.2	11.5	18.9	36.4	217
Cat	Ad-23	22.7	12.8	32.4	32.1	655	33	11.5	18.8	36.7	218
Assemblage B	Ad-28	22.6	12.8	32.3	32.3	654	33.5	11.5	18.8	36.2	218
Dog1	Ad-136	23.1	12.7	33	31.2	645	34	11.2	19.1	35.8	215
Dog2	Ad-148	23.9	15.1	31.8	29.2	623	34.8	12.1	17.4	35.7	207
<i>G. ardeae</i>		22.9	14.8	33.3	28.9	654	34.1	11.1	19.4	35.5	217
<i>G. muris</i>		24.1	18.2	28.5	29.2	643	35	11.2	18.7	35	214
<i>S. muris</i>		25.2	18.3	30.4	26.2	652	35	11.5	18.4	35	217
<i>H. inflata</i>		27.9	20.9	29.2	22	655	33	11.5	19.3	36.2	218
Diplomonad sp.		24.7	23.8	29.2	22.3	655	33.9	12.8	20.2	33	218
Lineage/species	Isolate	2 nd codon position					3 rd codon position				
		A	T	C	G	Total	A	T	C	G	Total
Assemblage A	Ad-2	34.9	27	22.3	15.8	215	0.5	0.9	54.6	44	216
Livestock	P-15	34.6	27.2	22.6	15.7	217	0.5	1.4	54.4	43.8	217
Cat	Ad-23	34.9	26.6	22.9	15.6	218	0.5	0.5	55.3	43.8	219
Assemblage B	Ad-28	34.4	27.1	22.5	16.1	218	0	0	55.5	44.5	218
Dog1	Ad-136	34.9	27	22.3	15.8	215	0.5	0	57.7	41.9	215
Dog2	Ad-148	33.7	27.4	23.6	15.4	208	3.4	5.8	54.3	36.5	208
<i>G. ardeae</i>		34.9	27.1	23.4	14.7	218	0	6.4	57.1	36.5	219
<i>G. muris</i>		34.6	27.1	22.4	15.9	214	2.8	16.3	44.2	36.7	215
<i>S. muris</i>		34.1	28.1	22.1	15.7	217	6.4	15.1	50.5	28	218
<i>H. inflata</i>		33	28.4	22.9	15.6	218	17.8	22.8	45.2	14.2	219
Diplomonad sp.		33	27.5	24.8	14.7	218	7.3	31.1	42.5	19.2	219

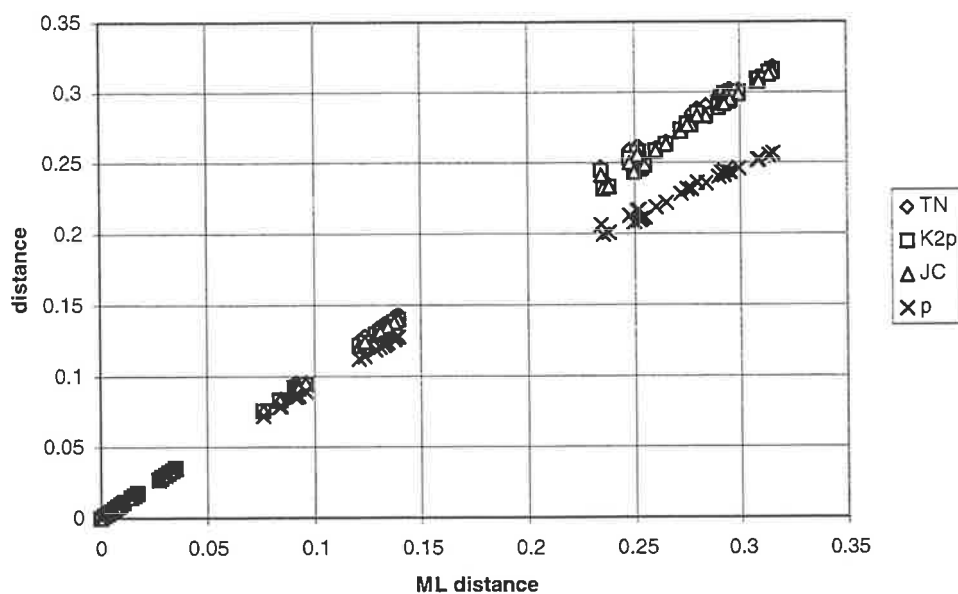


Figure 40. Dot plots of *eflα* ML distances versus TN, K2p, JC, & p-distances calculated for all pair-wise sequence comparisons using all classes of nt substitutions.

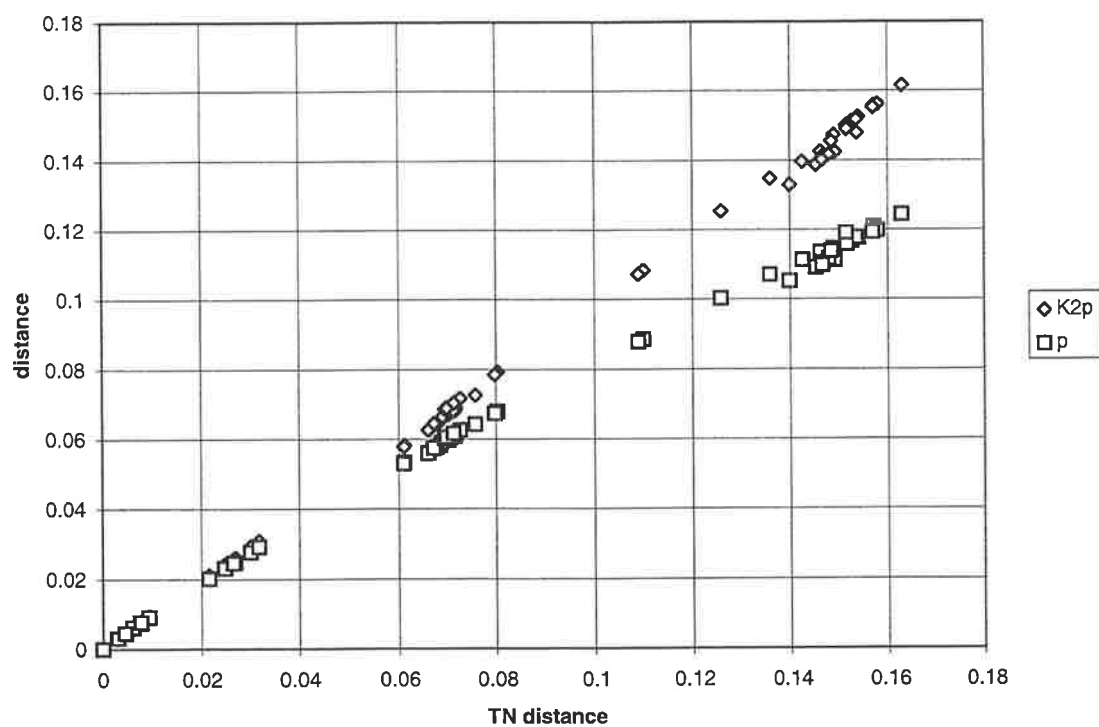


Figure 41. Dot plots of *eflα* TN distances versus K2p & p-distances calculated for all pair-wise sequence comparisons using transitions only.

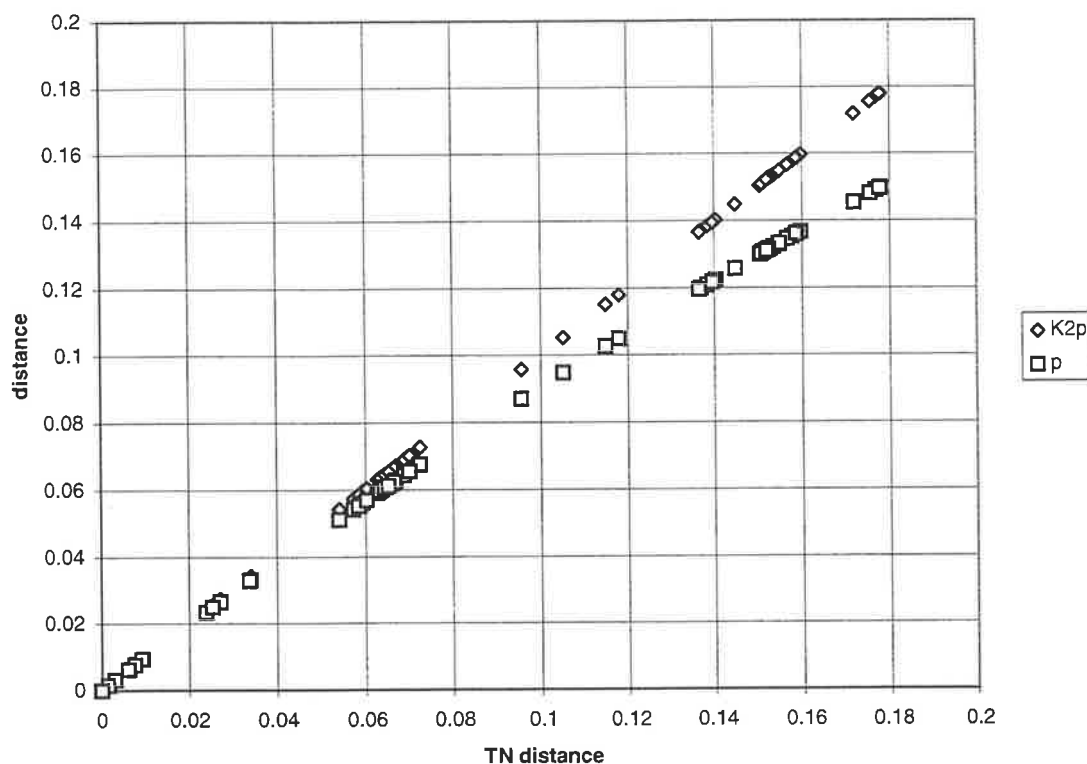


Figure 42. Dot plots of *eflα* TN distances versus K2p & p-distances calculated for all pair-wise sequence comparisons using transversions only.

(indicated by a flattening of the curve in Figure 44, p. 144). Closer examination of the distance plots showed that the graph was relatively linear for the points representing pair-wise distance comparisons of *Giardia* sequences, with no difference between the corrected and uncorrected distance estimates for the 1st and 2nd codon positions, and no evidence for saturation at the 3rd codon position. Examination of the corrected (by JC) and uncorrected distance estimates calculated using synonymous or nonsynonymous substitutions (estimated using MEGA) gave a similar result (Figures 45 & 46, p. 145). There was no evidence of multiple hits for any of the *G. intestinalis* pair-wise sequence comparisons for distances calculated using either synonymous or nonsynonymous substitutions (distances < 0.2 and < 0.04 respectively). However, there was evidence for saturation for change for the comparisons between the *Giardia* and outgroup sequences (JC distance > 1) for distances calculated using synonymous substitutions (ie., a flattening of the curve, Figure 45, p. 145). The finding of multiple synonymous substitutions was supported by analysis of corrected and uncorrected protein sequence distances, which showed no evidence for saturation (Figure 47, p. 146).

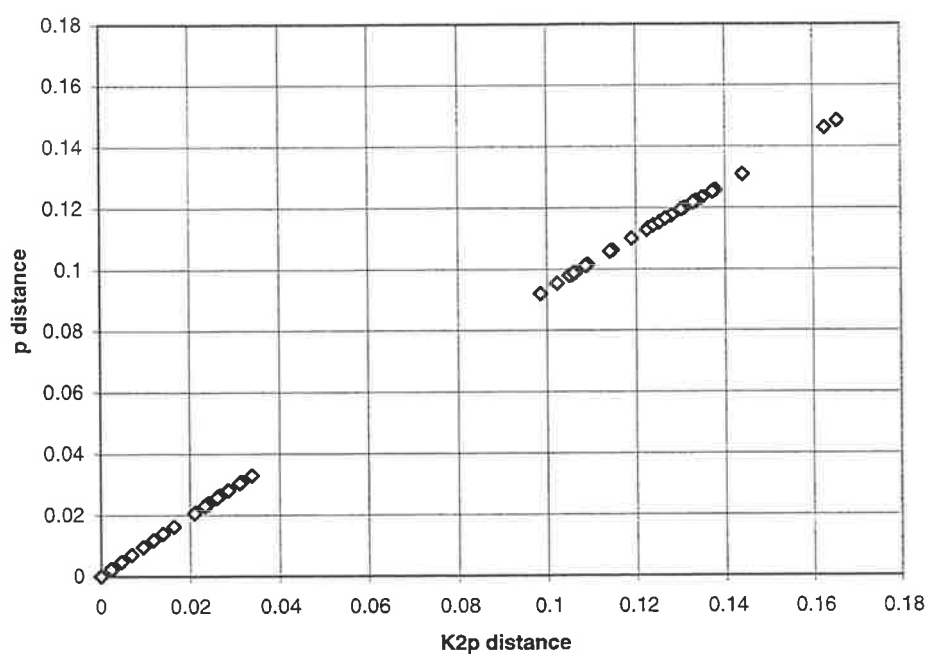


Figure 43. Dot plot of *eflα* p-distances versus K2p distances calculated for all pair-wise sequence comparisons using 1st and 2nd codon positions only.

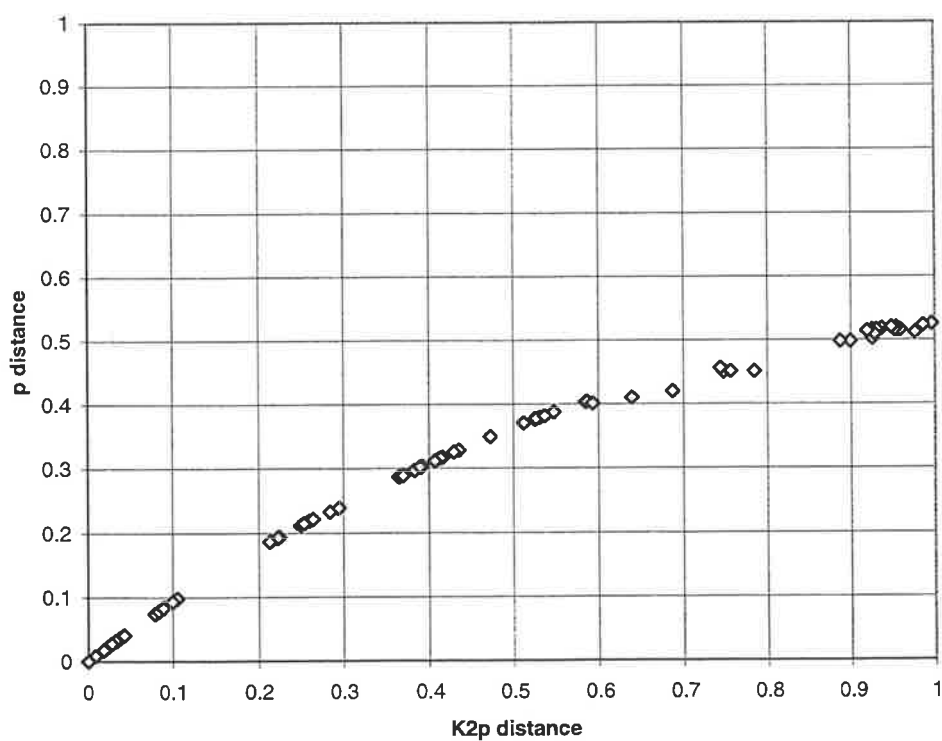


Figure 44. Dot plot of *eflα* p-distances versus K2p distances calculated for all pair-wise sequence comparisons using 3rd codon positions only.

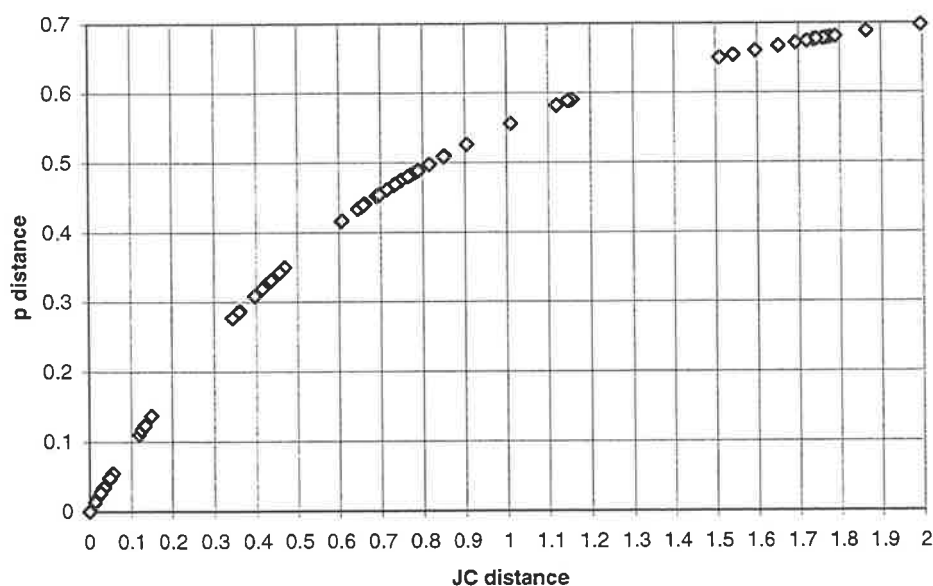


Figure 45. Dot plot of *eflα* p-distances versus JC distances calculated for all pair-wise sequence comparisons using synonymous substitutions only.

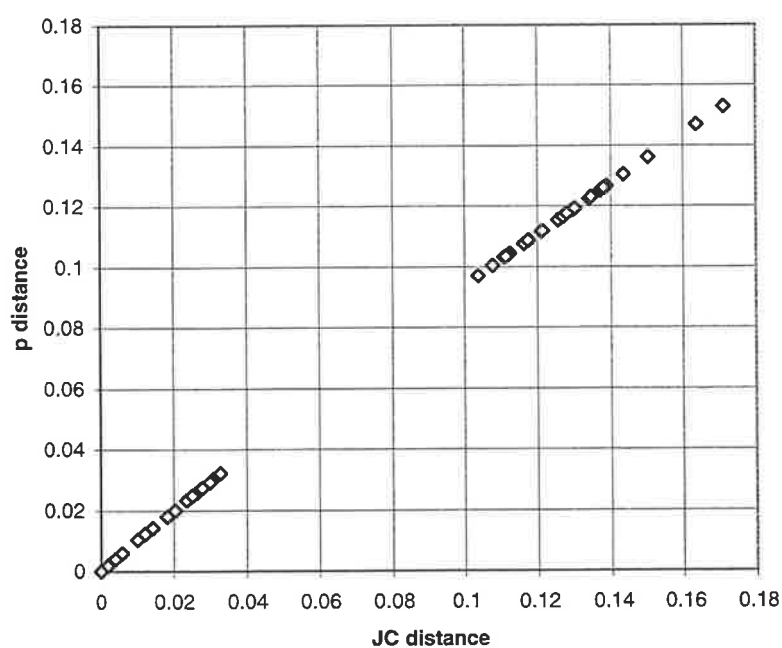


Figure 46. Dot plot of *eflα* p-distances versus JC distances calculated for all pair-wise sequence comparisons using non synonymous substitutions only.

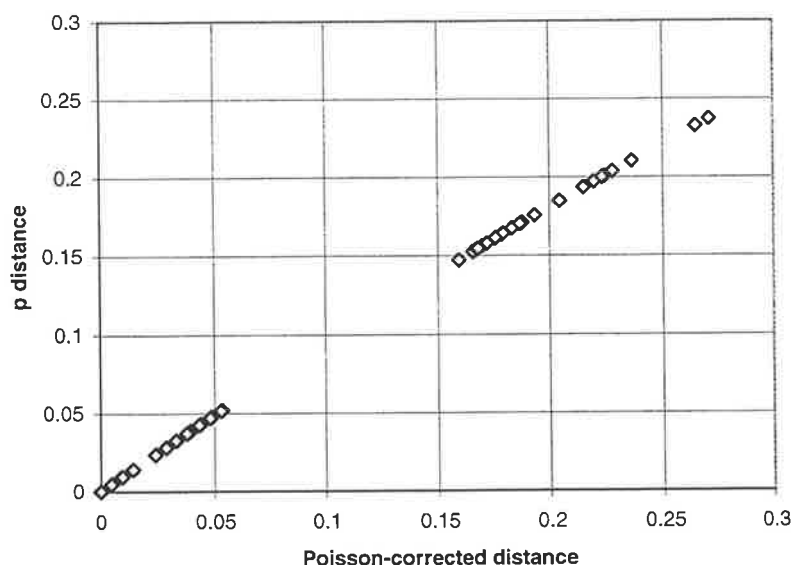


Figure 47. Dot plot of *eflα* p-distances versus Poisson-corrected distances calculated for all pair-wise sequence comparisons using deduced amino acid sequences.

Among-site rate variation was examined using the approach described in section 4.4.2.7. Due to the memory constraints of the DOS version of PUZZLE, the data set was limited to 8 taxa (*G. muris* (Ad-120), P-15, Ad-23, Ad-2, Ad-28, Ad-136, Ad-148 and *G. ardeae*). The α parameter calculated using all codon positions was 0.03, indicating that the estimation of α was suspect. The value of α was then determined for each codon position separately. This suggested that most 1st codon positions ($\alpha = 48.7$) had similar substitution rates, and 3rd codon positions ($\alpha = 0.62$) had moderate levels of among-site rate variation. The α parameter could not be properly estimated for 2nd codon positions ($\alpha = 0.02$), presumably due to the high level of conservation (90.5%) at these sites. Although some of the 3rd codon positions may have high rates of substitution based upon the estimation of α , there does not appear to be any evidence of saturation for change for the *G. intestinalis* comparisons (distances < 0.3 substitutions/site, Figure 44, p. 144). The differences in the 3rd codon position GC content between the *eflα* sequences of *G. muris* and the other *Giardia* species were sufficient to significantly alter the overall nt composition of *G. muris* compared to the model used by the ML analysis. The differences in GC content at the 3rd codon positions between the *eflα* sequences of different *G. intestinalis* isolates also became significant when

the ML analysis was restricted to the 3rd codon positions, potentially affecting the resulting tree (see section 4.4.3.4).

4.4.4.4 Distance-based phylogenetic analysis of *efl* α nucleotide sequences

The approach used here was similar to that used in section 4.4.3.5, except that deduced amino acid sequences were also analysed. In the latter case, distance estimates were made using the Poisson correction. Both the NJ and ME tree building methods recovered phylogenies with similar topologies irrespective of the distance estimate used (Figures 48 & 49, p. 148 & 149). The main difference was the placement of *G. ardeae*, which changed from being monophyletic with *G. muris* (Figure 49, p. 149) to the closest sister taxon to the *G. intestinalis* species complex (Figure 48, p. 148). The topologies of the *efl* α distance-based phylogenies were essentially the same as the 18S rRNA gene phylogenies. The Livestock, Cat and Assemblage A lineages were monophyletic, although the topology within this group was different compared to the 18S rRNA-inferred phylogeny. The placement of Assemblage B and the Dog1 and Dog2 lineages was also the same, with Dog1 and Dog2 being the earliest branching lineages within the *G. intestinalis* species complex (Figure 48, p. 148).

Reliability of the tree topologies was examined by bootstrap analysis. Good support was found for almost all of the nodes of the NJ tree (Figure 48, p. 148). The exceptions were the position of *G. ardeae* as sister taxon to *G. intestinalis*, and the branching order of the Dog1 and Dog2 lineages, which had poor support (66% & 65% respectively). Both of these parts of the tree received better support from the ME bootstrap analysis (Figure 48, p. 148). There was good bootstrap support (82%) for the grouping of the Livestock and Cat lineages, which was in conflict with the 18S rRNA and AE results (Figures 23 & 34, pp. 101 & 130). Bootstrap analysis using distances calculated from synonymous substitutions gave a similar result, except that the bootstrap support for all nodes was decreased by 15% - 20% (not shown).

Bootstrap analyses using other distance measures gave different results. Bootstrapping using distance estimates calculated from nonsynonymous substitutions found high support for the monophyly of *G. muris* and *G. ardeae* (90%, not shown), however this was not as strongly supported by the bootstrap analysis using distances estimated from the deduced amino acid sequences, which found only 69% or 75% for either ME or NJ trees respectively (Figure 49, p. 149). The monophyly of the *G. intestinalis* species complex was supported in 87% - 92% of

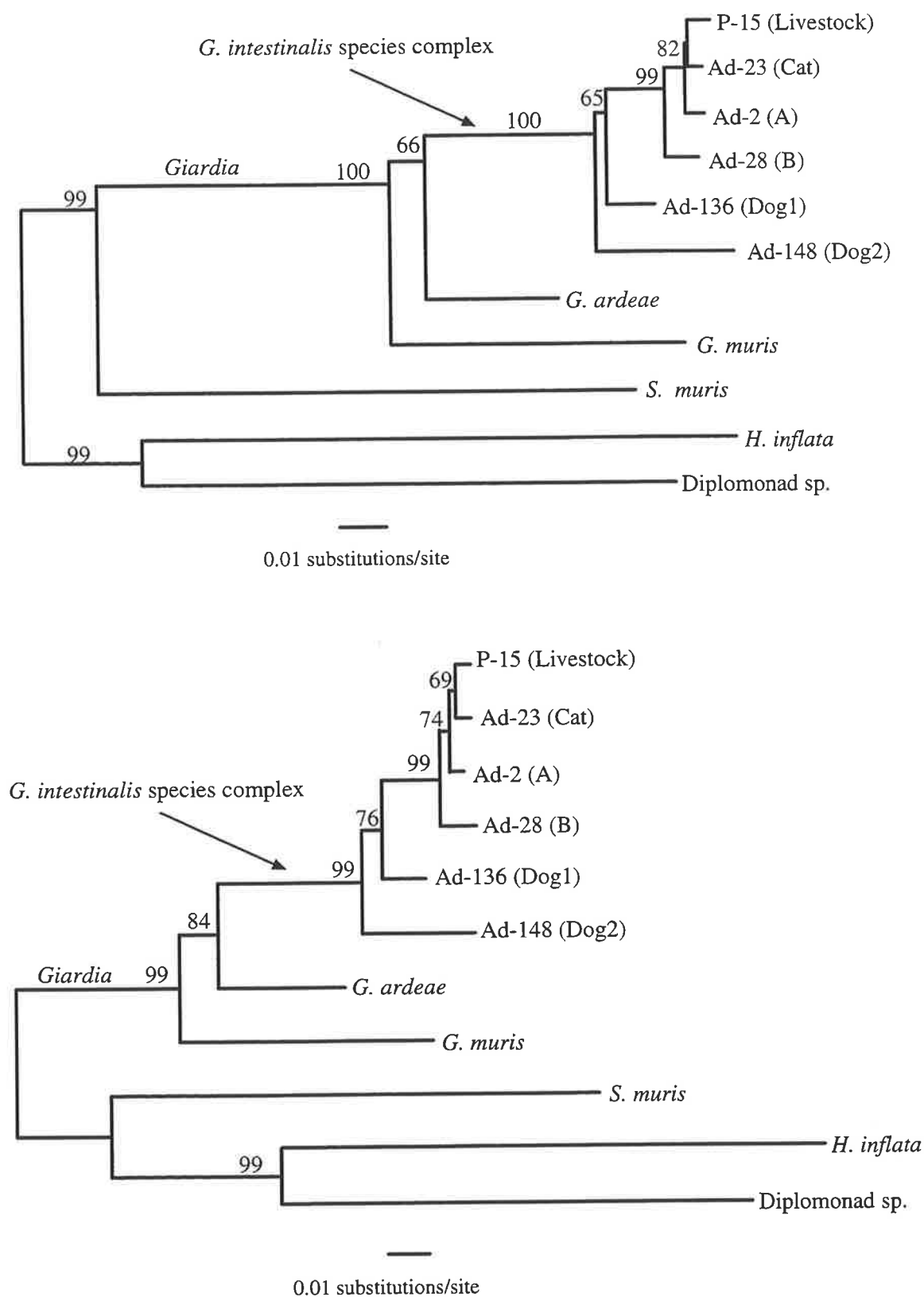


Figure 48. Phylogenetic relationships of isolates inferred from the *eflα* nt sequences using TN or K2p distances and the NJ (top) or ME (bottom) tree building methods. The allozyme defined lineage of an isolate is given in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (500 or 1,000 replicates) is indicated above the branches.

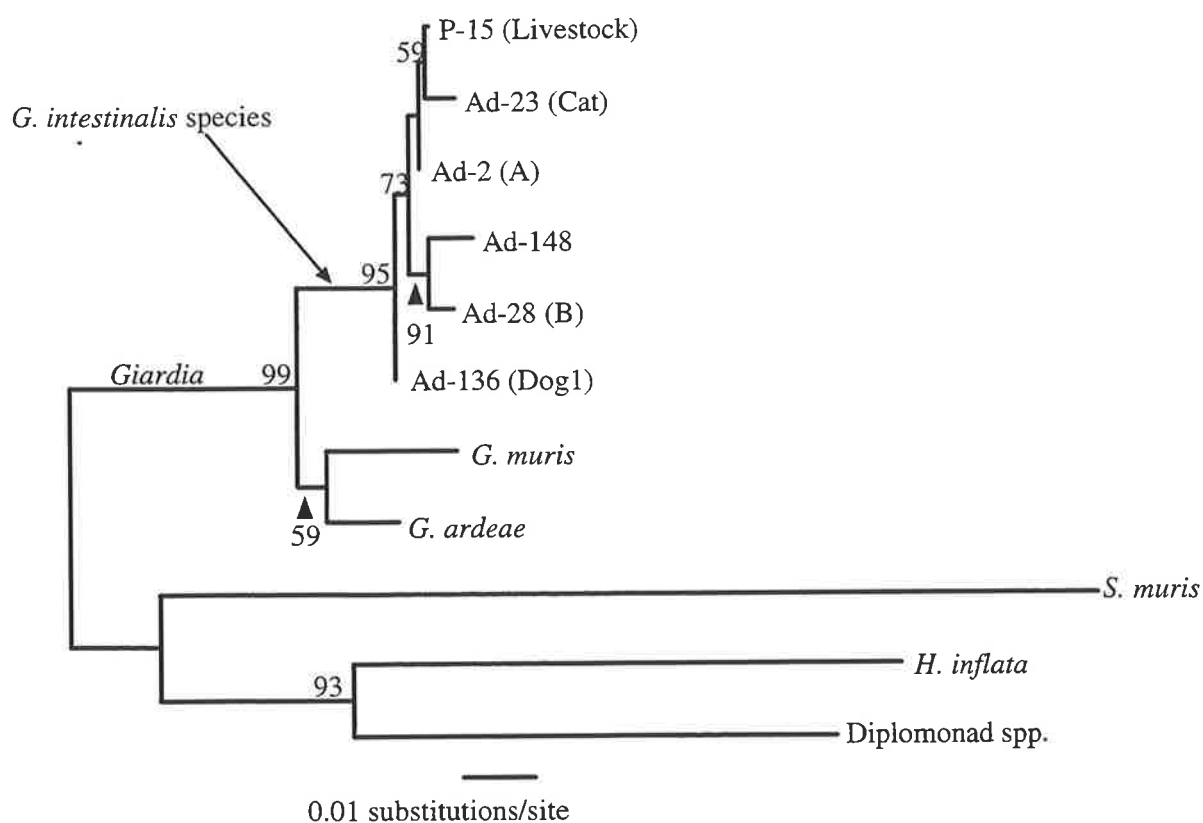
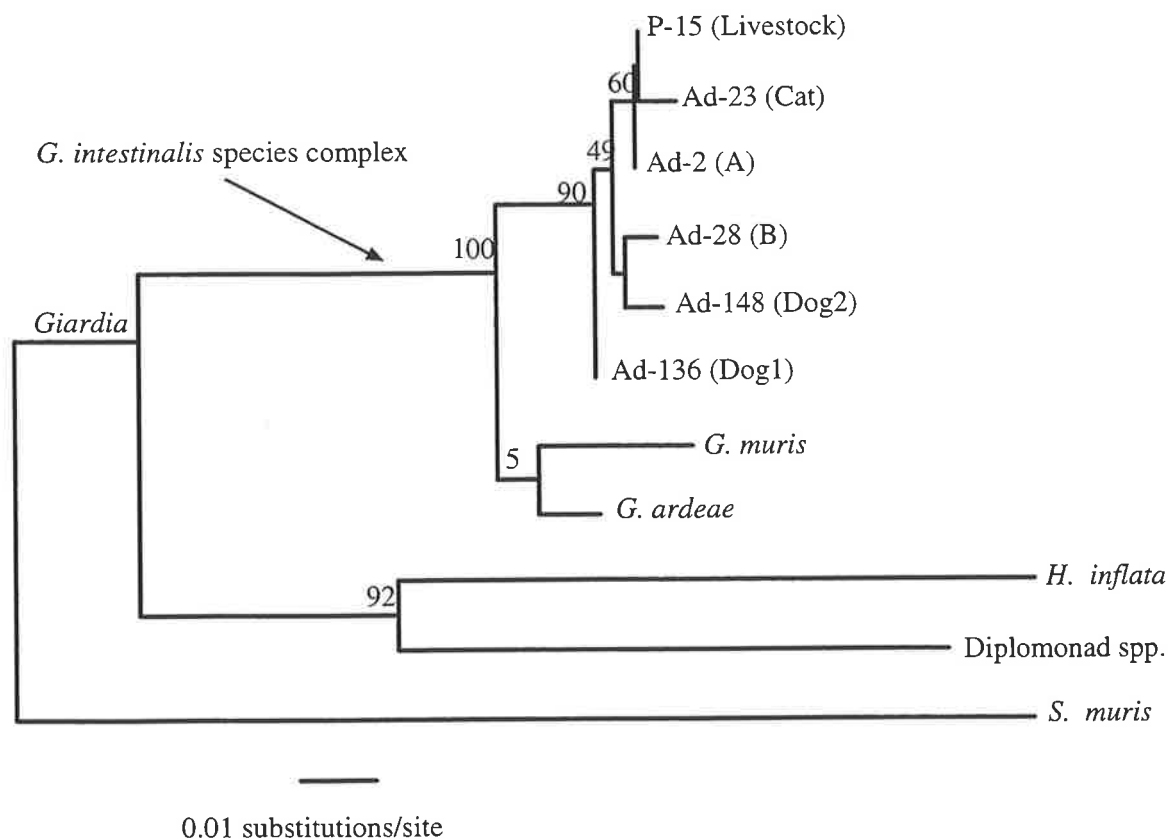


Figure 49. Phylogenetic relationships of isolates inferred from the *eflα* deduced amino acid sequences using Poisson corrected distances and the NJ (top) or ME (bottom) tree building methods. The allozyme defined lineage of an isolate is given in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (500 or 1,000 replicates) is indicated above the branches.

bootstrap replicates, but the support for any of the interior branches was much lower, resulting in an unresolved polytomy.

4.4.4.5 Maximum likelihood phylogenetic analysis of *eflα* nucleotide sequences

Maximum likelihood analyses were conducted using quartet puzzling for all codon base positions and each position separately, under both gamma and Poisson substitution models (Figures 50-53, pp. 151 - 154). The substitution model used did not appear to affect the tree topology, although as would be expected, some of the branch lengths were different. The topologies of the ML trees were very similar to the trees determined by the distance analyses. Examination of the separate codon base positions showed that all supported the monophyly of the *G. intestinalis* species complex. The 2nd codon positions were too conserved to resolve the relationships of the lineages within the *G. intestinalis* species complex (Figure 52, p. 153), while substitutions at the 1st codon positions supported the monophyly of the Assemblage A (Ad-2), Livestock (P-15) and Cat (Ad-23) lineages but could not determine the positions of the Assemblage B (Ad-28), Dog1 (Ad-136) or Dog2 (Ad-148) lineages (Figure 51, p. 152). The significant variation of the nt composition at the 3rd codon position among isolates made the results using these sites questionable, although the overall tree topology was similar to that of the NJ trees (Figure 48, p.148). ML analysis of the deduced amino acid sequences (Figure 54, p. 155) gave a tree with similar topology to analysis using 1st codon positions only (Figure 51, p. 152).

4.4.4.6 Cladistic analysis of *eflα* nucleotide sequences

Parsimony analyses were conducted using PAUP for the nt sequences, and PHYLIP (PROTPARS) for the deduced amino acid sequences. *G. muris* and *G. ardeae* were used as outgroups. An exhaustive search using PAUP found only 1 most parsimonious tree of 81 steps in length (Figure 55, p. 156). The topology of this tree was the same as those determined by distance and ML analyses (Figures 48 & 50, pp. 148 & 151). Bootstrap analysis using the branch and bound searching algorithm found high (96%) support for the clade containing the Assemblage A, Assemblage B, Cat and Livestock lineages, but poor (62%) support for the clade containing the Assemblage A, Cat and Livestock lineages. The positions of the Dog1 and Dog2 lineages could not be resolved. Although poorly supported, the clade containing the Assemblage A, Cat, and Livestock lineages was consistent with the distance-based analyses, and both the 18S rRNA sequence and AE analyses.

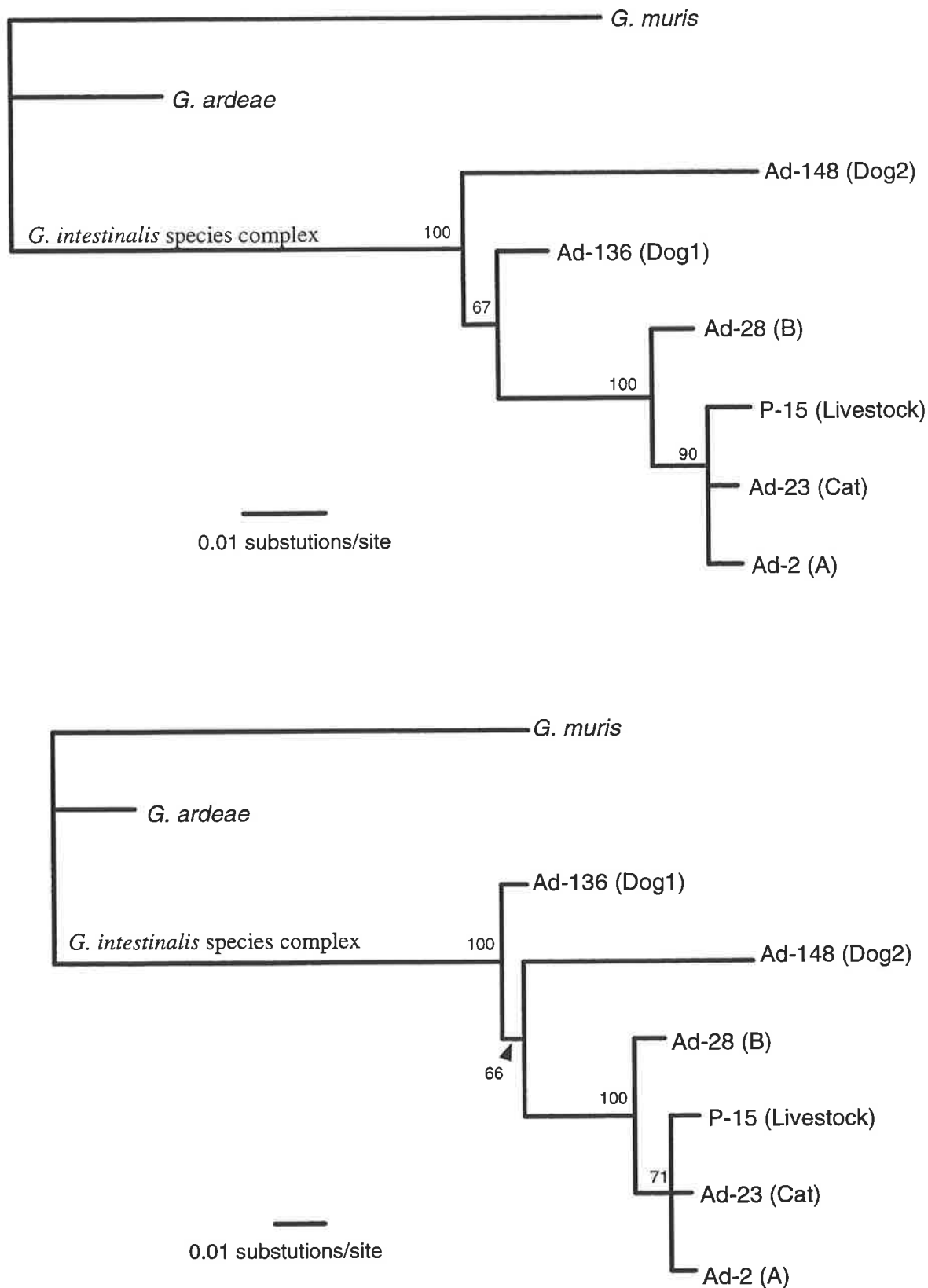


Figure 50. Phylogenetic relationships of *Giardia* isolates inferred from the *eflα* nt sequences using ML analysis assuming Poisson (top) or gamma (bottom) distributed substitution rates. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (1,000 quartet samples) is indicated above the branches.

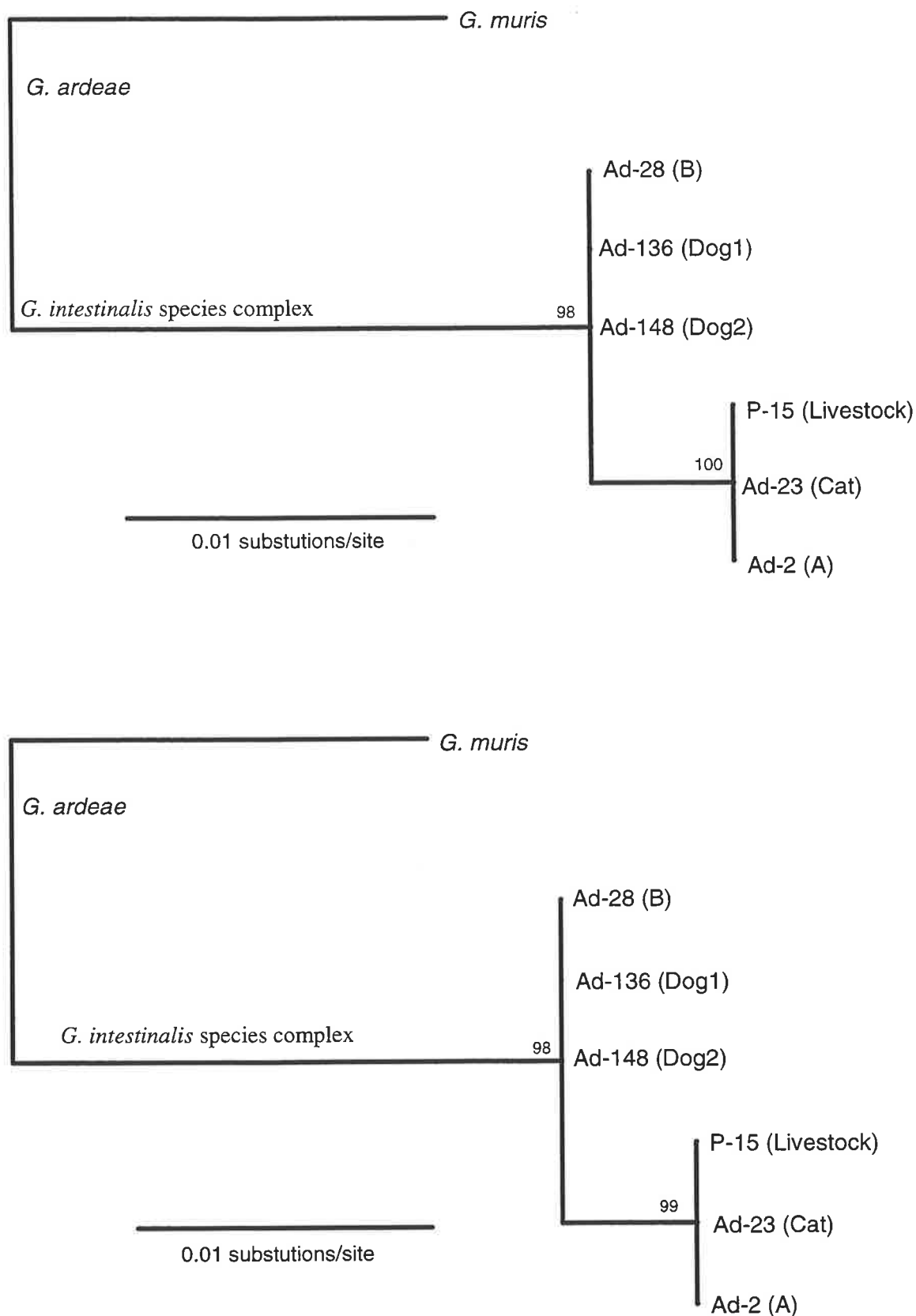


Figure 51. Phylogenetic relationships of *Giardia* isolates inferred from the *eflα* nt sequences (1st codon positions only) using ML analysis assuming Poisson (top) or gamma (bottom) distributed substitution rates. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (1,000 quartet samples) is indicated above the branches.

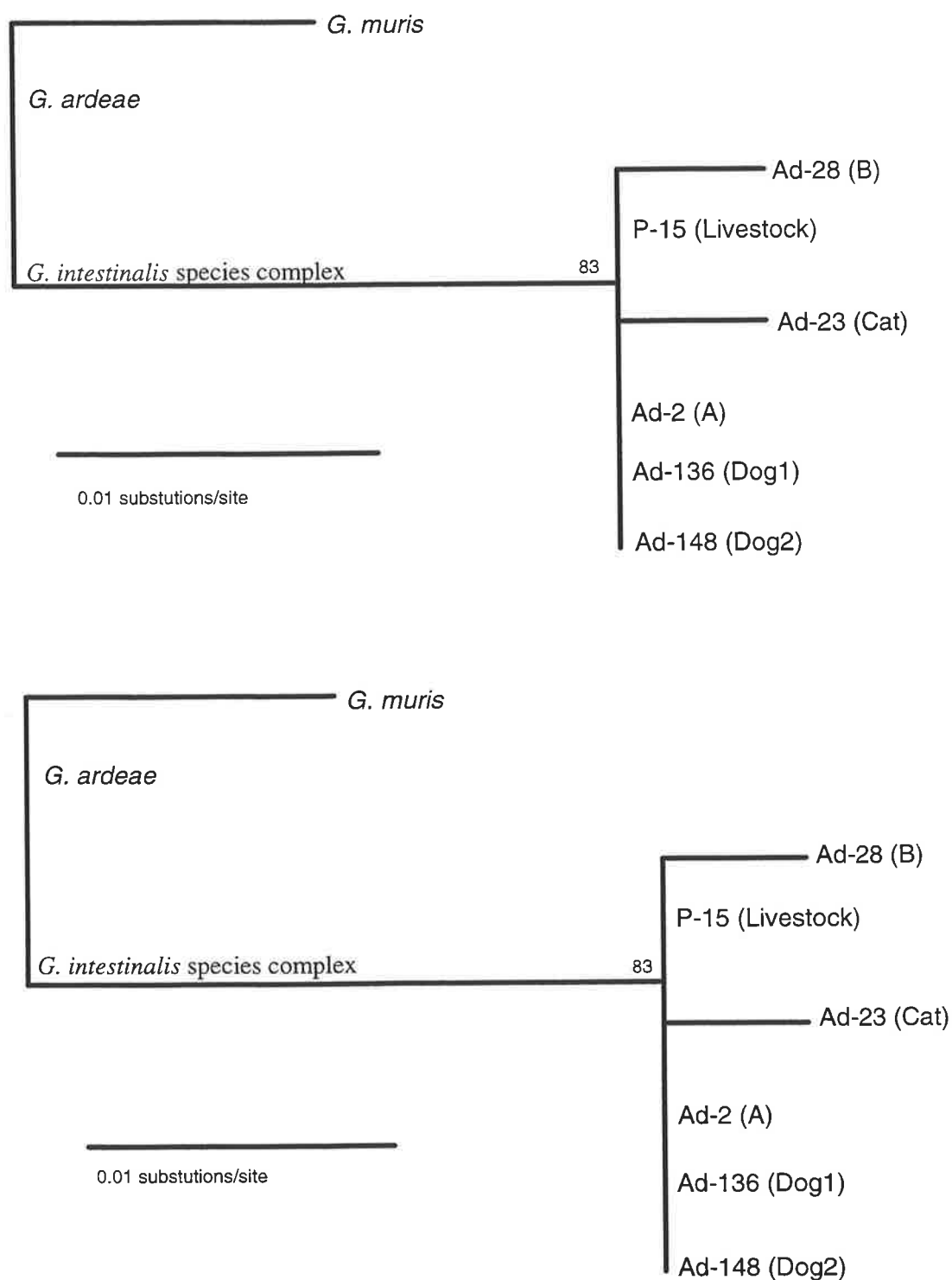


Figure 52. Phylogenetic relationships of *Giardia* isolates inferred from the *eflα* nt sequences (2nd codon positions only) using ML analysis assuming Poisson (top) or gamma (bottom) distributed substitution rates. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (1,000 quartet samples) is indicated above the branches.

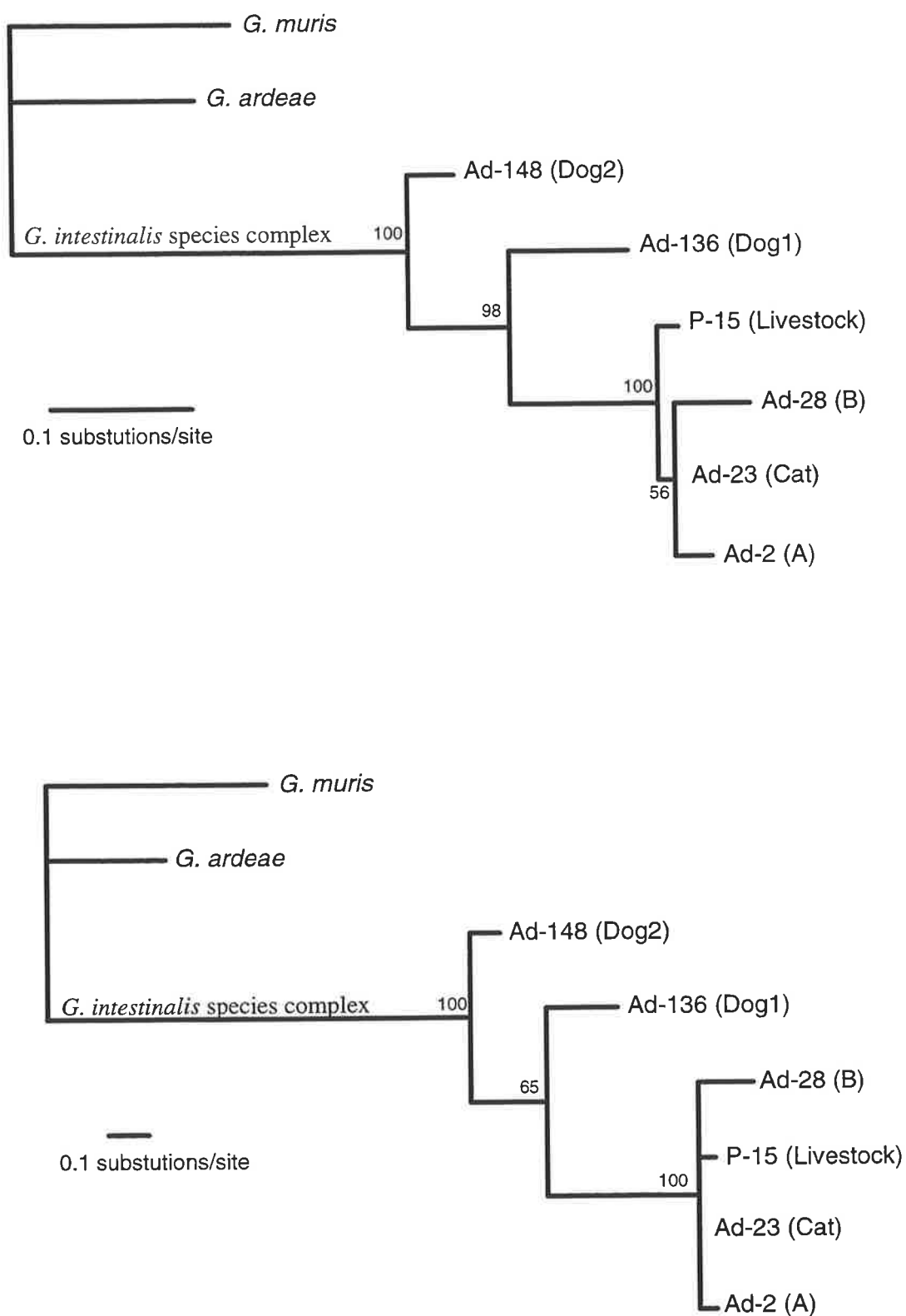


Figure 53. Phylogenetic relationships of *Giardia* isolates inferred from the *eflα* nt sequences (3rd codon positions only) using ML analysis assuming Poisson (top) or gamma (bottom) distributed substitution rates. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (1,000 quartet samples) is indicated above the branches.

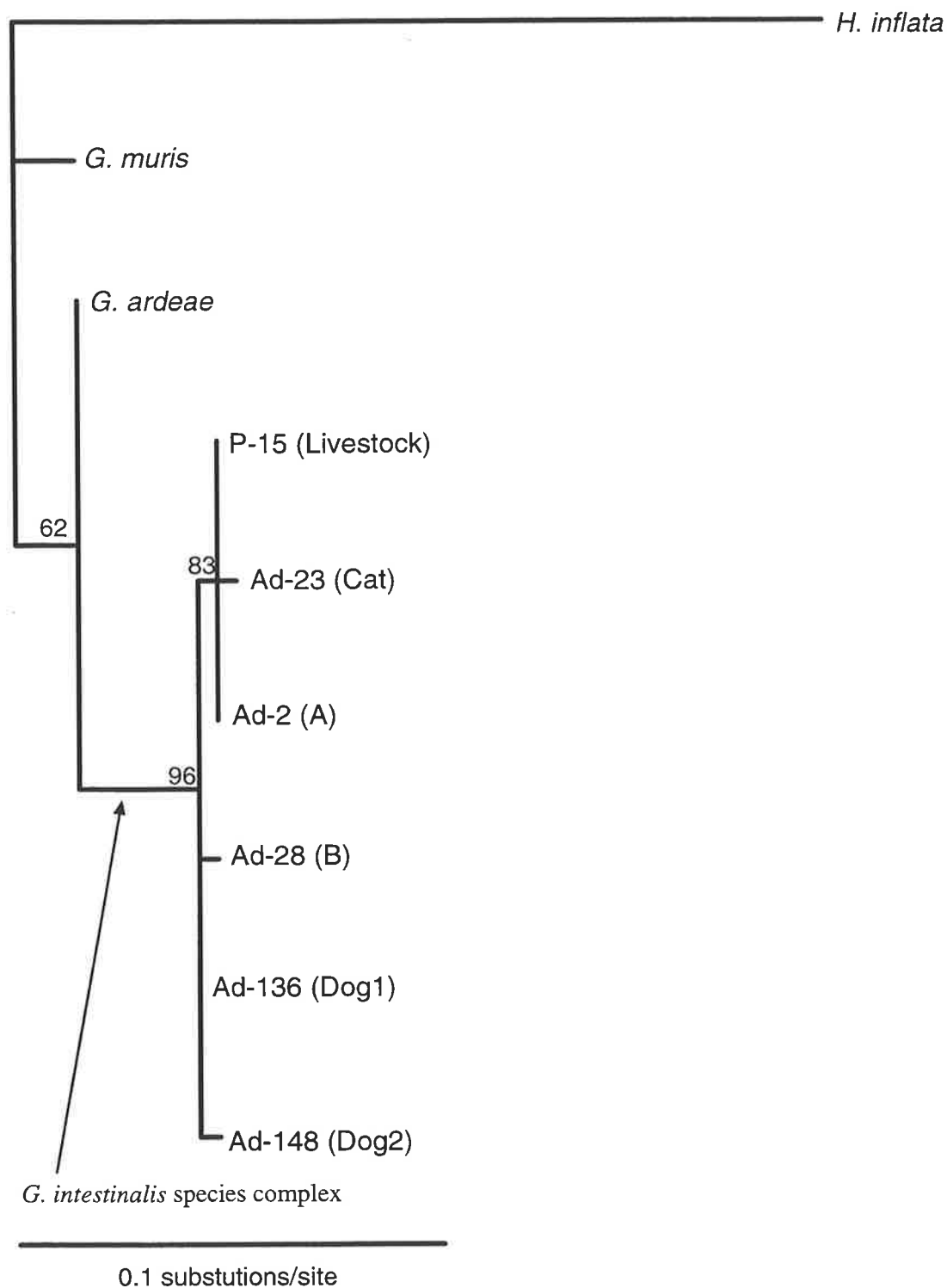


Figure 54. Phylogenetic relationships of *Giardia* isolates inferred from the *eflα* deduced amino acid sequences using ML analysis assuming Poisson substitution rates. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (1,000 quartet samples) is indicated above the branches.

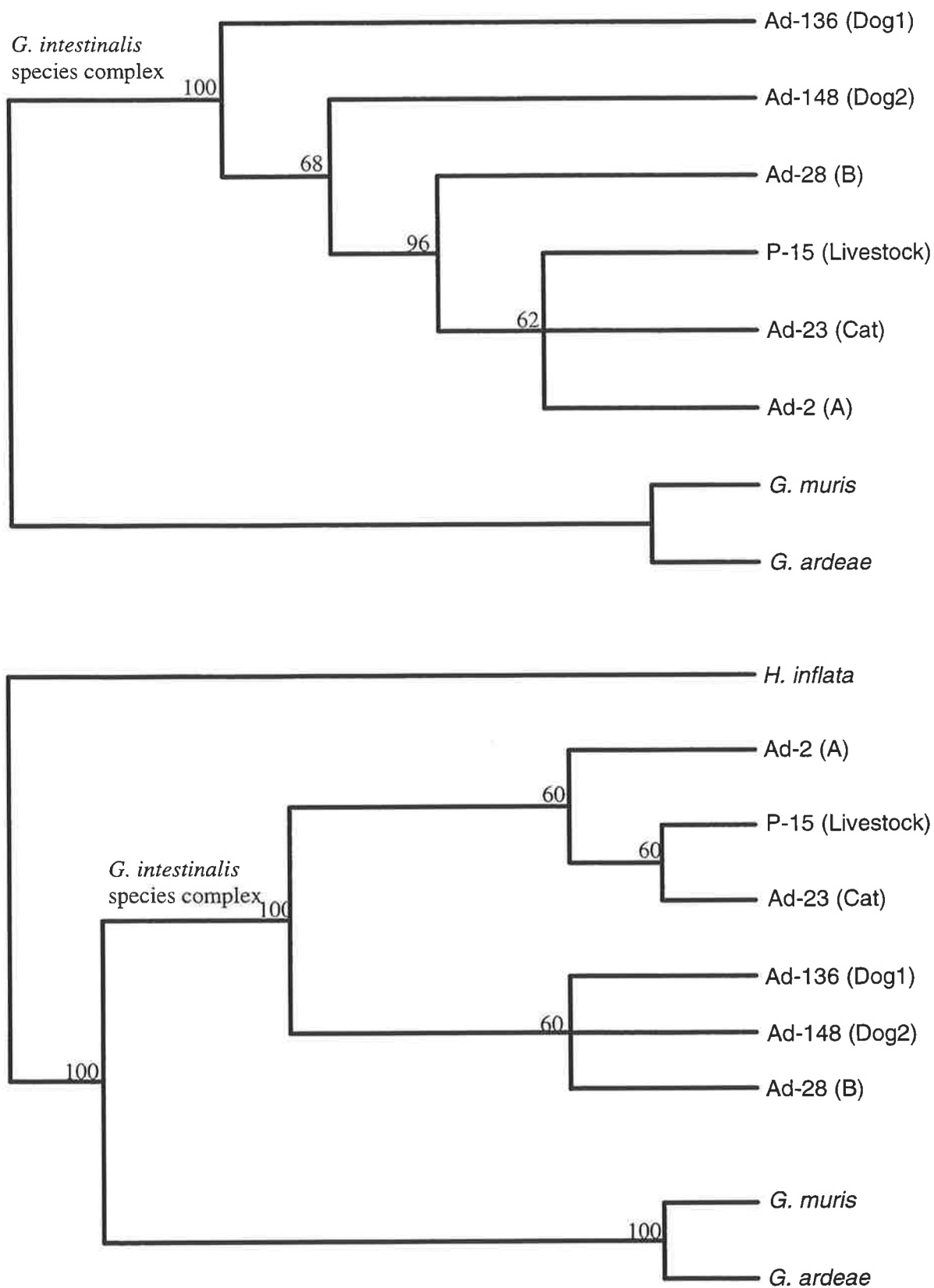


Figure 55. Cladistic analyses of *eflα* nt (top) & deduced amino acid (bottom) sequences. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). **Top:** Single most parsimonious tree identified using the exhaustive search of PAUP. Bootstrap values (% of 1,000) were calculated using the branch and bound algorithm. **Bottom:** Consensus of the 50 most parsimonious trees identified by parsimony analysis using PROTPARS. The % of trees in which particular branches occur is indicated above the branch.

Parsimony analysis using protein sequences found 50 equally most parsimonious trees (consensus tree Figure 55, p. 156). There was very little structure in the consensus tree, except for the monophyly of *Giardia*, the monophyly of the *G. intestinalis* species complex, and the monophyly of *G. muris* and *G. ardeae*.

The amount of phylogenetic signal in the data was tested by plotting the tree length distributions of the trees examined in the exhaustive search (Figure 56, p. 158). A total of 10,395 trees were examined. The mean tree length was 120.17 ± 9.54 steps, compared to the length of the shortest tree, which was 81 steps. The g_1 statistic was -1.45, which indicates significant skewness of the distribution, and hence phylogenetic signal. In order to determine the origin of the phylogenetic signal, the outgroups were removed and the analysis was repeated. The exhaustive search of the data containing only *Giardia* sequences evaluated 105 trees, and found a single shortest tree 20 steps in length. The average tree length was 28.77 ± 3.29 steps, and the g_1 statistic was -1.15, indicating a significant skewness of the distribution. Therefore, phylogenetic signal was also present within the *Giardia efla* nt sequence data.

4.4.5 Summary and conclusions

The phylogenetic analyses of segments of the genes encoding 18S rRNA and EF1 α provide support for the genetic model of *G. intestinalis* proposed in Figure 23 (p. 101), and validate the lineages identified by AE analysis. The topologies of the trees generated by all methods were very similar. A consensus tree for the lineages based upon the results of the three separate analyses is shown in Figure 57 (p. 159).

The analyses support the grouping of the Assemblage A and Livestock lineages, with the Cat lineage as a sister taxon to this clade. The grouping of these three taxa, as well as the branching order, was supported by the majority of the statistical tests employed. Assemblage B was the sister taxon to this clade, with the Dog1 and Dog2 lineages outside of this grouping. This branching order was supported by bootstrap analysis of the *efl* α sequences, but not the 18S rRNA or AE data. All of the analyses supported the monophyly of the *G. intestinalis* isolates.

All of the data suggested that *G. muris* was the most distantly related outgroup to the *G. intestinalis* species complex. The placement of *G. ardeae* could not be reliably determined, resulting in a polytomy containing all three morphological species. Analysis of

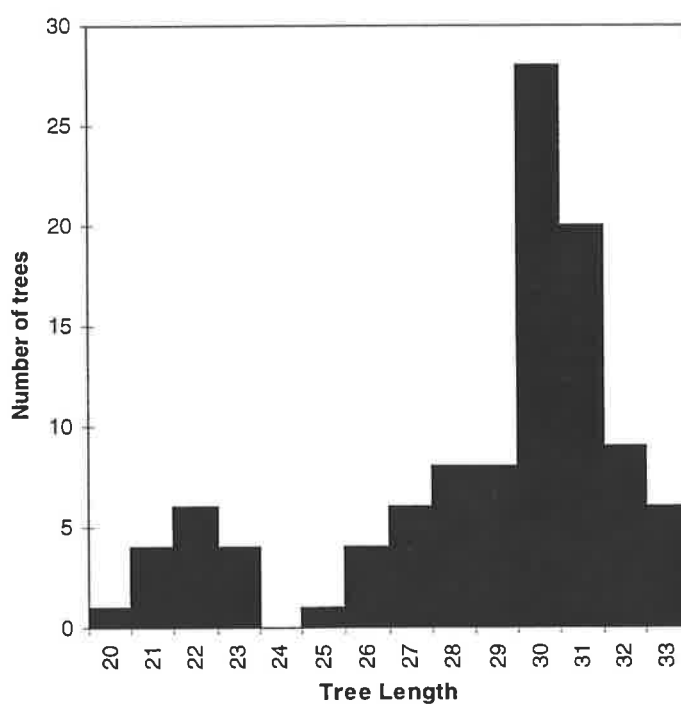
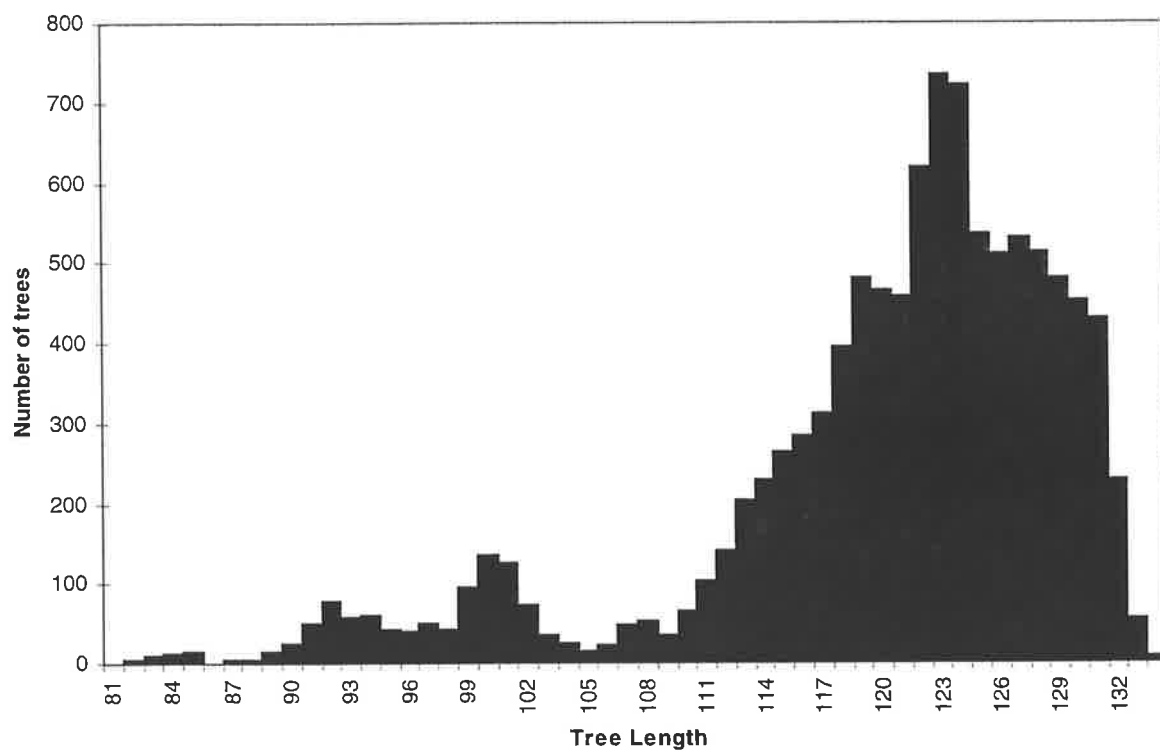


Figure 56. Distributions of tree lengths for trees examined in the *eflα* exhaustive searches with (top) or without (bottom) outgroup sequences.

the 18S rRNA sequences suggested that *G. psittaci* is a close sister taxon to the *G. intestinalis* species complex. Both the 18S rRNA and EF1 α gene sequences were too conserved to detect any significant intra-lineage variation. In contrast, AE could identify a number of distinct genetic groups within the major lineages, all of which were supported by bootstrap analysis (Chapter 3).

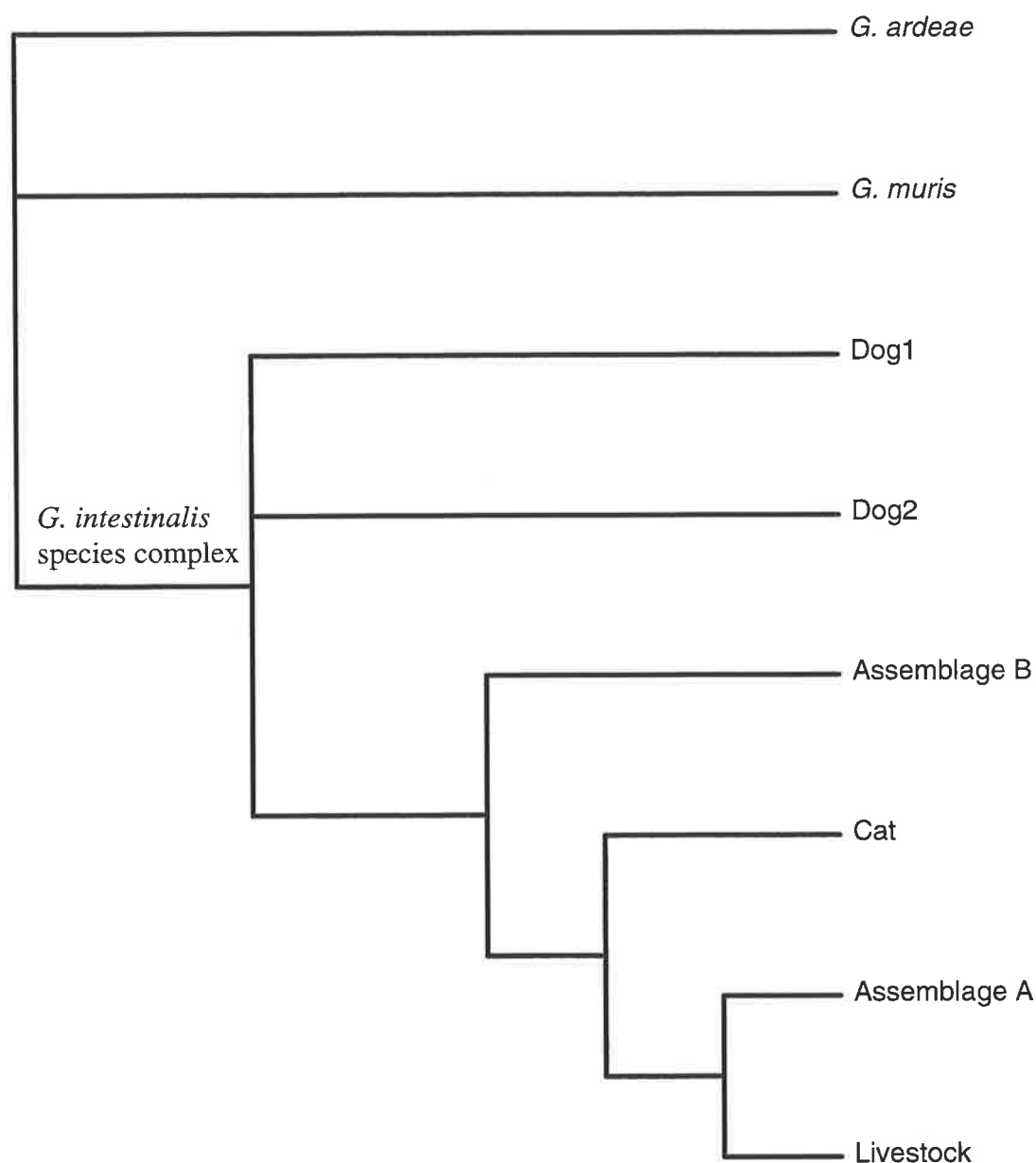


Figure 57. Consensus tree depicting evolutionary relationships of the major lineages of *G. intestinalis* inferred from analyses of the AE, 18S rRNA and *efl* α data.

4.5 Phylogenetic analyses of the *gdh* and *tpi* nucleotide sequences

Phylogenetic analyses of 18S rRNA and *efl* α nt sequences in the preceding section demonstrated the existence of discrete lineages within the *G. intestinalis* species complex, which supported the lineages identified by AE analysis. Unfortunately, the genes chosen were too conserved to allow very good resolution of the phylogenetic relationships between or within the lineages. To clarify these relationships, moderately conserved genes, such as those which encode housekeeping (eg., glycolytic enzymes), were examined. This approach was successful for Nelson *et al.* (1991) and Lawrence *et al.* (1991), who used glyceraldehyde-3-phosphate dehydrogenase (*gapd*) nt sequences to phylogenetically analyse species of *Salmonella* and *Escherichia* that exhibited levels of allozyme variation comparable to that detected among the *G. intestinalis* lineages.

4.5.1 Background

The glutamate dehydrogenase locus was the first *Giardia* gene encoding a metabolic enzyme to be completely sequenced and characterised. It was fortuitously cloned by Yee & Dennis (1993) while screening a cDNA library for ribosomal protein genes, and was found to be similar to the eubacterial and fungal genes encoding NADP-dependent glutamate dehydrogenase. Biochemical studies on the recombinant protein confirmed that the *Giardia* Gdh was NADP-dependent, and Southern analyses demonstrated that it was present as a single copy (Yee & Dennis, 1993). There was no evidence for the presence of a NAD-dependent GDH encoding gene (Yee & Dennis, 1993). Phylogenetic analyses of *gdh* sequences provided evidence for the existence of two paralogous *gdh* loci, one possessed by vertebrates and archaeobacteria, and the other possessed by the eubacteria and more primitive eukaryotes, such as *G. intestinalis* and fungi (Benachenhou-Lahfa *et al.*, 1993). The *Giardia* sequence was most similar to the *gdh* of *Chlorella sorokiniana* ($\approx$ 67% nt sequence similarity), followed by the eubacterial sequences ($\approx$ 60% nt sequence similarity). The availability of a published *G. intestinalis* *gdh* sequence, plus the fact that polymorphisms were detectable at this locus by AE analysis (Chapter 3), were the primary reasons for choosing to examine this locus.

The gene encoding triose phosphate isomerase (TPI) was originally cloned by Mowatt *et al.* (1994), who used PCR products amplified from *G. intestinalis* (isolate WB) as probes. The cloned *G. intestinalis* *tpi* could functionally complement an *E. coli* *tpi* deficient mutant

(Mowatt *et al.*, 1994), even though there was only minimal similarity, with the exception of small conserved domains, between the *G. intestinalis tpi* sequence and *tpi* from other organisms (prokaryotic or eukaryotic). *Tpi* sequences were available (courtesy of Dr. Michael Mowatt, NIH, Bethesda, USA) from two different isolates of *G. intestinalis* (WB and GS/M) which belong to Assemblages A and B respectively (P. Ey, unpub. data). Comparison of the nt sequences from isolates within Assemblages A and B for *tpi* (0.187 p-distance), *eflα* (0.014 p-distance), and 18S rRNA (0.021 p-distance) indicates that *tpi* is much less conserved than the latter two genes, and would therefore be an ideal candidate for determining the intra-lineage relationships within the *G. intestinalis* species complex.

4.5.2 Aims

The focus of this part of the project was to identify and utilise genes that were less conserved than either the 18S rRNA or *eflα* loci for the purpose of further resolving the phylogeny of the lineages comprising the *G. intestinalis* species complex. A secondary aim was to identify sequence motifs (eg., restriction endonuclease cleavage sites) which could be used in diagnostic assays to reliably identify the genotypes which comprise each lineage.

Hypotheses to test:

- The clusters identified by AE analysis represent distinct lineages.
- Less conserved genes will provide more information about the phylogeny of isolates within lineages.

4.5.3 Materials and methods

4.5.3.1 List of isolates

Isolates were chosen to represent the lineages identified by the earlier studies, and are listed in Table 22 (p. 162). Additional isolates for which only DNA was available were also characterised so that they could be placed within the current genetic scheme. DNA samples from the isolates listed in Table 22 as Polish or Belgian genotypes were provided by Dr. Wieger Homan (Parasitology Laboratory, National Institute of Public Health and Environmental Protection, The Netherlands). DNA samples from animal-derived isolates originating from the Czech Republic were provided by Prof. Jaroslav Kulda (Department of Parasitology, Charles University).

Table 22. List of *Giardia* isolates used in section 4.5.

Species	Isolate	Genotype	Host	Location	propagation method
<i>G. intestinalis</i>	Portland 1	A	human	Oregon, USA	axenic
	Ad-1	A	human	Adelaide, SA	axenic
	Ad-2	A	human	Adelaide, SA	axenic
	Bris136	A	Human	Brisbane, SA	axenic
	Ad-7	B	human	Adelaide, SA	s/m
	Ad-19	B	human	Adelaide, SA	s/m
	Ad-28	B	human	Adelaide, SA	s/m
	Ad-52	B	human	?	s/m
	Ad-121	?	human	SA	s/m
	AMC-3 ¹	Belgian	human	Africa	axenic
	LD-20 ¹	Belgian	human	Africa	axenic
	CP-117 ¹	Belgian	cuis	Poland	axenic
	GGPRP ¹	Belgian	pouched rat	Poland	axenic
	LSLP-116 ¹	Belgian	slow loris	Poland	axenic
	SP-115 ¹	Belgian	siamang	Poland	axenic
	Ad-156	B	marmoset	Adelaide, SA	s/m
	Ad-23	Cat	cat	Adelaide, SA	s/m
	Ad-142	Cat	cat	Adelaide, SA	s/m
	Ad-154	Cat	cat	Adelaide, SA	s/m
	WOOF	B	dog	Canada	axenic
	D47	?	dog	Czech Republic	axenic
	Ad-136	Dog1	dog	Alice Springs, NT	clinical
	Ad-137	Dog1	dog	Port Augusta, SA	s/m
	Ad-147	Dog1	dog	Alice Springs, NT	s/m
	Ad-141	Dog1	dog	Port Augusta, SA	s/m
	Ad-148	Dog2	dog	Ceduna, SA	s/m
	Ad-119	Livestock	calf	Adelaide, SA	s/m
	CH-105	?	chinchilla	Czech Republic	axenic
	P-15	Livestock	pig	Czech Republic	axenic
	P2-MER	?	pig	Czech Republic	axenic
	G-247	?	goat	Czech Republic	axenic
	Ad-159	?	horse	Adelaide, SA	s/m
	J-17	?	lamb	Czech Republic	axenic
	J-245	?	lamb	Czech Republic	axenic
	Ad-152	Livestock	sheep	Adelaide, SA	s/m
	Ad-165	?	nilgai	?	s/m
	Ad-149	?	alpaca	?	s/m
<i>G. ardeae</i>	<i>G. ardeae</i>		blue heron	Minnesota, USA	axenic
<i>G. muris</i>	Underdown		mouse	Canada	s/m
	Ad-120		mouse	Adelaide, SA	s/m
	Ad-160		mouse	Adelaide, SA	s/m

1. Homan *et al.* 1992

s/m = grown in suckling mice

SA = South Australia

? = not known

NT = Northern Territory

4.5.3.2 Design of primers for the *gdh* locus

Degenerate primers were constructed for regions of the *gdh* locus that were conserved between the published *G. intestinalis* (Portland 1, Yee & Dennis, 1993), *Escherichia coli*, *Chlorella sorokiniana*, and yeast sequences (GenBank accession numbers: K00565, X58832, & M11297; Figure 58). Internal sequencing primers were designed by selecting segments which contained codons possessing ≤ 2 -fold degeneracy, with sufficient spacing (approximately 350 bp) between primers to allow full overlap by cycle sequencing. Additional primers were synthesised as required, using sequence data accumulated during the course of this work to identify conserved regions (Figure 59, pp. 164 - 165).

Figure 58. Design of primers #578 & #579.

Portland 1	GAGAGGATGCTTGAGCCGGAGCGCGTCATC
<i>C. sorokiniana</i>	A..CA...CG.....T.....G...
<i>E. coli</i>	...C.TC..G....A.....G...
Yeast	CCA.TTG.TTC..TT..A..AA.AA....A
#578	→5' GAGAGGATCCTTGARCCNGAGCGCGTNATC 3'→
Portland 1	GGCGCGAACATCGCCGGGTTTCCTGAAGGT
<i>C. sorokiniana</i>	C.....G..C...ACC.....
<i>E. coli</i>	T.....T..TG.....
Yeast	..T..T..T.....AA.T...A.C.....
#579	←3' CCGCGNTTGTADCGNCCNAAGATCTTCCA 5'←

4.5.3.3 *Gdh* PCR conditions

Amplification of a 1.17-kb segment from *gdh*

Amplifications (94°C for 2 min, then 30 cycles of 94°C for 30 s, 56°C for 30 s and 72°C for 2 min, with a final 72°C extension for 7 min) were performed on a Perkin Elmer GeneAmp PCR System 2400, using 0.2 ml thin wall tubes, in 50 µl reaction volumes containing 1 x *Taq* reaction buffer (see section 4.4.2.3), 4 mM MgCl₂, 0.2 mM of each dNTP, 0.8 µM of each primer (#578 and #579), 5% DMSO, 1 unit of *Taq* DNA polymerase and *Giardia* DNA (50-200 ng).

Amplification of other segments from *gdh*

Primer pairs #1051/#1052 and GDHGENF/GDHGENR were used as described above, but with different cycle conditions (94°C for 2 min, then 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min, with a final 72°C extension for 7 min), and without adding DMSO to the reaction mixture.

G.ardeae CGC CCC GAG GCC ACG GGC TAT GGT GCC GTG TAC TAC CTC GAA GAG ATG TGC AAG GAC AAC GGC AGT CCG TCC CGT GGG AAG ACT ATC
 Portland1 AGG CCG GAG GCC ACG GGC TAT GGC GCT GTC TAC TTC CTG GAG GAG ATG TGC AAG GAC AAC AAC ACT GTG ATC AGG GGT AAG AAC GTC
 Ad-148 AGG CCG GAG GCC ACG GGC TAT GGT GCC GTC TAC TTC CTT GAG GAG ATG TGC AAG GAC AAC AAC ACC ATA ATC AGG GGC AAG AAC GTC
 Ad-147 AGG CCA GAG GCC ACC GGA TAT GGC GCT GTC TAC TTC CTC GAG GAG ATG TGC AAG GAC AAC AAC ACC ATA ATC AGG GGT AAG AAC GTC
 Ad-7 AGA CCA GAG GCC ACA GGG TAT GGA GCT GTC TAC TTC CTG GAG GAG ATG TGC AAG GAT AAC AAC ACC GTA ATC AGG GGC AAG AAC GTC
 GAGDH500F 5' GAG ATG TGC AAG GAG AAC 3' →

G.ardeae CTC CTC TCG GGC TCT GGG AAC GTC GCC CAG TTC GCA TGC CAG AAG CTC AAC GAG CTC GGT GCC CGC GTT ATC TCC TTC TCC GAC TCG
 Portland1 CTT CTT TCT GGC TCC GGC AAC GTT GCC CAG TTT GCT TGC GAG AAG CTC ATT CAG CTC GGC GCA AAG GTC CTC ACC TTC TCA GAC TCC
 Ad-148 CTG CTC TCT GGT TCT GGA AAC GTC GCT CAA TTC GCG TGC GAG AAA CTC CTT CAG CTA GGC GCA AAA GTG CTT ACA TTC TCT GAC TCT
 Ad-147 CTC CTC TCC GGG TCC GGC AAC GTT GCC CAG TTC GCG TGC GAG AAG CTC ATC CAG CTC GGC GCA AAG GTC CTC ACC TTC TCT GAC TCC
 Ad-7 CTC CTC TCT GGC TCT GGC AAC GTT GCT CAG TAC GCG TGC GAG AAG CTC CTC CAG CTC GGT GCG AAG GTC CTC ACC TTC TCG GAC TCC

G.ardeae AAC GGC ACA GTC ATC GAC GAG GAG GGC TTC AAC GAG GAG AAG CTC GCC CAC GTG ATG TAC ATC AAG AAC GAG AAG CGC GGC
 Portland1 AAC GGG ACC ATT GTC GAC AAG GAC GGG TTC AAC GAG GAG AAG CTG GCC CAC CTC ATG TAC CTC AAG AAC GAG AAG CGT GGG CGT GTT
 Ad-148 AAC GGA ACC ATC GTC GAT AAG GAT GGC TTC AAC GAG GAG AAA CTT ACT CAC CTC AAG TAC CTC AAG AAC GAG AAG CGT GGT
 Ad-147 AAC GGA ACC ATC GTC GAC AAG GAT GGC TTC AAC GAG GAG AAG CTT GCC CAC GTC AAG TAT CTC AAG AAC GAG AAG CGC GCT
 Ad-7 AAC GGG ACT ATC GTC GAT AAG GAT GGC TTC AAC GAG GAG AAG CTG GCC CAC CTC ATA CAC CTC AAG AAT GAG AAG CGC GGG
 #1052 3' G CTR TTC CTR CCS AAG TTG C 5' ←
 GDHGENR 3' CCS AAG TKG CTC CTC TTY GA 5' ←

Portland1 TCC GAG TTC AAG GAC AAG TAT CCC AGC GTC GCG TAC TAC GAG GGC AAG AAG CCT TGG GAG TGC TTC GAG GGC CAG ATG GAT
 #862 3' TC ATA GGG TCG CAG CGC ATG A 5' ←

1 position 1 is equal to nucleotide position 398 from the published Portland 1 sequence, GenBank Accession number M84604.

The primer pair #578/#862 could only be successfully used when amplifications (25 cycles of 95°C for 10 s, 55°C for 10 s, 72°C for 20 s, with a final 72°C extension of 7 min) were carried out on a Corbett Research FTS-1 thermal cycler, which used 20 µl capillary tubes and very short ramp times between steps. Otherwise, the reaction conditions were the same as those described in the preceding paragraph. These amplifications failed when a Corbett Research FTS-320 was used, even though the same program and PCR reaction mixtures were used. The main difference between the two machines were (1) the ramping times between steps, with the FTS-320 being slower, and (2) the reaction tubes, which were standard 0.5 ml PCR tubes in the FTS-320. The ability of #862 to act as a sequencing primer was not affected by the type of thermal cycler used.

4.5.3.4 Design of *tpi* primers

The full length *tpi* sequences, as well as primers oMM44 - oMM91 (see Figure 60, p. 167) were kindly provided by Dr. Michael Mowatt (Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, USA). Additional primers were also constructed (designated TPI16F and TPI533R), using either the two published sequences (Mowatt *et al.*, 1994), or the alignment of the sequences determined herein. All of the primers used are listed in Figure 60 (p. 167).

4.5.3.5 *Tpi* PCR conditions

Amplifications (94°C for 2 min, then 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min, with a final 72°C extension for 7 min) were performed on a Perkin Elmer GeneAmp PCR System 2400, using 0.2 ml thin wall tubes, in 50 µl reaction volumes containing 1 x *Taq* reaction buffer (see section 4.4.2.3), 4 mM MgCl₂, 0.2 mM of each dNTP, 0.8 µM of each primer, 1 unit of *Taq* DNA polymerase and *Giardia* DNA (50-200 ng). These conditions were adequate for all primer pair combinations, although the amplification of *tpi* from *G. ardeae* DNA required the inclusion of 5% DMSO.

Figure 60. Position of primer sequences within *tpi* sequenced from the isolate WB (Mowatt *et al.*, 1994).

The sequence begins at the start codon. Primers in bold were used for PCR and/or DNA sequencing.

```

ATG CCT GCT CGT CGC CCC TTC ATC GGC GGT AAC TTC AAG TGT AAC GGC TCT CTT GAC TTT ATC AAG AGC CAC GTG GCG GCA ATT GCT GCC 90
      TPI19F 5' TTY ATC GGY GGT AAC TTY AAR TGY AA 3' →
      TPIGENF 5' ATC GGY GGT AAY TTY AAR TG →
      T TPI16F 5' CCC TTC ATC GGY GGT AAC 3' →
oMM91 5' GCT CGT CGC CCC TTC ATC GG 3' →

CAT AAG ATC CCT GAT TCC GTG GAC GTC GTC ATT GCC CCT TCC GCC GTA CAC CTG TCA ACA GCC ATT GCG GCA AAC ACG TCA AAA CAG TTG 180
oMM58 3' GGA CTA AGG CAC CTG CAG 5' ←
      oMM78 3' G CAG CAG TAA CGG GGA 5' ←

AGG ATA GCA GCG CAG AAT GTG TAC CTA GAG GGG AAC GGG GCG TGG ACT GGC GAG ACA AGT GTT GAG ATG CTT CAG GAC ATG GGT TTG AAG 270
oMM57 5' GCG CAG AAT GTG TAC C 3' →

CAT GTG ATA GTA GGG CAC TCT GAA AGA CGC AGA ATC ATG GGG GAG ACC GAC GAG CAA AGC GCC AAG AAG GCT AAG CGT GCC CTG GAA AAG 360
      oMM47 3' T TCT GCG TCT TAG TAC C 5' ←
oMM54 3' CCC GTG AGA CTT TCT G 5' ←

GGG ATG ACG GTC ATC TTC TGC GTC GGA GAG ACC TTG GAC GAG CGC AAG GCC AAC CGC ACC ATG GAG GTG AAC ATC GCC CAG CTT GAG GCG 450
      oMM55 5' TG GAG GTG AAC ATC GC 3' →

CTT GGC AAG GAG CTC GGA GAG TCC AAG ATG CTC TGG AAG GAG GTT GTC ATT GCT TAC GAG CCC GTG TGG TCC ATT GGC ACG GGC GTG GTG 540
CTT GGC AAG GAG CTC GG 3' oMM48 →
      TPI533R 3' CAC ACC AGR TAR CCG TGC CC 5' ←
      oMM50 3' GA ATG CTC GGG CAC AC 5' ←

GCC ACG CCC GAG CAG GCA GAG GAG GTC CAT GTG GGG CTC CGA AAG TGG TTT GCG GAG AAG GTT TGT GCC GAG GGC GCA CAG CAT ATC CGT 630
TPI572R 3' CTC GTY CGY CTC CTY CAG GTR CA 5' ←
oMM45 3' CGT CTC CTC CAG GTC C 5' ←

ATC ATT TAC GGG GGA TCG GCC AAT GGA AGC AAC TGC GAG AAG CTT GGC CAG TGT CCG AAT ATT GAC GGC TTC CTT GTC GGT GGC GCT TCC 720
      TPIGENR 3' ACR CTY TTY GAA CCG GTC AC 5' ←
oMM82 3' TTG ACG CTC TTC GAA CCG GT 5' ←

```

4.5.4 Phylogenetic analysis using *gdh* nucleotide sequences

4.5.4.1 Optimisation of PCR conditions

Preliminary testing of primers #578 and #579 found that they would not work (ie., produce the expected sized amplification product) under the predicted conditions (based on primer melting temperatures), or using low or high annealing temperatures (data not shown). Analysis of the Portland 1 *gdh* sequence identified a potential stem-loop at the 3' end, just upstream of primer #579. Co-solvents (DMSO, glycerol) and a denaturant (formamide) were used in attempts to remove the presumed structure. The addition of 5% DMSO was found to be essential for the successful amplification of the 1.17-kb *gdh* segment from *G. intestinalis* DNA samples using these primers. Use of 10% DMSO appeared to decrease the efficiency of amplification, resulting in lower yields of product (Figure 61). The putative secondary structure also appeared to inhibit sequencing reactions using primer #579 (addition of 5% DMSO to the sequencing reaction had no effect). The primer pair #578 and #579 worked for all *G. intestinalis* isolates, as well as *G. ardeae* and a variety of bacteria, but not for yeast or *G. muris* DNA samples (data not shown).

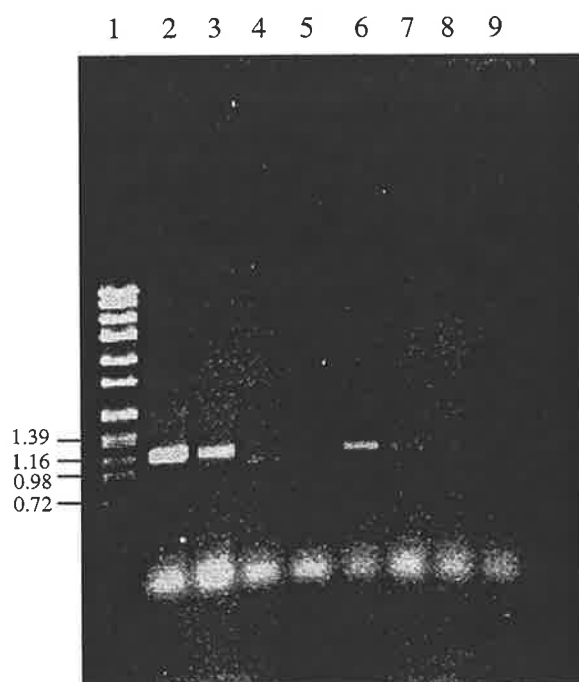


Figure 61. Amplification of a 1.17-kb segment from the *gdh* of *Giardia* DNA samples using primers #578 & #579 in the presence of different concentrations of DMSO (other reaction conditions given in section 4.5.3.3). Lane 1: *Eco*R1-digested bacteriophage SPP-1. Lanes 2-4: *G. intestinalis* (Ad-2, axenic), *G. intestinalis* (Ad-7, suckling-mouse derived) & *G. muris* (Underdown, suckling mouse derived) with 5% DMSO, and lanes 6 - 8: as for lanes 2 - 4 but with 10% DMSO in the reaction mixture. Lanes 5 & 9: -ve PCR controls using 5% or 10% DMSO.

All of the other primer sets that were utilised worked under standard conditions for most isolates. The exceptions were isolates of *G. muris*, from which the shorter *gdh* fragments could not be amplified under any conditions (not shown).

4.5.4.2 PCR controls

The amplification of contaminating host DNA in PCR reactions using DNA samples from isolates propagated in suckling-mice did not appear likely. The presence of introns within the mouse *gdh* would affect the size of the amplified fragment, and being paralogous to the *Giardia* *gdh*, was unlikely to bind the primers. However, there was a real possibility that enteric bacteria could contaminate trophozoites grown in suckling-mice. The expected 1.17-kb fragment could be amplified from a range of bacteria, including *E. coli*, *Salmonella typhi* and *Corynebacterium diphtheria* (data not shown). In order to examine the possibility of bacterial contamination of the suckling-mouse derived samples, uninfected specific pathogen free suckling mice were used as controls. Samples were collected from the uninfected suckling mouse intestines using the same protocol that was used to purify trophozoites from infected mice (section 2.2.3). DNA was extracted from both the pellets and supernatants and used in PCR reactions. DNA was detected in the pellet and supernatants, but it failed to amplify the 1.17-kb *gdh* segment (not shown). Although this indicated that bacterial contamination was unlikely to present a problem, it could be detected in some *Giardia* samples (eg., the *G. muris* sample used in Figure 61, p. 168). Numerous attempts were made to amplify *gdh* from *G. muris*, finally yielding a faint fragment of the expected size (not shown). However, sequencing of this fragment identified it as bacterial in origin, with a high similarity (> 90%) to the *gdh* of enteric Gram negative bacteria. There are several possible explanations for this result. *Giardia* infections in suckling-mice may promote bacterial growth above normal levels, or the original inoculum (usually purified from faecal material) may introduce bacteria which are not normally part of the suckling mouse intestinal flora. Alternatively, the presence of *Giardia* trophozoites may provide a mechanical vehicle to trap the bacteria. Such contamination could not be detected in any of the *G. intestinalis* samples, presumably because the *G. intestinalis* *gdh* was selectively amplified by the primers used in this assay.

4.5.4.3 Detection of samples comprising multiple genotypes using DNA sequencing

During the course of sequencing, some isolates consistently gave ambiguous results for specific nt positions from both strands, suggesting that the amplified DNA contained a

mixture of sequences derived from more than one genotype. This was tested by sequencing a 1:1 mixture of *gdh* PCR products amplified from Assemblage A (Ad-1) and Assemblage B (LD-20) DNA samples. The sequence of the individual products was unambiguous, but the mixed product clearly showed ambiguities, which appeared as two distinct peaks superimposed over each other (Figure 62, p. 171). The position and composition of each ambiguity were as expected from the known differences between the original sequences. However, because of differences in the relative efficiencies of incorporation of the different dye labelled di-terminators, not all of the predicted ambiguities were detected in both strands of the sequence. Two isolates were identified that either comprised mixed genotypes or for which the PCR mixture was contaminated; J-245, which appeared to be a mixture of Assemblage A and Livestock genotypes, and Ad-145, which appeared to be a mixture Dog1 and Dog2 genotypes.

4.5.4.4 DNA sequencing of *gdh* PCR products

PCR-amplified DNA from the majority of isolates were sequenced using the primers #578, #913, #914 and #862. The isolates from the Livestock lineage could not be sequenced using primer #862, so the full length sequence (690 nt) could not be determined. This prompted the development of new primers, which, linked with an improvement in the efficiency of the sequencing kit (by the manufacturer, section 2.1.1), allowed the sequencing of 500 nt with only two primers. The drawback of this approach was that the full 690 nt sequence could not be determined for all isolates. However, it provided almost the same amount of sequence information for half the cost.

4.5.4.5 Alignment of *gdh* nucleotide sequences

The *gdh* sequences could be readily aligned by eye. *G. ardeae* was used as the outgroup, as the *G. muris* sequence could not be determined. Although it would have been of interest to compare *gdh* from another morphological species, the amplification and sequencing of *G. muris* *gdh* was not pursued further as an alternative outgroup was available. The aligned sequences are shown in Appendix C. There was 76 - 79% nt sequence similarity between the *G. intestinalis* and *G. ardeae* sequences, which was similar to that observed between the published Assemblages A and B *tpi* sequences (Mowatt *et al.*, 1994). Nucleotide sequence similarity between the lineages comprising *G. intestinalis* ranged from 87% - 94%, with 99 - 100% similarity among isolates within each lineage.

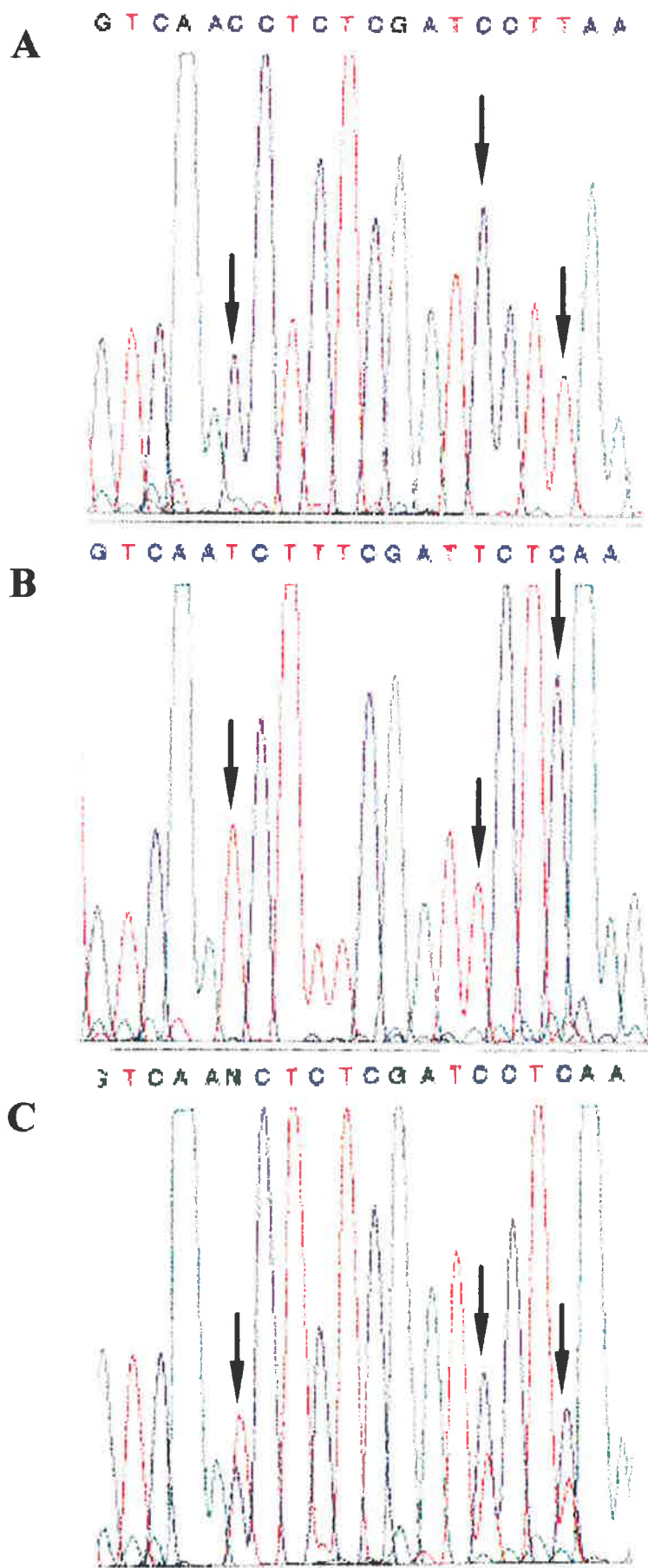


Figure 62. *gdh* DNA sequence chromatograms generated using primer #578 and purified #578/#579 amplification products from **A:** Ad-1 (Assemblage A), **B:** LD-20 (Assemblage B), & **C:** Ad-1/LD-20 1:1 mixture. Arrows indicate polymorphisms between Ad-1 & LD-20 which appeared as ambiguous sites in the mixed reaction.

4.5.4.6 RFLP analysis using the 1.17-kb *gdh* PCR product

Preliminary analysis of the *gdh* sequence data (Appendix C) indicated that some restriction endonuclease sites were diagnostic and could differentiate the different lineages comprising the *G. intestinalis* species complex. Additional diagnostic sites were identified by digestion of the 1.17-kb amplified DNA fragment with some of the more commonly available restriction enzymes. The resulting restriction site data is shown in Table 23. These differences formed the basis of a diagnostic system which was used to genetically characterise a panel of European isolates of *Giardia* which had been previously characterised as “Polish” or “Belgian” (Homan *et al.*, 1992). The major finding of this study was that the Polish group corresponded to Assemblage A, and the Belgian group corresponded to Assemblage B, suggesting that these two groupings have a world-wide distribution (Monis *et al.*, 1996; Appendix D).

Table 23. Restriction site data for the 1.1-kb *gdh* segment from different *Giardia* lineages.

Enzyme	AatII	Apal	Apal	EcoR1	EcoR1	HindIII	KpnI	RcaI	SacI	XhoI	XhoI
Site	1	2	1	2						1	2
Lineage/species											
Assemblage A	-	-	-	-	+	-	+	+	+	+	-
Assemblage B	-	-	+	-	-	-	-	-	+	+	-
Livestock	-	-	+	-	+	-	-	+	-	+	-
Cat	-	-	+	-	-	-	-	+	+	+	-
Dog1	-	-	+	-	-	+	-	+	+	-	+
Dog2	-	-	+	+	-	+	-	+	+	-	+
<i>G. ardeae</i>	+	+	+	-	-	-	-	+	+	?	-

+ = restriction site present, - = restriction site absent

4.5.4.7 Determination of *gdh* sequence parameters

The *gdh* sequences were examined using the same approach described in section 4.4.3.4. ML analyses indicated that the optimal s/v ratio was 2 (Table 24, p. 173). The nt composition was also determined (Table 25, p. 174). The sequences from isolates of *G. intestinalis* had GC contents of 58 - 62%, with a roughly equal ratio of G:C and A:T. The GC content of *G. ardeae* was almost 63%. The GC composition of 1st and 2nd codon positions was 56% and 38 - 39% respectively for *G. intestinalis*, with *G. ardeae* 2 - 3% higher in each case. The 3rd codon positions had a high GC content of 77 - 92% (including *G. ardeae*), as

well as at least 4 times as many A's than T's. The nt composition of the 3rd codon position was different between isolates from the different lineages.

Table 24. Log likelihood values for trees determined using different s/v ratios.

s/v ratio	log likelihood
0.5	-2643.71785
1	-2588.90199
2	-2569.84032 ← optimal
2.5	-2571.32310
3	-2574.57689
4	-2583.73780

The *gdh* sequences were examined for multiple substitutions at single sites and saturation for change by comparing the corrected and uncorrected distances estimated from each pair-wise sequence comparison. The different distance estimates were plotted against each other, and the resulting graphs are illustrated in Figures 63 - 66 (pp. 175 - 178). The slope of the p-distance curve in Figure 63 (p. 175), showed that the corrected and uncorrected distances were identical for pair-wise comparisons of isolates within the same lineage (distances < 0.05), and almost identical for comparisons involving isolates from different lineages (distances < 0.15). The p-distances between *G. intestinalis* and *G. ardeae* sequences were lower than the corrected distances, indicating that multiple substitutions at the same site had occurred. A similar trend was apparent for the plots of distance estimates using 1st and 2nd codon positions only, with no evidence for multiple hits for the *G. intestinalis* comparisons (slope = 1), but some evidence for multiple hits for the *G. intestinalis* - *G. ardeae* comparisons (Figure 64A, p. 176). Examination of distances calculated using 3rd codon positions only suggested that these sites rapidly approached saturation when the p-distances exceeded 0.2 substitutions/site (eg., distances between isolates from different lineages), and were saturated for change when the p-distances exceeded 0.8 substitutions/site (eg., distances between *G. intestinalis* and *G. ardeae* sequences; Figure 64B, p. 176). The distances estimated using only synonymous or nonsynonymous substitutions reflected the above result, with few multiple hits at nonsynonymous sites, but many at synonymous positions, especially for distances > 0.4 (Figure 65, p. 177). Comparison of TN and p-distances calculated using transitions or transversions found that neither class was saturated for change (Figure 66, p. 178). The numbers of transitions and transversions appeared to be equal for pair-wise comparisons of isolates from the same lineage (p-distance < 0.02 for each

Table 25. Nucleotide composition of partial *Giardia gdh* sequences
(for all sites and each individual codon base position).

Lineage/species	Isolate	All positions					1 st codon positions				
		A	T	C	G	Total	A	T	C	G	Total
Assemblage A	Portland 1	20.7	19.6	31.3	28.4	690	27	17	20.4	35.7	230
Assemblage A	Ad-1	20.7	19.6	31.3	28.4	690	27	17	20.4	35.7	230
Assemblage A	Ad-2	20.7	19.7	31.2	28.4	690	27	17	20.4	35.7	230
Livestock	Ad-165	22.3	21.1	28.9	27.7	654	28	17.4	20.2	34.4	218
Livestock	P-15	20.2	21.6	30.2	28	639	26.2	17.8	20.1	36	214
Cat	Ad-23	20.1	17.7	33.2	29	690	27.4	17.4	20	35.2	230
Cat	Ad-142	20.1	17.7	33.2	29	690	27.4	17.4	20	35.2	230
Dog1	Ad-136	21.2	19.9	32.2	26.8	690	27.8	16.5	20.4	35.2	230
Dog1	Ad-137	21	19.9	32.2	26.9	689	27.5	16.6	20.5	35.4	229
Dog1	Ad-141	21.2	19.9	32.2	26.7	689	27.9	16.6	20.5	34.9	229
Dog1	Ad-147	21	19.9	32.2	26.9	689	27.5	16.6	20.5	35.4	229
Dog2	Ad-148	22.8	23.3	28	25.9	690	27.8	16.5	21.3	34.3	230
Assemblage B	Ad-7	21	20.9	29.7	28.4	690	27.4	16.1	21.7	34.8	230
Assemblage B	Ad-28	21	20.9	29.7	28.4	690	27.4	16.1	21.7	34.8	230
Assemblage B	Ad-52	21	20.9	29.7	28.4	690	27.4	16.1	21.7	34.8	230
Assemblage B	LD-20	21.1	20.7	29.8	28.4	687	27.4	16.1	21.7	34.8	230
Assemblage B	WOOF	21	20.9	29.7	28.4	690	27.4	16.1	21.7	34.8	230
Assemblage B	D47	20	21.7	29.8	28.4	640	27.1	17.3	20.6	35	214
Assemblage B	CH105	20.9	20.6	30.2	28.2	678	27.4	16.4	22.1	34.1	226
Assemblage B	LSLP-116	20.2	21.9	30	27.9	639	26.6	18.7	20.6	34.1	214
Assemblage B	CP-117	20.3	21.2	30.1	28.4	631	27	17.1	20.9	35.1	211
Assemblage B	Ad-156	21.4	21.2	29.3	28.1	683	28.1	16.2	20.6	35.1	228
<i>G. ardeae</i>		19	18.6	32.7	29.8	689	23.9	17.4	21.7	37	230

		2 nd codon positions					3 rd codon positions				
		A	T	C	G	Total	A	T	C	G	Total
Assemblage A	Portland 1	32.2	29.1	17.4	21.3	230	3	12.6	56.1	28.3	230
Assemblage A	Ad-1	32.2	29.1	17.4	21.3	230	3	12.6	56.1	28.3	230
Assemblage A	Ad-2	32.2	29.1	17.4	21.3	230	3	13	55.7	28.3	230
Livestock	Ad-165	33.9	28.4	17	20.6	218	5	17.4	49.5	28	218
Livestock	P15	30	30	18.3	21.6	213	4.2	17	52.4	26.4	212
Cat	Ad-23	32.6	28.7	17.4	21.3	230	0.4	7	62.2	30.4	230
Cat	Ad-142	32.6	28.7	17.4	21.3	230	0.4	7	62.2	30.4	230
Dog1	Ad-136	32.2	29.1	17.8	20.9	230	3.5	13.9	58.3	24.3	230
Dog1	Ad-137	32.2	29.1	17.8	20.9	230	3.5	13.9	58.3	24.3	230
Dog1	Ad-141	32.2	29.1	17.8	20.9	230	3.5	13.9	58.3	24.3	230
Dog1	Ad-147	32.2	29.1	17.8	20.9	230	3.5	13.9	58.3	24.3	230
Dog2	Ad-148	32.2	29.1	17.4	21.3	230	8.3	24.3	45.2	22.2	230
Assemblage B	Ad-7	32.2	29.1	17	21.7	230	3.5	17.4	50.4	28.7	230
Assemblage B	Ad-28	32.2	29.1	17	21.7	230	3.5	17.4	50.4	28.7	230
Assemblage B	Ad-52	32.2	29.1	17	21.7	230	3.5	17.4	50.4	28.7	230
Assemblage B	LD-20	32.2	29.1	17	21.7	230	3.5	16.7	51.1	28.6	227
Assemblage B	WOOF	32.2	29.1	17	21.7	230	3.5	17.4	50.4	28.7	230
Assemblage B	D47	30	29.6	17.8	22.5	213	2.8	18.3	51.2	27.7	213
Assemblage B	CH105	31.9	29.2	17.3	21.7	226	3.5	16.4	51.3	28.8	226
Assemblage B	LSLP-116	30.5	29.6	17.8	22.1	213	3.3	17.5	51.9	27.4	212
Assemblage B	CP-117	30.3	29.4	18	22.3	211	3.3	17.2	51.7	27.8	209
Assemblage B	Ad-156	32.5	29.4	17.1	21.1	228	3.5	18.1	50.2	28.2	227
<i>G. ardeae</i>		31.3	27.8	18.7	22.2	230	1.7	10.5	57.6	30.1	229

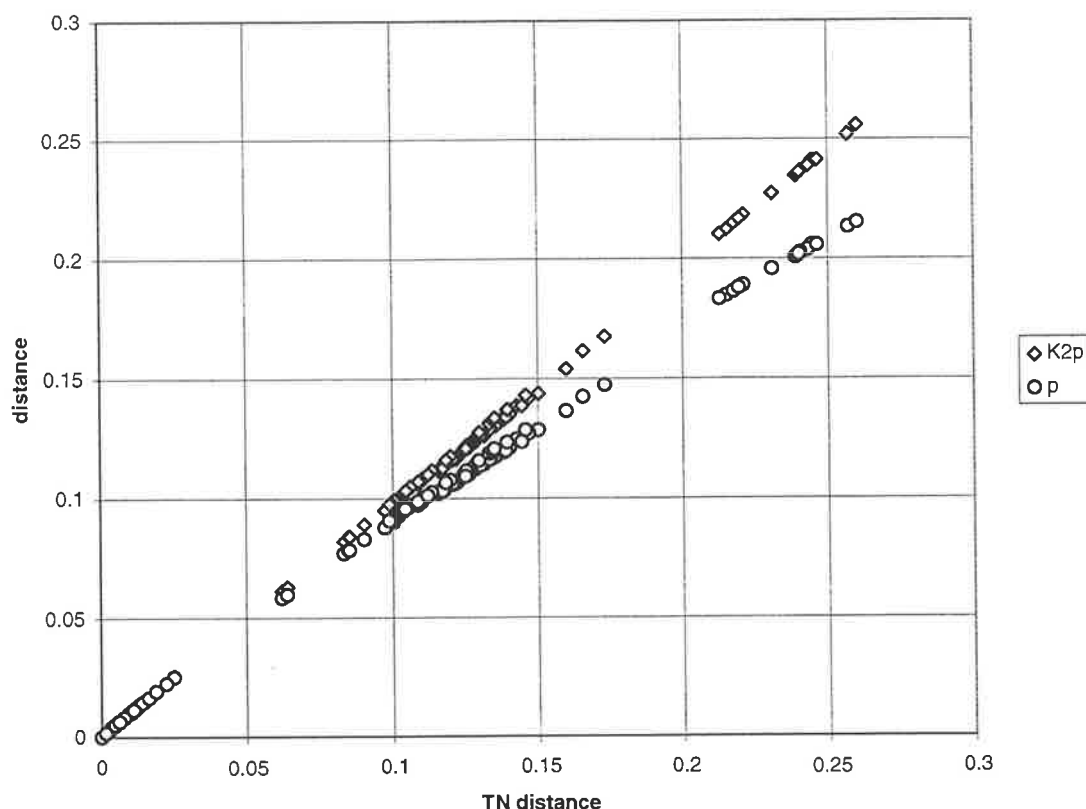
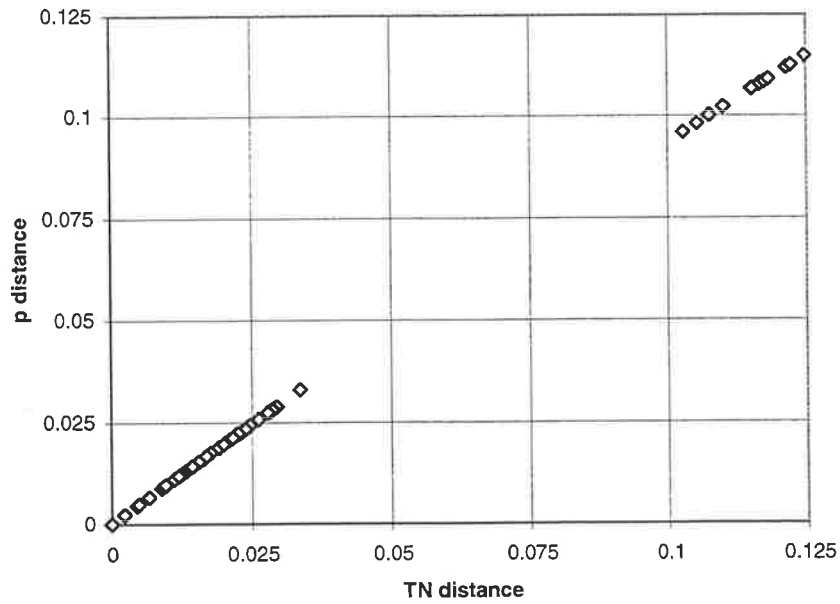


Figure 63. Dot plots of *gdh* TN distances versus K2p and p-distances calculated for all pairwise sequence comparisons using all classes of nt substitutions.

class), suggesting an s/v ratio of 0.5 (compared to the ML estimate of 2 calculated using the entire data set) for these sequence comparisons.

Among-site rate variation was examined as described previously in section 4.4.2.7. The α parameter estimate using all sites was 0.2, indicating extreme rate variation. This result would be expected for a protein coding gene. Each codon base position was then analysed separately. The 1st codon positions had an estimated α of 0.45, indicating substantial rate variation. When constant sites were treated as invariable (a program option in PUZZLE), $\alpha = 2.46$, which suggested that the apparent rate heterogeneity was due to a combination of low mutation rate and relatively recent divergence times, and not caused by a few sites with a high substitution rate (ie among-site rate variation). The 2nd codon positions had an estimated α of 1, which changed to $\alpha = 97$ when constant sites were treated as invariant. This suggested that most of the variant sites had similar substitution rates (the gamma distribution becomes bell-shaped for large α values), and that the lower α value was caused by the similarity between the sequences analysed (the presence of many conserved sites can bias the estimate of α).

A



B

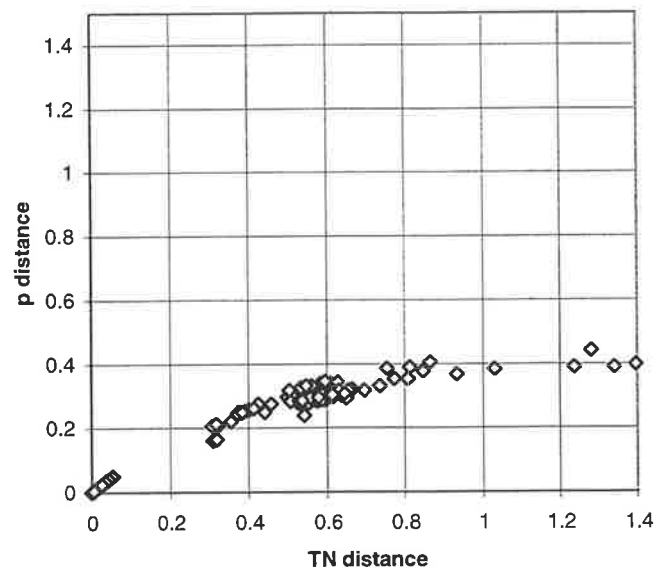


Figure 64. Dot plots of *gdh* p distances versus TN distances calculated for all pair-wise sequence comparisons using **A:** 1st and 2nd codon positions only; **B:** 3rd codon position only.

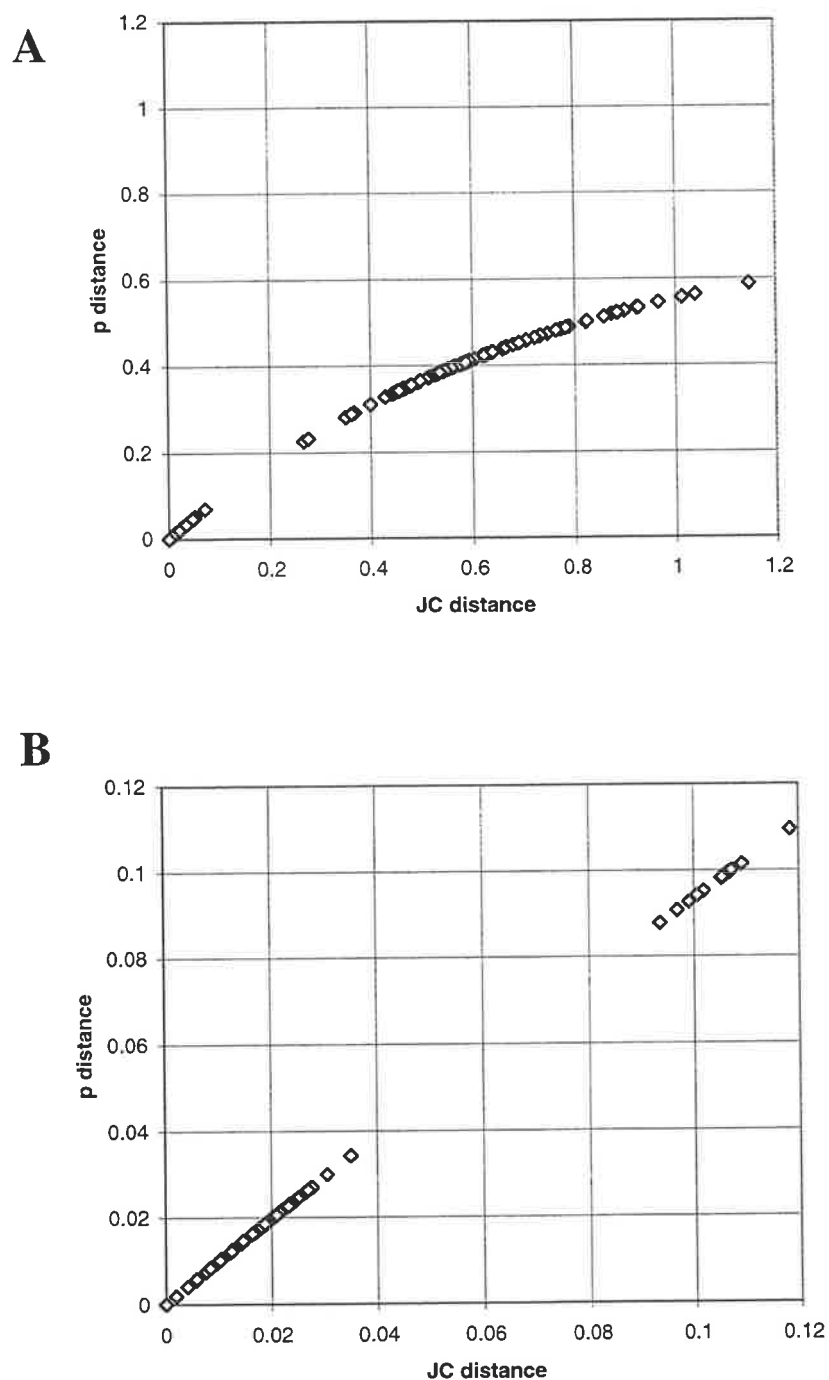


Figure 65. Dot plots of *gdh* p distances versus JC distances calculated for all pair-wise sequence comparisons using **A:** synonymous substitutions only; **B:** nonsynonymous substitutions only.

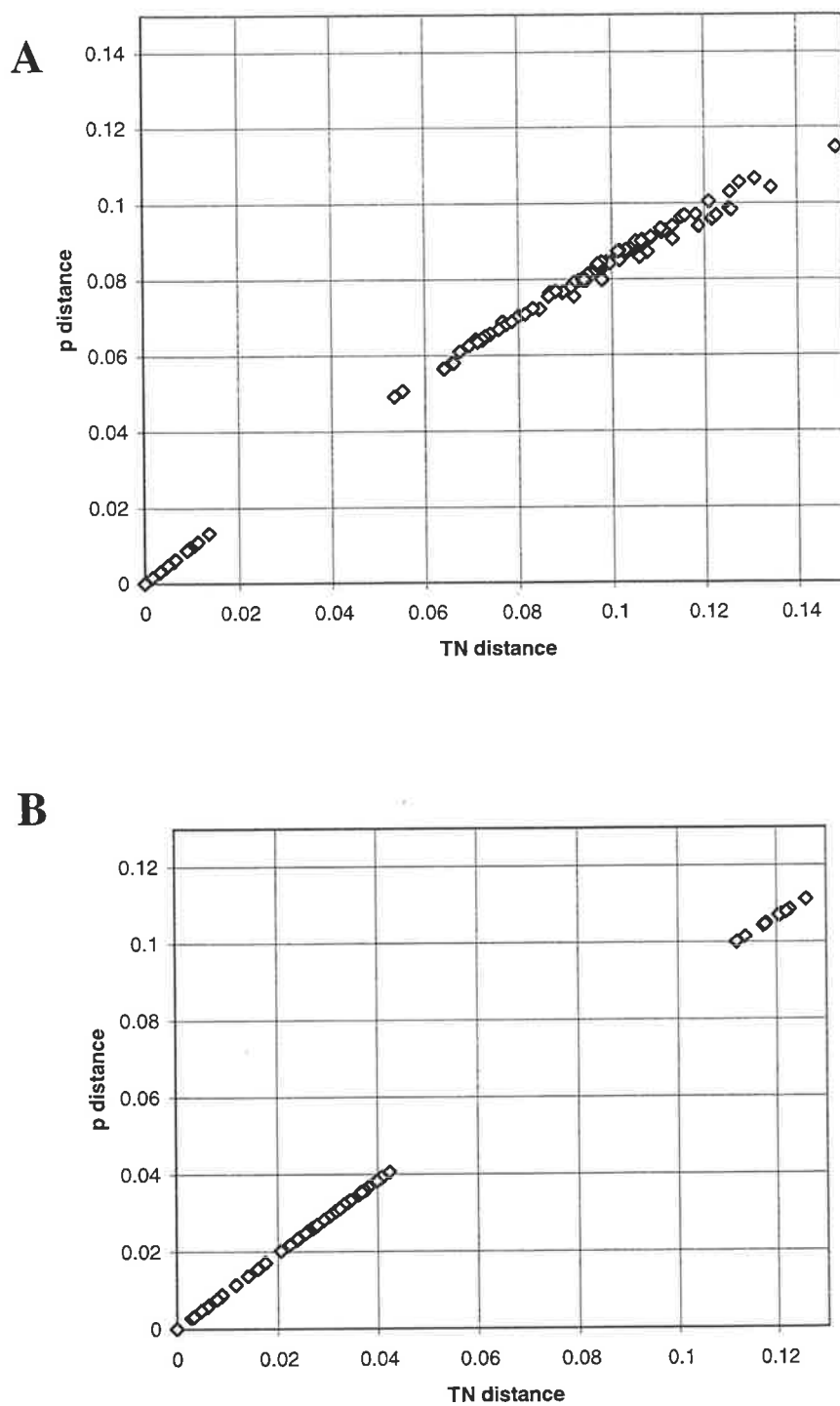


Figure 66. Dot plots of *gdh* p distances versus TN distances calculated for all pair-wise sequence comparisons using **A**: transitions only; **B**: transversions only.

Analysis of 3rd codon positions found no evidence for significant rate heterogeneity, with $\alpha = 1.22$. The results suggest that the low α value when all positions were used was caused by a combination of the conservation at the 2nd codon position and the difference in the rates at the three codon base positions.

4.5.4.8 Distance-based phylogenetic analysis of *gdh* nucleotide sequence data

Two separate data sets were analysed using distance methods. The first comprised all of the isolates sequenced (Appendix C), while the second was a reduced data set with only 1 or 2 representatives of each lineage (indicated in the relevant figures). This reduction was made for two reasons, (1) full length sequences could not be obtained for all isolates, and (2) practicality, because the speed and accuracy of some methods (eg. ML or parsimony) can be influenced detrimentally by the number of taxa analysed. The full data set was analysed using NJ with TN distances, resulting in the phylogram in Figure 67 (p. 180). There was complete agreement between the composition of the lineages identified by this analysis and those determined previously by AE. However, the topology of this tree was different. The grouping of the Livestock, Cat and Assemblage A lineages received good bootstrap support (85%), but the branching order of these lineages received lower support (74%). In contrast to the previous analyses (Figure 57, p. 159), Assemblage B was clustered with the Dog1 and Dog2 lineages, although the support for this was not very strong (71%). With the exception of Assemblage A, which received support for the division between genetic groups I and II, none of the intra-lineage relationships could be resolved. Although there was variation within these other lineages, it was insufficient to allow resolution of their phylogenies. The longer branches within the livestock lineage may be an artefact of the distance estimate, due to the missing sequence data for these isolates (see Appendix C), although they may also correspond to the divisions observed within the Livestock lineage by allozyme analysis (Figure 23, p. 101). Virtually all of the informative nt substitutions were synonymous substitutions, as analysis using distances estimated from these sites produced a tree (not shown) with almost identical topology to the one inferred using all sites (Figure 67. p. 180). Use of nonsynonymous substitutions for the distance estimate resulted in the loss of virtually all resolution within the tree, producing a polytomy with only Assemblage A, Cat and Assemblage B identifiable as discrete lineages (not shown). A similar result was also obtained using Poisson corrected amino acid distances (not shown).

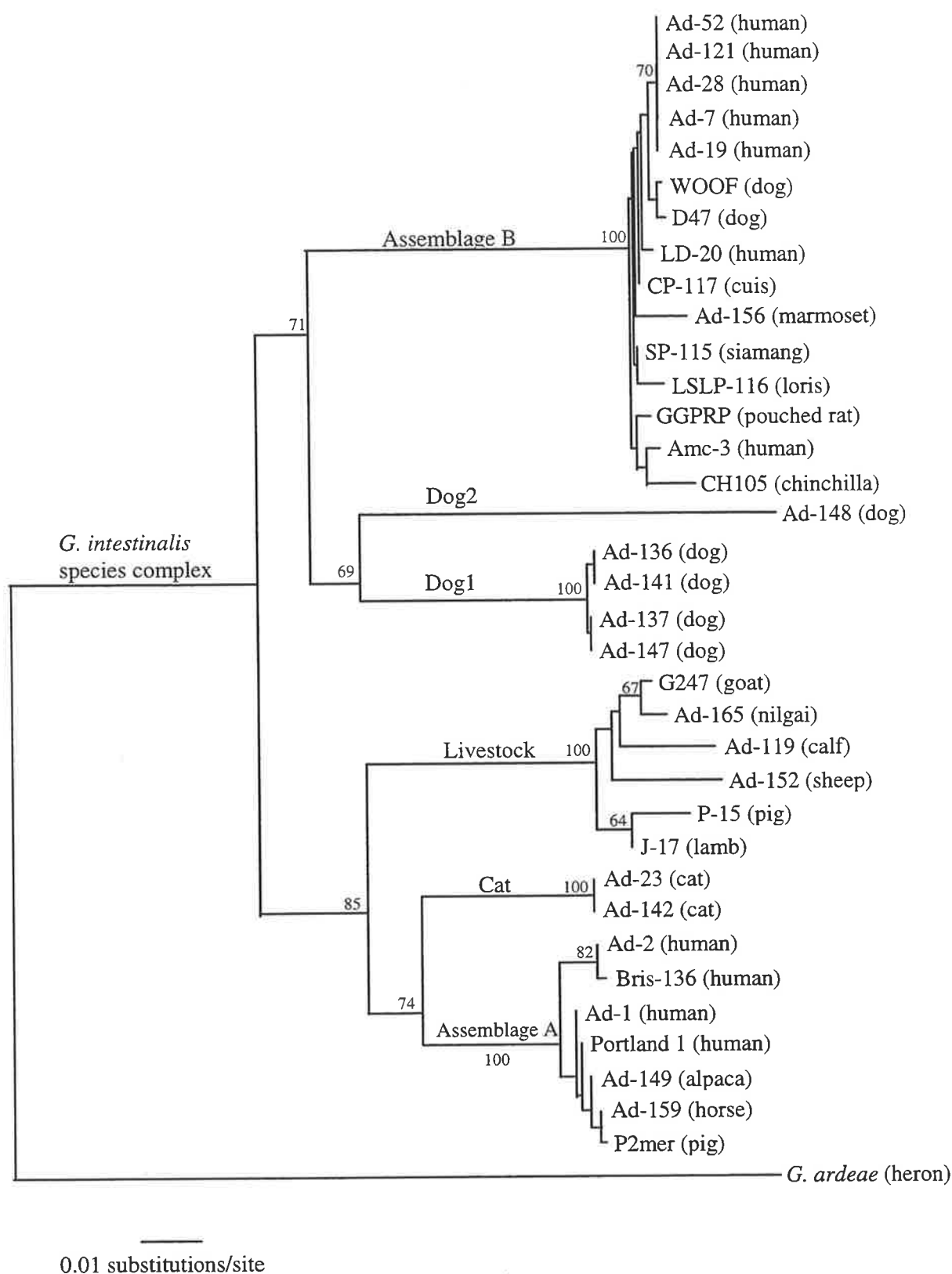


Figure 67. Phylogenetic relationships of *Giardia* isolates inferred from *gdh* nt sequences using TN distance estimates and the NJ tree building method. Host origin of each isolate is indicated in parentheses. Percentage bootstrap support (>50% from 500 replicates) is indicated above the supported branches.

Analysis of the restricted data set produced the same overall topology as the larger data set, for both ME and NJ tree building methods (Figure 68, p. 182). Both methods found good support ($\geq 90\%$) for the grouping of the Dog1, Dog2 and Assemblage B lineages, as well as the Assemblage A, Cat and Livestock lineages. Bootstrapping using Poisson corrected amino acid distance estimates and ME found no support for the grouping of Dog1, Dog2, and Assemblage B and good support (83%) for the grouping of Assemblage A, Cat, and Livestock (Figure 69, p. 183).

4.5.4.9 Maximum likelihood phylogenetic analysis of *gdh* nucleotide sequences

Maximum likelihood analysis fully supported the tree topologies determined by the distance methods, although it did not indicate the monophyly of the *G. intestinalis* species complex (Figure 70, p. 184). There was strong support for two main groups, the first comprising the Assemblage A, Cat and Livestock lineages, and the second comprising the Dog1, Dog2 and Assemblage B lineages. The lengths of the branches separating the lineages within the second group were longer than those in the first. Gamma distances were not used because the standard model could not properly account for the distribution of substitution rates within the data set. The phylogeny inferred by ML analysis of the deduced amino acid sequences (Figure 71, p. 185) was less resolved than the one inferred from nt sequences, with the only consistency being the support for the monophyly of the Assemblage A, Cat and Livestock lineages. In contrast to the results of the other analyses, the distance separating the Ad-156 and Ad-28 isolates within Assemblage B was similar to that between the Assemblage A and Cat lineages.

4.5.4.10 Cladistic analysis of *gdh* nucleotide sequences

Parsimony analysis was attempted on the full data set using the heuristic search option of PAUP. However, due to the very large numbers of trees ($> 3,800$) with an equal length of 305 steps which were found, the search was terminated due to insufficient memory. The consensus tree from the incomplete search is given in Figure 72 (p. 186). The topology of this tree is similar to those determined by distance methods, identifying two major clades,- one comprising the Livestock, Cat and Assemblage A lineages and the other comprising the Assemblage B, Dog1 and Dog2 lineages. The branching order of the Assemblage A, Livestock and Cat lineages was more consistent with the 18S rRNA and EF1 α analyses than

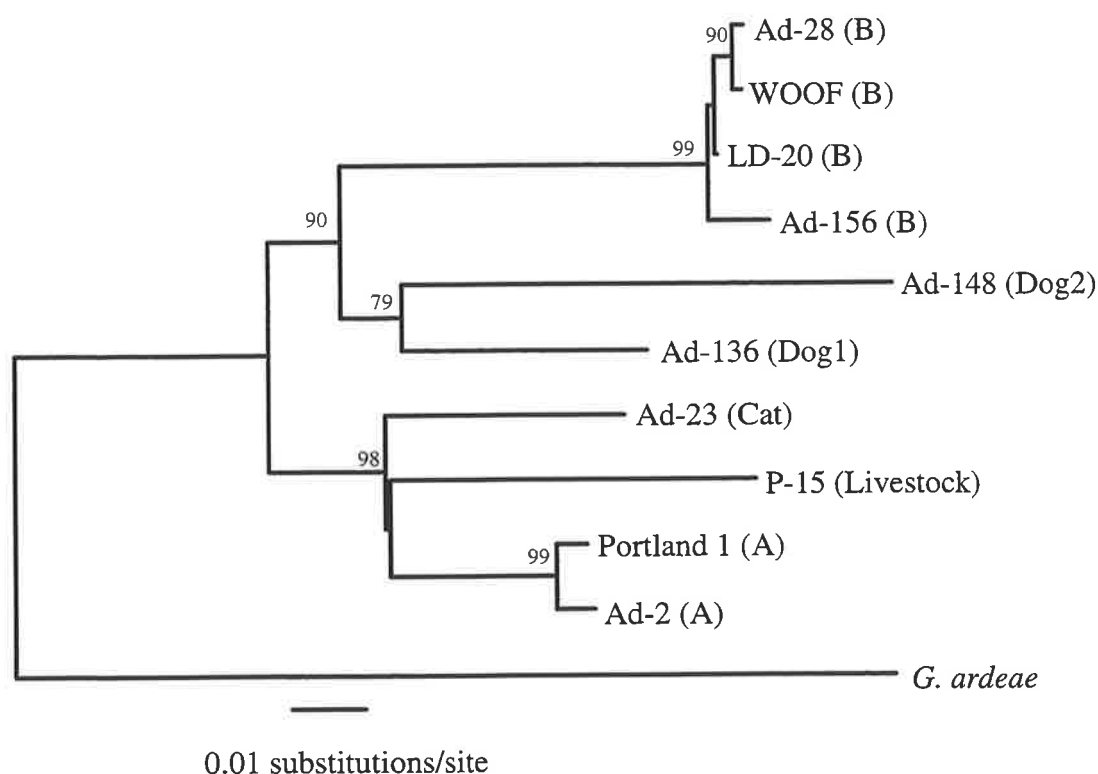
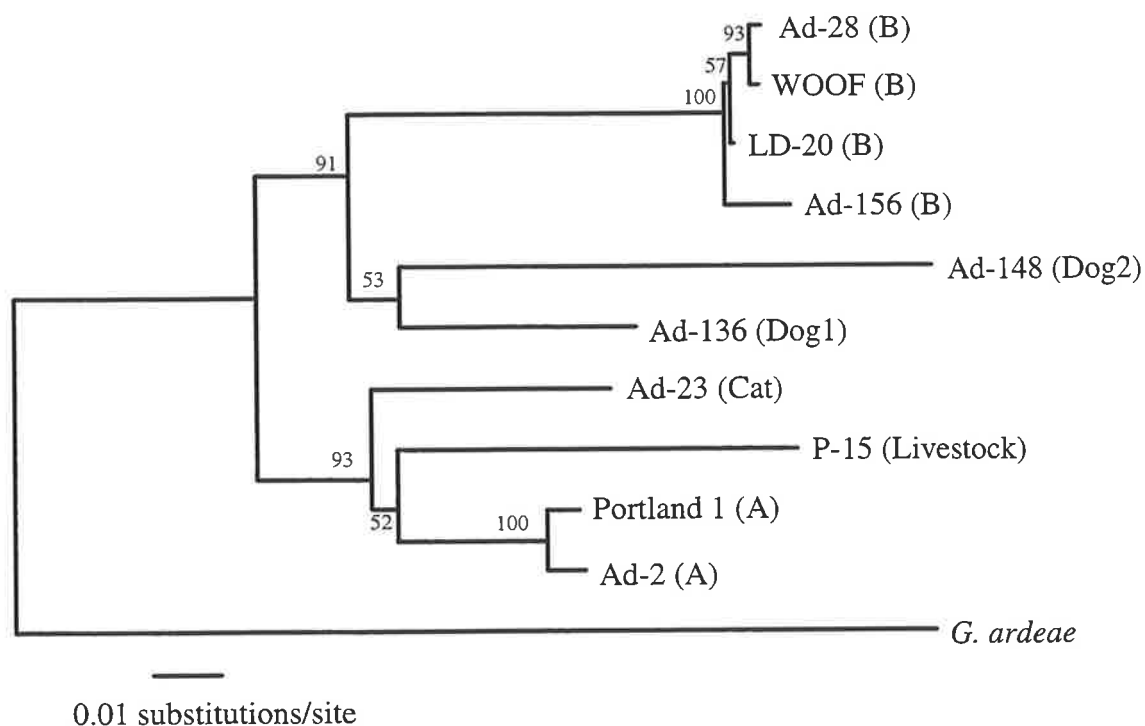


Figure 68. Phylogenetic relationships of *Giardia* isolates inferred from the *gdh* nt sequences using TN or K2p distances and the NJ (top) or ME (bottom) tree building methods. The allozyme defined lineage of an isolate is given in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (500 or 1,000 replicates) is indicated above the branches.

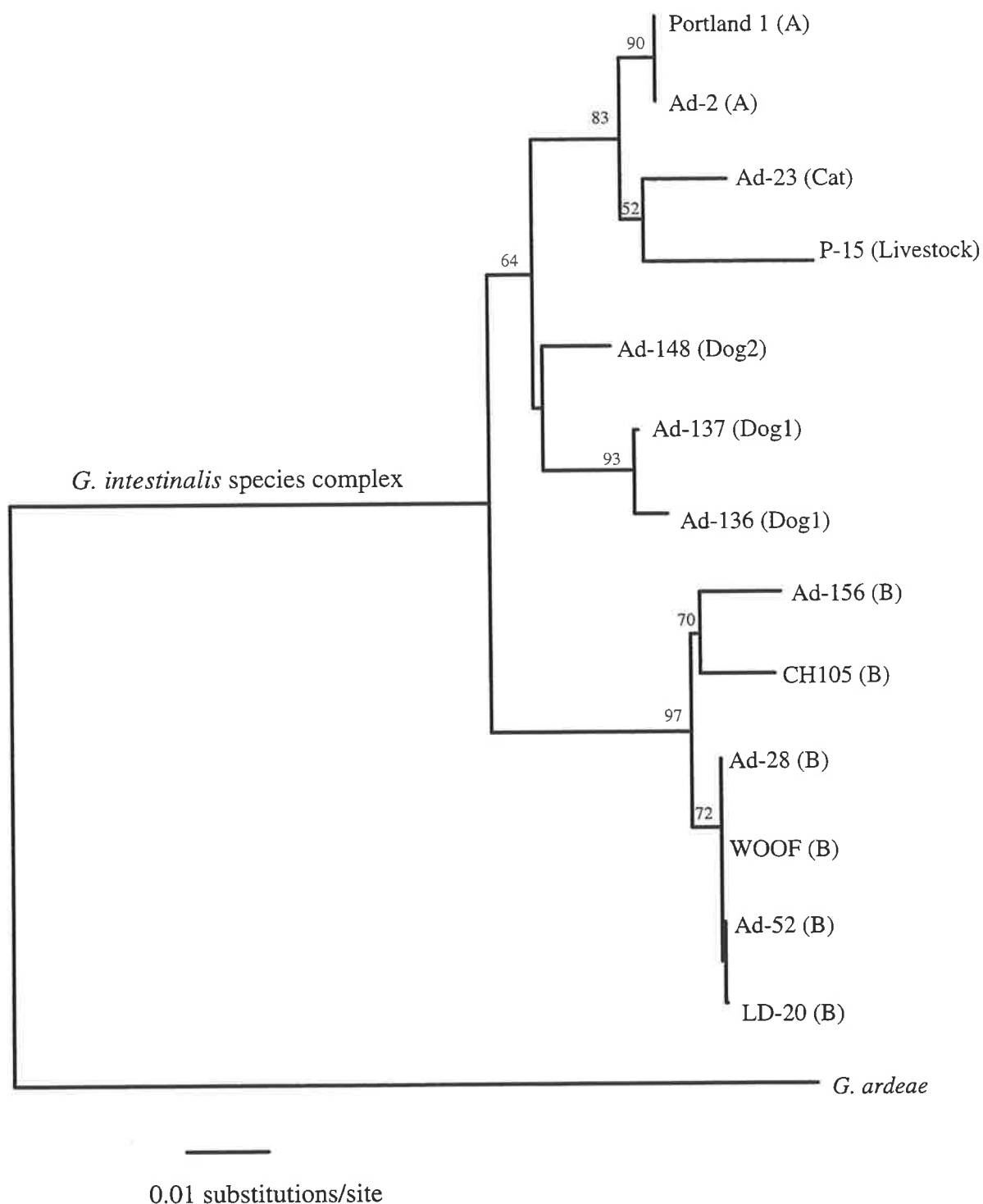


Figure 69. Phylogenetic relationships of *Giardia* isolates inferred from the *gdh* nt sequences using Poisson corrected amino acid distances and ME. The allozyme defined lineage of an isolate is given in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (1,000 replicates) is indicated above the branches. Analysis using NJ gave virtually identical results and is not shown.

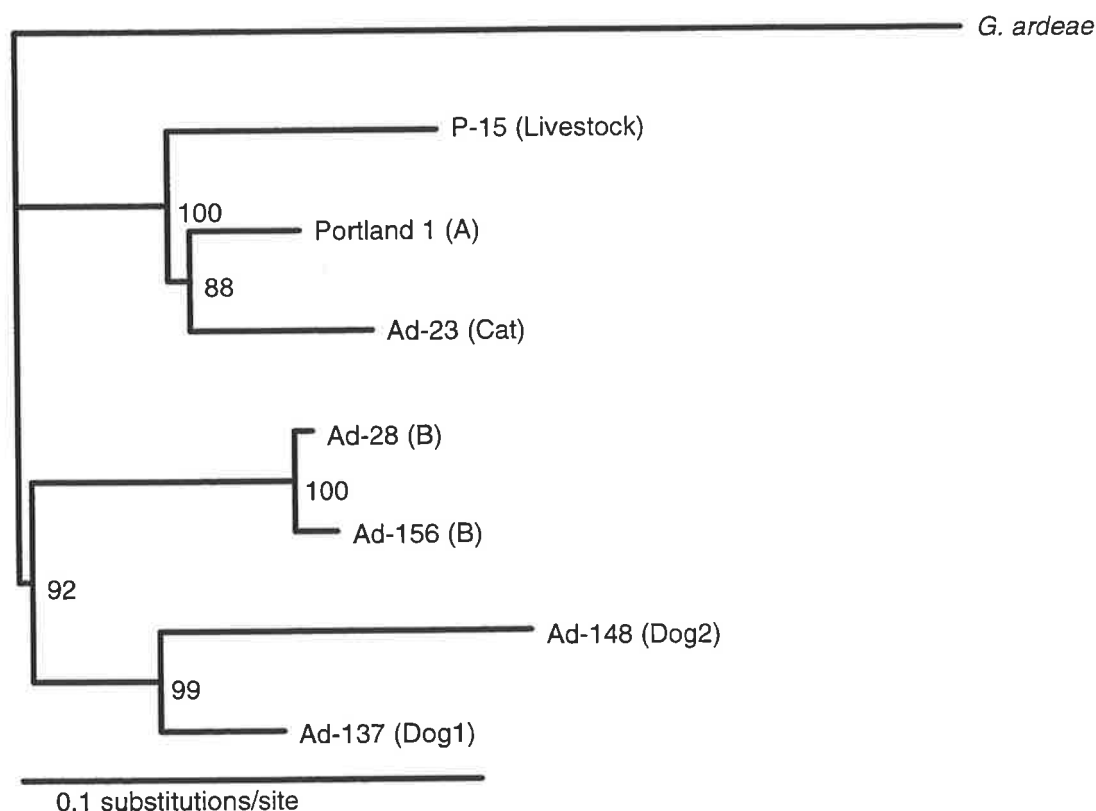


Figure 70. Phylogenetic relationships of *Giardia* isolates inferred from the *gdh* nt sequence sequences using ML and quartet puzzling. The allozyme defined lineage of an isolate is given in parentheses (A = Assemblage A, B = Assemblage B). Percentage support (1,000 quartets) is indicated at the supported node.

the *gdh* distance analyses. Analysis of a restricted *gdh* data set using the branch and bound searching algorithms of PAUP or DNAPENNY produced trees with identical topologies (PAUP tree shown, Figure 73, p. 187). These supported the topology of the consensus tree from the incomplete heuristic search (Figure 72, p. 186). Bootstrap analysis found good support (83%) for the monophyly of the Assemblage A, Cat and Livestock lineages, but weaker (67%) support for the clade containing the Assemblage B, Dog1 and Dog2 lineages. Parsimony analysis using deduced amino acid sequences produced an unresolved polytomy comprising all of the lineages (not shown).

The amount of phylogenetic signal in the data set was also investigated. Because of the number of taxa involved, the data was further reduced to 1 or 2 representatives for each lineage. An exhaustive search was conducted, which examined 2,027,025 trees and found a single most parsimonious tree of 243 steps in length. The tree length distributions are given in Figure 74 (p. 188).

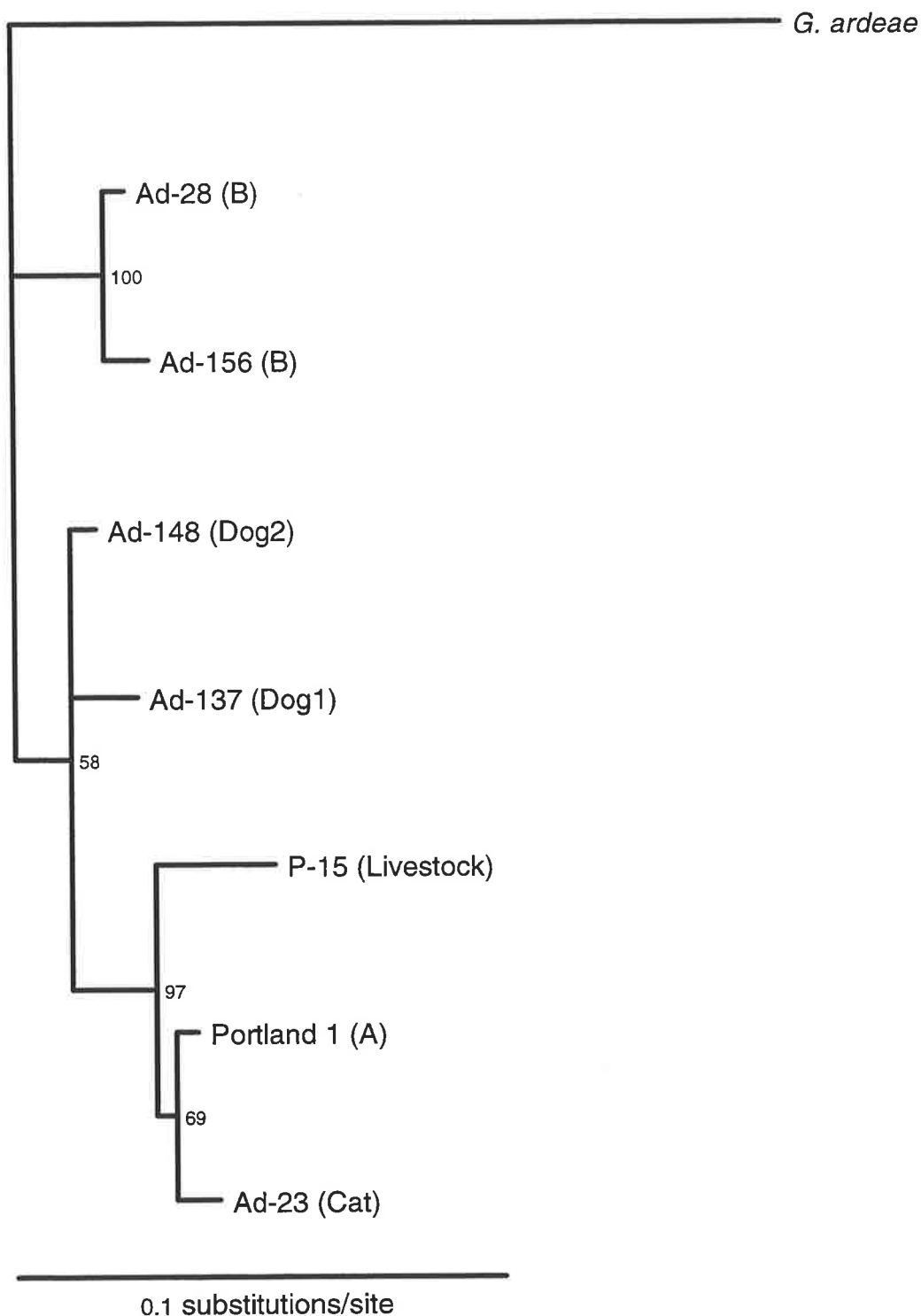


Figure 71. Phylogenetic relationships of *Giardia* isolates inferred from the *gdh* deduced amino acid sequences using ML and quartet puzzling. The allozyme defined lineage of an isolate is given in parentheses (A = Assemblage A, B = Assemblage B). Percentage support (1,000 quartets) is indicated at the supported node.

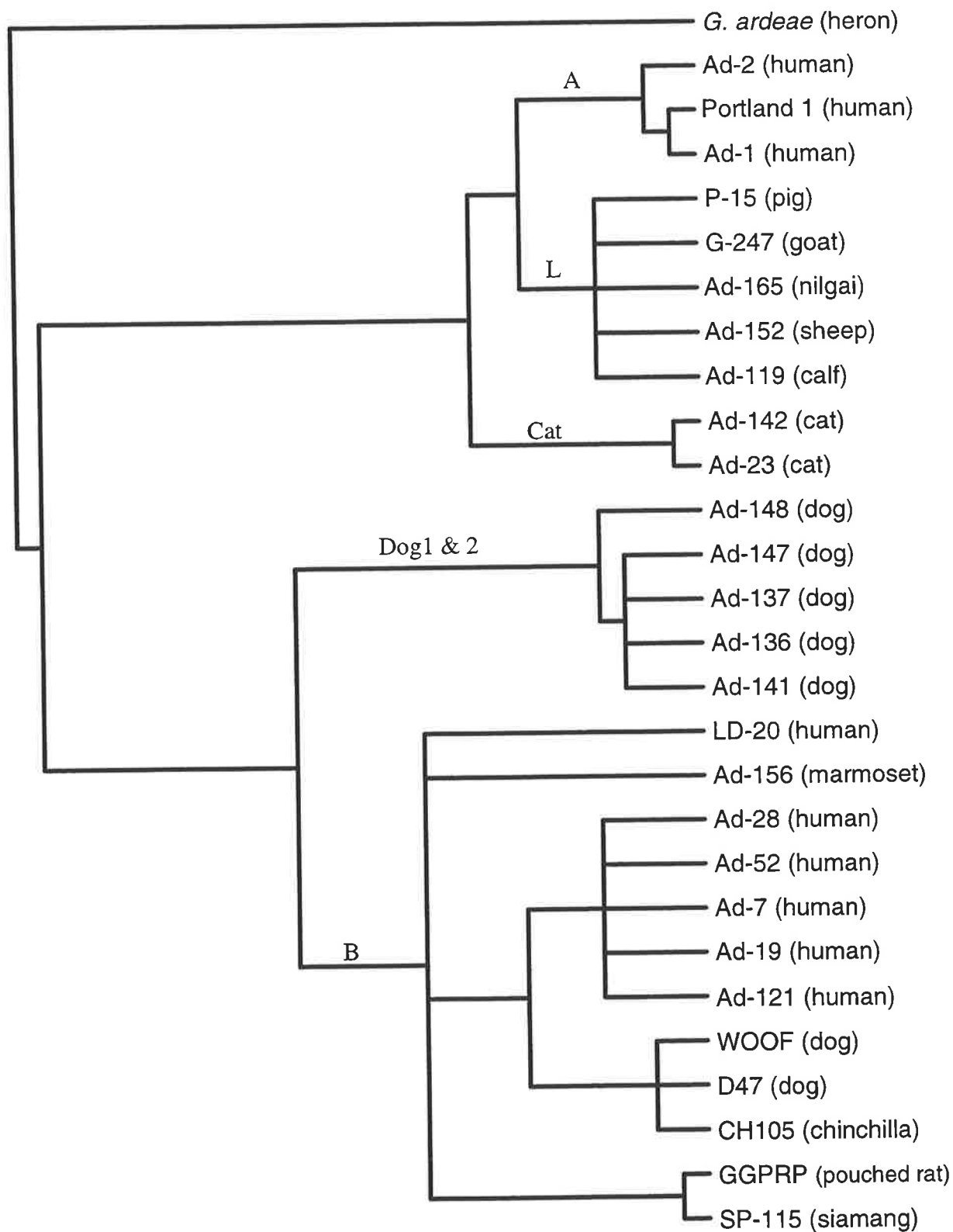


Figure 72. Strict consensus tree of the incomplete heuristic search analysis of the *gdh* nt sequences. Host origin of isolates is in parentheses (A = Assemblage A, B = Assemblage B, L = Livestock).

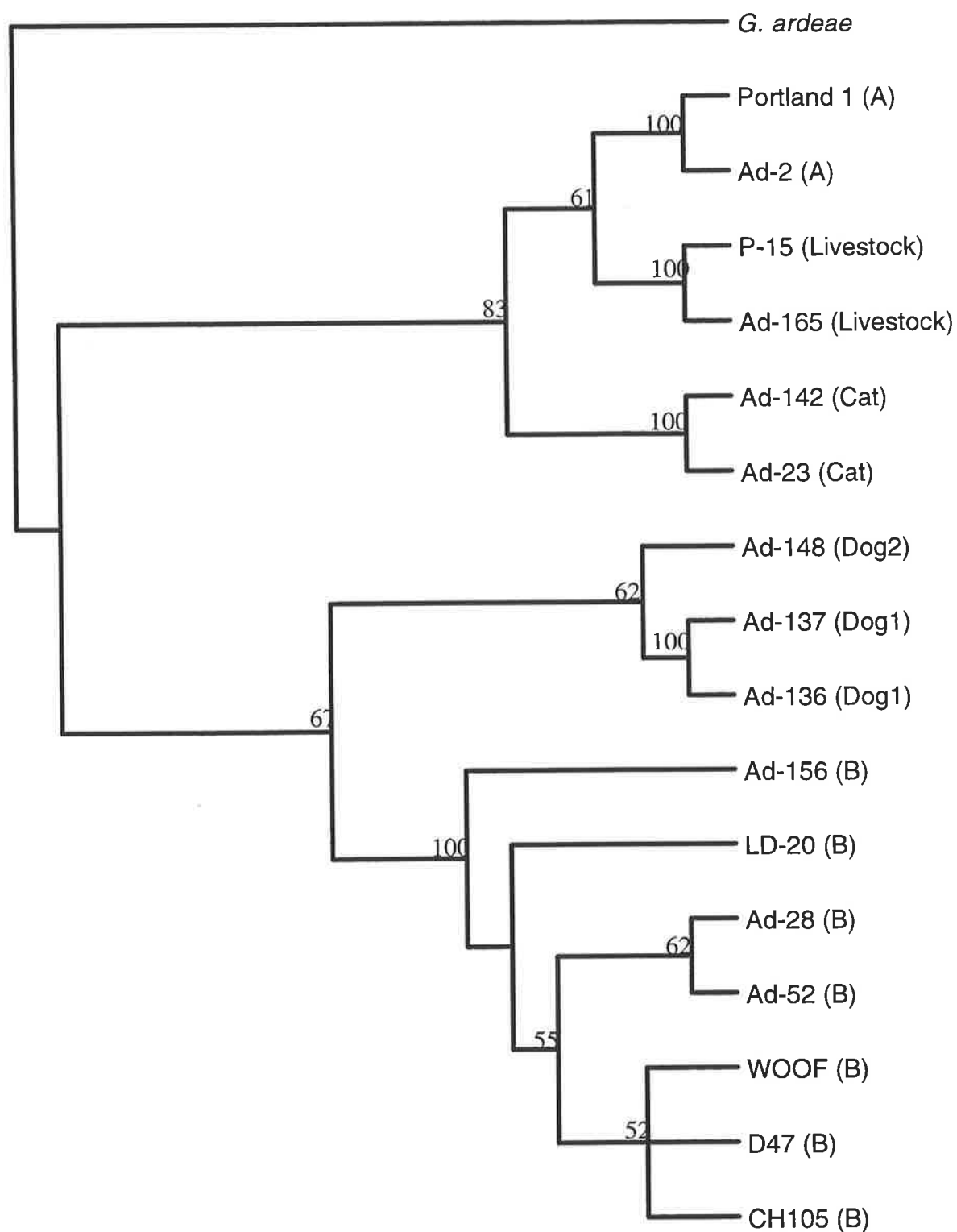


Figure 73. Strict consensus tree resulting from cladistic analysis of the *gdh* nt sequences using the branch and bound search algorithm. Two equally most parsimonious trees were found. The allozyme defined lineage of an isolate is given in parentheses (A = Assemblage A, B = Assemblage B). Bootstrap values (% of 500) are indicated above the branches.

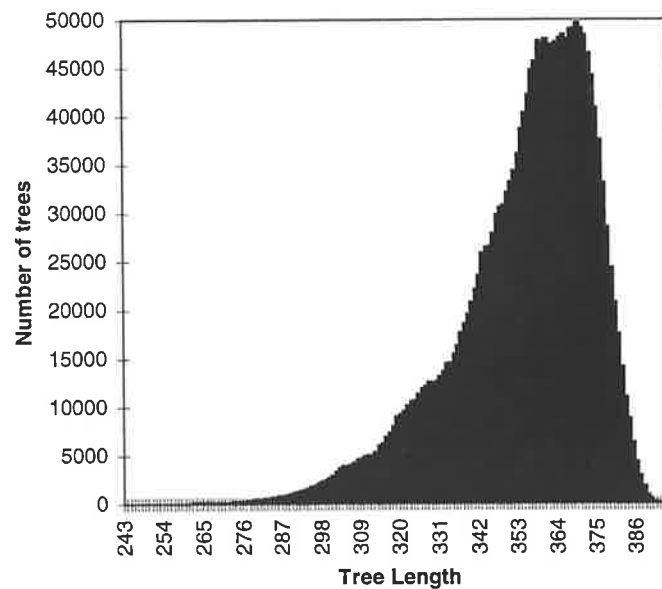


Figure 74. Tree length distributions of the trees examined during the exhaustive search analysis of the *gdh* DNA sequence data. Taxa used in the analysis were: Portland 1, Ad-2, P-15, Ad-16, Ad-23, Ad-148, Ad-147, Ad-28, Ad-156, and *G. ardeae*.

The average tree length was 354.62 ± 19.56 steps, and the g_1 statistic was -1.02, indicating significant skewness and hence phylogenetic signal. Removal of the outgroup did not alter the presence of phylogenetic signal.

4.5.5 Phylogenetic analysis of *tpi* nucleotide sequences

4.5.5.1 PCR amplification of *tpi*

The primers originally used to amplify DNA from the *tpi* locus were provided by Dr. Michael Mowatt (described in section 4.5.3.4). These were used in various combinations to amplify different segments of *tpi*, although they did not work with all DNA samples. New primers were constructed for conserved regions identified using the alignment of the two published *Giardia* sequences (Mowatt *et al.*, 1994). These primers (TPI16F & TPI533R) amplified DNA from *tpi* from all of the *G. intestinalis* isolates except Ad-148 (Dog2), but they did work using DNA from *G. muris* or *G. ardeae*. Degenerate primers (TPIGENF & TPIGENR) were therefore designed using the nt sequences of the TPI16F/TPI533R

amplification products derived from isolates comprising the different *G. intestinalis* lineages. These primers were used to amplify *tpi* from *G. muris* DNA. A different reverse primer (TPI572R), in combination with TPIGENF, was required to amplify *tpi* from *G. ardeae* DNA. Ad-148 DNA could not be made to amplify the targeted *tpi* segment using any combination of the available primers or by modifying the reaction conditions.

4.5.5.2 Alignment of *tpi* nucleotide sequences

The *G. intestinalis tpi* nt sequences could be readily aligned by eye, although ClustalW was used to confirm the alignment with the outgroup sequences. A single 3 nt indel was required to maintain the alignment at the 3' end (Figure 75, pp. 190 - 192). Direct comparison of the sequences showed that the isolates comprising different lineages were between 77 - 91% similar. There was almost no variation between sequences of isolates comprising the same lineage (> 99.5% similarity). This situation was similar to that observed for the *gdh* data (> 99% nt sequence similarity of isolates comprising the same lineage). The *G. muris* and *G. ardeae* sequences were only 62 - 67% similar to the *G. intestinalis* sequences, which makes their use as outgroups suspect. These large differences were quite remarkable, considering that the *G. ardeae* sequence is 58% similar to the *tpi* nt sequence of *Drosophila melanogaster* (result of an NCBI BLAST search of the GenBank database, not shown). This might suggest that the sequences originated from contaminating DNA in the *Giardia* samples. However, these sequences were most similar to the *G. intestinalis tpi* sequences, and had lower similarity (43 - 58%) with any of the published *tpi* nt sequences from plants, bacteria or animals (data not shown). An unexpected finding was the high (83%) similarity between the *G. muris* and *G. ardeae* nt sequences, which was larger than the similarities between some of the isolates of *G. intestinalis* from the different lineages.

4.5.5.3 Examination of *tpi* sequence parameters

The nt composition of the *tpi* sequences (Figure 75, pp. 190 - 192) was determined (Table 26, p. 193). Mowatt *et al.* (1994) had already reported the 3rd codon position GC bias in *tpi* codons, and it was of interest to see how the other *Giardia* species compared. The overall GC content of the *G. intestinalis* sequences ranged from 49 - 57%, which was a larger range than observed for the *eflα* (Table 21, p. 141) or *gdh* (Table 25, p. 174) nt sequences. The GC content of the *G. ardeae* and *G. muris tpi* sequences were 65% and 62% respectively. The GC content at the 1st codon position was 55 - 57% for the *G. intestinalis* sequences, and

Figure 75. Alignment of *Giardia*-derived *tpi* nucleotide sequences.

JH ¹	TTC	AAG	TGT	AAC	GGC	TCT	CTT	GAC	TTT	ATC	AAG	AGC	CAC	GTG	GCG	GCA	ATT	GCT	GCC	CAT	AAG	ATC	CCT	GAT
WB ²	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23	N.N	...	..C	...	...	..G	..C	...	...	...	...	..T	...	...	..N	..C	...	...	...	..C	...	...	...	...
Ad-142	N.T	...	..C	...	...	..G	..C	...	...	...	...	..T	...	...	..G	..C	...	...	...	..C	...	...	...	...
P-15	N.T	...	...	...	...	..N	...	..T	...	...	...	...	...	...	...	..C	...	...	...	..C	...	...	..C	...
Ad-7 ³	N..	..A	..C	..T	..A	..G	..C	...	..C	..T	...	...	...	..A	...	T.C	..C	..C	T..	T..	...	...	..C	..G
GSM ⁴	...	..A	..C	..T	..A	..G	..C	...	..C	..T	...	...	...	..A	...	T.C	..C	..C	T..	T..	...	...	..C	..G
Ad-147 ⁵	N.T	...	..CC	...	..G	..G	...	...	...	..A	...	..T	..A	...	..C	..C	..G	T..	..C	...	..T	..C	..C	...
<i>G. ardeae</i>	...	...	..C	..T	...	A.C	G.G	...	..C	...	...	..C	...	..C	..C	..G	...	...	GG	..C	..A	...	...	..G
<i>G. muris</i>	NNN	NNN	NNN	NNN	NNN	NNN	NNN	...	..GC	...	...	..C	...	..C	...	..G	..C	..G	..G	..C	...	...	..C	..G

72

JH ¹	TCC	GTG	GAC	GTC	GTC	ATT	GCC	CCC	TCC	GCC	GTA	CAC	CTG	TCA	ACA	GCC	ATT	GCG	GCA	AAC	ACG	TCA	AAA	CAG
WB ²	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23	...	...	...	..T	..T	G.C	...	..T	..T	...	..G	...	...	...	...	...	...	...	...	...	...	...	..G	...
Ad-142	...	...	...	..T	..T	G.C	...	..T	..T	...	..G	...	...	...	...	...	...	...	...	...	...	...	..G	...
P-15	...	A.A	...	..T	..T	G..	...	..T	..T	...	..T	T.A	...	...	..T	...	..A	...	...	...	...	...	...	...
Ad-7 ³	...	...	...	..T	..T	G..	..T	...	...	TTT	..G	...	..T	..T	...	..T	...	..G	..T	..T	..G	..G	TGT	...
GSM ⁴	...	...	...	..T	..T	G..	..T	...	...	TTT	..G	...	..T	..T	...	..T	...	..G	..T	..T	..G	..G	TGT	...
Ad-147 ⁵	..T	..T	..T	..G	A..	..C	...	...	..G	T..	..G	..T	...	..T	..G	...	..C	..A	..G	...	..A	..G	..G	...
<i>G. ardeae</i>	AA.	A.C	..T	...	...	G.C	...	...	G.G	ATG	CCG	...	..C	GTC	CTC	...	CAG	..C	..G	...	...	AGC	..G	GC.
<i>G. muris</i>	AG.	A.C	...	...	..C	G.C	...	...	G..	ATG	CCT	...	..C	CTC	CTC	...	CAG	..A	..TC	...	...	AAC	..G	GCA

144

JH ¹	TTG	AGG	ATA	GCA	GCG	CAG	AAT	GTG	TAC	CTA	GAG	GGG	AAC	GGG	GCG	TGG	ACT	GGC	GAG	ACA	AGT	GTT	GAG	ATG
WB ²	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23	...	..A	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	..C	..T	...	..G	..C	...	...	...
Ad-142	...	..A	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	..C	..T	...	..G	..C	...	...	...
P-15	...	..A	...	..G	...	...	...	...	...	...	...	..T	..A	...	...	..T	...	..G	...	...	...	...	...	...
Ad-7 ³	C..	..AA	...	...	..A	...	..C	...	..T	..G	..A	...	...	..T	..A	...	..C	...	...	...	..C	..C	...	...
GSM ⁴	C..	..AA	...	...	..A	...	..C	...	..T	..G	..A	...	...	..T	..A	...	..C	...	...	...	..C	..C	...	...
Ad-147 ⁵	C..	..A	...	...	...	...	...	...	..C	...	..A	...	..C	..A	...	..G	...	...	...	...	...	...	...	...
<i>G. ardeae</i>	C..	..A	..C	..G	T.C	...	..C	..C	..T	..C	C..	CCC	CC.	..T	..T	...	..C	...	..A	GTG	..C	..G	...	...
<i>G. muris</i>	C..C	..A	..C	T.G	T.T	...	..C	...	...	..C	C..	CC.	CCT	..C	..C	...	..C	..T	...	GTG	..C	..C	...	...

216

Figure 75 (cont). Alignment of *Giardia*-derived *tpi* nucleotide sequences.

JH ¹	CTT	CAG	GAC	ATG	GGT	TTG	AAG	CAT	GTG	ATA	GTA	GGG	CAC	TCT	GAA	AGA	CGC	AGA	ATC	ATG	GGG	GAG	ACC	GAC
WB ²	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23	..C	...	...	...	..C	C.A	G..	...	...	...	A..	...	...	...	...	..G	...	...	...	...	...	...	...	...
Ad-142	..C	...	...	...	..C	C.A	G..	...	...	...	A..	...	...	...	...	..G	...	...	...	...	...	...	...	...
P-15	...	...	...	...	..C	...	G..	T.C	...	...	A..	...	..T	...	...	..G	..T	...	...	...	...	...	...	...
Ad-7 ³	..G	.T.	...	...	..G	C..	.GC	...	..A	...	A..	..A	...	...	...	...	..T	...	...	...	..C	...	...	A.T
GSM ⁴	..G	.T.	...	...	..G	C..	.GC	...	..A	...	A..	..A	...	...	...	...	..T	...	...	...	..C	...	...	A.T
Ad-147 ⁵	...	...	...	...	..C	C..	.GT	..C	...	...	...	...	...	...	...	...	..T	..G	...	...	..C	...	...	A..
<i>G. ardeae</i>	..C	A..	...	C.C	..G	G.C	G..	GTG	ACC	C.C	..C	...	...	..C	..G	C.C	...	C.C	...	...	..C	...	..G	...
<i>G. muris</i>	..G	A..	...	C.C	..G	A.T	G..	ATG	ACC	C.T	...	..A	..T	..C	..G	C.C	..T	C.C	...	...	..C	...	...	...

288

JH ¹	GAG	CAA	AGC	GCC	AAG	AAG	GCT	AAG	CGT	GCC	CTG	GAA	AAG	GGG	ATG	ACG	GTC	ATC	TTC	TGC	GTC	GGA	GAG	ACC
WB ²	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23	...	..G	...	...	..G	...	..G	...	..C	..T	..A	...	...	...	...	...	...	...	...	...	..A	...	...	...
Ad-142	...	..G	...	...	..G	...	..G	...	..C	..T	..A	...	...	...	...	...	...	...	...	...	..A	...	...	...
P-15	...	..G	..T	...	...	...	...	...	...	..T	..A	...	...	...	...	...	..T	...	..T	..T	..T	...	...	...
Ad-7 ³	...	..G	..T	..T	...	...	..G	...	..T	...	..C	..A	..T	...	..T	..T	...	...	...	...	AC	...	...	...
GSM ⁴	...	..G	..T	..T	...	...	..G	...	..T	...	..C	..A	..T	...	..T	..T	...	...	...	...	AC	...	...	...
Ad-147 ⁵	...	..G	..T	...	...	...	...	...	..T	...	..G	...	..C	...	..T	...	...	...	...	...	ACT	..G	...	..A
<i>G. ardeae</i>	...	..G	...	...	C..	...	..C	G..	..C	..G	..C	..CG	GCC	...	...	...	...	C.A	...	...	A..	..C	...	..G
<i>G. muris</i>	...	..G	...	..A	C..	..A	..C	GCC	...	...	..T	..CC	TCC	..A	...	...	...	C..	...	...	A.T	...	...	...

360

JH ¹	TTG	GAT	GAG	CGC	AAG	GCC	AAC	CGC	ACC	ATG	GAG	GTG	AAC	ATC	GCC	CAG	CTT	GAG	GCG	CTT	GGC	AAG	GAG	CTC
WB ²	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23	C..	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..C	A..	...	...	..T
Ad-142	C..	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..C	A..	...	...	..T
P-15	C..T	...	...	...	...	...	...	...	...	...	...	..A	...	..T	..T	...	...	...	...	..C	A..	..A	...	...
Ad-7 ³	C..	...	..A	...	...	...	..T	AA.	..T	...	...	..A	..T	..T	..T	...	..C	...	..T	...	AAG	...	...	A..T
GSM ⁴	C..	...	..A	...	...	...	..T	AA.	..T	...	...	...	..T	..T	..T	...	..C	...	..T	...	AAG	...	...	A..T
Ad-147 ⁵	C..	..C	...	...	...	...	...	AAG	..T	...	..T	...	...	..T	..GA	...	..C	...	..C	...	AAG	...	..A	G..
<i>G. ardeae</i>	C..C	..C	...	...	..C	..G	GGG	AAG	..G	...	..C	..C	...	...	AAG	...	A..G	...	...	..G	C..	GCT	..CC	A..
<i>G. muris</i>	C..T	..C	...	...	..T	..G	GGG	AAG	..G	...	..C	..C	...	...	CAG	...	..C	...	..C	T..A	C..	GT.	..CC	A..

432

Figure75 (cont). Alignment of *Giardia*-derived *tpi* nucleotide sequences.

JH ¹	GGA	GAG	TCC	AAG	ATG	CTC	TGG	AAG	GAG	GTT	GTC	ATT	GCT	TAC	GAG	CCC
WB ²	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23	...	...	..G	...	...	...	...	...	.GA	...	..T	...	..C	...	...	...
Ad-142	...	...	..G	...	...	...	...	...	.GA	..C	..T	...	..C	...	...	...
P-15	..G	...	..T	...	C..	..A	...	...	A.A	..C	..T	...	...	...	...	...
Ad-7 ³	...	..A	..A	...	.A.	T.A	...	G..	A.C	...	..A	...	..C	..T	NNN	NNN
GSM ⁴	...	..A	..A	...	.A.	T.A	...	G..	A.C	...	..A	...	..C	..T	...	..G
Ad-147 ⁵	..T	..C	G.T	...	GC.	...	...	...	AGT	..C	...	..C	..C	...	...	...
<i>G. ardeae</i>	AAG	...	C..	TCT	---	.AT	...	G.C	CGC	..C	...	G.C	..C	...	...	..T
<i>G. muris</i>	AAG	..A	C.T	GGC	---	.AT	...	G.C	A.C	..C	..T	G..	..G	...	...	...

- 1 published sequence (Baruch *et al.*, 1996), identical to Ad-2 & Bris-136 *tpi* nt sequences.
2 published sequence (Mowatt *et al.*, 1994), identical to Ad-1 *tpi* nt sequence.
3 identical to Ad-19 & Ad-52 *tpi* nt sequences.
4 published sequence (Mowatt *et al.*, 1994).
5 identical to Ad-136 *tpi* nt sequence.

Table 26. Nucleotide composition (%) of the partial *tpi* sequences (for all sites and each individual codon base position).

Lineage/species	Isolate	All positions					1 st codon position				
		A	T	C	G	Total	A	T	C	G	Total
Assemblage A	JH	25.4	19.2	24.2	31.3	480	31.3	11.9	16.3	40.6	160
Assemblage A	WB	25.4	19.2	24.2	31.3	480	31.3	11.9	16.3	40.6	160
Livestock	P-15	26.6	22.6	20.9	29.9	478	32.1	11.9	16.4	39.6	159
Cat	Ad-142	24.6	17.5	24.8	33	479	31.4	10.1	17.6	40.9	159
Assemblage B	Ad-7	28.3	22.6	21.4	27.7	473	35	13.4	14.6	36.9	157
Assemblage B	GSM	27.9	22.5	21.5	28.1	480	34.4	13.8	15	36.9	160
Dog1	Ad-147	25.9	18.8	24.8	30.5	479	34.6	10.1	17	38.4	159
<i>G. ardeae</i>		19.9	14.7	33.3	32.1	477	27	7.5	22.6	42.8	159
<i>G. muris</i>	Ad-120	21.5	16	33.1	29.4	456	27	7.9	24.3	40.8	152

		2 nd codon position					3 rd codon position				
		A	T	C	G	Total	A	T	C	G	Total
Assemblage A	JH	32.5	28.8	23.1	15.6	160	12.5	16.9	33.1	37.5	160
Assemblage A	WB	32.5	28.8	23.1	15.6	160	12.5	16.9	33.1	37.5	160
Livestock	P-15	32.1	28.9	22.6	16.4	159	10.7	13.8	34	41.5	159
Cat	Ad-142	31.9	28.8	22.5	16.9	160	10.6	13.8	34.4	41.3	160
Assemblage B	Ad-7	32.9	29.1	22.8	15.2	158	17.1	25.3	26.6	31	158
Assemblage B	GSM	33.1	28.8	23.1	15	160	16.3	25	26.3	32.5	160
Dog1	Ad-147	33.1	28.1	23.8	15	160	10	18.1	33.8	38.1	160
<i>G. ardeae</i>		30.8	28.9	25.2	15.1	159	1.9	7.5	52.2	38.4	159
<i>G. muris</i>	Ad-120	31.6	28.9	24.3	15.1	152	5.9	11.2	50.7	32.2	152

64% for *G. muris* and *G. ardeae* sequences (Table 26). The GC content at the 2nd codon position was similar for all three species, ranging from 38 - 40%. The high GC content of the 3rd codon positions of the *G. muris* and *G. ardeae* sequences (83% and 90% respectively) supported the authenticity of these sequences, as this is a characteristic of many *Giardia* protein coding genes (Alonso & Peattie, 1992; Mowatt *et al.*, 1994). The GC content of the 3rd codon position observed for the *tpi* sequences of the *G. intestinalis* isolates was not as biased as that observed for the *eflα* (> 98%) or *gdh* (77 - 92%) sequences, and it also varied to a greater extent, ranging from 58 - 75%.

Table 27. Comparison of log likelihoods and s/v ratio used.

s/v ratio	log likelihood
0.7	-2386.34412
1	-2361.74157
1.5	-2347.59943
2	-2346.55157 ← optimal
2.5	-2350.69009
3	-2357.11244
4	-2372.40127

The optimal transition/transversion ratio of the *tpi* sequences was determined as described previously (section 4.4.3.7), and was found to be 2 (Table 27, p. 193).

The sequences were examined for multiple hits and saturation for change in the same manner described earlier (section 4.4.3.7). The situation was very similar to that observed for *gdh*. Distances of < 0.2 (ie., comparisons of sequences from isolates comprising the same lineage, or from isolates comprising the Assemblage A, Livestock or Cat lineages) showed no evidence of multiple hits when all codon base positions were analysed (slope of line = 1, Figure 76, p. 195). Comparison of corrected and uncorrected distances estimated using 1st or 2nd codon positions suggested that multiple hits had not occurred at these sites for any of the sequence comparisons involving the isolates comprising the *G. intestinalis* lineages (slope of line = 1 for distances < 0.1 , Figure 77, p. 196). There was evidence for multiple hits for both transition and transversion substitutions (Figure 78, p. 197). Plots of TN and p-distances estimated using transitions suggested that multiple hits had occurred for all sequence comparisons involving isolates from different lineages (distances > 0.05 , Figure 78A). For TN distances > 0.15 , there appeared to be two separate curves (Figure 78A). The lower curve was comprised of pair-wise distances from sequence comparisons between *G. intestinalis* isolates and the *G. muris* and *G. ardeae*, while the upper curve comprised *G. intestinalis* pair-wise sequence comparisons. In the case of distances estimated using transversions, multiple hits only occurred for comparisons of the *G. intestinalis* sequences with either *G. muris* or *G. ardeae* sequences (distances > 0.2 , Figure 78B). Comparison of the different corrected distance measures (Figure 76, p. 195) showed that the TN distance was very similar to the distance estimates of the simpler models (JC, K2p), although there was a slight difference between the TN and ML distance estimates, with the ML distances slightly larger for the comparisons of *G. intestinalis* with either *G. ardeae* or *G. muris* (Figure 76, p. 195). This may have been caused by differences in the base composition of some of the sequences, which violates an assumption of DNAML. Examination of distances estimated using 3rd codon positions showed that saturation occurred once the p-distance exceeded 0.5 substitutions per site (Figure 77, p. 196). Comparisons of corrected and uncorrected distances estimated using synonymous or nonsynonymous substitutions suggested that the former was saturated for change (Figure 79A, p. 198), and that few multiple hits had occurred for the latter (Figure 79B, p. 198). Multiple hits were not evident for any of the comparisons involving *G. intestinalis* sequences when corrected and uncorrected pair-wise distances inferred from the deduced amino acid sequences were plotted against each other (Figure 80, p. 199).

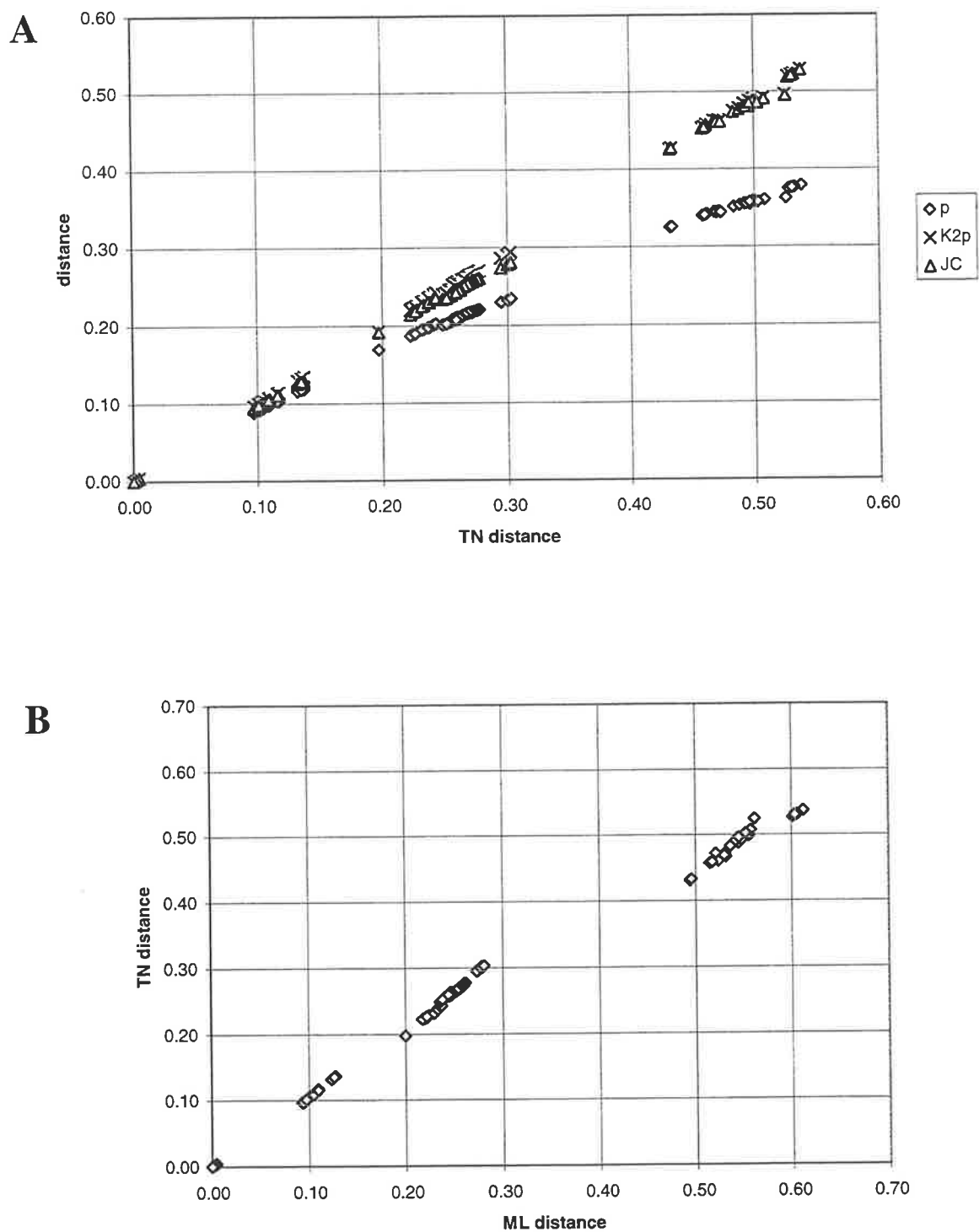


Figure 76. Dot plots of *tpi* TN distances versus **A:** p, K2p and JC distances and **B:** ML distances calculated for all pair-wise sequence comparisons using all classes of nt substitutions.

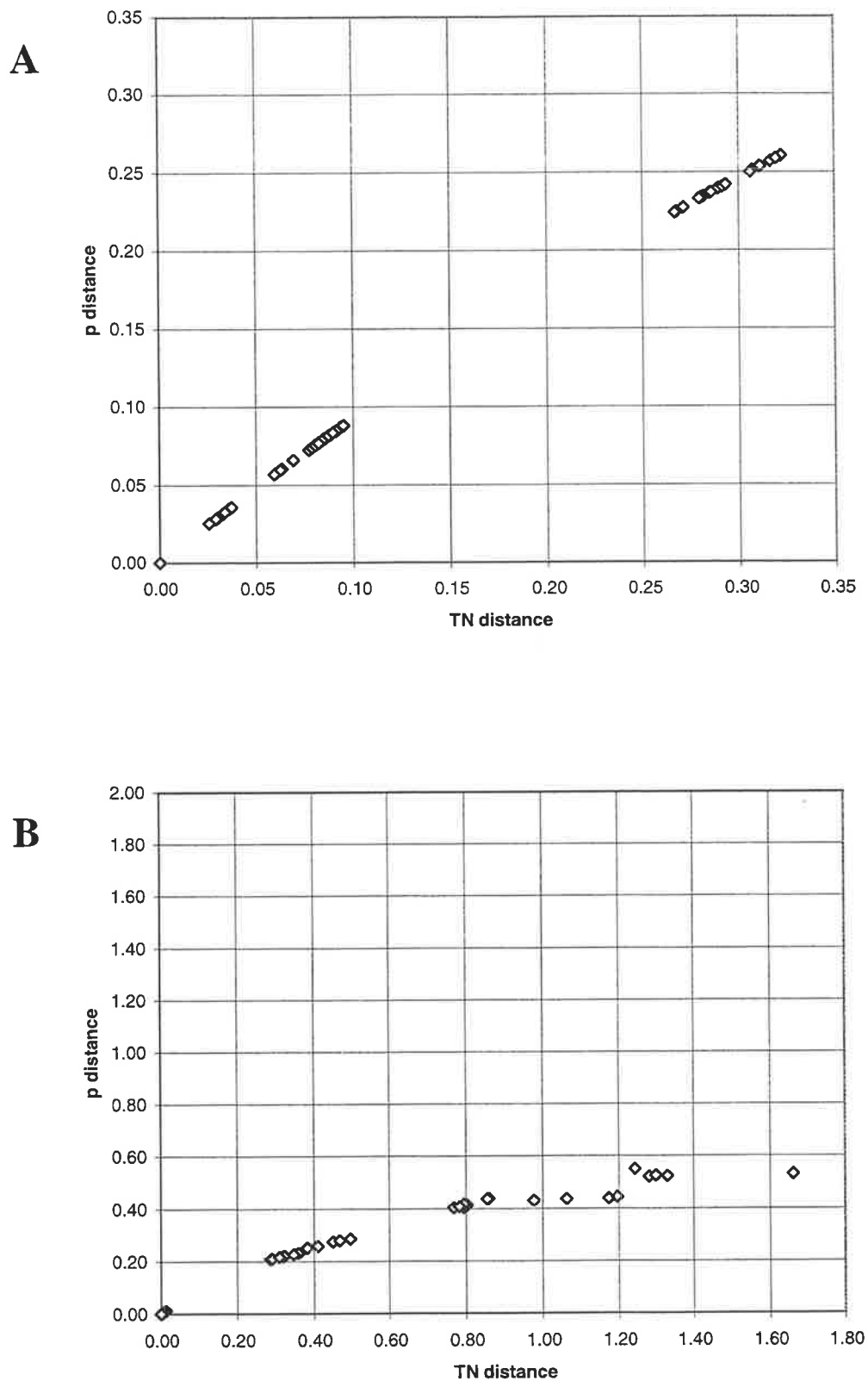


Figure 77. Dot plots of *tpi* TN distances versus p-distances calculated for all pair-wise sequence comparisons using **A:** 1st & 2nd codon positions only or **B:** 3rd codon positions only.

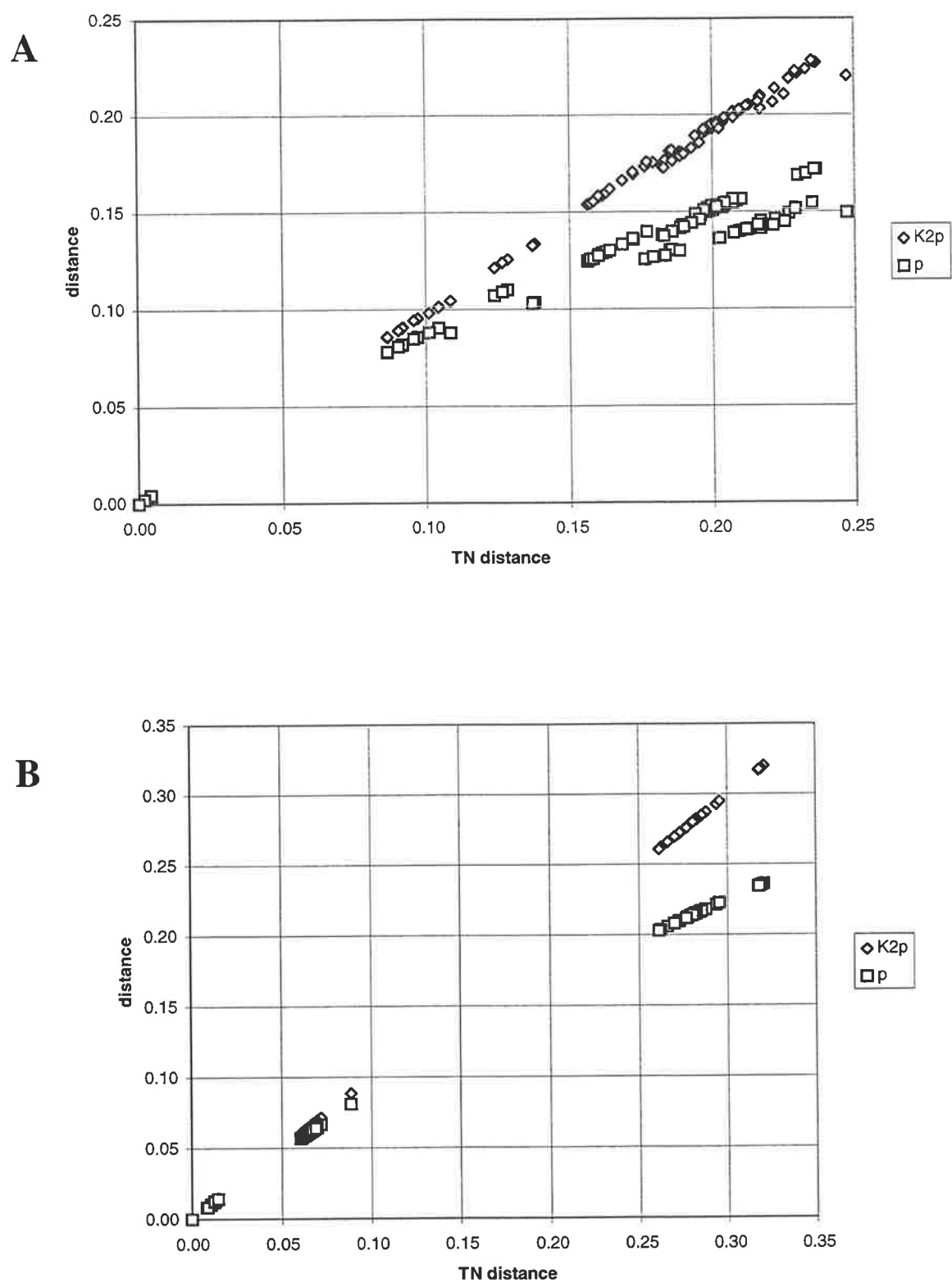


Figure 78. Dot plots of *tpi* TN distances versus K2p and p-distances calculated for all pairwise sequence comparisons using **A**: transitions only or **B**: transversions only.

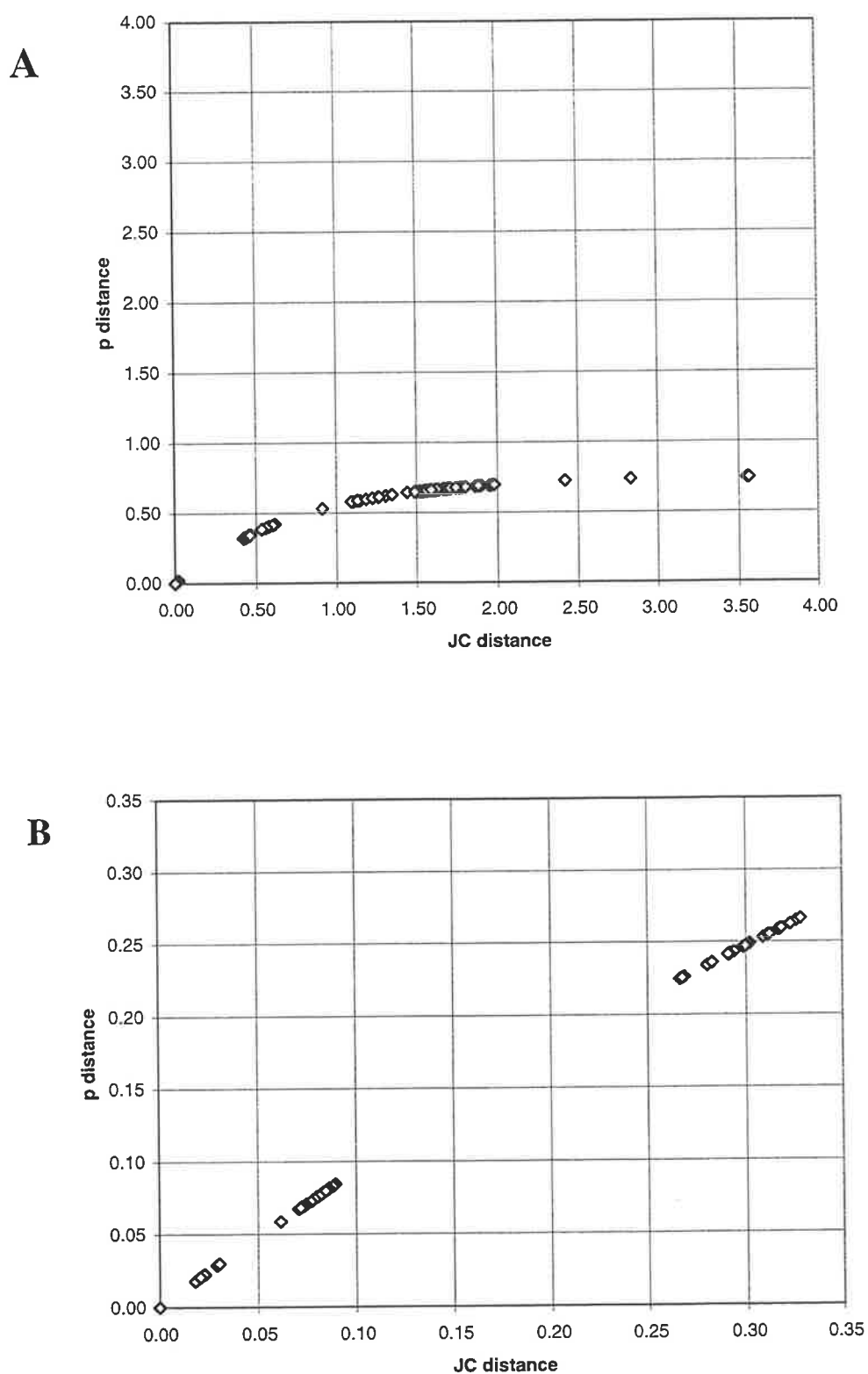


Figure 79. Dot plots of *tpi* p-distances versus JC distances calculated for all pair-wise sequence comparisons using **A:** synonymous substitutions only or **B:** nonsynonymous substitutions only.

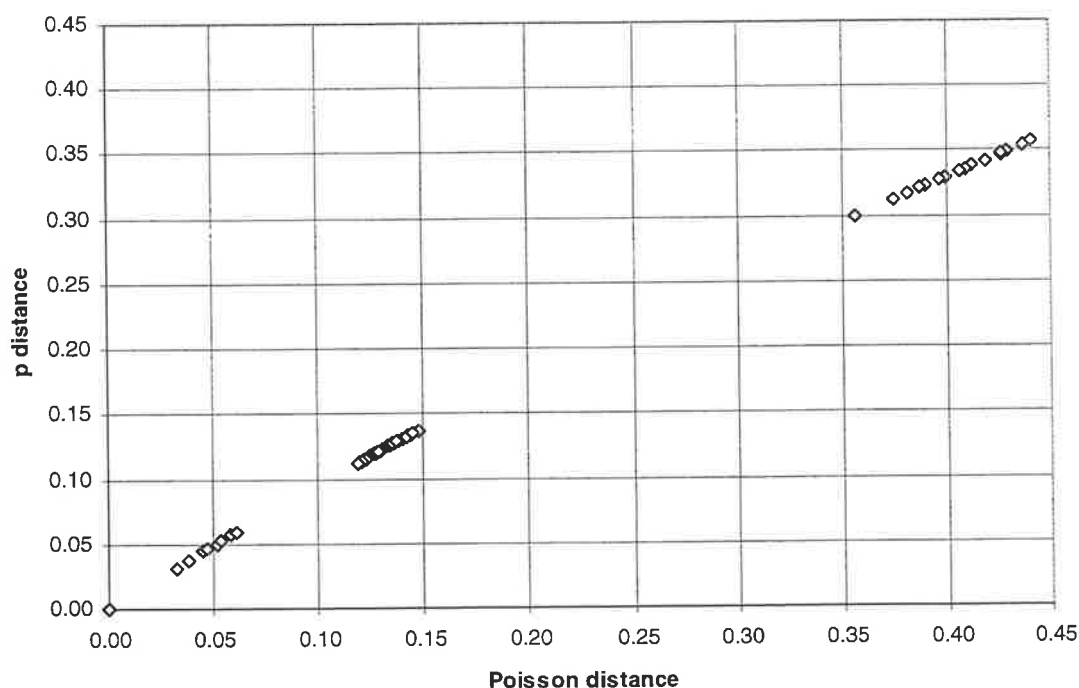


Figure 80. Dot plot of *tpi* Poisson-corrected distances and p-distances calculated for all pairwise sequence comparisons using predicted amino acid sequences.

Among-site rate variation was examined as described in section 4.4.2.7. Sequences from the following isolates were used: *G. muris*, *G. ardeae*, Ad-7, GSM, Ad-147, WB, Ad-142 & P-15. The *G. muris* sequence had to be excluded to allow the estimation of α for all positions (due to a program error that only became apparent when this sequence was included). There was evidence for among site rate variation when all codon positions were compared ($\alpha = 0.39$). However the estimate may have been affected by significant differences in the nt composition of the *G. ardeae* sequence compared to the other sequences ($p < 0.05$ as determined by PUZZLE ML analysis). The α parameter for 1st or 2nd codon positions was 0.33 for each. Examination of variable sites (by excluding invariant sites from analysis) yielded $\alpha = 0.78$ and $\alpha = 1.16$ respectively, indicating moderate rate heterogeneity at the 1st codon positions. Surprisingly, there was no evidence for extensive rate heterogeneity at the 3rd codon positions ($\alpha = 2.63$), although the estimate should be treated with caution because of significant differences ($p < 0.05$) in the nt composition at these sites between some of the *G. intestinalis* sequences (eg., isolates from Assemblages A & B, Table 26, p. 193).

4.5.5.4 Distance-based phylogenetic analysis of *tpi* nucleotide sequences

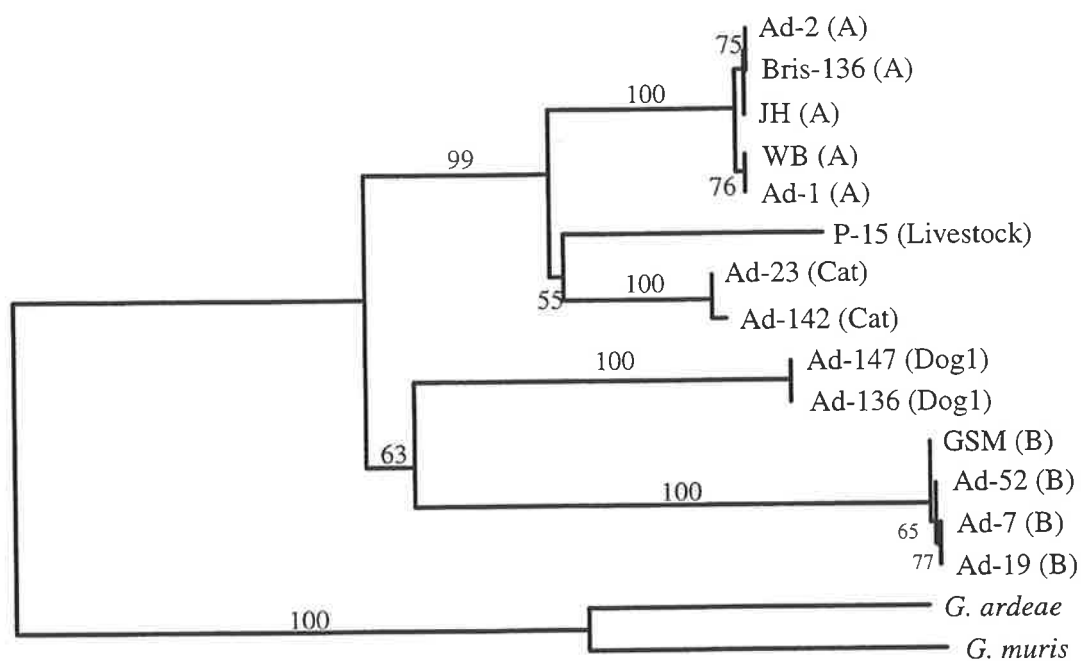
The aligned *tpi* nt sequences (Figure 75, pp. 190 - 192) were analysed by distance methods. Distances were estimated using the ML, TN, K2p, synonymous, nonsynonymous, and Poisson corrected amino acid distance measures, and trees were built using either the ME or NJ algorithms. Examples of the resulting phylograms are given in Figures 81 & 82 (pp. 201 & 202). Methods using DNA-based distances produced trees which had almost identical topologies, and which were similar to the trees inferred by analysis of the *gdh* sequences; ie., placing the Livestock, Assemblage A and Cat lineages into one group, and Assemblage B and Dog1 lineages into another. Bootstrap analyses strongly supported the former grouping (99%) but not the latter (63 - 70%) (Figure 81, p. 201). Analyses using distances estimated from the deduced amino acid sequences supported the monophyly of the Assemblage A, Cat and Livestock lineages, but could not resolve the positions of the Assemblage B and Dog1 lineages (Figure 82, p. 202). The NJ tree using TN distances was also able to distinguish genetic groups I & II, both of which had moderate bootstrap support (Figure 81, p. 201). All of the *tpi* trees examined provided strong support for the monophyly of *G. muris* and *G. ardeae*, which were separated by a distance that was no larger than those separating the *G. intestinalis* lineages from each other. This result was in conflict with the *eflα* and 18S rRNA results and may provide evidence for introgression (discussed in section 5.3).

4.5.5.5 Maximum likelihood phylogenetic analysis of *tpi* nucleotide sequences

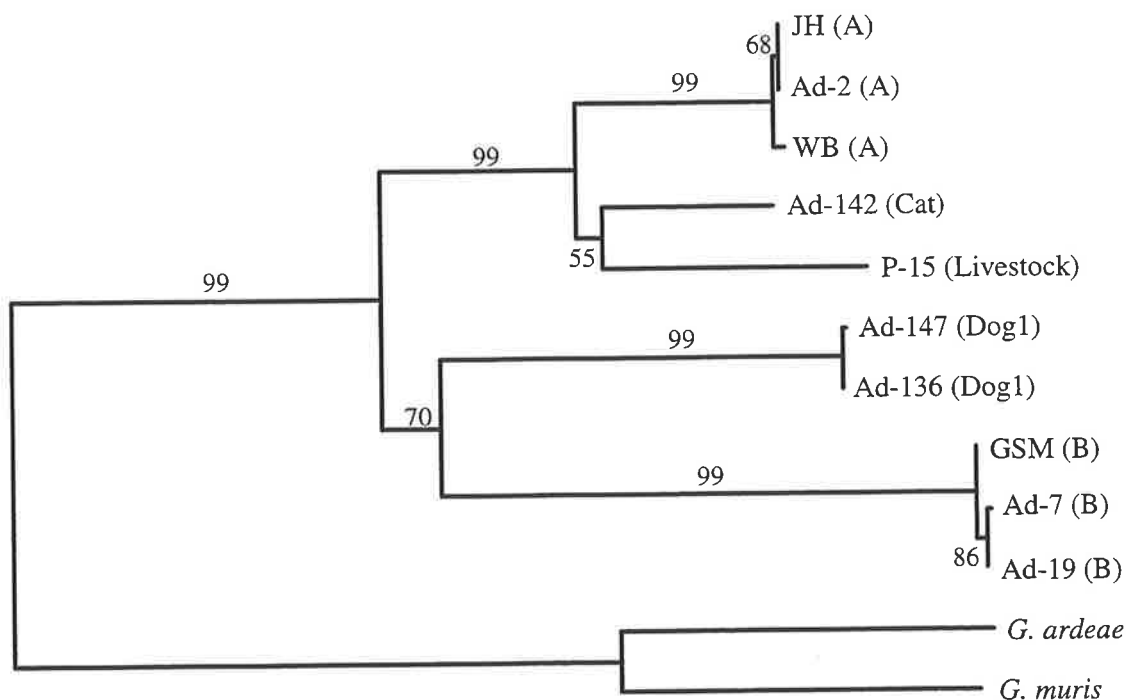
Maximum likelihood analyses, using either quartet puzzling (PUZZLE) or the original method described by Felsenstein (PHYLIP, DNAML), produced similar results (PUZZLE tree shown, Figure 83, p. 203). The tree topologies were consistent with the distance analyses, suggesting that the slight differences in nt composition detected between the ingroup (*G. intestinalis*) and the outgroup (*G. muris*, *G. ardeae*) sequences did not affect the ML analysis. ML analysis of the deduced amino acid sequences using the substitution model of Jones *et al.* (1992) produced a tree which fully supported the DNAML and distance analyses (Figure 83, p. 203).

4.5.5.6 Cladistic analysis of *tpi* nucleotide sequences

Parsimony analyses of the *tpi* sequences supported the tree topologies determined by the distance and ML methods. The exhaustive search of PAUP recovered a single tree of 342



0.01 substitutions/site



0.01 substitutions/site

Figure 81. Phylogenetic relationships of *Giardia* isolates inferred from the *tpi* nt sequences using TN or K2p distances and the NJ (top) or ME (bottom) tree building methods. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (500 or 1,000 replicates) is indicated above the branches.

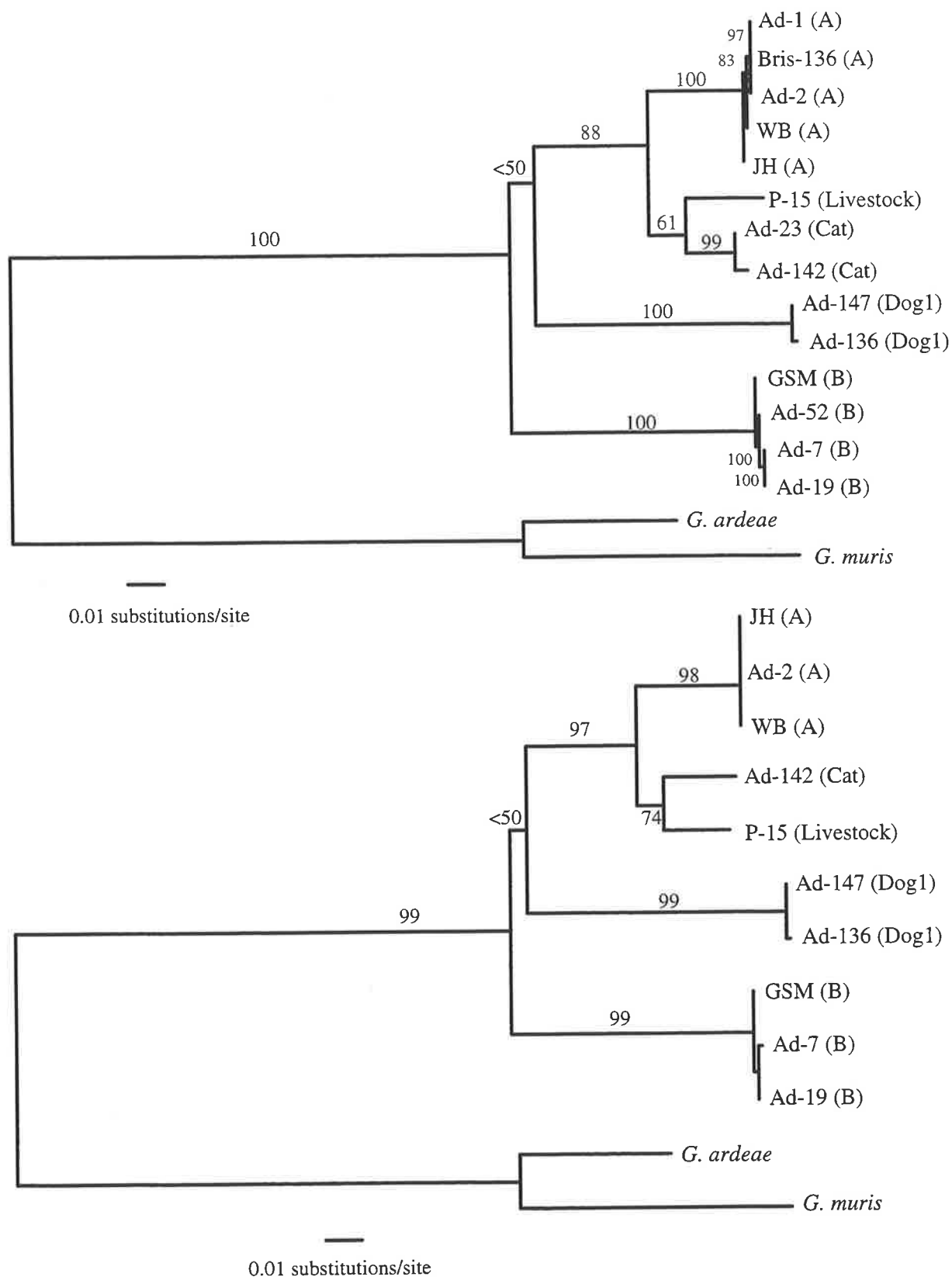


Figure 82. Phylogenetic relationships of *Giardia* isolates inferred from the *tpi* deduced amino acid sequences using Poisson corrected distances and the NJ (top) or ME (bottom) tree building methods. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). Percentage bootstrap support (500 or 1,000 replicates) is indicated above the branches.

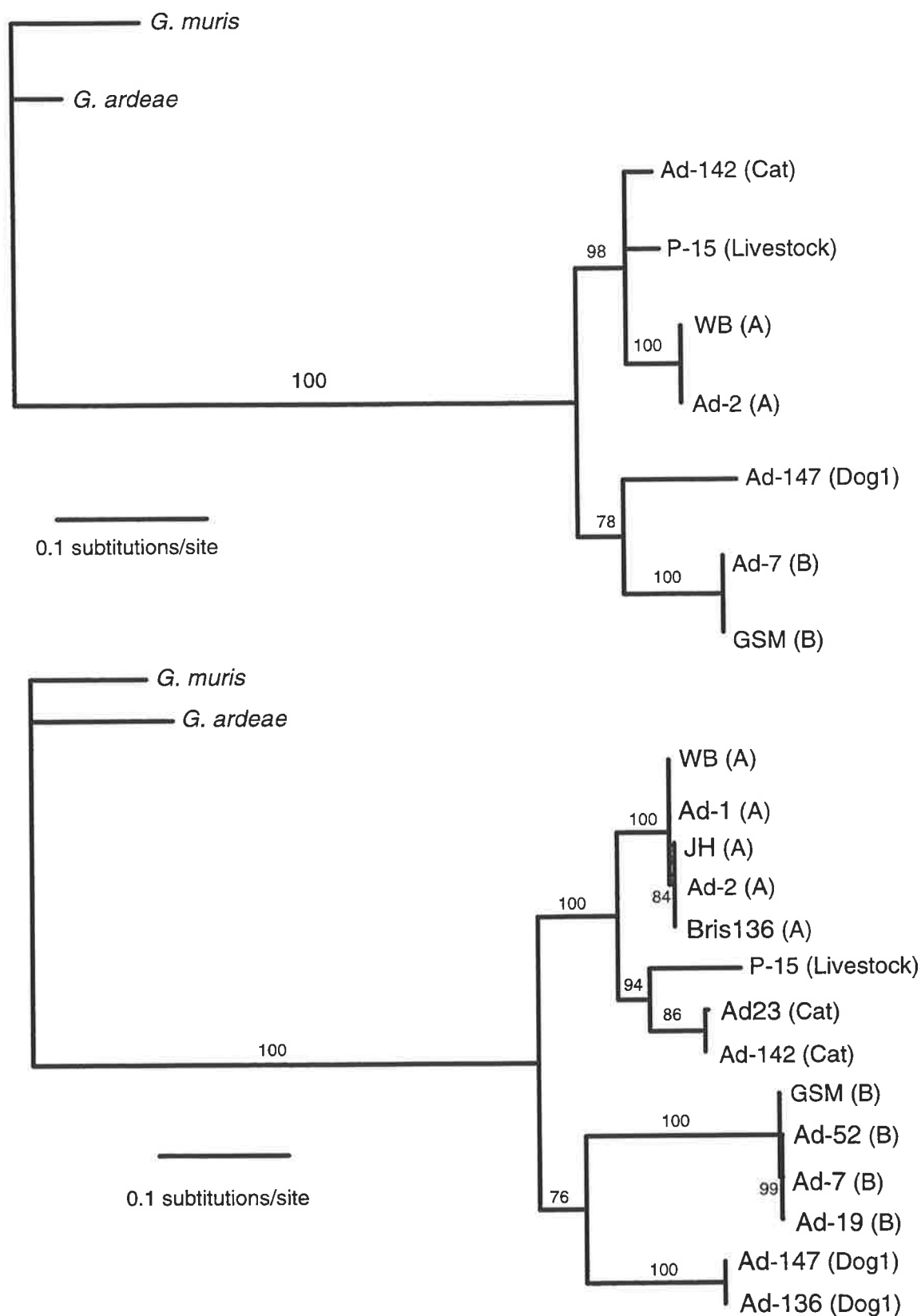


Figure 83. Phylogenetic relationships of *Giardia* isolates inferred by ML quartet puzzling of the *tpi* deduced amino acid (top) or nt (bottom) sequences. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). Percentage support (1,000 replicates) is indicated above the branches.

steps in length (Figure 84, p. 205). Bootstrapping (using the branch and bound search option) found good support for the clade containing the Cat, Livestock and Assemblage A isolates (99%), but none for any of the interior branching order ($< 70\%$). Parsimony analysis placed the Dog1 lineage as the sister taxon to the Cat, Livestock and Assemblage A clade, followed by Assemblage B, which is different to the branching order determined by the distance-based analyses (section 4.5.5.2). As was the case for the distance-based analyses, the branching order of the Assemblage B, Dog1, and the clade containing the other three lineages could not be resolved, producing an unresolved polytomy. Parsimony analysis of the deduced amino acid sequences gave a slightly different result, producing a tree topology which was almost identical to the distance-based trees (Figure 84, p. 205). The *G. intestinalis* lineages were divided into two major clades, the first containing the Assemblage A, Cat and Livestock lineages, and the second containing the Assemblage B and Dog1 lineages. There was strong bootstrap support for the former clade (93%), and good support for the latter clade (85%).

The amount of phylogenetic signal relative to random noise in the data was also examined. The tree length distributions of the trees examined by the exhaustive search were graphed (Figure 85, p. 206), and the g_1 statistic determined. A total of 135,135 trees were examined, with a mean of 532.24 ± 35.34 steps compared to the single shortest tree of 342 steps. The g_1 statistic was -1.43, which indicated significant skewness and therefore the presence of phylogenetic signal. Removal of the outgroups had no significant effect on the skewness of the tree distributions ($g_1 = -1.50$ for 7 taxa, Figure 85, p. 206).

4.5.6 Summary and conclusions

Analysis of the sequence data from the *tpi* and *gdh* genes provides support for the model of genetic relationships presented in section 3.5. The consensus tree (topology only) for the *gdh* and *tpi* results is given in Figure 86 (p. 207). The composition of the lineages identified was consistent with the clusters determined by AE analysis. Although both *gdh* and *tpi* appear to be evolving more rapidly than either 18S rRNA or *eflα*, there was no improvement in the resolution of the relationships between the lineages. Furthermore, there was very little intra-lineage variation detected from either gene, which was surprising considering the level of inter-lineage divergence. The intra-lineage differences allowed reliable resolution of genetic groups I and II within Assemblage A, but not of any of the other groupings identified within the other lineages by AE.

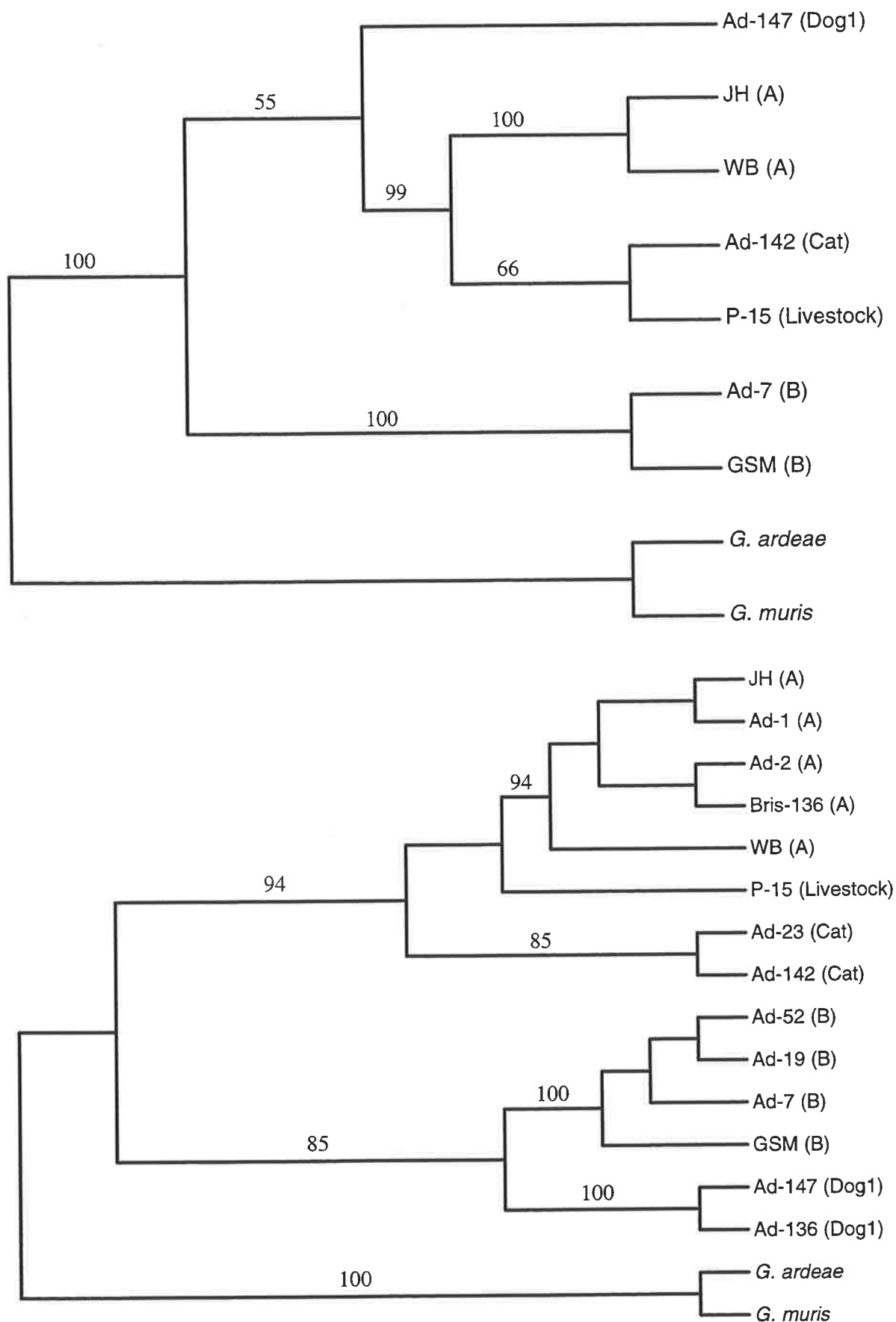


Figure 84. Evolutionary relationships of *Giardia* isolates inferred by parsimony analysis of the *tpi* nt (top) or deduced amino acid (bottom) sequences. The allozyme defined lineage of an isolate is in parentheses (A = Assemblage A, B = Assemblage B). Percentage support (>50% from 1,000 replicates) is indicated above the branches.

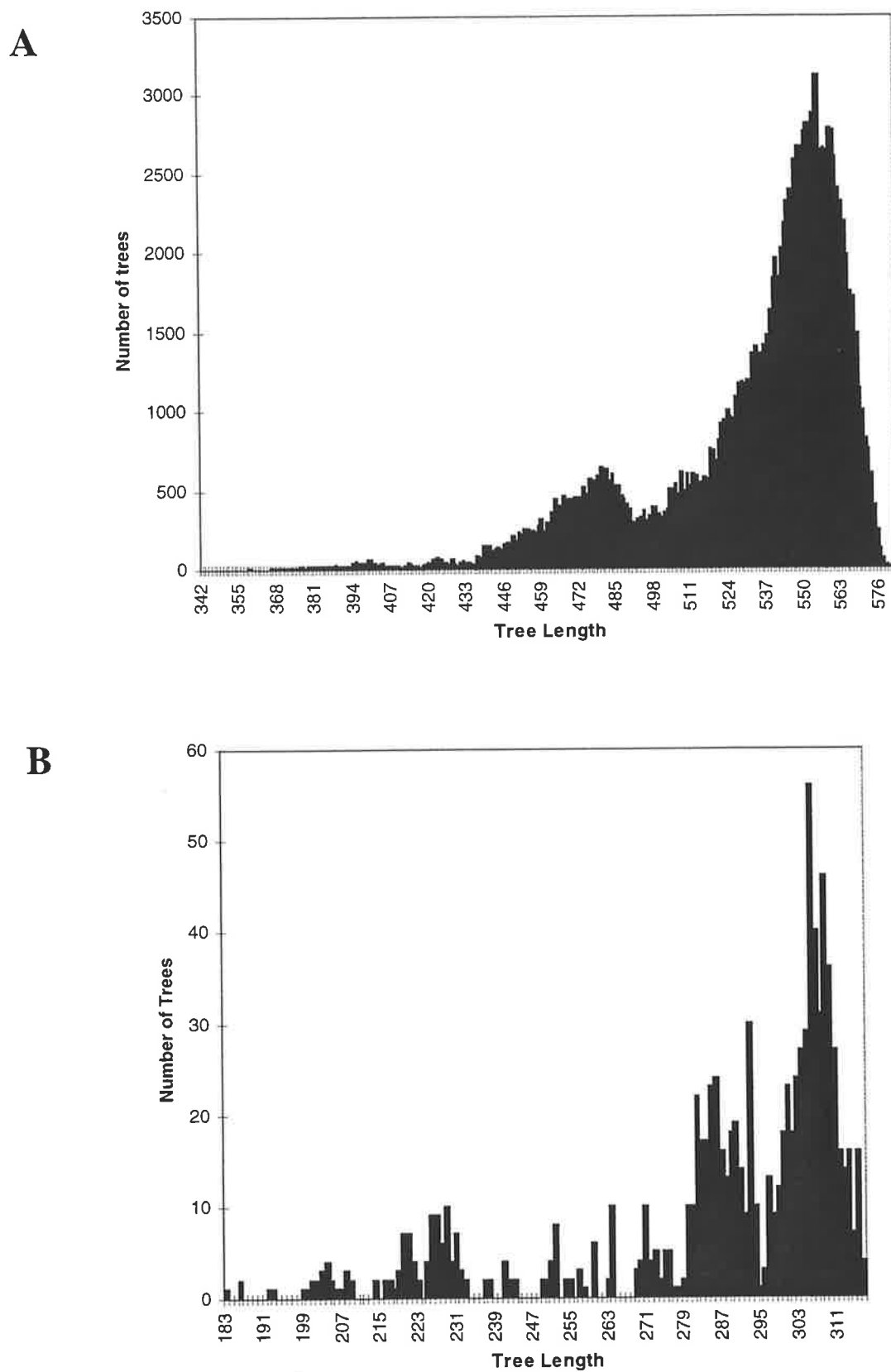


Figure 85. Length distributions of the trees examined during the exhaustive search analysis of the *tpi* nucleotide sequences (Figure 84, p. 205, top cladogram). **A:** all taxa, **B:** only *G. intestinalis* sequences analysed.

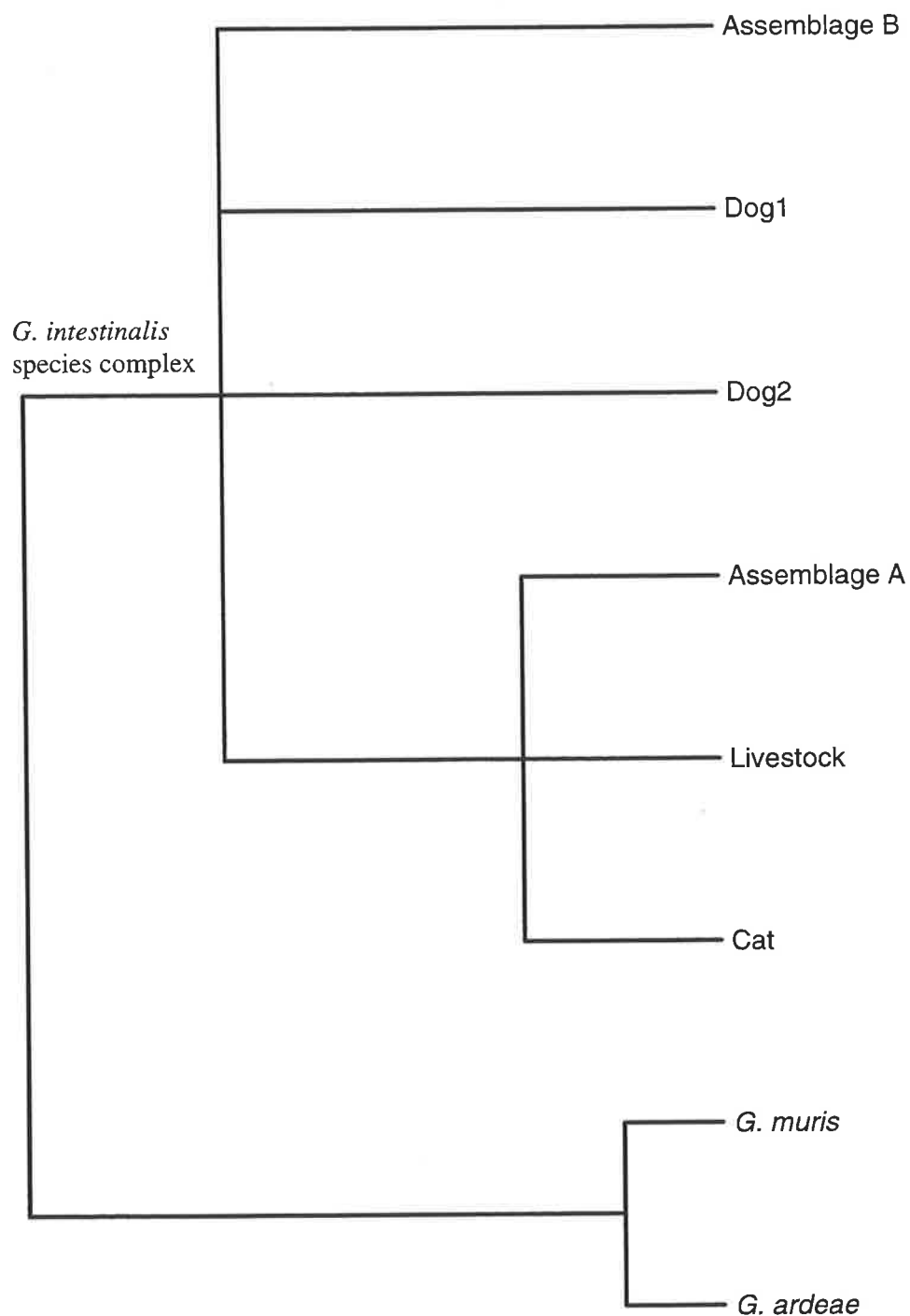


Figure 86. Strict consensus tree of evolutionary relationships of the lineages comprising the *G. intestinalis* species complex determined by phylogenetic analyses of nt sequences from the *gdh* and *tpi* genes.

A more detailed analysis of the *gdh* nt sequences identified restriction site polymorphisms that were diagnostic for the separate lineages of *G. intestinalis* and *G. ardeae*. These polymorphisms have subsequently been used in PCR-RFLP assays to characterise samples for which only DNA was available, demonstrating that the “Polish” and “Belgian” groupings of Homan *et al.* (1992) were equivalent to Assemblages A and B (Monis *et al.*, 1996; Appendix D), and that the genotypes belonging to the livestock lineage were present in Europe and Australia (Ey *et al.*, 1997; Appendix F).

An interesting and perhaps significant finding was the unexpectedly high similarity between the *tpi* sequences of *G. ardeae* and *G. muris*. Both sequences were very different from the *G. intestinalis* sequences (62 - 67% similar), but they exhibited a high GC content (>80%) at the 3rd codon positions - a characteristic of *G. intestinalis* genes (Alonso & Peattie, 1992; Mowatt *et al.*, 1994). In comparison, the GC content of the 3rd codon position of yeast or bacterial *tpi* was less than 60% (by inspection of sequences acquired from GenBank, accession numbers M14432 & X00617, not shown). Contamination of the *G. ardeae* PCR by other *Giardia* DNA sources could be effectively ruled out, as this sequence was distinct from any of the other *tpi* sequences generated in this study. Possible explanations for the relatively high similarity between *G. muris* and *G. ardeae* *tpi* sequences could be; (1) recombination between these two species, (2) the presence of paralogous genes, or (3) culture contamination.

Despite the similarities of the phylogenies inferred using nt data obtained from 18S rRNA, *eflα*, *gdh*, and *tpi*, there were sufficient differences to question the validity of combining the data for further analyses. Aside from the uncertainty regarding the origin of the *G. ardeae* and *G. muris* *tpi* nt sequences, the four genes, not surprisingly, appear to be evolving at significantly different rates (see Discussion, section 5.2). Differences in the nt composition of each gene could also conceivably invalidate the assumptions of some of the inferencing methods, such as maximum likelihood. For these reasons, a combined analysis was not conducted.

Chapter 5

DISCUSSION

The research conducted in this project was directed towards examining the systematics of the protozoan parasite *Giardia*, specifically the members of the *G. intestinalis* species complex. The techniques of allozyme electrophoresis and nucleotide sequence analysis were used to determine phylogenetic relationships among isolates of *G. intestinalis* of different host origin. The resulting predictions provide a framework around which the systematics of the *G. intestinalis* species complex can be built, and they allow a comparison of the relative utility of these two disparate but complementary techniques in systematics.

5.1 Comparison of allozyme and nucleotide sequence analyses

There was clear support from all of the data presented herein for the monophyly of the *G. intestinalis* species complex, which comprised six distinct lineages. The phylogenetic relationships inferred by analysis of the 18S rRNA, *efl* α , *gdh*, and *tpi* nucleotide sequences were in agreement with the major groupings identified by allozyme analysis. A new model of the relationships of these lineages, based upon the consensus of the results, can now be proposed (Figure 87, p. 210). This model illustrates the relationships between the defined lineages, with the data sets supporting each particular branch indicated in Figure 87. All of the branch points within the model have good support from at least one of the data sets. In some cases the different data sets produced conflicting results; eg., the branching order of the Livestock, Cat and Assemblage A lineages. However, in all cases there was only one topology which had good statistical support, while the rest were reduced to unresolved polytomies, indicating that the particular data sets in question contained insufficient information to resolve the relationships adequately.

A seventh lineage, comprising the rat-derived isolates Ad-155 and Ad-157, was identified by AE analysis (Table 14, p. 87). The Rat lineage was separated from all of the other *G. intestinalis* isolates by fixed differences at 50 - 79% of loci, but its placement within the proposed model requires additional sequence data. The magnitude of the allelic differences however, suggests that the branch-point would be at a level comparable to that separating the Assemblage B, Dog1 and Dog 2 lineages. The levels of divergence separating

the lineages identified herein are similar to those separating recognised species in other genera of parasitic protozoa (eg., Adams *et al.*, 1989; Andrews *et al.*, 1988) and bacteria (eg., Lawrence *et al.*, 1991). This suggests that each of these lineages warrant recognition as distinct phylogenetic species.

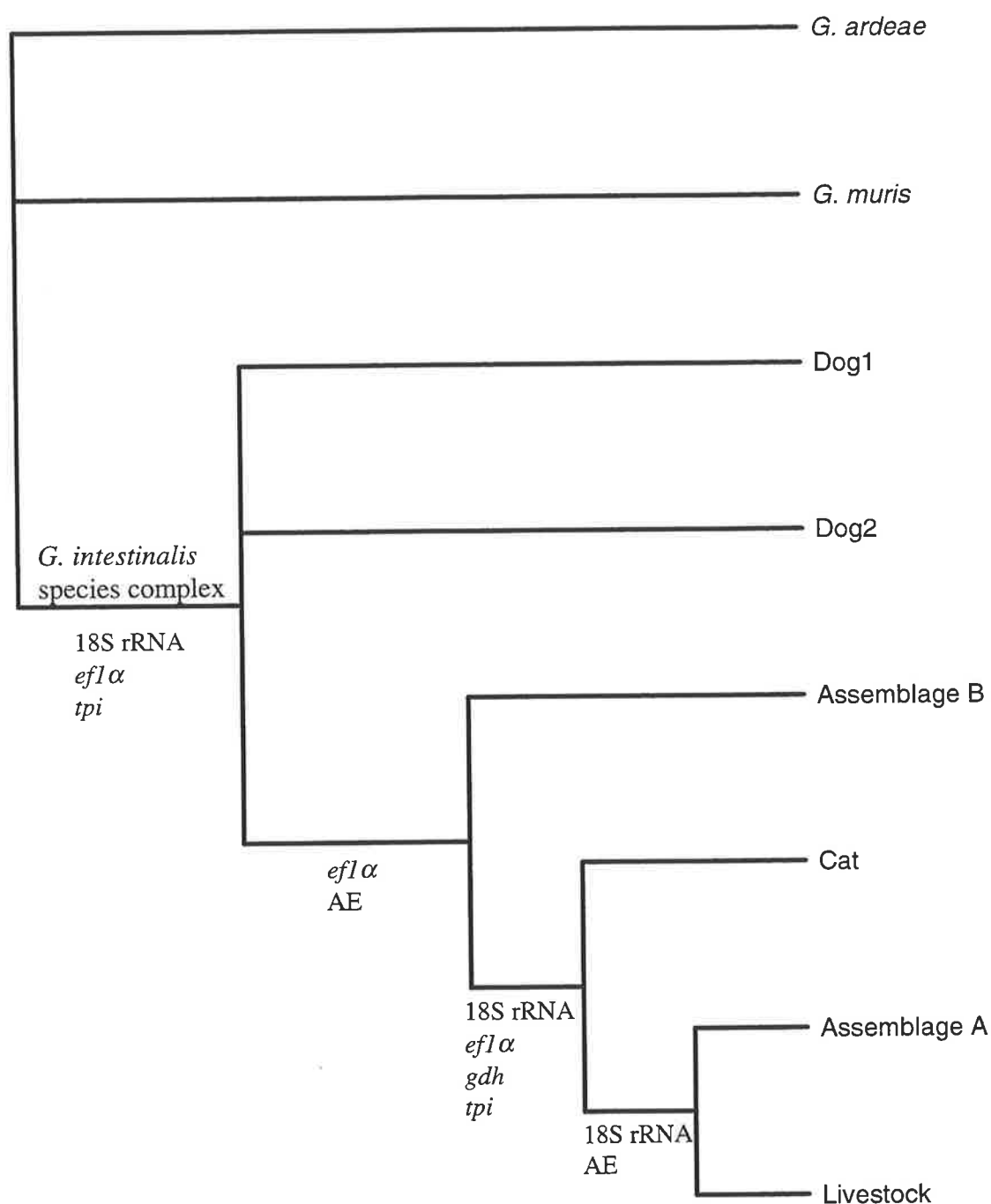


Figure 87. Consensus tree depicting the proposed evolutionary relationships of the major lineages of *G. intestinalis* inferred from analyses of the AE, 18S rRNA, *efl*α, *gdh* and *tpi* data. The relevant data sets supporting particular nodes are indicated under the branches.

The relationships between recognised *Giardia* species are less certain. There appears to be little doubt that the greatest evolutionary distance lies between the *G. intestinalis* species complex and *G. muris*. Furthermore, analysis of the 18S rRNA gene sequences strongly suggested that *G. psittaci* is a close sister taxon of the *G. intestinalis* species complex. The position of *G. ardeae* is more tentative. The *eflα* data set was the only one with outgroups for *Giardia* (namely; *H. inflata*, *S. muris*, and an unnamed diplomonad species). Although a *Hexamita* 18S rRNA sequence was available, it was not suitable for use as an outgroup because it was less than 60% similar to the *Giardia* sequences. Parsimony analyses suggested that *G. ardeae* is a sister taxon of the *G. intestinalis* species complex, with *G. muris* forming the outermost taxon, but the bootstrap support for this model was poor. Both maximum likelihood and DNA distance measures supported this branch order, with good bootstrap support for the DNA distance based tree topologies. Midpoint rooting of the 18S rRNA phylogenies also supported the branching order inferred by the *eflα* analyses, so this was the order used for the model in Figure 87 (p. 210). The phylogeny inferred by analysis of the *tpi* data had an alternative arrangement, suggesting that *G. muris* and *G. ardeae* were sister taxa. However, this result may be due to other mechanisms such as recombination (as discussed in section 5.3).

5.2 Resolution of allozyme and nucleotide sequence analysis

The results unambiguously demonstrate that AE and nucleotide sequence analyses provide complementary and supportive data. The amount and type of information provided by each data set was different, thus giving resolution for different parts of the phylogeny of the lineages comprising the *G. intestinalis* species complex. The AE data were able to determine the genetic relationships between all of the isolates, identifying two levels of organisation: (1) *genetic groups*, which comprise isolates differing at less than 25% of loci on average, and (2) *lineages*, which comprise genetic groups that have fixed genetic differences at 25% to 47% of loci examined. The lineages were distinguished by differences at 56% to 84% of loci. To put these levels of genetic differentiation into perspective, it should be noted that the maximum level of intra-specific variation within *G. muris* (which included isolates from different continents) was only 13% fixed genetic differences. The relationships between these lineages cannot be reliably determined by AE because of the increased likelihood of homoplasy in the data as the distance between samples exceeds 60% fixed allelic differences (section 3.4.3.3 & Richardson *et al.*, 1986).

The analyses of the 18S rRNA, *efl* α , *gdh*, and *tpi* sequences were unable to resolve the relationships between isolates at the level of the genetic group, but they did provide information about the relationships between the lineages. Each of the gene segments examined appears to be evolving at different rates, as evidenced by the relative branch lengths of the phylogenies inferred from each data set. As a result, each nt sequence contains information that is capable of resolving different parts of the phylogeny (shown in Figure 87, p. 210). All of the data supported the grouping of the Cat, Livestock and Assemblage A lineages. These appear to be the three most recently separated lineages, although paradoxically only the 18S rRNA data were capable of resolving the relationships between them. Similarly, analysis of all of the data sets identified the lineages, but only the *efl* α data enabled the order of the branches containing the Assemblage B, Assemblage A, Cat and Livestock lineages to be resolved. Apart from corroborating the composition of the lineages determined by AE, the *gdh* and *tpi* data sets provided little additional information regarding the phylogeny of the *G. intestinalis* species complex. However, the restriction site polymorphisms identified within the 1.17 kb *gdh* segment do provide diagnostic markers for the lineages of *G. intestinalis* described herein, and for *G. ardeae* (Monis *et al.*, 1996; Appendix D).

The identification of discrete lineages within the *G. intestinalis* species complex conflicts with the findings of Meloni *et al.* (1995). These authors identified a large number of genetic groups using the criteria originally proposed by Andrews *et al.* (1989) but found no higher order relationships, ie., lineages. Furthermore, they reported a much greater level of genetic variation (20% or more) when comparing some isolates that were analysed herein (Table 28, p. 213). For instance, isolates BAH1, BAH2 and BAH4 differ allozymically at 68 - 78% of loci from the Group III representative (BAH12), but they were assessed as 100% different by Meloni *et al.* (1995). However, in comparing these studies it is important to note that within vertebrate taxa (Richardson *et al.*, 1986) and micro-organisms (Selander *et al.*, 1986), the accuracy of genetic distance measures is directly proportional to the number of independent characters examined. In the present comparison, the discrepancy in the distance estimates may be attributed largely to differences in the number of enzyme loci examined in the respective studies, ie., 13 by Meloni *et al.* (1995) compared to 23 herein. Indeed, when the number of loci in the AE data set are decreased (by excluding part of the data set) to that used by Meloni *et al.* (1995), the distance estimates increased, although only by approximately 10% fixed differences (Table 28, p. 213). This was despite the inclusion of almost all of the loci sampled by Meloni *et al.* (1995). In this case, the differences in observed distances may lie in the different methods used to interpret the electrophoretic banding patterns. A zymodemic

interpretation of the isozyme banding patterns is likely to over-estimate distances, especially for isolates exhibiting multiple banding patterns (see: Figure 7, p. 63). The strength of the allozyme analysis conducted herein lies in the analysis of an unparalleled number of independent enzyme loci (23 - 27, compared with ≤ 14 by the other investigators listed in Table 10, p. 61).

Table 28. Comparison of the levels of genetic difference as determined by different investigators for the same isolates.

comparison	Meloni <i>et al.</i> (1995)	23 loci	13 loci
BAH12 vs BAH1, 2, 4	100%	68 - 78%	78 - 85%
BAH1, 2 vs BAH4	46%	22 - 29%	23 - 38%

BAH1, BAH2 are equivalent to Ad-1, BAH4 equivalent to Ad-2 as shown by Andrews *et al* (1989).

An unexpected finding was the limited degree of intra-lineage polymorphism detected using the sequence-based approach, especially at the less conserved *tpi* and *gdh* loci. Assuming rate constancy along lineages, it would seem reasonable to expect that the relative levels of inter- and intra-lineage divergence would be similar for the different loci. As can be seen from Table 29 (p. 214), the *efl α* nt sequence-based p-distances between isolates of *G. intestinalis* and *G. ardeae* are almost double compared to the estimates based on the 18S rRNA nt sequences. However, the inter-lineage distances estimated from these two loci are almost the same. Comparison of the 18S rRNA and *gdh* data showed that the distances between *G. ardeae* and *G. intestinalis* were only 30% larger for *gdh* than for 18S rRNA, but for the inter-lineage comparisons they were up to 5-fold larger than those estimated from the 18S rRNA data. Another inconsistency can be seen in the comparison of the *gdh* and *tpi* distances, which shows that the *tpi* distances are larger for inter-lineage and species comparisons than the equivalent *gdh* distances, yet the intra-lineage variation of *tpi* is less than *gdh*. To my knowledge, this unusual type of rate inconsistency, which appears to affect all of the taxa examined and all of the loci analysed, has not been reported previously in any organism. The lack of intra-lineage polymorphism detected at these loci contrasts with the sensitivity of AE, which is highly discriminatory at this level. The possible causes for these phenomena are not known.

Table 29. Summary of uncorrected p distances (%) determined for different loci.

Locus	Intra-lineage	Inter-lineage	<i>G. intestinalis</i> vs. <i>G. ardeae</i>	<i>G. intestinalis</i> vs. <i>G. muris</i>
<i>efl</i> α	0 - 0.3	0.8 - 3.4	7.1 - 8.9	11.2 - 12.7
18S rRNA	0	0 - 3.1	15.1 - 16.5	27.8 - 28.9
<i>gdh</i>	0 - 2.5	5.8 - 14.7	18.3 - 21.5	nd
<i>tpi</i>	0 - 0.4	8.8 - 23.4	32.5 - 37.6	34.4 - 36

nd = not determined

A potential problem in using AE data to determine phylogenetic relationships, especially by cladistic methods, is that the amount of homoplasy (or noise) in the data tends to increase as the evolutionary distance between taxa increases (Richardson *et al.*, 1986). This can be illustrated by comparing the nt sequence data for *gdh* and *tpi* with the allelic profiles for each allozyme locus. In the case of GDH and TPI, isolates from both Assemblages A and B exhibited the same electrophoretic mobilities, yet the sequence data suggest that members of the Cat and Livestock lineages (which differed electrophoretically at one or both loci from Assemblages A & B; Table 12, pp. 73 - 76), were more closely related to Assemblage A (eg., Figure 67, p. 180). The underlying basis for this phenomenon (ie., character convergence versus true homology) cannot be determined because the entire *gdh* or *tpi* sequences from these isolates are not available. However, this example demonstrates the conservative nature of AE (ie., it can only detect mutations affecting the overall charge of a protein), as well as the potential conflict that may arise between characters. In practice, character conflicts are overcome by using as many characters as possible in an analysis. The effect of this can be seen from the analysis of the entire AE data set, which supports the relationships inferred by the nt sequence analyses.

5.3 Evidence for introgression between *G. muris* and *G. ardeae*

One of the reasons for analysing nt sequences from more than one locus is that the phylogeny of a gene may not reflect the phylogeny of the organism (Moritz & Hillis, 1996). Such discrepancies can be caused by the presence of paralogous genes, or by genetic exchange between species (introgression). Introgression (as defined in the glossary) can occur either through hybridisation in the case of sexual species, or recombination in the case of asexual species. Many examples of introgression can be found in bacterial systematics, especially when the characters used to determine the phylogeny are located on mobile elements such as

plasmids (eg., *Rhizobium*, Gordon *et al.*, 1995). Currently, there is no record of a sexual cycle in *Giardia*, nor has recombination between species or isolates been reported. The detection of multiple banded allozyme genotypes, suggestive of diploid heterozygotes, along with the presumed homozygotes for the same alleles (Andrews *et al.*, 1989; Meloni *et al.*, 1995), has been used to suggest that occasional recombination or sex has occurred (Meloni *et al.*, 1995). Definitive experiments to test these hypotheses have not been conducted.

Differences in the observed genetic distances between *G. muris* and *G. ardeae* deduced from the *tpi*, 18S rRNA, and *efl* α sequence data (relative to the *G. intestinalis* isolates) suggests that introgression may have occurred between the former two species. The results of AE and nt sequence analysis of 18S rRNA and *efl* α all confirm that *G. ardeae* is distinct from both *G. intestinalis* and *G. muris*. In the case of *efl* α , *G. ardeae* is separated from each of these species by similar distances, but appears to be more closely related to *G. intestinalis* based on both AE and 18S rRNA sequence analysis. On the other hand, results using the *tpi* sequence data showed *G. ardeae* to be much more similar to *G. muris* than in the previous comparisons, with the separation similar in magnitude to the distances separating the lineages of *G. intestinalis*. Alternative explanations for this result; eg., culture or PCR contamination, can be effectively excluded because the *G. ardeae tpi* sequence differed from all of the known *Giardia tpi* sequences, including those from the three *G. muris* isolates. It was not found to be similar to any of the sequences available from the GenBank database, excepting the published *G. intestinalis tpi* sequences. Further evidence that the sequence originated from a *Giardia* species was: (1) the high GC content at 3rd codon positions, which appears to be a signature of *Giardia* protein coding genes, and; (2) the lack of introns within the amplified *tpi* segment, which effectively excludes most other eukaryotes from being potential sources of the sequence.

Alternative hypotheses that might explain the *tpi* phylogeny are that the *G. muris* and *G. ardeae* sequences are derived from a paralogous *tpi* locus, or that the original *G. ardeae* isolate was established from a mix of genotypes. If paralogous copies of *tpi* exist in the *Giardia* genome, then only one appears to be expressed because only single enzyme bands are detected by AE. On the basis of Southern analysis, Mowatt *et al.* (1994) reported that *tpi* was present as a single copy in *G. intestinalis*. However, it is unlikely that Southern hybridisation would detect sequences with less than 70% sequence similarity, especially if moderate or high stringencies are used. If the *G. muris* and *G. ardeae tpi* sequences are from a paralogous

locus, then this locus would appear to be evolving at a much slower rate than the *G. intestinalis tpi* gene, at a rate on par with that of 18S rRNA and *efl* α .

The presence within the *G. ardeae* culture of a second *Giardia* species seems unlikely, although it cannot be excluded. The detailed morphological assessment of this isolate by Erlandsen *et al.* (1990a) did not reveal the presence of any other *Giardia*. However, the morphological similarities between *G. muris* and *G. ardeae* could mask the presence of a *G. muris*-like organism, especially if it was a minor contaminant. AE analysis did not detect other genotypes within the *G. ardeae* culture, although this would be expected if the genotype detected by PCR was indeed a minor contaminant. Further evidence against the presence of a *G. muris*-like genotype is that the *G. ardeae* DNA sample failed to produce a *Giardia* 18S rRNA amplification product under conditions which favour the amplification of this segment from *G. muris* DNA. The presence of a novel *G. muris*-like species would be an interesting finding, as *G. muris* have proven refractory to axenisation. An appropriate way to test these competing hypotheses would be to produce cloned strains from the existing *G. ardeae* culture, which could then be sequenced to test the mixed-genotype hypothesis, or used in Southern analyses to determine whether paralogous copies of *tpi* are present in a single clone. Additionally, genetic mapping of the *G. intestinalis*, *G. muris*, and *G. ardeae tpi* loci would be required to test the homology or paralogy of the sequences.

The finding of introgression between *G. muris* and *G. ardeae* would provide a valuable insight into the genesis of *G. ardeae*, as well as highlighting a new area of *Giardia* genetics which is worthy of investigation. If a molecular clock is assumed, then the timing of the recombination or hybridisation event would be similar to that of the initial radiation of the *G. intestinalis* lineages (based on the similar divergences of the *G. intestinalis* lineages from each other compared to the divergence between *G. muris* and *G. ardeae* from the *tpi* sequences). Furthermore, if introgression between *G. muris* and *G. ardeae* has occurred, it may provide a plausible explanation for the morphology of the median body in *G. ardeae*. In other *Giardia* species, the median body morphology is uniform between isolates and usually diagnostic for the species. For instance, *G. muris* has a rounded median body and *G. intestinalis* a claw-hammer shaped median body (Figure 4, p. 23). *G. ardeae* trophozoites, which appear to have a pyriform cell shape similar to *G. intestinalis*, appear to possess either (but not both in the same cell) rounded or claw hammer median bodies (Erlandsen *et al.*, 1990a). If the *G. ardeae* used by Erlandsen *et al.* (1990a) represent a pure culture, then the ability to produce either type of median body may lie in the differential expression of separate genes which produce

rounded or claw hammer median bodies, or in differences in the regulation of median body assembly which produce either type of morphology. The orientation of the claw hammer in *G. ardeae* is different to that of *G. intestinalis* (Erlandsen *et al.*, 1990a), which suggests differences in the assembly/synthesis of the median body between the species.

5.4 Molecular epidemiology of *G. intestinalis*

The results from both the allozyme and DNA based studies presented herein have important implications for the study of the epidemiology of *G. intestinalis*. There are three issues which have been raised as a result of this work.

The first arises from the fact that *G. intestinalis* can now be considered, almost indisputably, a species complex, and that the only lineages (species) which have been detected in other laboratories are Assemblages A and B. As a result, all of the PCR based diagnostic methods described in the literature have been designed to differentiate organisms belonging to either of these two Assemblages. In some instances, such as the 18S rRNA 3-nt signature sequence described by van Keulen *et al.* (1995), the diagnostic marker will not be specific. Examination of the data described herein (Figure 88) shows that the characteristic GCG signature sequence of the “Polish” (Assemblage A) strains is also shared by members of the Livestock lineage, as well as by *G. ardeae*. The ATC signature sequence of the “Belgian” (Assemblage B) strains is also shared with members of both Dog1 and Dog2 lineages, and with *G. psittaci*. The use of this signature sequence as an epidemiological marker would therefore provide misleading information because all *G. intestinalis* isolates (and *G. ardeae* and *G. psittaci*) would be typed as either Assemblage A or Assemblage B genotypes, and the presence of the other lineages would remain undetected.

Figure 88. The 3 nucleotide signature sequences (in bold) from the 5' end of 18S rRNA.

Assemblage A	GGAG GCG CGA
(Polish)	GGAG GCG CGA
Livestock	GGAG GCG CGA
<i>G. ardeae</i>	GGAG GCG CGA
Assemblage B	GGA ATC CGA
(Belgian)	GGA ATC CGA
Dog1	GGA ATC CGA
Dog2	GGA ATC CGA
<i>G. psittaci</i>	GGA ATC CGA
<i>G. muris</i>	GGAG TACTA

The second issue concerns the resolving power (ie., ability to differentiate between genotypes) of nt sequences, and the hazards of using conserved nt sequence data in epidemiology. None of the gene sequences examined herein were as discriminatory as AE (although AE cannot reliably determine the phylogeny of the lineages identified herein, it can readily discriminate between them). Employing nt sequences from conserved loci for molecular epidemiological purposes in the absence of a sound systematic framework can adversely affect the interpretation of data. For example, Baruch *et al.* (1996), who used *tpi* nt sequences and PCR-RFLP of *tpi* to examine the epidemiology of *G. intestinalis*, incorrectly assumed that nucleotide sequence identity for a segment of *tpi* signified genetic identity between isolates. While their approach would be ideal to examine the epidemiology of the organisms which comprise each lineage of the *G. intestinalis* species complex, the data presented herein suggests that *tpi* nt sequences will not be useful to identify the genotypes of isolates within each lineage. To illustrate this, consider the dog-derived *G. intestinalis* isolate "WOOF", which was characterised allozymically (herein) and by nt sequencing of the *tpi* locus (Baruch *et al.*, 1996). The finding of sequence identity between WOOF and human-derived isolates of *G. intestinalis* prompted Baruch *et al.* (1996) to conclude that they were genetically identical, therefore providing supportive evidence for zoonosis. If this approach were to be adopted, then the same conclusion could also be drawn from the *gdh* sequence data presented herein. In contrast, allozyme analysis found that WOOF was genetically distinct from all of the human-derived isolates within Assemblage B (Figure 23, p. 101) at approximately 25% of the loci examined. This result indicates that further prevalence and cross-transmission data would be required to determine the zoonotic status of this isolate. The major failing of the approach used by Baruch *et al.* (1996) was the absence of a sufficiently detailed and comprehensive genetic model from which to work. If a model such as the one provided by this study had been available, it would have shown that *tpi* is too conserved to examine variation within the lineages of *G. intestinalis*.

The third issue relates to the potential of host tissue or enteric flora contamination of *G. intestinalis* samples, which could affect the results of PCR-based epidemiological studies. The method which would be most affected by such contamination would be RAPD-PCR. Morgan *et al.* (1993) have already noted that RAPD-PCR would not be applicable to faecal samples because of the variability of the quality and content of the samples, and the results presented in section 4.4.3.2 show that the same concern applies to *G. intestinalis* isolates grown in suckling mice. In some cases (eg., isolates from the dog lineages), the use of the suckling mouse method is unavoidable as it is the only means by which the isolates can be

grown. The presence of contaminating non-*Giardia* DNA's would be almost impossible to monitor because of the complex and undefined nature of the RAPD reaction, which could be affected by slight changes in the relative abundance of different DNA molecules. This would effectively exclude the use of the RAPD-PCR technique for analysis of this sort of material, unless a method was developed to purify trophozoites from other micro organisms.

5.5 Molecular systematics of *G. intestinalis*

The model of phylogenetic relationships which I am proposing (Figure 87, p. 209) extends the genetic relationships described previously (Andrews *et al.*, 1989; Ey *et al.*, 1997, Appendix F; Mayrhofer *et al.*, 1995) and it provides a strong basis for redefining the systematics of the *G. intestinalis* species complex. The need for a revision of the taxonomy of this species group has long been recognised (eg., Kulda and Nohýnková, 1978; Erlandsen *et al.*, 1990a; Meloni *et al.*, 1995) and the results from this project provide, for the first time, sufficient information to begin this task. From the data presented herein, there is enough information to invoke the phylogenetic species concept. However, the naming of species without first considering any of the biological consequences is of little practical utility. It would be more realistic and useful to adopt the approach used in systematics, which is to produce a classification system which reflects natural relationships (Whitehead, 1990). In the past, this has not been possible for *G. intestinalis*, largely due to the lack of an appropriate genetic model which could be used as a starting point, but also because of the failure by many researchers to recognise the significance of the results of Andrews *et al.* (1989) and Andrews *et al.* (1992b), which originally suggested the existence of cryptic species within *G. intestinalis*, and demonstrated that culture conditions could allow the preferential growth of specific genotypes.

Apart from the methodology and techniques which have been used in this study, a critical difference between this and other studies which have examined the genetic diversity within *G. intestinalis* has been the use of isolates grown by infection of suckling mice (Mayrhofer *et al.*, 1992). Although not tested exhaustively, there is evidence to suggest that the use of suckling mice to propagate *G. intestinalis* isolates allows a wider range of genotypes to grow than does axenic culture (Mayrhofer *et al.*, 1995; Monis *et al.*, 1997, Appendix D). The main support for this conclusion comes from the evidence that only two broad genetic groups, equivalent to Assemblage A ($\equiv$ Polish, Homan *et al.*, 1992; groups 1 & 2, Nash *et al.*, 1992; group 17 & 18, Meloni *et al.*, 1995) and Assemblage B ($\equiv$ Belgian,

Homan *et al.*, 1992; group 3, Nash *et al.* 1992; groups 1 - 14, Meloni *et al.* 1995), have been detected among axenic isolates of *G. intestinalis*. The only isolates external to these groupings published by other workers (Meloni *et al.*, 1992) are samples which have either not been cultured (eg., BAD1), or samples which have been propagated in suckling mice (eg., BAR1). Attempts to axenically culture *G. intestinalis* from infected dogs have had a very low success rate (eg., Meloni *et al.*, 1987), compared to the high success rate using the suckling mouse model (see Monis *et al.*, 1997, Appendix E). A consequence of this sample bias is that an oversimplified genetic model is currently being used to study the biology of *Giardia* (eg., Baruch *et al.*, 1996).

The basis for much of the biological heterogeneity exhibited by different *G. intestinalis* isolates has not been understood to date. It is therefore a worthwhile exercise to re-examine what is known, and to determine how these independent biological data fit with the proposed phylogeny of the *G. intestinalis* species complex derived from the research described in this study.

5.6 Host-preference of *Giardia*: re-incarnation of a taxonomic marker?

One of the criteria originally used in the taxonomy of *Giardia* species was presumed host specificity. This resulted in a plethora of *Giardia* species being described (reviewed by Kulda & Nohýnková, 1978). The criterion of host specificity has since been rejected on the basis of cross-infectivity studies (eg., Hewlett *et al.*, 1982). Instead, the morphological criteria of Filice (1952) have been adopted.

A variety of cross-transmission trials have shown that some human *G. intestinalis* isolates are not rigidly host specific, and that the cross-infectivity of the isolates can be highly variable (Table 5, pp. 29 - 30). Few of these studies used genetically characterised isolates of *Giardia* and none of them validated the genetic authenticity of the isolates following experimental infection. The results of these experiments were varied, ranging from a failure to cross-infect to the establishment of low or high levels of infection, depending on the combination of the host species and the *Giardia* isolate tested (Table 5, pp. 29 - 30). In most cases, larger numbers of cysts ($> 10^5$) were required to establish infections than are likely to cause natural infections. Such numbers of cysts ($> 10^5$) are only likely to occur in faeces or raw sewage (Jakubowski, 1984), but not in surface waters (Ongerth *et al.*, 1995). The only report of unambiguous animal to human cross-transmission has been by Majewska (1994),

who successfully infected a human volunteer using 10^5 trophozoites from an axenic *G. intestinalis* isolate derived from a Gambian giant pouched rat (GGPRP-114).

The variable success observed in such cross-transmission trials begins to make sense when viewed against the major genetic divisions that are now apparent within the *G. intestinalis* species complex. Although it has been known for many years that *G. intestinalis* isolates are genetically heterogeneous, efforts to demonstrate a connection between known genotypes and differences in cross-infectivity have been unproductive (eg., de Jonckheere *et al.*, 1990; Nash & Mowatt, 1992). Nonetheless, differences in host-preference provide a very plausible explanation for the conflicting results of cross-infection experiments. For instance, Hewlett *et al.* (1982) reported successful cross-infection of dogs with cysts and trophozoites of human origin, whereas Woo and Paterson (1986) reported the opposite, ie., failure to cross-infect. Chaudhuri *et al.* (1991) suggested that the genetic heterogeneity of *Giardia* isolates from humans may cause differences in clinical manifestations and host responses, although this does not seem to have been pursued.

Based on the results presented herein, host-preference should not be discounted as a taxonomic marker. For instance, the isolates comprising the Dog1 and Livestock lineages appear to exhibit distinct host-preferences (Ey *et al.*, 1997; Monis *et al.*, 1997; see Appendices E & F), as do some of the genetic groups within these lineages; eg., the 3 groups comprising either sheep, cattle or pig-derived isolates within the Livestock lineage, and the genetic groups II and IV, which exclusively comprise human-derived isolates (data presented herein and Mayrhofer *et al.*, 1995). Furthermore, the majority of the animal-derived *G. intestinalis* isolates within Assemblages A and B were in genetic groups which do not contain human-derived *G. intestinalis*, also raising the possibility of host-preference within these groups. Currently, the numbers of isolates from the different lineages or genetic groups are too small to rigorously test this hypothesised host-preference, with the possible exception of the isolates comprising the Dog1 and Livestock lineages. However, it is worth noting that all of the human-derived *G. intestinalis* isolates (>150 in total characterised by Homan *et al.* (1992); van Keulen *et al.* (1995); Mayrhofer *et al.* (1995); Meloni *et al.* (1995)) grown either *in vitro* or in suckling mice belong exclusively to either Assemblages A or B.

The hypothesis of host-preference can be tested in a number of ways. One approach would be to use the data presented herein to develop markers to examine the prevalence of the *G. intestinalis* lineages in different hosts (eg., using the *gdh* PCR-RFLP system described in

section 4.5.4.5 as a diagnostic assay). The use of PCR-based techniques would allow the direct characterisation of samples (environmental and faecal), and would avoid the pitfalls inherent in culturing isolates (eg., selection of specific genotypes). However, the application of PCR-based techniques would have to be used with caution, because: (1) they may exhibit biases if mixes of genotypes are present (eg., preferential amplification of DNA from a specific genotype in an environmental sample comprising many different genotypes); and, (2) they may not be able to detect or discriminate all *Giardia*. Infectivity trials using genetically defined isolates of *G. intestinalis* could also be used to test the predicted host-preference of an isolate. The demonstration of host-preference would provide an additional independent taxonomic marker for species delineation within *Giardia*.

5.7 Correlation between growth characteristics and genotype

Other biological parameters which can be used to provide additional information about the systematics of *G. intestinalis* are the growth characteristics and metabolisms of isolates. This may overlap with host-preference, which is likely to be affected to some degree by the metabolic requirements of the isolate. A limitation of this approach would be obtaining data on the metabolism of isolates which cannot be adapted to axenic culture (eg., *G. muris* and isolates comprising the Dog1 lineage). Metabolic differences have been reported by Binz *et al.* (1992) and Hall *et al.* (1992), who found that the isolates BAH12 (Assemblage B genotype) and Portland 1 (Assemblage A genotype) had very different *in vitro* growth requirements and energy metabolisms.

Some information is available concerning the comparative growth rates of isolates from the different lineages. A comparison of the *in vitro* growth rates of the Portland 1 (Assemblage A) and BAH12 (Assemblage B) *G. intestinalis* isolates found the mean generation time of the Assemblage A isolate to be only half that of the Assemblage B isolate (Binz *et al.*, 1992). Similarities in growth characteristics have been observed between Assemblage A and Livestock genotypes (Koudela *et al.*, 1991; J. Kulda, pers. comm.). Isolates from the Cat, Dog1 and Dog2 lineages have not been cultured axenically, and numerous attempts at establishing dog-derived *G. intestinalis* in axenic culture have met with little success. Meloni *et al.* (1987) failed to establish axenic cultures from 24 dog-derived isolates of faecal or intestinal origin. A low success rate of 1/7 was also reported by Kulda (see: Monis *et al.*, 1997, Appendix E). Only 4 dog-derived axenic isolates have been characterised world-wide (3 herein), and these are either Assemblage A or B genotypes. *G.*

intestinalis isolated from Norway rats also fail to grow axenically (Kulda & Nohýnková, 1976; Meloni *et al.*, 1989). In contrast, a high success rate has been reported for the establishment of human-derived (58/64) and livestock-derived (18/21) isolates of *G. intestinalis* in axenic culture (J. Kulda, pers. comm.). A consequence of growth rate differences between isolates is that competition occurs when they are co-cultured, resulting in overgrowth by the faster growing genotype (Andrews *et al.*, 1992b; Thompson *et al.*, 1996).

Inoculation of suckling mice with mixed samples comprising Assemblage A and B genotypes has demonstrated that Assemblage B genotypes have a growth advantage in this host (Andrews *et al.*, 1992b). Comparisons of the yields of Cat and Livestock genotypes from infected suckling mice indicate that they grow well in this environment (G. Mayrhofer & J. Mackrill, pers. comm.). In contrast, isolates comprising the Dog1 or Dog2 lineages grew erratically in suckling mice, perhaps indicating much larger metabolic differences between these lineages and the others (discussed in Monis *et al.*, 1997, Appendix E).

5.8 *Giardia* ecology

The genetic heterogeneity exhibited by isolates comprising the *G. intestinalis* species complex is not fully compatible with the proposal of Tibayrenc *et al.* (1991) that *G. intestinalis* has a clonal population structure. According to Thompson and Meloni (1993), clonal competition/selection should limit genetic heterogeneity within clonal lineages. This does not appear to be the case for many of the lineages examined herein. In a study of human-derived *G. intestinalis* isolated from residents of an Aboriginal community, Thompson and Meloni (1993) proposed that a possible mechanism for the maintenance of genetic heterogeneity was exposure of trophozoites to sub-lethal concentrations of chemotherapeutics, which might interfere with any competitive interactions between clones. This can occur when drug therapy regimes are not followed through to completion, and is apparently a frequent occurrence in these communities (Thompson & Meloni, 1993). Although this hypothesis has been experimentally verified using an *in vitro* model (Thompson *et al.*, 1996), it cannot account for the genetic heterogeneity exhibited by isolates of *G. intestinalis* derived from animals, eg., among isolates comprising the Dog1 lineage.

The mechanisms responsible for the generation or maintenance of genetic heterogeneity could alternatively be examined by studying the ecology of *Giardia* infections. Unlike the *in vitro* experimental model used by Thompson *et al.* (1996), the host gut contains

a microbial flora and fauna which may include other parasites. These are likely to compete with any infecting *G. intestinalis* trophozoites, and affect the dynamics of clonal competition. Another possibility worth consideration is that different genotypes of *G. intestinalis* niche partition. The different lineages identified in this study appear to differ in host-preference and growth (or at least culture) requirements. Adaptation to specific ecological niches within the host would help to explain how different lineages could co-exist within a single host (eg., the mixed infections from humans described by Andrews *et al.*, 1992b), thus maintaining the genetic diversity. This might also explain the apparent broad host range of some genotypes (eg., genetic group I isolates) and the narrow host range of others (isolates comprising the Dog1 lineage). Such ecological data would also provide useful systematic information.

5.9 Implications for zoonosis

One of the most controversial and topical issues concerning *Giardia* has been the potential for zoonotic transmission of *G. intestinalis*, and the ability of animals to act as reservoirs for human infection. As discussed in section 1.9.4, there is a high prevalence rate of *G. intestinalis* in domestic pets and livestock, as well as in wild animals which inhabit watersheds. One of the main assumptions of the prevalence studies has been that *G. intestinalis* is a single species. This is clearly not true based upon previous studies (eg., Andrews *et al.*, 1989) and the data from this study. While human derived *G. intestinalis* can be infective to animals under certain conditions (Table 5, pp. 29-30), it would be an oversimplification to then extend this behaviour to include natural infections. This is illustrated by the cross-infection studies of Erlandsen *et al.* (1988). They found that muskrats, which have a high prevalence rate for *Giardia* (having *G. microti* cyst morphology) infection, could only be infected by either beaver- or human-derived *Giardia* when high cyst doses were used. On the other hand, beavers could be readily infected by human-derived *Giardia*, but were not susceptible to infection by muskrat-derived *Giardia*. These results show that prevalence *per se* is not a good indicator of potential reservoirs for human infection, and that isolates can vary markedly in host infectivity.

The identification herein of potentially dog-specific lineages of *G. intestinalis* may explain the findings of epidemiological studies which found no correlation between *Giardia* infections in pets and their owners (Asano *et al.*, 1991; Castor & Lindqvist, 1990; Pospisilova & Svobodova, 1992). Regardless of the cross-infectivity potential of an isolate, epidemiological data on genotypes from each lineage should ultimately indicate whether

animals play a significant role as a zoonotic reservoir for human giardiasis. This type of study requires a sound systematic framework and reliable taxonomic characters to differentiate the cryptic species comprising *G. intestinalis* and allow the detection of mixed infections.

5.10 Comparison with the systematics of other asexual organisms

The significance of the evolutionary divisions detected within the *G. intestinalis* species complex can be put into perspective by comparing the phylogeny of *G. intestinalis* with the systematic relationships among other groups of asexual organisms. The examples chosen were the enteric bacterial genera *Salmonella* and *Escherichia*, and the amoeba genus *Naegleria*. These have all been well characterised by AE (eg., Adams *et al.*, 1989; Selander *et al.*, 1991) or DNA sequence analysis (Lawrence *et al.*, 1991; Nelson *et al.*, 1991).

In the case of the *gdh* locus, the lineages comprising the *G. intestinalis* species complex are separated by levels of divergence that are similar to those that separate different genera of bacteria. For example, a 13% sequence divergence separates the Assemblage A and Dog1 lineages, compared to a 17% sequence divergence between *Escherichia coli* and *Salmonella typhimurium* for homologous regions of the *gdh* locus (using GenBank sequence accession numbers X58832 and M24021). A similar observation can be made for the 18S rRNA sequence data, with 3% sequence divergence separating Assemblage A and Dog1, and 2 - 3% sequence divergence separating *E. coli* and *S. typhimurium* (Lawrence *et al.*, 1991).

A comparison of the phylogeny of *Escherichia* and *Salmonella* species inferred from *gapA* (Figure 89, p. 226) and the *G. intestinalis* *gdh* phylogeny (Figure 67, p. 180) supports the contention that the lineages within the *G. intestinalis* species complex are distinct species. The structure of the trees from the two different phylogenies are very similar. Most of the bacterial species are separated by long branch lengths with relatively little intra-species variation, and a similar trend is evident among the *G. intestinalis* lineages. Lawrence *et al.* (1991) recognised that the genus *Escherichia* was not monophyletic, and proposed that a molecular systematic approach be used in bacterial taxonomy. With the exception of *Salmonella* sp. group V, the *Salmonella* species/sub-species were monophyletic and more closely related to each other compared to the observed relationships among the *Escherichia* species. The patterns of relationships among the lineages of the *G. intestinalis* species complex more closely resemble those between distinct bacterial species and genera than those between more closely related bacterial species or sub-species.

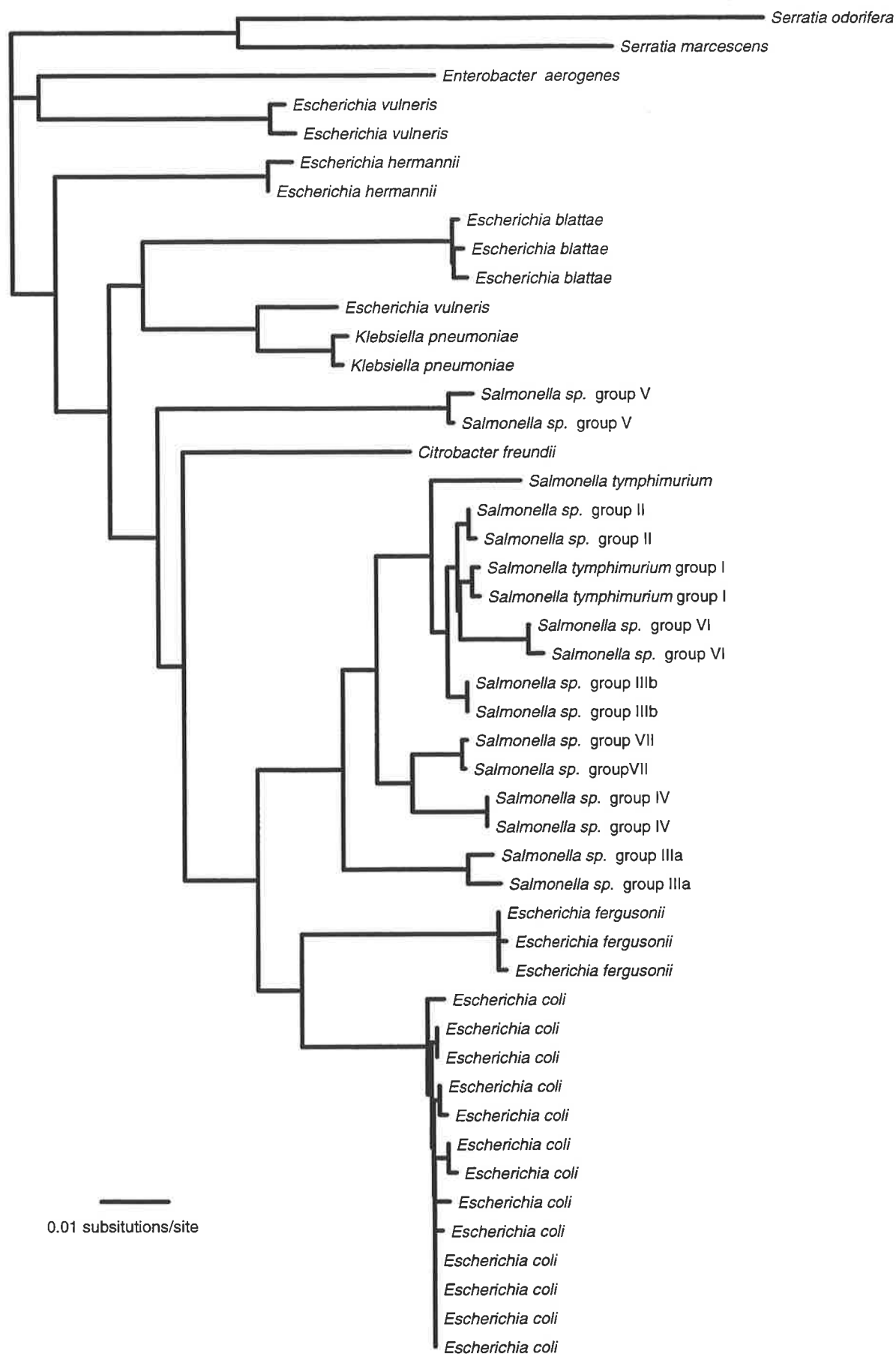


Figure 89. Evolutionary relationships of Gram negative enteric bacterial species inferred by phylogenetic analysis (Tamura Nei distances with Neighbour Joining) of glyceraldehyde-3-phosphate dehydrogenase (*gapA*) nt sequences (obtained from GenBank, described in Lawrence *et al.* (1991) or Nelson *et al.* (1991)).

The uniqueness of each *G. intestinalis* lineage can be further emphasised by comparison of the AE results presented herein with similar studies by Selander *et al.* (1991) on the evolutionary genetics of *Salmonella* and Adams *et al.* (1989) on species identification within the genus *Naeglaria*. In the case of *Salmonella*, fixed genetic differences at 20 - 30% of loci were observed for intra-species comparisons (eg., between different *Salmonella typhi* strains), while polymorphism within *Naeglaria* species generally occurred at less than 21% fixed differences. These levels of intra-specific variation are comparable to the observed intra-lineage heterogeneity detected within the *G. intestinalis* species complex (25 - 47% between genetic groups). The divergence between species of *Salmonella* or *Naeglaria* (55 - 80% and 52 - 90% fixed differences respectively), is similar to that observed between the *G. intestinalis* lineages (56 - 84% fixed differences). In the case of *Salmonella*, differences in biological parameters such as host range and pathogenicity occur at the subspecies level (Selander *et al.*, 1991). Similar levels of divergence (as assessed by AE) also separate different species of other protozoan parasites (eg., *Leishmania*, Andrews *et al.*, 1988).

Chapter 6

CONCLUSIONS

The primary aim of this study was to construct a framework around which to build the systematics of *G. intestinalis*. This has been achieved by the determination of the phylogenetic relationships of *G. intestinalis* isolates from different host species, using both contemporary and traditional molecular genetic techniques, namely DNA sequence analysis and allozyme electrophoresis. The molecular data generated were used to produce the model of phylogenetic relationships presented in Figure 87 (p. 210).

The existence of distinct lineages of *G. intestinalis* is compatible with the hypothesis of Andrews *et al.* (1989) that *G. intestinalis* comprises a species complex. Examination of the available biological data for different *G. intestinalis* isolates (characterised herein) indicates that the lineages consistently differ from each other in terms of their apparent host-preference and growth characteristics. Furthermore, the evolutionary distances separating these lineages were similar to those separating morphologically distinct species of protozoa (eg., *Leishmania*, *Naeglaria*) and genera of enteric bacteria. These levels of divergence are commonly associated with differences in virulence, drug-resistance, host-preference, etc. commonly observed among other asexual microorganisms (Adams *et al.*, 1989).

The data presented herein provide strong evidence to support the proposal that the separate lineages of *G. intestinalis* represent distinct species. The data support this in terms of the phylogenetic species concept, and the sparse biological data currently available supports this division from a systematic perspective. The data also make sense in an evolutionary context. Given the apparent age of the Diplomonadida and *Giardia*, which has been suggested by analyses of highly conserved genes (eg., rRNA genes, van Keulen *et al.*, 1993; elongation factor genes, Hashimoto *et al.*, 1994, 1995), it seems reasonable to expect a high level of taxic diversity similar to that observed in other ancient lineages of organisms (eg., *Naeglaria*). The divisions between the lineages appear to be biologically meaningful, thereby producing a natural classification if they are recognised as separate species. Finally, the phylogenetic species identified herein fulfil the criteria of Meloni *et al.* (1995), who proposed that a lineage of related isolates of *Giardia* that behave similarly would represent a “natural” group worthy of formal taxonomic recognition.

The formal description of these lineages will require adherence to the appropriate code of practice for nomenclature, and will therefore require consultation with “older and wiser heads” more knowledgeable in this area. Because these species are morphologically cryptic and different species may infect the same host (eg., Livestock and Assemblage A genotypes have both been recovered from sheep; Assemblage A and Assemblage B genotypes both infect humans), it is difficult to envisage how priority using previous taxonomic descriptions (eg., Kulda & Nohýnková, 1978) of *Giardia* species could be applied. A practical approach would be to use the host-preference of a species as a guide to naming it, although difficulties may arise if different species (eg., Dog1 and Dog2 lineages) have the same host preference. Some of the sub-groupings within the lineages may also warrant formal recognition because they appear to exhibit host-preferences (eg., the pig-derived group of isolates within the Livestock lineage). However, further detailed epidemiological studies regarding the prevalence of the genotypes comprising the lineages and genetic groups will be required to further test the cohesiveness of these groupings, and to confirm that host-preference is a valid taxonomic marker.

The combination of traditional and contemporary molecular systematic techniques has provided a framework for the further investigation of the systematic relationships of isolates within *Giardia*. The results presented herein demonstrate both the utility of a holistic approach to biological investigation, and the importance of understanding the systematic relationships among organisms when attempting to interpret biological data. Attempting to investigate the biology of an organism without first determining its systematic relationships can provide misleading results. As some of the examples discussed herein demonstrate, it is not possible to examine the prevalence or epidemiology of a pathogen using morphological criteria or molecular techniques if these methods cannot discriminate between the species present. Similarly, it is not possible to properly investigate issues such as zoonosis, or select markers with diagnostic utility, without a full understanding of the genetic relationships or extent of genetic heterogeneity within the organism of interest. The identification and naming of species as biologically relevant units forms an integral part of systematics, which is central to all of the biological sciences. After all;

“no biological study is worth the paper it is written on unless informed, if not by systematics, then by its subdiscipline taxonomy, while systematics as a whole provides the unifying structure around which all biological exploration must be built”

Whitehead, 1990

Chapter 7

FUTURE DIRECTIONS

A number avenues may be pursued as a logical extension of the research conducted in this study. Parts of the phylogeny of the *G. intestinalis* species complex remain unresolved, and additional genes will need to be analysed to clarify these relationships. This includes the development of markers to examine the intra-species phylogenetic relationships of isolates, which could only be resolved by AE. It would also be of biological interest to determine the position of the Rat lineage within the current DNA-based phylogenies, as well as the relative positions of *G. psittaci* and *G. microti*, which only differ from the *G. intestinalis* species complex by a single morphological character in each case. Another aspect requiring clarification is the origin of the *G. ardeae tpi* sequence. This could be due to either introgression, gene paralogy or contamination by a novel species of *Giardia*, each of which would be of interest.

The identification of the cryptic species uncovered herein raises a number of issues which need to be addressed. Firstly, very few *Giardia* isolates with the *G. intestinalis* morphology have been characterised ultrastructurally. The identification of polymorphic ultrastructural characters would have little practical value in a diagnostic context, but they could further strengthen the taxonomic ranking of the lineages and provide additional phylogenetic characters. Similarly, detailed investigation of the growth characteristics and metabolisms of isolates from each lineage should provide additional systematic information, as well as provide an insight into the mechanisms responsible for host-preference.

Perhaps most importantly, in pragmatic terms, issues regarding zoonoses and epidemiology can now be addressed because an appropriate phylogenetic model is now available. This should facilitate the selection of appropriate markers for epidemiological studies (eg., strain variation using AE), or the prevalence of particular *Giardia* species, which could be investigated using the *gdh* PCR-RFLP system described in section 4.5.4.6. The application of PCR techniques in epidemiology should enable the direct typing of samples, thereby minimising bias by the selection of particular genotypes via different methods of culture. The population structure of *Giardia* can now also be addressed. This will require more extensive sampling of isolates from each phylogenetic species.

GLOSSARY

Term	Definition	Definition from:
anagenesis	accumulated change within a lineage	Avise (1994)
anology	the component of biological similarity due to convergence from different ancestors	Avise (1994)
apomorph	a derived or newly evolved character state, not present in the common ancestor of the taxa under study	Avise (1994)
autapomorph	a character state unique to a single taxon	Avise (1994)
clade	a branch of a phylogenetic tree containing the set of all organisms descended from a particular common ancestor which is not an ancestor of any non-member of the group - usually used within the context of cladistics	1
cladistics	method of classification of living organisms that makes use of lines of descent only, rather than phenotypic similarities, to deduce evolutionary relationships, and which groups organisms strictly on the relative recency of common ancestry. Cladistic methods of classification only permit taxa in which all the members share a common ancestor who is also a member of the taxon and which include all the descendants of that common ancestor, regardless of their different degrees of divergence	1
cladogenesis	branch splitting component of a phylogeny	Avise (1994)
cladogram	tree-like diagram showing the evolutionary descent of any group of organisms - used to describe a dendrogram constructed using cladistic methodology	1
dendrogram	general term for any tree-like diagram representing a phylogeny (from the Greek <i>dendron</i> meaning tree)	Marcus (1991)
homology	the component of biological similarity due to descent from common ancestry	Avise (1994)
homoplastic (homoplasy)	the component of overall similarity due to convergence from unrelated ancestors	Avise (1994)
housekeeping genes	genes that are expressed in most cell types and which are concerned with basic metabolic activities common to all cells.	1
ingroup	An assumed monophyletic group , usually comprising the taxa of primary interest.	Hillis <i>et al.</i> (1996)
introgression	the gradual diffusion of genes from the gene pool of one species into another where there is some hybridisation between them as a result of incomplete genetic isolation	1
monophyletic group	an evolutionary assemblage that includes a common ancestor and all of its descendants	Avise (1994)
orthology	synonymous with homolgy, common ancestry by speciation	Miyamoto & Cracraft (1991)

outgroup	One or more taxa assumed to be phylogenetically outside the ingroup.	
parapatric	adjacent but non-overlapping distributions	Hillis <i>et al.</i> 1996
paraphyletic group	an artificial assemblage that includes a common ancestor and some but not all of its descendants	Avise, (1994)
parology	ancestry by gene duplication	Miyamoto & Cracraft (1991)
parsimony	a criterion for estimating a parameter from observed data based on the principle of minimising the number of events needed to explain the data.	Hillis <i>et al.</i> (1996)
patristic	similarity due to common ancestry the component of overall similarity due to shared ancestry	1 Avise (1994)
phenetic	classification purely based on similarities in phenotypic characters, not necessarily reflecting relationships by evolutionary descent	1
phenogram	overall resemblance between organisms a branching tree diagram or dendrogram, but it records at each branching level only the overall numerical value of similarity or difference at that level. The phenogram was not designed to depict hypotheses of shared ancestry.	Avise (1994) Marcus (1991)
phyletic	a branching diagram that links entities by estimates of overall similarity. Usually constructed using UPGMA cluster analysis. a phylum or a major branch of an evolutionary lineage; a group of species related to each other by common descent; a line of direct descent	Hillis <i>et al.</i> (1996) 1
phylogeny	the evolutionary history and line of descent of a species or higher taxonomic group	1
phylogram	A tree that depicts inferred historical relationships among entities. Differs from a cladogram in that the branches are drawn proportional to the amount of inferred character change.	Hillis <i>et al.</i> (1996)
plesiomorph	an ancestral character state (one present in the common ancestor of the taxa under study)	Avise (1994)
polyphyletic	an artificial assemblage derived from two or more distinct ancestors	Avise (1994)
polytomy	syn. multifurcation A node in a tree that connects more than three branches.	Hillis <i>et al.</i> (1996)
sympatric	occurring in the same place	Hillis <i>et al.</i> (1996)
symplesiomorph	an ancestral character state shared by two or more descendant taxa	Avise (1994)
synapomorph	a derived character state shared by two or more descendant taxa	Avise (1994)

systematics	the scientific discipline that not only classifies and names diversity of the organic world (through taxonomy), but seeks to understand the processes that produce and sustain that diversity. Its results are classifications which aim to show historically derived relationships between organisms at all taxonomic levels and in so doing, to point to genetic lineages, since these underlie all other forms of relationship. The method is comparative and the materials are all living and extinct organisms and all biological sources of characters that illuminate relationships and their history.	Whitehead (1990)
	the science of comparative biology the primary goal of systematics is to describe taxic diversity and to reconstruct the hierarchy, or phylogenetic relationships, of those taxa.	Miyamoto & Cracraft (1991)
	the study of the identification, taxonomy and nomenclature of organisms, including the classification of living things with regard to their natural relationships and the study of variation and the evolution of taxa	1
	Taxonomy is generally taken as a synonym of systematics or biosystematics and is traditionally divided into three parts: (i) classification.....(ii) nomenclature.....and (iii) identification of unknown organisms	Vandamme <i>et al.</i> (1996)
taxon (plural taxa)	any formal unit in the taxonomic system, at any level or rank. It is used as a substitute for any group of organisms, for example, species, genus, or family, when one does not want to specify rank.	Marcus (1991).
taxonomy	the theory and practice of classifying organisms into taxa within a hierarchy	Whitehead (1990).
	The science of classification, especially of living organisms, which arranges organisms into hierarchical groupings	1
	the analysis of an organism's characteristics for the purpose of classification	1

1. Henderson's Dictionary of Biological Terms, 11th edition, Eleanor Lawrence (ed.) Longman House, London

Appendix A

Prevalence of *Giardia intestinalis*

Appendix A: Prevalence of *G. intestinalis*.

Reference	Location	Host
		domestic dogs
Arashima <i>et al.</i> , 1992	Japan kennels	18.6% of 366
Arashima <i>et al.</i> , 1992	Japan general public	9.3% of 1811
Arashima <i>et al.</i> , 1992	Japan research labs	4.9% of 42
Asano <i>et al.</i> , 1991	Japan	16.4% of 354
Bemrick, 1961	Minnesota, USA	7.6% of 2083
Burgu, 1979	Turkey	3.37% of 237
Castor & Lindqvist, 1990	Sweden	21% of 62
Catcott, 1946	Ohio	17.7% of 113
Choquette, 1950	Canada	9% of 155
Collyer <i>et al.</i> , 1992	Brisbane, Australia strays	7.2% of 400
Collyer <i>et al.</i> , 1992	Brisbane, Australia, pets	0 of 54
Craig, 1948	USA	15% of 160
Davies & Hibler, 1979	Colorado, USA	12.8% of 78
Diaz Saez <i>et al.</i> , 1987	Granada, Spain	10.2% of 157
Epe, 1993	Germany 1984 - 1991	6% of 3329
Johnston & Gasser, 1993	Melbourne, Australia, 1991 - 1992	2.3 - 14.2%
Hahn <i>et al.</i> , 1988	USA pups	35.9% of 117
Jungmann <i>et al.</i> , 1986	Berlin, Germany	3.5% of 141
Kasprzak <i>et al.</i> , 1983	Poland 1978-1980	<1% of 724
Kirkpatrick, 1988	England	7.2% of 2294
Lewis, 1988	Lethbridge, Canada	33.3% of 81
Lewis, 1988	Calgary, Canada	11.9% of 310
Lewis, 1988	Red Deer, Canada	6.9% of 246
Lewis, 1988	Edmonton, Canada	5.4% of 368
Meloni <i>et al.</i> , 1993	Kimberly, Australia	17% of 182
Meneses Marcel <i>et al.</i> , 1994	Mexico	10.5% of 86
Milstein & Goldsmid, 1995	Hobart	14.5% of 55
Nolan & Smith, 1995	Pennsylvania, USA	4.7% of 8077
Pfeiffer & Supperer, 1976	Austria pups < 6 mths	53% of 45
Pfeiffer & Supperer, 1976	Austria pups 6 - 24 mths	10% of 59
Pfeiffer & Supperer, 1976	Austria dogs	1.7% of 116
Pospisilova & Svobodova 1992	Czechoslovakia	7.3% of 123
Rahman, 1990	Penang, Malaysia	21.9% of 237
Rahman, 1990	Malaysia, 6-12 mths	39% of 41
Rahman, 1990	Malaysia 2-4 yrs	8.1% of 86
Seiler <i>et al.</i> , 1983	Switzerland in 1981	6.5% of 662
Seiler <i>et al.</i> , 1983	<1 year	16%
Seiler <i>et al.</i> , 1983	1 - 3 years	7%
Sogayar <i>et al.</i> , 1984	Batucatu, Brazil	6.7% of 450
Stehr Green 1987	Atlanta puppies	34% of 143
Sugano <i>et al.</i> , 1989	Japan breeding kennel	72% of 25
Swan & Thompson 1986	Australia	21% of 333
Sykes & Fox, 1989	London 1984-86	14.5% of 255
Tonks <i>et al.</i> , 1991	Palmerston North, New Zealand	7.7% of 142
Tonks <i>et al.</i> , 1991	Hamilton, New Zealand	24.7% of 81
Wallis <i>et al.</i> , 1984	Alberta, Canada	4.5% of 22
Winsland <i>et al.</i> , 1989	Essex	20% of 80
Wolff & Eckert, 1979	Switzerland 1978 - 1979	1.6% of 2414
Zimmer & Burrington, 1986	USA dog kennel	87% of 76

Reference	Location	Host
		coyotes
Davies & Hibler, 1979	Colorado, USA	5.9% of 34
		cats
Asano <i>et al.</i> , 1991	Japan	8.8% of 34
Davies & Hibler, 1979	Colorado, USA	25% of 4
Epe, 1993	Germany 1984 - 1991	2.4% of 1147
Kasprzak <i>et al.</i> , 1983	Poland 1978 - 1980	0 of 88
Kirkpatrick, 1988	England	3.5% of 452
Nolan & Smith, 1995	Pennsylvania, USA	2.4% of 2000
Seiler <i>et al.</i> , 1983	Switzerland in 1981	5.3% of 94
Swan & Thompson, 1986	Australia	14% of 226
Tonks <i>et al.</i> , 1991	Palmerston North, New Zealand	6.7% of 120
Tonks <i>et al.</i> , 1991	Hamilton, New Zealand	2.8% of 72
Winsland <i>et al.</i> , 1989	Essex	3.5% of 20
Wolff & Eckert, 1979	Switzerland 1978 - 1979	0.6% of 927
		beavers
Isaac-Renton <i>et al.</i> , 1987	British Columbia, Canada	14.7% of 299
Davies & Hibler, 1979	Colorado, USA	18% of 244
Erlandsen & Bemrick, 1988	New England, USA	17%
Erlandsen <i>et al.</i> , 1990	Minnesota, USA, kill trapped	9.2% of 662
Erlandsen <i>et al.</i> , 1990	Minnesota, USA, live trapped	13.7% of 302
		cattle
Buret <i>et al.</i> , 1990	Alberta, Canada	10.4%
Buret <i>et al.</i> , 1990	Alberta, Canada, calves	27.7%
Davies & Hibler, 1979	Colorado, USA	10.3% of 58
Meneses Marcel <i>et al.</i> , 1994	Mexico, calves	18.2% of 137
Xiao, 1994	Cuba, calves	11% of 506
Xiao, 1994	India, calves	51.6% of 157
Xiao, 1994	Italy, calves	19.3% of 247
Xiao, 1994	Soviet Union, calves	34.3% of 417
Xiao, 1994	Switzerland, calves	26.6% of 815
		humans
Davies & Hibler, 1979	Colorado, USA	6.3% of 32
Diaz Saez <i>et al.</i> , 1987	Granada, Spain children	8.6% of 175
Engbaek, 1978	Denmark, children	5.1% of 411
Islam <i>et al.</i> , 1983	Bangladesh, India, mothers	82% of 33
Islam <i>et al.</i> , 1983	Bangladesh, India, infants	42% (?33)
Joyce <i>et al.</i> , 1996	Kenya, Africa, children < 5yrs	31%
Loewenson <i>et al.</i> , 1986	Zimbabwe	17.4% of 1813
Meloni <i>et al.</i> , 1993	Kimberly, Australia, children	32% of 361
Pickering <i>et al.</i> , 1984	Houston, USA	21% & 26%
Pospisilova & Svobodova, 1992	Czechoslovakia	13.7% of 1140
Wallis <i>et al.</i> , 1984	Alberta, Canada	1.2% of 85
Wolff & Eckert, 1979	Switzerland 1973-1978	3.6% of 9832
Woo & Paterson, 1986	Ontario, Canada children 2 - 5yrs, 1982	8% of 97 all asymptomatic
Woo & Paterson, 1986	Ontario, Canada children 2 - 5yrs, 1983	6% of 147

Reference	Location	Host
Wallis <i>et al.</i> , 1984	Alberta, Canada	<i>Clethrionomys gapperi</i> 95% of 21
Grant & Woo, 1978 Wallis <i>et al.</i> , 1984	Ontario, Canada Alberta, Canada	<i>Peromyscus maniculatus</i> 98% of 49 10% of 50
Erlandsen & Bemrick, 1988 Erlandsen <i>et al.</i> , 1990b Erlandsen <i>et al.</i> , 1990b	New England, USA Minnesota, USA, kill trapped Minnesota, USA, live trapped	muskrats 100% 36.6% of 790 95.9% of 219
Wallis <i>et al.</i> , 1984	Alberta, Canada	Castor canadensis 3% of 58
Grant & Woo, 1978 Wallis <i>et al.</i> , 1984	Ontario, Canada Alberta, Canada	<i>Microtus pennsylvanicus</i> 98.8% of 326 33% of 6
Buret <i>et al.</i> , 1990 Buret <i>et al.</i> , 1990 Meneses Marcel <i>et al.</i> , 1994 Xiao, 1994 Xiao, 1994	Alberta , Canada Alberta , Canada, lambs Mexico Switzerland, lambs New Zealand	sheep 17.7% 35.6% 7% of 115 29.8% of 382 6% of 100
Meneses Marcel <i>et al.</i> , 1994 Xiao, 1994	Mexico Ohio, USA	pigs 3% of 105 7.1% of 238
Xiao, 1994	Romania	goats 50% of 200
Xiao, 1994 Xiao, 1994	Kentucky, USA Ohio, USA, foals	horses 17.2% of 29 32.4% of 37
Meneses Marcel <i>et al.</i> , 1994	Mexico	rabbits 0 of 45
Meneses Marcel <i>et al.</i> , 1994	Mexico	chickens 0 of 87
Diaz Saez <i>et al.</i> , 1987	Granada, Spain	rats 14.7% of 34

Appendix B

**Allele frequencies (%) for
operational taxonomic units of
*Giardia***

Appendix B: Allele frequencies (%) for operational taxonomic units of *Giardia*.

Locus	all	Assemblage A	Livestock	Cat	Assemblage B	Dog	Ad-148	Rat	<i>G. muris</i>
Acp	f	11	100	100	0	0	0	100	100
Acp	e	89	0	0	0	0	0	0	0
Acp	d	0	0	0	0	0	100	0	0
Acp	c	0	0	0	0	100	0	0	0
Acp	b	0	0	0	90	0	0	0	0
Acp	a	0	0	0	10	0	0	0	0
Ald	h	0	0	0	0	0	0	0	100
Ald	g	20	0	0	0	0	0	0	0
Ald	f	30	0	100	88	0	0	100	0
Ald	e	0	0	0	6	0	0	0	0
Ald	d	50	100	0	0	100	0	0	0
Ald	c	0	0	0	6	0	0	0	0
Ald	b	0	0	0	0	0	90	0	0
Ald	a	0	0	0	0	0	10	0	0
Dia-1	i	12	0	50	0	0	0	0	0
Dia-1	h	0	0	50	0	0	0	0	0
Dia-1	g	0	0	0	0	0	0	0	100
Dia-1	f	13	0	0	100	0	20	0	0
Dia-1	e	0	0	0	0	0	80	0	0
Dia-1	d	12	0	0	0	0	0	0	0
Dia-1	c	63	37	0	0	100	0	0	0
Dia-1	b	0	38	0	0	0	0	0	0
Dia-1	a	0	25	0	0	0	0	100	0
Dia-2	e	12	0	0	0	0	0	0	0
Dia-2	d	0	25	50	0	0	0	0	100
Dia-2	c	88	0	50	25	100	70	100	0
Dia-2	b	0	25	0	75	0	20	0	0
Dia-2	a	0	50	0	0	0	10	0	0
Enol	g	0	100	0	0	0	0	0	0
Enol	f	0	0	0	0	0	0	0	100
Enol	e	100	0	100	0	100	0	0	0
Enol	d	0	0	0	100	0	0	0	0
Enol	c	0	0	0	0	0	0	100	0
Enol	b	0	0	0	0	0	89	0	0
Enol	a	0	0	0	0	0	11	0	0
Est	h	0	0	0	0	0	0	0	50
Est	g	0	0	0	0	0	0	0	50
Est	f	0	0	0	0	0	20	0	0
Est	e	0	0	0	10	0	0	100	0
Est	d	0	12	0	30	0	0	0	0
Est	c	100	75	0	60	0	80	0	0
Est	b	0	13	0	0	0	0	0	0
Est	a	0	0	100	0	100	0	0	0

Appendix B (cont). Allele frequencies (%) for operational taxonomic units of *Giardia*.

Locus	all	Assemblage A	Livestock	Cat	Assemblage B	Dog	Ad-148	Rat	<i>G. muris</i>
Gdh	e	0	25	0	0	0	0	0	0
Gdh	d	0	0	0	0	0	0	0	100
Gdh	c	50	75	0	5	100	70	100	0
Gdh	b	50	0	0	95	0	30	0	0
Gdh	a	0	0	100	0	0	0	0	0
Got-1	f	0	0	0	0	0	0	0	0
Got-1	e	0	25	0	6	0	0	0	100
Got-1	d	0	0	0	44	0	0	100	0
Got-1	c	10	37	100	33	0	0	0	0
Got-1	b	90	38	0	17	100	0	0	0
Got-1	a	0	0	0	0	0	100	0	0
G6pd	g	0	0	0	0	100	75	0	0
G6pd	f	0	0	0	50	0	25	50	0
G6pd	e	0	0	0	50	0	0	50	0
G6pd	d	10	0	0	0	0	0	0	0
G6pd	c	90	0	0	0	0	0	0	100
G6pd	b	0	100	0	0	0	0	0	0
G6pd	a	0	0	100	0	0	0	0	0
Gpt	h	0	0	0	0	0	0	0	100
Gpt	g	20	0	0	0	0	0	0	0
Gpt	f	60	12	0	0	0	0	0	0
Gpt	e	20	88	50	0	0	0	100	0
Gpt	d	0	0	50	90	0	0	0	0
Gpt	c	0	0	0	10	0	0	0	0
Gpt	b	0	0	0	0	100	90	0	0
Gpt	a	0	0	0	0	0	10	0	0
Hk	g	10	0	0	60	0	0	0	0
Hk	f	0	0	0	40	0	0	0	0
Hk	e	90	0	100	0	0	0	0	0
Hk	d	0	100	0	0	100	0	0	0
Hk	c	0	0	0	0	0	100	0	0
Hk	b	0	0	0	0	0	0	0	75
Hk	a	0	0	0	0	0	0	100	25
Mdh	e	0	0	0	0	100	80	0	0
Mdh	d	0	0	0	0	0	20	0	0
Mdh	c	0	0	100	0	0	0	0	100
Mdh	b	0	25	0	100	0	0	50	0
Mdh	a	100	75	0	0	0	0	50	0

Appendix B (cont). Allele frequencies (%) for operational taxonomic units of *Giardia*.

Locus	all	Assemblage A	Livestock	Cat	Assemblage B	Dog	Ad-148	Rat	<i>G. muris</i>
Me	j	0	0	0	0	0	0	0	0
Me	i	0	0	0	0	0	30	0	25
Me	h	0	0	0	0	0	20	0	0
Me	g	0	0	0	0	100	10	0	75
Me	f	50	0	0	0	0	20	100	0
Me	e	40	100	100	33	0	0	0	0
Me	d	10	0	0	0	0	0	0	0
Me	c	0	0	0	28	0	20	0	0
Me	b	0	0	0	33	0	0	0	0
Me	a	0	0	0	6	0	0	0	0
Ndpk	c	0	0	25	0	0	0	0	100
Ndpk	b	0	0	75	0	0	0	0	0
Ndpk	a	100	100	0	100	100	100	100	0
Np-1	h	0	0	0	0	0	0	0	100
Np-1	g	0	0	0	0	0	0	100	0
Np-1	f	0	0	0	0	0	14	0	0
Np-1	e	0	0	0	0	0	86	0	0
Np-1	d	0	0	0	75	0	0	0	0
Np-1	c	0	0	0	12	0	0	0	0
Np-1	b	22	100	100	0	100	0	0	0
Np-1	a	78	0	0	13	0	0	0	0
Np-2	h	0	0	0	0	0	0	0	100
Np-2	g	0	0	0	0	0	0	100	0
Np-2	f	0	0	0	0	0	14	0	0
Np-2	e	0	0	0	0	0	86	0	0
Np-2	d	10	0	0	90	0	0	0	0
Np-2	c	10	100	100	0	0	0	0	0
Np-2	b	10	0	0	10	100	0	0	0
Np-2	a	70	0	0	0	0	0	0	0
Pep-A	d	0	0	0	0	0	10	0	0
Pep-A	c	0	12	0	0	0	80	100	100
Pep-A	b	78	88	100	11	100	0	0	0
Pep-A	a	22	0	0	89	0	10	0	0
Pep-D	d	20	0	100	62	0	0	100	0
Pep-D	c	80	100	0	38	0	0	0	0
Pep-D	b	0	0	0	0	100	80	0	100
Pep-D	a	0	0	0	0	0	20	0	0
Pgam	d	89	87	0	0	0	0	0	0
Pgam	c	0	13	0	5	0	0	0	0
Pgam	b	11	0	100	95	100	100	100	0
Pgam	a	0	0	0	0	0	0	0	100

Appendix B (cont). Allele frequencies (%) for operational taxonomic units of *Giardia*.

Locus	all	Assemblage A	Livestock	Cat	Assemblage B	Dog	Ad-148	Rat	<i>G. muris</i>
6Pgd	d	0	0	0	5	0	0	0	0
6Pgd	c	0	0	0	5	100	20	0	100
6Pgd	b	0	0	0	90	0	80	0	0
6Pgd	a	100	100	100	0	0	0	100	0
Pgk	c	0	0	0	33	0	0	100	0
Pgk	b	100	100	100	67	0	0	0	100
Pgk	a	0	0	0	0	100	100	0	0
Pgm	e	30	0	0	0	0	0	0	0
Pgm	d	20	0	100	0	0	0	0	0
Pgm	c	50	100	0	69	0	0	0	0
Pgm	b	0	0	0	31	0	0	100	100
Pgm	a	0	0	0	0	100	100	0	0
Tpi	d	0	0	25	0	0	0	0	100
Tpi	c	0	67	75	10	100	70	0	0
Tpi	b	0	0	0	0	0	30	0	0
Tpi	a	100	33	0	90	0	0	100	0

Appendix C

**Aligned *Giardia gdh* nucleotide
sequences**

Appendix C: Aligned *Giardia* *gdh* nucleotide sequences.

(Isolates marked # were sequenced using primers #1051 & #1052 only)

72

Portland1	ATC	TTC	CGC	GTG	CCC	TGG	ATG	GAT	GAC	GCT	GGA	CGC	ATC	AAC	GTC	AAC	CGC	GGC	TTC	CGT	GTC	CAG	TAC	AAC
Ad-1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Bris-136 [#]	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-159 [#]	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Alpaca [#]	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
P2mer [#]	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
P15	...	...	...	...	...	...	...	...	...	..C	..G	...	...	...	...	..T	...	...	...	...	..T	...	...	...
J17	...	...	...	...	...	...	...	...	...	..C	..G	..T	...	...	...	..T	...	...	...	...	..T	...	...	...
G247 [#]	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-165	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-152 [#]	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-119 [#]	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-23	...	...	...	...	...	...	...	..C	..T	..C	..G	...	..T	...	...	...	...	...	...	...	...	...	...	...
Ad-142	...	...	...	...	...	...	...	..C	..T	..C	..G	...	..T	...	...	...	...	...	...	...	...	...	...	...
Ad-136	...	...	..T	..C	..A	...	...	...	..T	..C	..G	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-137	...	...	..T	..C	..A	...	...	...	..T	..C	..G	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-141	...	...	..T	..C	..A	...	...	...	..T	..C	..G	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-148	...	...	..T	..T	..A	...	...	..C	..T	...	...	...	...	...	...	...	..T	..T	...	...	...	...	...	...
Ad-147	...	...	..T	..C	..A	...	...	...	..T	..C	..G	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-7	..G	...	..T	..C	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
Ad-19	NNN	NNN	NNN	..C	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
Ad-28	..G	...	..T	..C	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
Ad-52	..G	...	..T	..C	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
Ad-121	..G	...	..T	..C	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
Amc-3 [#]	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
LD-20	..G	...	..T	..C	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
WOOF	..G	...	..T	..C	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
D47	..G	...	..T	..C	...	...	...	..C	...	..C	...	A..	...	...	...	...	...	...	...	...	A..	...	...	...
CH105	..G	..G	...	..C	...	...	...	..C	...	C..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
GGPRP	..G	..G	..NT	..C	...	...	...	..C	A..	..C	...	N..	...	N..	...	...	...	...	...	...	A..	...	...	...
LSLP-116	..G	...	..T	..C	...	...	...	..C	...	T..C	...	A..	T..	N..	T..	...	...	...	...	...	A..	...	...	...
CP-117	..G	...	..T	..C	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
SP-115	..G	...	..T	..C	...	...	...	..C	...	..C	...	N..	N..	...	...	...	...	...	...	...	A..	...	...	...
Ad-156	..G	...	G..T	..C	A..	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	A..	...	...	...
<i>G. ardeae</i>	..G	..G	...	..T	..T	G..	...	...	...	..C	..C	GAG	...	C..A	...	..T	...	...	...	..C	...	...	...	...

Appendix C (cont.). Aligned *gdh* nucleotide sequences.

144

Portland1	TCT	GCT	CTC	GGC	CCC	TAC	AAG	GGT	GGC	CTC	CGC	TTC	CAC	CCC	TCT	GTC	AAT	CTT	TCG	ATT	CTC	AAG	TTC	CTC
Ad-1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Bris-136	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NN.
Ad-159	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NN.
Alpaca	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NN.	...
P2mer	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NN.	...
P15	..C	...	...	..A	...	...	...	..C	..G	...	...	...	...	...	...	...	...	...	..A	..C	...	...	...	...
J17	..C	...	...	..A	...	...	...	..C	..G	...	...	...	...	...	...	...	...	...	..A	..C	...	...	...	...
G247	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NN.
Ad-165	N.C	N..	...	N.A	...	...	...	..C	..G	...	...	...	...	...	...	...	...	...	..A	..C	...	...	...	...
Ad-152	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-119	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-23	..C	..C	...	..G	...	...	...	..C	...	...	...	...	...	...	..C	..T	...	..C	..C	..C	...	...	...	...
Ad-142	..C	..C	...	..G	...	...	...	..C	...	...	...	...	...	...	..C	..T	...	..C	..C	..C	...	...	...	...
Ad-136	..C	...	...	..G	...	...	...	..C	...	..T	...	...	...	...	...	...	..C	...	..A	..C	...	...	...	..T
Ad-137	..C	...	...	..G	...	...	...	..C	...	..T	...	...	...	...	...	...	..C	...	..A	..C	...	...	...	..T
Ad-141	..C	...	...	..G	...	...	...	..C	...	..T	...	...	...	...	...	...	..C	...	..A	..C	...	...	...	..T
Ad-148	...	...	...	..G	...	...	..A	...	...	..T	...	...	...	...	...	...	..C	...	...	..C	..T	...	...	..T
Ad-147	..C	...	...	..G	...	...	...	..C	...	..T	...	...	...	...	...	...	..C	...	..A	..C	...	...	...	..T
Ad-7	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
Ad-19	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
Ad-28	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
Ad-52	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
Ad-121	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
Amc-3	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	..C	...	..C	...	...	...
LD-20	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
WOOF	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
D47	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
CH105	..C	...	...	..G	...	...	..A	...	..T	...	...	...	...	...	..G	...	..C	..C	...	..C	..T	...	...	...
GGPRP	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
LSLP-116	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
CP-117	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
SP-115	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
Ad-156	..C	...	...	..G	...	...	...	...	..T	...	...	...	...	...	...	...	..C	..C	...	..C	..T	...	...	...
<i>G ardeae</i>	..C	..TC	...	..G	...	...	...	..C	..G	...	...	...	..T	..G	A..G	...	..C	..G	...	..C	...	...	...	...

Appendix C (cont.). Aligned *gdh* nucleotide sequences.

288

Portland1	GAC	CCA	AAG	GGC	AAG	TCC	GAC	AAC	GAG	GTC	ATG	CGC	TTC	TGC	CAG	TCC	TTC	ATG	ACC	GAG	CTC	CAG	AGG	CAC
Ad-1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Bris-136	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-159	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Alpaca	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P2mer	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P15	...	..G	...	...	...	..T	...	...	...	...	...	..T	...	...	...	...	...	...	..T	...	..T	...	...	...
J17	...	..G	...	...	...	..T	...	...	...	...	...	..T	...	...	...	...	...	...	..T	...	..T	...	...	...
G247	...	..G	...	...	...	..T	...	...	...	...	...	..T	...	...	...	...	...	...	..T	...	..T	...	...	...
Ad-165	...	..G	...	...	...	..T	...	...	...	...	...	..T	...	...	...	...	...	...	..T	...	..T	...	...	...
Ad-152	...	...	...	...	...	..T	...	...	...	...	...	..T	..C	...	...	...	...	...	..T	...	..T	...	...	...
Ad-119	...	..G	...	...	...	..T	...	...	...	..T	..T	...	...	...	...	...	...	...	..T	...	..T	..N	...	...
Ad-23	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-142	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-136	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-137	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-141	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-148	...	..C	...	...	...	..T	...	..T	...	..T	...	...	...	...	..A	...	..T	...	...	...	..T	...	...	...
Ad-147	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-7	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
Ad-19	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
Ad-28	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
Ad-52	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
Ad-121	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
Amc-3	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	..N	...	...	...	...	...
LD-20	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
WOOF	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
D47	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
CH105	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
GGPRP	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
LSLP-116	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
CP-117	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
SP-115	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
Ad-156	..T	..T	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
<i>G ardeae</i>	...	..T	...	..G	...	...	...	..T	...	..T	...	...	...	...	...	...	...	...	..G	...	..T	...	C.C	...

Appendix C (cont.). Aligned *gdh* nucleotide sequences.

360

Portland1	GTC	GGC	GCC	GAC	ACT	GAC	GTT	CCT	GCC	GGC	GAC	ATC	GGC	GTC	GGC	GCC	CGC	GAG	ATC	GGG	TAC	CTG	TAC	GGA
Ad-1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Bris-136	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-159	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Alpaca	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P2mer	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P15	..T	..G	..T	...	...	..T	...	...	...	...	...	...	...	...	...	..T	...	...	...	..T	...	..T	...	...
J17	..T	..G	..T	...	...	..T	...	...	...	...	...	...	...	...	...	..T	...	...	...	..T	...	..T	...	...
G247	..T	..G	..T	...	...	..T	...	...	...	...	...	...	...	...	...	..T	...	...	...	..T	...	..T	...	...
Ad-165	..T	..G	..T	...	...	..T	...	...	...	...	...	...	...	...	...	..N	...	...	...	..T	...	..T	...	...
Ad-152	...	..G	..T	...	...	..T	...	...	...	...	...	...	...	...	...	..T	...	..G	...	...	...	..T	...	...
Ad-119	..T	..G	..T	...	...	..T	...	...	...	...	..C	...	...	...	...	..T	...	...	...	..T	...	..T	...	...
Ad-23	...	...	...	..T	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	..T	..C	...	...	...	..G
Ad-142	...	...	...	..T	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	..T	..C	...	...	...	..G
Ad-136	...	...	..T	...	..C	...	...	...	..T	...	...	..T	..T	...	...	..T	...	...	...	..C	...	...	..TT	..G
Ad-137	...	...	..T	...	..C	...	...	...	..T	...	...	..T	..T	...	...	..T	...	...	...	..C	...	...	..TT	..G
Ad-141	...	...	..T	...	..C	...	...	...	..T	...	...	..T	..T	...	...	..T	...	...	...	..C	...	...	..TT	..G
Ad-148	...	...	..T	...	...	...	...	...	..T	..T	...	..T	...	...	..A	...	...	...	...	..T	...	...	..TT	..C
Ad-147	...	...	..T	...	..C	...	...	...	..T	...	...	..T	..T	...	...	..T	...	...	...	..C	...	...	..TT	..G
Ad-7	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
Ad-19	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
Ad-28	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
Ad-52	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
Ad-121	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
Amc-3	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
LD-20	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	..N	..T	..T	...	..TT	...
WOOF	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
D47	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
CH105	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
GGPRP	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
LSP-116	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
CP-117	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
SP-115	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
Ad-156	...	..G	..T	...	..C	...	...	...	..T	...	..T	..T	...	...	...	..GT	...	...	...	..T	..T	...	..TT	...
<i>G ardeae</i>	A..	..T	...	...	..G	...	..C	..C	..T	...	...	...	...	...	..G	...	...	...	...	...	..T	..C	..T	..C

Appendix C (cont.). Aligned *gdh* nucleotide sequences.

432

Portland1	CAG	TAC	AAG	CGC	CTG	AGG	AAC	GAG	TTC	ACA	GGC	GTC	CTC	ACA	GGC	AAG	AAC	GTC	AAG	TGG	GGC	GGG	TCT	TTC
Ad-1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Bris-136	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-159	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Alpaca	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P2mer	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P15	...	...	...	...	...	...	...	...	T	G	...	...	...	G	...	...	...	...	...	...	...	...	C	...
J17	...	...	...	...	...	...	...	...	T	G	...	...	...	G	...	A	...	...	...	...	...	...	C	...
G247	...	...	...	...	...	...	...	...	T	G	...	...	...	G	...	...	...	...	...	...	...	...	C	...
Ad-165	...	...	...	...	...	...	...	A	T	G	...	...	...	G	...	...	...	...	...	...	...	...	C	...
Ad-152	...	...	...	...	...	...	...	...	T	G	...	...	...	G	...	A	...	...	...	...	...	...	C	...
Ad-119	...	T	...	...	...	...	...	C	T	G	...	...	...	G	...	NA	...	...	...	...	...	...	C	...
Ad-23	...	...	...	...	...	...	...	...	...	C	...	...	...	G	...	...	...	...	...	...	...	...	C	...
Ad-142	...	...	...	...	...	...	...	...	...	C	...	...	...	G	...	...	...	...	...	...	...	...	C	...
Ad-136	...	...	...	...	C	...	...	...	...	...	G	...	...	T	T	...	...	...	...	...	...	T	C	C...
Ad-137	...	...	...	...	C	...	...	...	...	...	G	...	...	T	T	...	...	N	...	...	...	T	C	C...
Ad-141	...	...	...	...	C	...	...	...	...	...	G	...	...	T	T	...	...	N	...	...	...	T	C	C...
Ad-148	...	...	...	...	C	...	...	...	...	...	A	T	...	T	...	...	...	A	...	...	...	A	...	C...
Ad-147	...	...	...	...	C	...	...	...	...	...	G	...	...	T	T	...	...	N	...	...	...	T	C	C...
Ad-7	...	T	...	...	C	...	...	...	T	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
Ad-19	...	T	...	...	C	...	...	...	T	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
Ad-28	...	T	...	...	C	...	...	...	T	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
Ad-52	...	T	...	...	C	...	...	...	T	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
Ad-121	...	T	...	...	C	...	...	...	T	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
Amc-3	...	T	...	...	C	...	...	...	...	G	...	...	...	G	...	...	...	A	...	...	T	...	...	C...
LD-20	...	N	...	...	C	...	...	...	...	G	...	...	...	N	...	...	...	A	...	...	T	...	...	C...
WOOF	...	T	...	...	C	...	...	...	T	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
D47	...	T	...	...	C	...	...	...	T	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
CH105	...	T	...	...	C	...	...	...	T	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
GGPRP	...	T	...	...	C	...	...	...	...	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
LSLP-116	...	T	...	...	C	...	...	...	...	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
CP-117	...	T	...	...	C	...	...	...	...	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
SP-115	...	T	...	...	C	...	...	...	...	G	...	...	...	G	...	...	...	A	...	...	...	...	...	C...
Ad-156	...	T	...	...	C	...	...	...	...	G	...	T	...	G	...	...	...	A	GCN	...	...	...	...	C..T
G ardeae	...	...	...	...	T	GC	...	TCC	...	GAG	T	...	...	G	...	...	T	TA	GCN	...	T	..T	..C	C...

Appendix C (cont.). Aligned *gdh* nucleotide sequences.

504

Portland1	ATC	AGG	CCG	GAG	GCC	ACG	GGC	TAT	GGC	GCT	GTC	TAC	TTC	CTG	GAG	GAG	ATG	TGC	AAG	GAC	AAC	AAC	ACT	GTG
Ad-1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-2	...	...	...	...	..T	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Bris-136	...	...	...	...	..T	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-159	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Alpaca	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P2mer	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P15	...	..C	...	...	...	..A	..A.	...	...	...	...	...	...	...	..A	...	...	...	...	...	...	...	...	..A
J17	...	..C	...	...	...	..A	..AN	...	...	...	...	...	...	...	..A	...	...	...	...	...	...	...	...	..A
G247	...	...	...	...	...	..A	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..A
Ad-165	...	...	...	...	...	..A	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..A
Ad-152	...	...	...	...	...	..A	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..A
Ad-119	...	...	...	...	...	..A	...	...	...	...	...	...	...	...	...	..N	...	...	...	...	...	...	...	..A
Ad-23	...	...	...	...	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	..C	...
Ad-142	...	...	...	...	...	...	...	..C	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	..C	...
Ad-136	...	...	..A	...	...	..C	..A	...	...	...	...	...	...	..C	...	...	...	...	...	...	...	...	..C	A..A
Ad-137	...	...	..A	...	...	..C	..A	...	...	...	...	...	...	..C	...	...	...	...	...	...	...	...	..C	A..A
Ad-141	...	...	..A	...	...	..C	..A	...	...	...	...	...	...	..C	...	...	...	...	...	...	...	...	..C	A..A
Ad-148	...	...	...	...	...	...	...	..T	..C	...	...	...	...	..T	...	...	...	...	...	...	...	...	..C	A..A
Ad-147	...	...	..A	...	...	..C	..A	...	...	...	...	...	...	..C	...	...	...	...	...	...	...	...	..C	A..A
Ad-7	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
Ad-19	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
Ad-28	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
Ad-52	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
Ad-121	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	..N.	...	...	..T	...	...	..C	..A
Amc-3	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
LD-20	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
WOOF	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
D47	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
CH105	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
GGPRP	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
LSLP-116	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
CP-117	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
SP-115	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
Ad-156	...	..A	..A	...	...	..A	..G	...	..A	...	...	...	...	...	...	...	...	...	...	..T	...	...	..C	..A
<i>G ardeae</i>	...	C..C	..C	...	...	...	...	...	..T	..C	..G	...	..A.	..C	..A	...	...	...	...	...	...	GG.	..G.	CC.

Appendix C (cont.). Aligned *gdh* nucleotide sequences.

576

Portland1	ATC	AGG	GGT	AAG	AAC	GTC	CTT	CTT	TCT	GGC	TCC	GGC	AAC	GTT	GCC	CAG	TTT	GCT	TGC	GAG	AAG	CTC	ATT	CAG
Ad-1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-2	...	...	...	...	...	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Bris-136	...	...	...	...	...	...	..C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-159	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Alpaca	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P2mer	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-23	...	...	..C	...	...	...	..C	...	..C	...	..T	..G	...	...	...	...	..C	...	...	...	...	...	...	...
Ad-142	...	...	..C	...	...	...	..C	...	..C	...	..T	..G	...	...	...	...	..C	...	...	...	...	...	...	...
P15	...	...	..C	...	...	...	..C	...	..C	...	...	...	..C	..G	..C	..C	...	...	..T	...	...	...	..C	...
J17	...	...	N..N	...	...	...	..C	N..	...	...	...	..N	...	N..C	..T	..N	...	..N	..T	..N	...	N..	N..	...
G247	...	...	..C	...	...	...	..C	...	..C	...	...	...	..C	..G	..A	...	...	..T	...	...	...	...	..C	...
Ad-165	...	...	..C	...	...	...	..C	...	..C	...	...	...	..C	..T	..A	...	...	..T	...	...	...	...	..C	...
Ad-152	...	...	..C	...	...	...	..C	...	..C	...	...	...	..C	..T	..A	...	...	..T	...	...	...	...	N..	...
Ad-119	...	...	..C	...	...	...	..C	...	..C	...	...	...	..C	..T	..A	...	...	..T	...	...	...	...	..C	...
Ad-136	...	...	...	...	...	...	..C	..C	..C	..G	...	...	...	...	...	...	..C	..G	...	...	...	...	..C	...
Ad-137	...	...	...	...	...	...	..C	..C	..C	..G	...	...	...	...	...	...	..C	..G	...	...	...	...	..C	...
Ad-141	...	...	...	...	...	...	..C	..C	..C	..G	...	...	...	...	...	...	..C	..G	...	...	...	...	..C	...
Ad-148	...	...	..C	...	...	...	..G	..C	...	..T	..T	..A	...	..C	..T	..A	..C	..G	...	...	..A	...	..C	...
Ad-147	...	...	...	...	...	...	..C	..C	..C	..G	...	...	...	...	...	...	..C	..G	...	...	...	...	..C	...
Ad-7	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
Ad-19	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
Ad-28	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
Ad-52	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
Ad-121	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
Amc-3	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	...	...	..AC	..G	...	...	...	...	..C	...
LD-20	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
WOOF	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
D47	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
CH105	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
GGPRP	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
LSLP-116	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
CP-117	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
SP-115	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
Ad-156	...	...	..C	...	...	...	..C	..C	...	...	..T	...	...	...	..T	...	..AC	..G	...	...	...	...	..C	...
<i>G ardeae</i>	TC.	C..T	..G	...	..CT	A..	..C	..C	..G	...	..T	..G	...	..C	...	...	..C	..A	...	C..	...	...	..AC	G..

Appendix C (cont.). Aligned *gdh* nucleotide sequences.

648

Portland1	CTC	GGC	GCA	AAG	GTC	CTC	ACC	TTC	TCA	GAC	TCC	AAC	GGG	ACC	ATT	GTC	GAC	AAG	GAC	GGG	TTC	AAC	GAG	GAG
Ad-1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-2	..T	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..T	...	...	...	...	...
Bris-136	..T	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..NN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-159	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..N	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Alpaca	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
P2mer	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-23	...	...	..G	...	...	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-142	...	...	..G	...	...	...	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
P15	...	..G	...	...	...	..T	...	..T	..T	..C	...	..N	..A	...	...	...	...	...	...	...	..T	..NN	NNN	NNN
J17	...	..G	..N	...	...	..T	...	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
G247	...	...	..G	...	...	..T	...	..T	...	...	...	..T	..A	...	..N	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-165	...	...	..G	...	...	..T	...	..T	...	...	...	..T	..A	...	...	...	..T	...	...	..T	..T	...	..A	...
Ad-152	...	...	...	...	...	..T	...	..T	...	..C	..NN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-119	...	..N	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-136	...	...	...	...	...	...	...	...	..T	...	...	...	..A	...	..C	...	...	...	..T	..C	...	...	...	...
Ad-137	...	...	...	...	...	...	...	...	..T	...	...	...	..A	...	..C	...	...	...	..T	..C	...	...	...	...
Ad-141	...	...	...	...	...	...	...	...	..T	...	...	...	..A	...	..C	...	...	...	..T	..C	...	...	...	...
Ad-148	..A	...	...	..A	..G	..T	..A	...	..T	...	..T	...	..A	...	..C	...	..T	...	..T	..C	...	...	...	...
Ad-147	...	...	...	...	...	...	...	...	..T	...	...	...	..A	...	..C	...	...	...	..T	..C	...	...	...	...
Ad-7	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..T	..C	...	...	...	...
Ad-19	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..T	..C	...	...	...	...
Ad-28	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..T	..C	...	...	...	...
Ad-52	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..T	..C	...	...	...	...
Ad-121	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..T	..C	...	...	...	...
Amc-3	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
LD-20	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..T	..C	...	...	...	...
WOOF	...	..T	..G	...	...	...	...	...	..G	...	...	..T	...	..T	..C	...	..T	...	..T	..C	...	...	...	...
D47	...	..T	..G	...	...	...	...	...	..G	...	...	..T	...	..T	..C	...	..T	...	..T	..C	...	..NN	NNN	NNN
CH105	...	..T	..G	...	...	...	...	...	..G	...	...	..T	...	..T	..C	...	..T	...	..T	..C	...	...	...	...
GGPRP	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..T	..C	...	..NN	NNN	NNN
LSLP-116	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..T	..C	...	..NN	NNN	NNN
CP-117	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..N	NNN	NNN	NNN	NNN	NNN
SP-115	...	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..A	..T	..C	...	NNN	NNN
Ad-156	..T	..T	..G	...	...	...	...	...	..G	...	...	...	...	..T	..C	...	..T	...	..T	..C	...	...	...	...
<i>G ardeae</i>	...	..T	..C	CGC	..T	A..	T..	...	..C	...	..G	...	..C	..A	G..C	A..	...	G..	..G	..C	...	...	...	...

Appendix C (cont.). Aligned *gdh* nucleotide sequences.

690

Portland1	AAG	CTG	GCC	CAC	CTC	ATG	TAC	CTC	AAG	AAC	GAG	AAG	CGT	GGG
Ad-1	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Ad-2	...	...	..T	...	...	...	...	...	...	...	...	...	..C	...
Bris-136	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-159	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Alpaca	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
P2mer	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-23	...	...	A..	T..	...	.A.	...	...	...	...	...	..A	..C	...
Ad-142	...	...	A..	T..	...	.A.	...	...	...	...	...	..A	..C	...
P15	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
J17	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
G247	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-165	...	...	.A.	...	...	.A.	..T	...	...	...	..A	...	...	...
Ad-152	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-119	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-136	...	..T	...	...	A..	.A.	..T	...	...	...	...	...	..C	.CT
Ad-137	...	..T	...	...	G..	.A.	..T	...	...	...	...	...	..C	.CT
Ad-141	...	..T	...	...	A..	.A.	..T	...	...	...	...	...	..C	.CT
Ad-148	..A	..T	A..T	...	...	.A.	...	...	...	...	...	...	...	..T
Ad-147	...	..T	...	...	G..	.A.	..T	...	...	...	...	...	..C	.CT
Ad-7	...	...	...	...	...	..A	C..	...	...	..T	...	...	..C	...
Ad-19	...	...	...	...	...	..A	C..	...	...	..T	...	...	..C	...
Ad-28	...	...	...	...	...	..A	C..	...	...	..T	...	...	..C	...
Ad-52	...	...	...	...	...	..A	C..	...	...	..T	...	...	..C	...
Ad-121	...	...	...	...	...	..A	C..	...	...	..T	...	...	..C	...
Amc-3	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
LD-20	...	...	...	...	...	..A	C..	...	...	..T	...	...	..C	...
WOOF	...	...	...	...	...	..A	C..	...	...	..T	...	...	..C	...
D47	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
CH105	...	...	...	...	...	...	C..	...	...	..T	NNN	NNN	NNN	NNN
GGPRP	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
LSLP-116	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
CP-117	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
SP-115	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN
Ad-156	..A	...	...	...	...	...	C..	...	...	..T	...	..N	NNN	NNN
<i>G ardeae</i>	...	..C	...	...	G..G	...	...	A..	...	...	...	...	..C	..C

Appendix D

Molecular genetic analysis of *Giardia intestinalis* isolates at the glutamate dehydrogenase locus

**P.T. Monis, G. Mayrhofer, R.H. Andrews,
W.L. Homan, L. Limper and P.L. Ey**

Parasitology (1996), **112**, 1-12

Monis, P.T., Mayrhofer, G., Andrews, R.H., Homan, W.L., Limper, L., and Ey, P.L., (1996) Molecular genetic analysis of *Giardia intestinalis* isolates at the glutamate dehydrogenase locus.
Parasitology, v. 112 (1), pp. 1-12.

NOTE:

This publication is included in the print copy of the thesis held
in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1017/S0031182000065021>

Appendix E

**Novel lineages of *Giardia intestinalis*
identified by genetic analysis of
organisms isolated from dogs in
Australia**

**P.T. Monis, R.H. Andrews, G. Mayrhofer,
J. Mackrill, J. Kulda, J.L. Isaac-Renton
and P.L. Ey**

Parasitology, accepted, July 1997

Monis, P.T., Andrews, R.H., Mayrhofer, G., Mackrill, J., Kulda, J., Isaac-Renton, J.L., and Ey, P.L., (1998) Novel lineages of *Giardia intestinalis* identified by genetic analysis of organisms isolated from dogs in Australia. *Parasitology*, v. 116 (1), pp. 7-19.

NOTE:

This publication is included in the print copy of the thesis held
in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1017/S0031182097002011>

Appendix F

Genetic Analysis of *Giardia* from Hoofed Farm Animals Reveals Artiodactyl-Specific and Potentially Zoonotic Genotypes

**Peter L. Ey, Mandana Mansouri,
Jaroslav Kulda, Eva Nohýnková,
Paul T. Monis, Ross H. Andrews
and Graham Mayrhofer**

Journal of Eukaryotic Microbiology,
accepted, July 1997

Ey, P.L., Mansouri, M., Kulda, J., Nohynkova, E., Monis, P.T., Andrews, R.H., and Mahrhofer., (1997) Genetic analysis of *Giardia* from hoofed farm animals reveals artiodactyl-specific and potentially zoonotic genotypes.
Journal of Eukaryotic Microbiology, v. 44 (6), pp. 626-635.

NOTE:

This publication is included in the print copy of the thesis held
in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1111/j.1550-7408.1997.tb05970.x>

BIBLIOGRAPHY

- Adam, R.D. (1991). The biology of *Giardia* spp. *Microbiological Reviews*, **55**: 706-732.
- Adam, R.D. (1992). Chromosome-size variation in *Giardia lamblia*: The role of rDNA repeats. *Nucleic Acids Research*, **20**: 3057-3061.
- Adam, R.D., Nash, T.E. & Wellems, T.E. (1988). The *Giardia lamblia* trophozoite contains sets of closely related chromosomes. *Nucleic Acids Research*, **16**: 4555-4567.
- Adam, R.D., Nash, T.E. & Wellems, T.E. (1991). Telomeric location of *Giardia* rDNA genes. *Molecular and Cellular Biology*, **11**: 3326-3330.
- Adams, M., Andrews, R.H., Robinson, B., Christy, P., Baverstock, P.R., Dobson, P.J. & Blackler, S.J. (1989). A genetic approach to species criteria in the amoeba genus *Naegleria* using allozyme electrophoresis. *International Journal for Parasitology*, **19**: 823-834.
- Adamson, M.L. (1986). Modes of transmission and evolution of life histories in zooparasitic nematodes. *Canadian Journal of Zoology*, **64**: 1375-1384.
- Aggarwal, A., Romans, P., de la Cruz, V.F. & Nash, T.E. (1988). Conserved sequences of the HSP gene family in *Giardia lamblia*. pp.173-175, in: *Advances in Giardia research*. (P.M. Wallis and B.R. Hammond eds.) Calgary, Alberta, Canada.
- Alonso, R.A. & Peattie, D.A. (1992). Nucleotide sequence of a second alpha giardin gene and molecular analysis of the alpha giardin genes and transcripts in *Giardia lamblia*. *Molecular and Biochemical Parasitology*, **50**: 95-104.
- Anderson, R.M. & May, R.M. (1982). Coevolution of hosts and parasites. *Parasitology*, **85**: 411-426.
- Andrews, R.H., Adams, M., Boreham, P.F.L., Mayrhofer, G. & Meloni, B.P. (1989). *Giardia intestinalis*: electrophoretic evidence for a species complex. *International Journal for Parasitology*, **19**: 183-190.

- Andrews, R.H. & Beveridge, I. (1990). Apparent absence of genetic differentiation in *Teladorsagia* species. *The Journal of Helminthology*, **64**: 290-294.
- Andrews R.H. & Chilton, N.B. (1997). What parasite is that?: Uses and misuses of multilocus enzyme electrophoresis for parasite systematics. *International Journal for Parasitology*, (in prep.)
- Andrews, R.H., Chilton, N.B., Beveridge, I., Spratt, D. & Mayrhofer, G. (1992a). Markers for the identification of three Australian tick species at various stages in their life cycle. *Journal of Parasitology*, **78**: 366-368
- Andrews, R.H., Chilton, N.B., Ey, P.L. & Mayrhofer, G. (1993). Additional enzymes for the genetic characterisation of *Giardia* from different host species. *Parasitology Research*, **79**: 337-339.
- Andrews, R.H., Chilton, N.B. & Mayrhofer, G. (1992b). Selection of specific *Giardia intestinalis* by different growth conditions. *Parasitology*, **105**: 375-386.
- Andrews, R.H., Handman, E., Adams, M., Baverstock, P.R. & Mitchell, G.F. (1988). Genetic characteristics of *Leishmania* isolates at 37 enzyme loci. *International Journal for Parasitology*, **18**: 445-452.
- Arashima, Y., Kumasaka, K., Kawano, K., Asano, R., Hokari, S., Murasugi, E., Iwashita, E., Nishikawa, S. & Matsuo, K. (1992). Studies on the giardiasis as the zoonosis. III. Prevalence of *Giardia* among the dogs and the owners in Japan. *Kansenshogaku Zasshi*, **66**: 1062-1066.
- Arroyave, R.J., Ayala, D.E. & Hermida, R.C. (1990). Differences in circannual characteristics of the incidences of amebiasis and giardiasis. pp. 717-727, in: *Progress in clinical and biological research*. (D.K. Hayes, J.E. Pauly and R.J. Reiter eds.) Bethesda, MD, USA.
- Asano, R., Hokari, S., Murasugi, E., Arashima, Y. Kubo, N. & Kawano, K. (1991). Studies on the giardiasis as the zoonosis. II. Giardiasis in dogs and cats. *Kansenshogaku Zasshi*, **65**: 157-61.

- Avise, J.C. (1974). Systematic value of electrophoretic data. *Systematic Zoology*, **23**: 465-481.
- Avise, J.C. (1983). Protein variation and phylogenetic reconstruction. pp 103-130 in: *Protein Variation: Adaptive and Taxonomic Significance*, G. Oxford and D. Rollinson (eds.), Systematics Association, British Museum Natural History, London.
- Avise, J.C. (1994). *Molecular markers, natural history & evolution*. Chapman & Hall Inc.
- Avise, J.C. & Ball, R.M. Jr. (1990). Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology*, **7**: 45-67.
- Avise, J.C., Lansman, R.A. & Shade, R.O. (1979). The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. *Genetics*, **92**: 279-295.
- Bachmann, K. (1995). Progress and pitfalls in systematics: cladistics, DNA and morphology. *Acta Botanica Neerlandica*, **44**: 403-419.
- Backeljau, T., De Bruyn, L., De Wolf, H., Jordaens, K., Van Dongen, S., Verhagen, R. & Winnepenninckx, B. (1995). Random Amplified Polymorphic DNA (RAPD) and Parsimony Methods. *Cladistics*, **11**: 119-130.
- Barker, I.K., O'Callaghan, M.G. & Beveridge, I. (1989). Host-parasite associations of *Eimeria* spp. (Apicomplexa: Eimeriidae) in kangaroos and wallabies of the genus *Macropus* (Marsupialia: Macropodidae). *International Journal for Parasitology*, **19**: 241-263.
- Baruch, A.C., Isaac-Renton, J. & Adam, R.D. (1996). The molecular epidemiology of *Giardia lamblia*: a sequence based approach. *The Journal of Infectious Diseases*, **174**: 233-236.
- Baveja, U.K., Jyoti, A.S., Kaur, M., Aggarwal, D.S., Anand, B.S. & Nanda, R. (1986). Isoenzyme studies of *Giardia lamblia* isolated from symptomatic cases. *Australian Journal of Experimental Biology and Medical Science*, **64**: 119-126.
- van Belkum, A., Homan, W., Limper, L. & Quint, W.G.V. (1993). Genotyping isolates and clones of *Giardia duodenalis* by polymerase chain reaction: implications for the

- detection of genetic variation among protozoan parasite species. *Molecular and Biochemical Parasitology*, **61**: 69-77.
- Bemrick, W.J. (1961). A note on the incidence of three species of *Giardia* in Minnesota. *Journal of Parasitology*, **47**: 87-89.
- Benachenhou-Lahfa, N. & Forterre, P. & Labedan, B. (1993). Evolution of glutamate dehydrogenase genes: evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life. *Journal of Molecular Evolution*, **36**: 335-346.
- Bertram, M.A., Meyer, E.A., Lile, J.D. & Morse, S.A. (1983). A comparison of isoenzymes of five axenic *Giardia* isolates. *Journal of Parasitology*, **69**: 739-801.
- Bingham, A.K., Harroll, E.L. Jr. & Meyer, E.A. (1979). *Giardia* sp.: physical factors of excystation *in vitro*, and excystation vs eosin exclusion as determinants of viability. *Experimental Parasitology*, **47**: 284-291.
- Binz, N., Thompson, R.C.A., Lymbery, A.J. & Hobbs, R.P. (1992). Comparative studies on the growth dynamics of two genetically distinct isolates of *Giardia duodenalis in vitro*. *International Journal for Parasitology*, **22**: 195-202.
- Boothroyd, J., Wang, A., Campbell, D. & Wang, C. (1987). An unusually compact ribosomal DNA repeat in the protozoan *Giardia lamblia*. *Nucleic Acid Research*, **15**: 4065-4083.
- Brogden, R.N., Heel, R.C., Speight, T.M. & Avery, G.S. (1978). Metronidazole in anaerobic infections: a review of its activity, pharmacokinetics and therapeutic use. *Drugs*. **16**: 387-417.
- Brown, A. H. D., Feldman, M. W. & Nevo, E. (1980). Multilocus structure of natural populations of *Horeum spontaneum*. *Genetics* **96**: 523-536.
- Brugerolle, G. (1975). Contribution à l'étude cytologique et phylétique des Diplozaïres (Zoomastigophorea, Diplozoa Dangeard, 1910). V. Nouvelle interprétation de l'organisation cellulaire de *Giardia*. *Protistologica*, **11**: 99-109.

- Brugerolle, G. & Taylor, F.J.R. (1977). Taxonomy, cytology and evolution of the Mastigophora. *Proceedings of the Fifth International Congress of Protozoology-1977*, pp. 14-28.
- Buret, A., DenHollander, N., Wallis, P.M., Befus, D. & Olson, M.E. (1990). Zoonotic potential of giardiasis in domestic ruminants. *Journal of Infectious Diseases*, **162**: 231-237.
- Burgu, A. (1979). Distribution of *Giardia canis* in dogs in Ankara and its public health significance. (Ankara'da kopeklerde *Giardia canis*'in yayilisi ve bunun halk sagligi yonunden onemi.) *Ankara Universitesi, Veteriner Fakultesi Dergisi*. **26**: 3184-194.
- Burns, T.P. (1994). On the fittedness of organisms and the ascendancy of ecosystems: toward a hierarchical model of network development. *Journal of Theoretical Biology*, **170**: 115-127.
- Bush, A.O., Caira, J.N., Minchella, D.J., Nadler, S.A. & Seed, J.R. (1995). Parasitology Year 2000. *Journal of Parasitology*, **81**(5): 835-842
- Bushek, D. & Allen, S.K.Jr. (1996) Races of *Perkinsus marinus*. *Journal of Shellfish Research*, **15**: 103-107.
- Butcher, P.D., Cevallos, A.M., Carnaby, S., Alstead, E.M., Swarbrick, E.T. & Farthing, M.J.G. (1994). Phenotypic and genotypic variation in *Giardia lamblia* isolates during chronic infection. *Gut*, **35**: 51-54.
- Buth. D.G. (1984). The application of electrophoretic data in systematic studies. *Annual Review of Ecological Systematics*, **15**: 501-522.
- Campbell, S.R., van Keulen, H., Erlandsen, S.L., Senturia, J.B. & Jarroll, E.L. (1990). *Giardia* sp.: Comparison of Electrophoretic Karyotypes. *Experimental Parasitology*, **71**: 470-482.

- Carnaby, S., Butcher, P.D., Summerbell, C.D., Naeem, A. & Farthing, M.J.G. (1995). Minisatellites corresponding to the human polycore probes 33.6 and 33.15 in the genome of the most 'primitive' known eukaryote *Giardia lamblia*. *Gene*, **166**: 167-172.
- Castor S.B. & Lindqvist K.B. (1990). Canine giardiasis in Sweden: no evidence of infectivity to man. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **84**: 249-250.
- Catcott, E.J. (1946). The incidence of intestinal protozoa in the dog. *Journal of the American Veterinary Medical Association*, **108**: 34-36
- Cavalier-Smith, T. (1987). Eukaryotes with no mitochondria. *Nature*, **326**: 332-333.
- Cavalier-Smith, T. (1993). Kingdom Protozoa and its 18 phyla. *Microbiological Reviews*, **57**: 953-994.
- Cavalli-Sforza, L.L. & Edwards, A.W.F. (1967). Phylogenetic analysis: Models and estimation procedures. *Evolution*, **32**: 550-570.
- Cedillo Rivera, R., Enciso Moreno, J.A., Martinez Palomo, A. & Ortega Pierres, G. (1989). *Giardia lamblia*: isoenzyme analysis of 19 axenic strains isolated from symptomatic and asymptomatic patients in Mexico. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **83**: 644-646.
- Cerva, L. & Nohýnková, E. (1992). A light microscopic study of the course of cellular division of *Giardia intestinalis* trophozoites grown *in vitro*. *Folia Parasitologica*, **39**: 97-104.
- Chaudhuri, P., De, A., Bhattacharya, A., Pal, S.C. & Das, P. (1991). Identification of heterogeneity in human isolates of *Giardia lamblia* by isoenzyme studies. *International Journal of Medical Microbiology*, **274**: 490-495.
- Chaudhuri, P.P., Sil, A.K., Das, D., Sengupta, K., Bhattacharya, S., Chattopadhyay, D.J. & Das, P. (1994). Comparative biochemical and restriction endonuclease digestion pattern

- analysis of Indian and Portland-1 strains of *Giardia lamblia*. *Journal of Protozoology Research*, **4**: 29-39.
- Chen, L.M., Chern, Y., Ong, S.J. & Tai, J.H. (1994b). Molecular cloning and characterization of a ras-related gene of ran/tc4/spi1 subfamily in *Giardia lamblia*. *Journal of Biological Chemistry*, **269**: 17297-17304.
- Chen, N., Upcroft, J.A. & Upcroft, P. (1994a). Physical map of a 2 Mb chromosome of the intestinal protozoan parasite *Giardia duodenalis*. *Chromosome-Research*, **2**: 307-313.
- Cheng, C., Hay, D., Sullivan, T.P. & Isaac-Renton, J.L. (1993). Ability of a small wild rodent (*Peromyscus maniculatus*) to host strains of human-type *Giardia duodenalis*. *International Journal of Environmental Health Research*, **3**: 203-208.
- Chilton, N.B., Beveridge, I. & Andrews, R.H. (1992). Detection by allozyme electrophoresis of cryptic species of *Hypodontus macropi* (Nematoda: Strongyloidea) from macropodid marsupials. *International Journal for Parasitology*, **22**: 271-279.
- Choquette, L.P.E. & de Gelinas, L.G. (1950). The incidence of intestinal nematodes and protozoa in dogs in the Montreal district. *Canadian Journal of Comparative Medicine and Veterinary Science*, **14**: 33-38.
- Cole, F.J. (1926). *The History of Protozoology*. Univ. London Press, Ltd., London.
- Collins, F.H. & Paskewitz, S.M. (1996) A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic Anopheles species. *Insect Molecular Biology*, **5**: 1-9.
- Collyer, R., Lim, K.H., Tang, R. & Prociv, P. (1992). Suburban dogs - a reservoir of human giardiasis? *Medical Journal of Australia*, **156**: 814-815.
- Connor, R.J. (1994). The impact of nagana. *Onderstepoort Journal of Veterinary Research*, **61**: 379-383.
- Corliss, J.O. (1960). Comment on the systematics and phylogeny of the Protozoa. *Systematic Zoology*, **8**: 169-190.

- Crabtree, K.D., Ruskin, R.H., Shaw, S.B. & Rose, J.B. (1996). The detection of *Cryptosporidium* oocysts and *Giardia* cysts in cistern water in the U.S. Virgin Islands. *Water Research Oxford*, **30**: 208-216.
- Cracraft, J. (1983). Species concepts and speciation analysis, pp. 159-189. in: *Current Ornithology*. (R.F. Johnston ed.), Plenum Press, New York.
- Cracraft, J. (1989). Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation, pp. 28-59. in: *Speciation and Its Consequences*. (D. Otte and J.A. Endler eds.) Sinauer Associates, Sunderland, MA.
- Craige, J.E. (1948). Differential diagnosis and specific therapy of dysenteries in dogs. *Journal of the American Veterinary Medical Association*, **113**: 343-347.
- Crossley, R. & Holberton, D. (1983). Characterisation of proteins from the cytoskeleton of *Giardia lamblia*. *Journal of Cell. Science* **59**: 81-103.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. Murray, London.
- Davey, R.A., Mayrhofer, G. & Ey, P.L., (1992). Identification of a broad-specificity nucleoside transporter with affinity for the sugar moiety in *Giardia intestinalis* trophozoites. *Biochimica et Biophysica Acta*, **1109**: 172-178.
- Davies, R.B. & Hibler, C.P. (1979). Animal reservoirs and cross-species transmission of *Giardia*. pp.104-126. in: *Waterborne transmission of giardiasis*. (W. Jakubowski and J.C. Hoff eds.) Environmental Protection Agency, Cincinnati, USA.
- Davis, J.I. & Nixon, K.C. (1992). Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology*, **41**: 421-435.
- Diaz Saez, V., Verdejo Torralba, M.J., Campos Bueno, M., Manas Almendros, I. & Lozano Maldonado, J. (1987). Epidemiological aspects of giardiasis in the province of Granada. (Aspectos epidemiológicos de la giardiasis en la provincia de Granada.) *Revista Iberica*

de Parasitologia. 1987, Vol. Extraordinario: Enero, 25-29; IV Congreso Nacional de la Asociacion de Parasitologos Espanoles (A.P.E.).

Dobson, A.P. & Hudson, P.J. (1992). Regulation and stability of a free-living host-parasite system: *Trichostrongylus tenius* in red grouse. II. Population models. *Journal of Animal Ecology*, **61**: 487-498.

Dobzhansky, T. (1937). *Genetics and the Origin of Species*. Columbia University Press, New York.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher*, **35**: 125-129.

Donnellan, S.C. & Aplin, K.P. (1989). Resolution of cryptic species in the New Guinea lizard *Sphenomorphus jobiensis* (Scincidae) by electrophoresis. *Copeia*, **1989**: 81-88.

Dowling, T.E., Moritz, C., Palmer, J.D. & Riesberg, L.H. (1996). Chapter 8. Nucleic Acids III: Analysis of Fragments and Restriction Sites. in: *Molecular Systematics*. 2nd edition. (David M. Hillis, Craig Moritz and Barbara K. Mable eds.), Sinauer Associates Inc. Publishers..

Edlind, T.D. and Chakraborty, P.R. (1987). Unusual ribosomal RNA of the intestinal parasite *Giardia lamblia*. *Nucleic Acids Research*, **15**: 7889-7901.

Edlind, T.D., Li, J., Visvesvara, G.S., Vodkin, M.H., McLaughlin, G.L. & Katiyar, S.K. (1996). Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa. *Molecular Phylogenetics and Evolution*, **5**: 359-367.

Edwards, M.R., Knodler, L.A., Wilson, J.R. & Schofield, P.J. (1993). The transport and metabolism of alanine by *Giardia intestinalis*. *Molecular and Biochemical Parasitology*, **61**: 49-57.

Ellis, T.J., Luton, K., Baverstock, P.R., Whitworth, G., Tenter, A.M. & Johnson, A.M. (1995). Phylogenetic relationships between *Toxoplasma* and *Sarcocystis* deduced from a comparison of 18S rDNA sequences. *Parasitology*, **110**: 521-528.

- Engbaek, K. (1978). The incidence and significance of *Giardia lamblia* and other intestinal parasites in children in day-institutions. (*Giardia lamblia* og andre tarmparasitter hos boern i daginstitutioner. Forekomst og betydning.) *Ugeskrift for Laeger*, **140**: 14-17.
- Epe, C., Ising Volmer, S. & Stoye, M. (1993). Results of parasitological examinations of faecal samples from horses, donkeys, dogs, cats and hedgehogs between 1984 and 1991. (Ergebnisse parasitologischer Kotuntersuchungen von Equiden, Hunden, Katzen und Igeln der Jahre 1984-1991.) *Deutsche Tierarztliche Wochenschrift*, **100**: 426-428.
- Erlandsen, S.L., & Bemrick, W.J. (1987). SEM evidence for a new species, *Giardia psittaci*. *Journal of Parasitology*, **73**: 623-629.
- Erlandsen, S.L., & Bemrick, W.J. (1988). Waterborne giardiasis: sources of *Giardia* cysts and evidence pertaining to their implication in human infection. pp. 227-236. in: *Advances in Giardia research*. (P.M. Wallis & B.R. Hammond eds.) University of Calgary Press, Calgary.
- Erlandsen, S.L., Bemrick, W.J., Wells, C.L, Feely, D.E., Knudson, L., Campbell, S.R., Van Keulen, H. & Jarroll, E.L. (1990a). Axenic culture and characterisation of *Giardia ardeae* from the great blue heron(*Ardeae herodias*). *Journal of Parasitology*, **76**: 717-724.
- Erlandsen, S.L. & Rasch, E.M. (1994). The DNA content of trophozoites and cysts of *Giardia lamblia* by microdensitometric quantitation of feulgen staining and examination by laser scanning confocal microscopy. *Journal of Histochemistry and Cytochemistry*, **42**: 1413-1416.
- Erlandsen, S.L., Sherlock, L.A., Bemrick, W.J., Ghobrial, H. & Jakubowski, W. (1990b). Prevalence of *Giardia* spp. in beaver and muskrat populations in northeastern states and Minnesota (USA): Detection of intestinal trophozoites at necropsy provides greater sensitivity than detection of cysts in fecal samples. *Applied and Environmental Microbiology*, **56**: 31-36.
- Erlandsen, S.L., Sherlock, L.A., Janushka, M., Schupp, D.G., Schaefer, F.W. III, Jakubowski, W. & Bemrick, W.J. (1988). Cross-species transmission of *Giardia* spp.: Inoculation of

- beavers and muskrats with cysts of human, beaver, mouse and muskrat origin. *Applied and Environmental Microbiology*, **54**: 2777-2785.
- Escalante, A.A. & Ayala, F.J. (1994). Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proceedings of the National Academy of Sciences*, **91**: 11373-11377.
- Ey, P.L., Andrews, R.H. & Mayrhofer, G. (1993a). Differentiation of major genotypes of *Giardia intestinalis* by polymerase chain reaction analysis of a gene encoding a trophozoite surface antigen. *Parasitology*, **106**: 347-356.
- Ey, P.L., Bruderer, T., Wehrli, C. & Kohler, P. (1996). Comparison of genetic groups determined by molecular and immunological analyses of *Giardia* isolated from animals and humans in Switzerland and Australia. *Parasitology Research*, **82**: 52-60.
- Ey, P.L., Darby, J.M., Andrews, R.H. & Mayrhofer, G. (1993b). *Giardia intestinalis*: detection of major genotypes by restriction analysis of gene amplification products. *International Journal for Parasitology*, **23**: 591-600.
- Ey, P.L., Khanna, K., Andrews, R.H., Manning, P.A. & Mayrhofer, G. (1992). Distinct genetic groups of *Giardia intestinalis* distinguished by restriction fragment length polymorphisms. *Journal of General Microbiology*, **138**: 2629-2637.
- Ey, P. L., Mansouri, M., Kulda, J., Nohýnková, E., Monis, P. T., Andrews, R. H. A. & Mayrhofer, G. (1997). Genetic analysis of *Giardia* from hoofed farm animals reveals artiodactyl-specific and potentially zoonotic genotypes. *Journal of Eukaryotic Microbiology* (submitted). **Appendix F**.
- Fallis, A.M. (1980) Arthropods pests and vectors of disease. *Veterinary Parasitology* **6**: 47-73.
- Farthing, M.J.G., Mata, L., Urrutia, J.J. & Kronmal, R.A. (1986). Natural history of *Giardia* infection of infants and children in rural Guatemala and its impact on physical growth. *American Journal of Clinical Nutrition*, **43**: 395-405.

- Faubert, G.M. (1996). The immune response to *Giardia*. *Parasitology Today*, **12**: 140-145.
- Faubert, G.M., Belosevic, M., Walker, T.S., MacLean, J.D. & Meerovitch, E. (1983). Comparative studies on the pattern of infection with *Giardia* spp. in Mongolian gerbils. *Journal of Parasitology*, **69**: 802-805.
- Faust, E.C., Russell, P.F. & Jung, R.C. (1976). Important Landmarks in Parasitology. in: *Craig and Faust's Clinical Parasitology*. Eighth edition, Lea & Febiger, Philadelphia.
- Feely, D.E. (1986). A simplified method for in vitro excystation of *Giardia muris*. *Journal of Parasitology*, **72**: 474-475.
- Feely, D.E. (1988). Morphology of the cyst of *Giardia microti* by light and electron microscopy. *Journal of Protozoology*, **35**: 52-54.
- Feely, D.E., Schollmeyer, J.V. & Erlandsen, S.L. (1982). *Giardia* spp.: distribution of contractile proteins in the attachment organelle. *Experimental Parasitology*, **53**: 145-154.
- Felsenstein, J. (1978). Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology*, **27**: 401-410.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**: 783-791.
- Felsenstein, J. (1989). PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**: 164-166.
- Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
- Felsenstein, J. & Kishino, H. (1993). Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Systematic Biology*, **42**: 193-200.
- Ferguson, H.W. (1979). Scanning and transmission electron microscopical observations on *Hexamita salmonis* (Moore, 1922) related to mortalities in rainbow trout fry *Salmo gairdneri* Richardson. *Journal of Fish Diseases*, **2**: 57-67.

- Filice, F.P. (1952). Studies on the cytology and life history of a *Giardia* from the laboratory rat. *University of California Publications in Zoology*, **57**: 53-146.
- Finelle, P. (1974). African animal trypanosomiasis. Part IV. Economic Problems. *World Animal Review*, **10**: 15-18.
- Foster, W.D. (1965). *A History of Parasitology*. E. & S. Livingstone, Ltd., London.
- Gardner, S.L. & Campbell, M.L. (1992). Parasites as probes for biodiversity. *Journal of Parasitology*, **78**: 596-600.
- Gasser, R.B., Eckert, J. & Rohrer, L. (1987). Infectivity of Swiss *Giardia* isolates to jirds and mice, and *in vitro* cultivation of trophozoites originating from sheep. *Parasitology Research*, **74**: 103-111
- Gillin, F.D., Reiner, D.S. & Boucher, S.E. (1988). Small-intestinal factors promote encystation of *Giardia lamblia* *in vitro*. *Infection and Immunity*, **56**: 705-707.
- Gillin, F.D., Reiner, D.S., Gault, M.J., Douglas, H., Das, S., Wunderlich, A. & Sauch, J.F. (1987). Encystation and expression of cyst antigens by *Giardia lamblia* *in vitro*. *Science*, **235**: 1040-1043.
- Good, D.A. (1989). Hybridisation and cryptic species in *Dicamptodon* (Caudata: Dicamptodontidae). *Evolution*, **7**: 203-219.
- Gordon, D.M., Wexler, M., Reardon, T.B. & Murphy, P.J. (1995). The genetic structure of *Rhizobium* populations. *Soil Biology and Biochemistry*, **27**: 491-495.
- Grant, D.R. & Woo, P.T.K. (1978). Comparative studies of *Giardia* spp. in small mammals in southern Ontario. II. Host specificity and infectivity of stored cysts. *Canadian Journal of Zoology*, **56**: 1360-1366.
- Hafner, M.S. & Nadler, S.A. (1988). Phylogenetic trees support the coevolution of parasites and their hosts. *Nature*, **332**: 258-259.

- Hafner, M.S. & Nadler, S.A. (1990). Cospeciation in host-parasite assemblages: Comparative analysis of rates of evolution and timing of cospeciation events. *Systematic Zoology*, **39**: 192-204.
- Hafner, M.S. & Page, R.D.M. (1995). Molecular phylogenies and host-parasite cospeciation: Gophers and lice as a model system. *Philosophical Transactions of the Royal Society of London*, **349**: 77-83
- Hafner, M.S., Sudman, P.D., Villablanca, F.X., Spradling, T.A., Demastes, J.W. & Nadler, S.A. (1994). Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science*, **265**: 1087-1090.
- Hahn, N.E., Glaser, C.A., Hird, D.W. & Hirsh, D.C. (1988). Prevalence of *Giardia* in the faeces of pups. *Journal of the American Veterinary Medical Association*, **192**: 1428-1429.
- Hall, M.L., Costa, N.D., Thompson, R.C.A., Lymbery, A.J., Meloni, B.P. & Wales, R.G. (1992). Genetic variants of *Giardia duodenalis* differ in their metabolism. *Parasitology Research*, **78**: 712-714.
- Harris, H. & Hopkinson, D.A. (1976). *Handbook of Enzyme Electrophoresis in Human Genetics*. North-Holland, Amsterdam.
- Hashimoto, T., Adachi, J. & Hasegawa, M. (1992). Phylogenetic place of *Giardia lamblia* a protozoan that lacks mitochondria. *International Journal on Endocytobiosis and Cell Research*, **9**: 59-69.
- Hashimoto, T., Nakamura, Y., Kamaishi, T., Nakamura, F., Adachi, J., Okamoto, K.I., Hasegawa, M. (1995). Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia*, inferred from amino acid sequences of elongation factor 2. *Molecular Biology and Evolution*, **12**: 782-793.
- Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K.I. & Hasegawa, M. (1994). Protein phylogeny gives a robust estimation for early

- divergences of eukaryotes: Phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Molecular Biology and Evolution* **11**: 65-71.
- Hennig, W. (1966). *Phylogenetic Systematics*. Univ. of Illinois Press, Urbana.
- Herlich, H. (1978). The importance of helminth infections in ruminants. *World Animal Review*, **26**: 22-26.
- Hewlett, E.L., Andrews, J.S. Jr., Ruffier, J. & Schaefer, F.W. III. (1982). Experimental infection of mongrel dogs with *Giardia lamblia* cysts and cultured trophozoites. *Journal of Infectious Diseases*, **145**: 89-93.
- Heyworth, M.F. (1988). Immunology of *Giardia* Infections. In: *Advances in Giardia Research*, p.45-48, (P.M. Wallis & B.R. Hammond eds.), University of Calgary Press, Calgary, 1988.
- Hillis, D.M. (1987). Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics*, **18**: 23-42.
- Hillis, D.M. (1991). Discriminating between phylogenetic signal and random noise in DNA sequences. Chapter 13 in: *Phylogenetic analysis of DNA sequences*. (Michael M. Miyamoto and Joel Cracraft eds.) Oxford University Press.
- Hillis, D.M. & Bull, J.J. (1991). Of genes and genomes. *Science*, **254**: 528.
- Hillis, D.M., Moritz, C. & Mable, B.K. (1996). Glossary in: *Molecular Systematics* 2nd edition, Sinauer Associates, Inc. Sunderland, Massachusetts.
- Hoeppli, R. (1959). *Parasites and Parasitic Infections in Early Medicine and Science*. Univ. Malaya Press, Singapore.
- Hoff, J.C., Rice, E.W. & Schaefer, F.W. 3rd. (1985). Comparison of animal infectivity and excystation as measures of *Giardia muris* cyst inactivation by chlorine. *Applied and Environmental Microbiology*, **50**: 1115-1117.

- Holberton, D.V. & Marshall, J. (1995). Analysis of consensus sequence patterns in *Giardia* cytoskeleton gene promoters. *Nucleic Acids Research*, **23**: 2945-2953.
- Homan, W.L., van Enkevort, F.H.J., Limper, L., van Eys, G.J.J.M., Schoone, G.J., Kasprzak, W., Majewska, A.C. & van Knapen, F. (1992). Comparison of *Giardia* isolates from different laboratories by isoenzyme analysis and DNA probes. *Parasitology Research*, **78**: 316-323.
- Hou, G., Le Blancq, S.M., E, Y., Zhu, H. & Lee, M.G.S. (1995). Structure of a frequently rearranged rRNA-encoding chromosome in *Giardia lamblia*. *Nucleic Acids Research*, **23**: 3310-3317.
- Huelsenbeck, J.P., Bull, J.J. & Cunningham, C.W. (1996). Combining data in phylogenetic analysis. *Tree*, **11**: 152-157.
- Hunter, R.L. & Market, C.L. (1957). Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science*, **125**: 1294-1295.
- Hyman, B.C. (1990). Molecular diagnosis of *Meloidogyne* species. *Journal of Nematology*, **22**: 24-30.
- Isaac-Renton, J., Moorehead, W. & Ross, A. (1996). Longitudinal studies of *Giardia* contamination in two community drinking water supplies: cyst levels, parasite viability, and health impact. *Applied and Environmental Microbiology*, **62**: 47-54.
- Isaac-Renton, J.L., Moricz, M.M. & Proctor, E.M. (1987). A *Giardia* survey of fur-bearing water mammals in British Columbia, Canada. *Journal of Environmental Health*, **50**: 80-83.
- Isaac-Renton, J.L., Cordeiro, C., Sarafis, K. & Shahriari, H. (1993). Characterization of *Giardia duodenalis* isolates from a waterborne outbreak. *Journal of Infectious Disease*, **167**: 431-440.

- Isaac-Renton, J.L., Shahriari, H. & Bowie, W.R. (1992). Comparison of an *in vitro* method and an *in vivo* method of *Giardia* excystation. *Applied and Environmental Microbiology*, **58**: 1530-1533.
- Islam, A., Stoll, B.J., Ljungstrom, I., Biswas, J., Nazrul, H. & Hultdt, G. (1983). *Giardia lamblia* infections in a cohort of Bangladeshi mothers and infants followed for one year. *Journal of Paediatrics*, **103**: 996-1000.
- Jakubowski, W. (1984). Detection of *Giardia* Cysts in Drinking Water. Chapter 16 in: *Giardia and Giardiasis* (S.L. Erlandsen & E.A. Meyer eds) Plenum Press, New York.
- Jarroll, E.L., Muller, P.J., Meyer, E.A. & Morse, S.A. (1981). Lipid and carbohydrate metabolism of *Giardia lamblia*. *Molecular and Biochemical Parasitology*, **2**: 187-196.
- Jeffrey, C. (1973). *Biological Nomenclature*. Arnold, London.
- Jeffreys, A.J., Royle, N.J., Wilson, V. & Wong, Z. (1988). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature*, **332**: 278-281.
- Jian Bing Fan., Korman, S.H., Cantor, C.R. & Smith, C.L. (1991). *Giardia lamblia*: haploid genome size determined by pulsed field electrophoresis is less than 12Mb. *Nucleic Acids Research*, **19**: 1905-1908.
- Johnson, D.C., Reynolds, K.A., Gerba, C.P., Pepper, I.L. & Rose, J.B. (1995) Detection of *Giardia* and *Cryptosporidium* in marine waters. *Water Science and Technology*, **31**: 439-442.
- Johnston, J. & Gasser, R.B. (1993). Copro-parasitological survey of dogs in southern Victoria. *Australian Veterinary Practitioner*, **23**: 127-131.
- de Jonckheere, J.F., Majewska, A.C. & Kasprzak, W. (1990). *Giardia* isolates from primates and rodents display the same molecular polymorphism as human isolates. *Molecular and Biomedical Parasitology*, **39**: 23-30.

- Jones, D. T., Taylor W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *CABIOS*, **8**: 275-282.
- Joyce, T., McGuigan, K.G., Elmore Meegan, M. & Conroy, R.M. (1996). Prevalence of enteropathogens in stools of rural Maasai children under five years of age in the Maasailand region of the Kenyan Rift Valley. *East African Medical Journal*, **73**: 59-62.
- Jungmann, R., Hiepe, T. & Scheffler, C. (1986). Parasitic intestinal fauna of dogs and cats, particularly *Giardia* infection. (Zur parasitären Intestinalfauna bei Hund und Katze mit einem speziellen Beitrag zur *Giardia*-Infektion.) *Monatshefte für Veterinärmedizin*, **41**: 309-311.
- Jyothisri, A. & Baveja, U.K. (1988). Cytopathogenicity of *Giardia lamblia* in HeLa and Vero Cell Monolayers. pp.33-37 in: *Advances in Giardia Research*. (P.M. Wallis & B.R. Hammond eds.) University of Calgary Press, Calgary.
- Kabnick, K.S. & Peattie, D.A. (1990) *In situ* analyses reveal that the two nuclei of *Giardia lamblia* are equivalent. *Journal of Cell Science*, **95**: 353-360.
- Kasprzak, W., Mazur, T. & Rauhut, W. (1983). The role of dogs and cats in the transmission of giardiasis. (Uwagi o roli psow i kotow w transmisji giardiozy.) *Przegląd Epidemiologiczny*, **37**., 359-362.
- Keeling, P.J. & Doolittle, W.F. (1996). A non-canonical genetic code in an early diverging eukaryotic lineage. *The EMBO Journal*, **15**: 2285-2290.
- Keister, D.B. (1983). Axenic culture of *Giardia lamblia* in TYI-S33 medium supplemented with bile. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **77**: 487-488.
- van Keulen, H., Campbell, S.R., Erlandsen, S.L. & Jarroll, E.L. (1991). Cloning and restriction enzyme mapping of ribosomal DNA of *Giardia duodenalis*, *Giardia ardeae* and *Giardia muris*. *Molecular and Biochemical Parasitology*, **46**: 275-284.

- 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 subunit ribosomal RNA sequence of *Giardia ardeae*, *G. muris*, *G. duodenalis* and *Hexamita* sp. *FASEB Journal*, **7**: 223-231.
- van Keulen, H., Homan, W.L., Erlandsen, S.L. & Jarroll, E.L. (1995). A three nucleotide signature sequence in small subunit rRNA divides human *Giardia* in two different genotypes. *Journal of Eukaryotic Microbiology*, **42**: 392-394.
- Kirk-Mason, K.E., Turner, M.J. & Chakraborty, P.R. (1988). Cloning and sequencing of β tubulin cDNA from *Giardia lamblia*. *Nucleic Acids Research*, **16**: 2733.
- Kirk-Mason, K.E., Turner, M.J. & Chakraborty, P.R. (1989). Evidence for unusually short tubulin mRNA leaders and characterisation of tubulin genes in *Giardia lamblia*. *Molecular and Biochemical Parasitology*, **36**: 87-100.
- Kirkpatrick, C.E. (1988). Epizootiology of endoparasitic infections in pet dogs and cats presented to a veterinary teaching hospital. *Veterinary Parasitology*, **30**: 113-124.
- Kirkpatrick, C.E. & Green, G.A. (1985). Susceptibility of domestic cats to infections with *Giardia lamblia* cysts and trophozoites from human sources. *Journal of Clinical Microbiology*, **21**: 678-680.
- Klassen, G.J. (1992). Coevolution: A history of the macroevolutionary approach to studying host-parasite associations. *Journal of Parasitology*, **78**: 573-587.
- Knodler, L.A., Schofield, P.J. & Edwards, M.R. (1995). L-Arginine transport and metabolism in *G. intestinalis* support its position as a transition between the prokaryotic and eukaryotic kingdoms. *Microbiology*, **141**: 2063-2070.
- Koudela, B., Nohýnková, E., Vitovec, J., Pakandl, M. & Kulda, J. (1991). *Giardia* infection in pigs: detection and *in vitro* isolation of trophozoites of the *Giardia intestinalis* group. *Parasitology*, **102**: 163-166.

- Krebber, H., Woestmann, C. & Bakker Grunwald, T. (1994). Evidence for the existence of a single ubiquitin gene in *Giardia lamblia*. *FEBS (Federation of European Biochemical Societies) Letters* **343**: 234-236.
- Kuhner, M.K. & Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution*, **11**: 459-468.
- Kulda, J. & Nohýnková, E. (1978). Flagellates of the Human Intestine and of Intestines of Other Species, pp. 83-89 in: *Parasitic Protozoa*. (ed. P.P. Kreier) Vol. II, , Academic Press, New York.
- Kulda, J. & Nohýnková, E. (1996). *Giardia* in humans and animals. pp. 225-422 in: *Parasitic Protozoa*; 2nd edition (ed. P.P. Kreier), Vol. 10, Academic Press, San Diego.
- Kumar, S., Tamura, K. & Nei, M. (1993). *MEGA: Molecular Evolutionary Genetics Analysis*, version 1.0. The Pennsylvania State University, University Park, PA 16802, U.S.A..
- Kumar, S., Tamura, K. & Nei, M. (1994). MEGA: molecular evolutionary genetics analysis software for microcomputers. *Computers in the Applied Biosciences*, **10**: 189-91.
- Kumazaki, T., Hori, H. & Osawa, S. (1983). Phylogeny of protozoa deduced from 5S rRNA sequences. *Journal of Molecular Evolution*, **19**: 411-419.
- Ladle, R.J., Johnstone, R.A. & Judson, O.P. (1993). Coevolutionary dynamics of sex in a metapopulation: Escaping the Red Queen. *Proceedings of the Royal Society of London*, **253**: 155-160.
- Lalova, I. (1976). Application of the immunofluorescent technique in detecting *Lambli*a infection in humans. *Problems of Infectious and Parasitic Diseases*, **4**: 149-151.
- Lanzer, M., Fischer, K. & Le Blancq, S.M. (1995). Parasitism and chromosome dynamics in protozoan parasites: Is there a connection? *Molecular and Biochemical Parasitology*, **70**: 1-8.

- Lapage, G. (1956). Parasitism; the kinds of animals parasitic in farm stock and their host-parasite relationships. Chapter 1 in: *Veterinary Parasitology*. Oliver & Boyd Publishers, London.
- Lawrence, J.G., Ochman, H. & Hartl, D.L. (1991). Molecular and evolutionary relationships among enteric bacteria. *Journal of General Microbiology*, **137**: 1911-1921.
- Le Blancq, S.M., Korman, S.H. & Van der Ploeg, L.H.T. (1991). Frequent rearrangements of rRNA-encoding chromosomes in *Giardia lamblia*. *Nucleic Acids Research*, **19**: 4405-4412.
- Lessa, E.P. & Applebaum, G. (1993). Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology*, **2**: 119-129.
- Lewis P.D. Jr. (1988). Prevalence of *Giardia* sp. in dogs from Alberta. pp. 61-64 in: *Advances in Giardia research*. (P.M. Wallis and B.R. Hammond eds.). Calgary, Alberta, Canada.
- Lichtenfels, J.R., Pilitt, P.A. & Lancaster, M.B. (1988). Systematics of the nematodes that cause ostertagiasis in cattle, sheep and goats in North America. *Veterinary Parasitology*, **27**: 3-12.
- Lindmark, D.G. (1980). Energy metabolism of the anaerobic protozoon *Giardia lamblia*. *Molecular and Biochemical Parasitology*, **1**: 1-12.
- Lindmark, D.G. & Jarroll, E.L. (1982). Pyrimidine metabolism in *Giardia lamblia* trophozoites. *Molecular and Biochemical Parasitology*, **5**: 291-296.
- Loaharanu, P. & Sornmani, S. (1991). Preliminary estimates of economic impact of liver fluke infection in Thailand and the feasibility of irradiation as a control measure. *Southeast Asian Journal of Tropical Medicine and Public Health*, **22**: 384-390.
- Loewenson, R., Mason, P.R. & Patterson, B.A. (1986). Giardiasis and the nutritional status of Zimbabwean schoolchildren. *Annals of Tropical Paediatrics*, **6**: 73-78.

- Lujan, H.D., Marotta, A., Mowatt, M.R., Sciaky, N., Lippincott-Schwartz, J. & Nash, T.E. (1995). Developmental induction of Golgi structure and function in the primitive eukaryote *Giardia lamblia*. *Journal of Biological Chemistry*, **270**: 4612-4618.
- Lujan, H.D., Mowatt, M.R. & Nash, T.E. (1996). Lipid requirements and lipid uptake by *Giardia lamblia* trophozoites in culture. *Journal of Eukaryotic Microbiology*, **43**: 237-242.
- Lykins, B.W. Jr. & Clark, R.M. (1994). U.S. drinking-water regulations: treatment technologies and cost. *Journal of Environmental Engineering*, **120**: 783-802.
- Majewska, A.C. & Kasprzak, W. (1990). Axenic *Giardia* isolates from primates and rodents. *Veterinary Parasitology*, **35**: 169-174.
- Majewska, A.C. (1994). Successful experimental infections of a human volunteer and Mongolian gerbils with *Giardia* of animal origin. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **88**: 360-362.
- Manchenko, G.P. (1994). *Handbook of detection of enzymes on electrophoretic gels*. CRC Press, Inc.
- Maniatis, T., Sambrook, J. and Fritsch, E.F. (1989). *Molecular Cloning. A Laboratory Manual* (2nd edn). Cold Spring Harbour Laboratory Press, New York.
- Marcus, L.F. (1991). The Goals and Methods of Systematic Biology. in: *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision* (Renaud Fortuner ed.) Baltimore: Johns Hopkins University Press.
- Mariaux, J. (1996). Cestode systematics: Any progress? *International Journal for Parasitology*, **26**: 231-243.
- Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B.G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Science*, **90**: 4384-4388.
- Mayr, E. (1969). *Principles of Systematic Zoology*. McGraw-Hill, New York.

- Mayrhofer, G., Andrews, R.H., Ey, P.L., Albert, M.J., Grimmond, T.R. & Merry, D.J. (1992). The use of suckling mice to isolate and grow *Giardia* from mammalian faecal specimens for genetic analysis. *Parasitology*, **105**: 255-263.
- Mayrhofer, G., Andrews, R.H., Ey, P.L. & Chilton, N.B. (1995). Division of *Giardia* isolates from humans into two genetically distinct assemblages by electrophoretic analysis of enzymes encoded at 27 loci and comparison with *Giardia muris*. *Parasitology*, **111**: 11-17.
- Mayrhofer, G. & Waight Sharma, A. (1988) The Secretory Immune Response in Rats Infected with Rodent *Giardia duodenalis* Isolates and Evidence for Passive Protection with Immune Bile. in: *Advances in Giardia Research*, p.49-54, (P.M. Wallis & B.R. Hammond eds), University of Calgary Press, Calgary.
- Medlin, L.K., Kooistra, W.H., Gersonde, R. & Wellbrock, U. (1996). Evolution of the diatoms (Bacillariophyta). II. Nuclear-encoded small-subunit rRNA sequence comparisons confirm a paraphyletic origin for the centric diatoms. *Molecular Biology and Evolution*, **13**: 67-75.
- Meloni, B.P., Lymbery, A.J. & Thompson, R.C.A. (1988). Isoenzyme electrophoresis of 30 isolates of *Giardia* from humans and felines. *American Journal of Tropical Medicine and Hygiene*, **38**: 65-73.
- Meloni, B.P., Lymbery, A.J. & Thompson, R.C.A. (1989). Characterization of *Giardia* isolates using a non-radiolabeled DNA probe, and correlation with the results of isoenzyme analysis. *American Journal of Tropical Medicine and Hygiene*, **40**: 629-637.
- Meloni, B.P., Lymbery, A.J. & Thompson, R.C.A. (1995). Genetic characterisation of isolates of *Giardia duodenalis* by enzyme electrophoresis: implications for reproductive biology, population structure, taxonomy, and epidemiology. *Journal of Parasitology*, **81**: 368-383
- Meloni, B.P. & Thompson, R.C.A. (1987). Comparative studies on the axenic in vitro cultivation of *Giardia* of human and canine origin: evidence for intraspecific variation. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **81**: 637-640.

- Meloni, B.P., Thompson, R.C.A., Hopkins, R.M., Reynoldson, J.A. & Gracey, M. (1993). The prevalence of *Giardia* and other intestinal parasites in children, dogs and cats from Aboriginal communities in the Kimberley. *The Medical Journal of Australia*, **158**: 157-159.
- Meloni, B. P., Thompson, R. C. A., Stranden, A. M., Kohler, P. & Eckert, J. (1992). Critical comparison of *Giardia duodenalis* from Australia and Switzerland using isoenzyme electrophoresis. *Acta Tropica* **50**: 115-124.
- Meneses Marcel, A. Olazabal Manso, E., Serrano Perez, H., Gonzalez Hernandez, O. & Salinas Melendez, J.A. (1994). Frequency of giardiasis in some species of domestic animals from Villa Clara province, Cuba. (Frecuencia de giardiasis en algunas especies de animales domesticos de la provincia Villa Clara, Cuba.) *Veterinaria Mexico*, **25**: 337-340.
- Milstein, T.C. & Goldsmid, J.M. (1995). The presence of *Giardia* and other zoonotic parasites urban dogs in Hobart, Tasmania. *Australian Veterinary Journal*, **72**: 154-155.
- Minchella, D.J. & Scott, M.E. (1991). Parasitism: A cryptic determinant of animal community structure. *Trends in Ecology and Evolution*. **6**: 250-254.
- Mishler, B.D. (1994). Cladistic analysis of molecular and morphological data. *American Journal of Physical Anthropology*, **94**: 143-156.
- Mitchell, G. (1991). Co-evolution of parasites and adaptive immune responses. *Parasitology Today*, **7**: A2-A5.
- Miyamoto, M.M. & Cracraft J. (1991). Phylogenetic Inference, DNA Sequence analysis, and the Future of Molecular Systematics. in: *Phylogenetic analysis of DNA sequences*. (Michael M. Miyamoto and Joel Cracraft eds.) Oxford University Press.
- Monis, P.T., Mayrhofer, G., Andrews, R.H., Homan, W.L., Limper, L. & Ey, P.L. (1996). Molecular genetic analysis of *Giardia intestinalis* isolates at the glutamate dehydrogenase locus. *Parasitology*, **112**: 1-12. **Appendix D**.

- Monis, P.T., Andrews, R.H., Mayrhofer, G., Kulda, J., Isaac-Renton, J.L. & Ey, P.L (1997). Novel lineages of *Giardia intestinalis* identified by genetic analysis of organisms isolated from dogs in Australia. *Parasitology*, (submitted). **Appendix E**.
- Montanez, C., Cervantes, L., Ovando, C. and Ortega-Pierres, M.G. (1988) Properties of *Giardia lamblia* RNAs. in: *Advances in Giardia Research* pp.181-185, (P.M. Wallis & B.R. Hammond eds.) University of Calgary Press, Calgary.
- Morand, S., Manning, S.D. & Woolhouse, M.E.J. (1996). Parasite-host coevolution and geographic patterns of parasite infectivity and host susceptibility. *Proceedings of the Royal Society of London*, **263**: 119-128.
- Morgan, U.M., Constantine, C.C., Greene, W.K. & Thompson, R.C.A. (1993). RAPD (random amplified polymorphic DNA) analysis of *Giardia* DNA and correlation with isoenzyme data. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **87**: 702-705.
- Moritz, C. & Hillis, D.M. (1996). Molecular Systematics: context and controversies. Chapter 1 in: *Molecular Systematics*. 2nd edition. (David M. Hillis, Craig Moritz and Barbara K. Mable eds.) Sinauer Associates Inc Publishers.
- Mowatt, M.R., Weinbach, E.C., Howard, T.C. & Nash, T.E. (1994). Complementation of an *Escherichia coli* glycolysis mutant by *Giardia lamblia* triosephosphate isomerase. *Experimental Parasitology*, **78**: 85-92.
- Munstermann, L.E. (1995). Mosquito systematics: current status, new trends, associated complications. *Journal of Vector Ecology*, **20**: 2, 129-138.
- Muralidharan, K. & Wakeland, K.E. (1993). Concentration of primer and template qualitatively affects products in random-amplified polymorphic DNA PCR. *BioTechniques*, **14**: 362-364.
- Murphy, R.W., Sites, J.W. Jr., Buth, D.G. & Haufler, C.H. (1996). Chapter 4. Proteins: Isozyme Electrophoresis. in: *Molecular Systematics*. 2nd edition. (David M. Hillis, Craig Moritz and Barbara K. Mable eds.), Sinauer Associates Inc Publishers.

- Nash, T.E., Herrington, D.A., Losonsky, G.A. & Levine, M.M. (1987). Experimental human infections with *Giardia lamblia*. *Journal of Infectious Disease*, **156**: 974-984.
- Nash, T.E., McCutchan, T., Keister, D., Dame, J.B., Conrad, J.D. & Gillin, F.D. (1985). Restriction-endonuclease analysis of DNA from 15 *Giardia* isolates obtained from humans and animals. *Journal of Infectious Diseases*, **152**: 64-73.
- Nash, T.E. & Mowatt, M.R. (1992). Identification and characterisation of a *Giardia lamblia* group-specific gene. *Experimental Parasitology*, **75**: 369-378.
- Nei, M. (1972). Genetic distance between populations. *American Naturalist*, **106**: 283-292.
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**: 583-590.
- Nei, M. (1991). Relative efficiencies of different tree making methods for molecular data. in *Phylogenetic analysis of DNA sequences*. (Michael M. Miyamoto and Joel Cracraft eds.) ,Oxford University Press.
- Nelson, K., Whittam, T.S. & Selander, R.K. (1991). Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (gapA) in natural populations of *Salmonella* and *Escherichia coli*. *Proceedings of the National Academy of Science*, **88**: 6667-6671.
- Noble, E.R. (1960). Fishes and their parasite-mix as objects of ecological studies. *Ecology*, **9**: 593-596.
- Nolan, T.J. & Smith, G. (1995). Time series analysis of the prevalence of endoparasitic infections in cats and dogs presented to a veterinary teaching hospital. *Veterinary Parasitology*, **59**: 87-96.
- Noller, H.F. (1984). Structure of Ribosomal RNA. *Annual Review of Biochemistry*, **53**:119-162.

- Ongerth, J.E., Hunter, G.D. & Dewalle, F.B. (1995). Watershed use and *Giardia* cyst presence. *Water Research* **29**: 1295-1299.
- Ortega-Pierres, G., Alonso, R., Cervantes, L. & Montanez, C. (1990). Characterisation of two DNA populations of *Giardia lamblia*. *Molecular Microbiology*, **4**: 1985-1991.
- Page, R.D.M. (1990). Temporal congruence and cladistic analysis biogeography and cospeciation. *Systematic Zoology*, **39**: 205-226.
- Page, R.D.M. (1991). Clocks, clades and cospeciation: comparing rates of evolution and timing of cospeciation events in host-parasite assemblages. *Systematic Zoology*, **40**: 188-198.
- Page, R.D.M. (1993). Parasites, phylogeny and cospeciation. *International Journal for Parasitology*, **23**: 499-506.
- Page, R.D.M. (1994) Parallel phylogenies: Reconstructing the history of host-parasite assemblages. *Cladistics*, **10**: 155-173.
- Page, R.D.M. (1996). Treeview - an application to display phylogenetic trees on personal computers. *Computers in the Applied Biosciences*, **12**: 357-58.
- Paget, T.A., Jarroll, E.L., Manning, P., Lindmark, D.G. & Lloyd, D. (1989). Respiration in the cysts and trophozoites of *Giardia muris*. *Journal of General Microbiology*, **135**: 145-154.
- Paterson, H.E.H., (1985). The recognition concept of species. pp. 21-29. in: *Species and speciation*. (E.S. Vrba, ed.), Transvaal Museum Monograph No. 4, Pretoria, South Africa.
- Patton, J.C. & Avise, J.C. (1983). An empirical evaluation of qualitative Hennigian analyses of protein electrophoretic data. *Journal of Molecular Evolution*, **19**: 244-254.
- Peattie, D.A., Alonso, R.A., Hein, A. & Caulfield, J.P. (1989). Ultrastructural Localisation of Giardins to the Edges of Disk Microribbons of *Giardia lamblia* and the Nucleotide and

Deduced Protein Sequence of Alpha Giardin. *The Journal of Cell Biology*, **109**: 2323-2335.

Pfeiffer, H. & Supperer, R. (1976). *Giardiasis* in dogs and its occurrence in Austria. (Über den Giardiabefall der Hunde und sein Auftreten in Österreich.) *Wiener Tierärztliche Monatsschrift*, **63**: 1-6.

Pickering, L.K., Woodward, W.E., DuPont, H.L. & Sullivan, P. (1984). Occurrence of *Giardia lamblia* in children in day care centers. *Journal of Paediatrics*, **104**: 522-526.

Poley, J.R. (1983). The small bowel mucosa in disease states characterised by chronic diarrhoea: observations by scanning electron microscopy. *Scanning Electron Microscopy*, (Pt 3): 1293-1306.

Pospisilova, D. & Svobodova, V. (1992). Problems of giardiasis in people keeping dogs and cats. (Problematika giardiozy u chovatelů psů a koček.) *Ceskoslovenska Epidemiologie, Mikrobiologie, Imunologie*, **41**: 106-108.

Price, P.W. (1977). General concepts on the evolutionary biology of parasites. *Evolution*. **31**: 405-420

Price, P.W. (1980). Evolutionary biology of parasites. in: *Monographs in population biology* (R.M. May ed.) Princeton University Press, Princeton, N.J.

Proctor, E.M., Isaac-Renton, J.L., Boyd, J., Wong, Q. & Bowie, W.R. (1989). Isoenzyme analysis of human and animal isolates of *Giardia duodenalis* from British Columbia, Canada. *American Journal of Tropical Medicine and Hygiene*, **41**: 411-415.

Quintans, J. (1994). Immunity and inflammation: the cosmic view. *Immunology and Cell Biology*, **72**: 262-6.

Rahman, W.A. (1990) Prevalence of *Giardia* in dogs in Malaysia: survey of a residential housing estate. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **84**: 805.

- Reiner, D.S., Hetsko, M.L. & Gillin, F.D. (1995). A lipoprotein-cholesterol-albumin serum substitute stimulates *Giardia lamblia* encystation vesicle formation. *Journal of Eukaryotic Microbiology*, **42**: 622-627.
- Reiner, D.S., McCaffery, M. & Gillin, F.D. (1990). Sorting of cyst wall proteins to a regulated secretory pathway during differentiation of the primitive eukaryote, *Giardia lamblia*. *European Journal of Cell Biology*, **53**: 142-153.
- Richardson, B.J., Baverstock, P.R. & Adams, M. (1986). *Allozyme Electrophoresis: A Handbook for Animal Systematics and Population Studies*. Sydney: Academic Press.
- Rijpstra, A.C. (1975). Results of duplicated series of stool-examinations for all intestinal parasites by five different methods in school-children in East Africa with remarks on serological aspects of amoebiasis and schistosomiasis. *Annales de la Societe Belge de Medecine Tropicale*, **55**: 415-424.
- Riley, I.T., Reardon, T.B. & McKay, A.C. (1989) Electrophoretic resolution of species boundaries in seed-gall nematodes, *Anguina* spp. (Nematoda: Anguinidae), from some graminaceous hosts in Australia and New Zealand. *Nematologica*, **34**: 401-411.
- Roberts-Thomson, I.C., Stevens, D.P., Mahmoud, A.A. & Warren, K.S. (1976). Giardiasis in the mouse: an animal model. *Gastroenterology*, **71**: 57-61.
- Rogers, J.S. (1972). Measures of genetic similarity and genetic distance. *Studies in Genetics VII Univ. Texas Publ. 7213*, pp. 145-153.
- Rohde, K. (1990). Phylogeny of platyhelminthes, with special reference to parasitic groups. *International Journal for Parasitology*, **20**: 979-1007.
- Rozario, C., Morin, L., Roger, A.J., Smith, M.W. & Muller, M. (1996). Primary structure and phylogenetic relationships of glyceraldehyde-3-phosphate dehydrogenase genes of free-living and parasitic diplomonad flagellates. *Journal of Eukaryotic Microbiology*, **43**: 33-340.

- Rzhetskii, Aiu, Rodin, S.N. & Ratner, V.A. (1987). Adaptive significance of the limited gene expression of the immune system (Ob adaptivnom znachenii ogranichennoi ekspressii genov immunnoi sistemy). *Genetika*, **23**: 2175-82.
- Rzhetsky, A. & Nei, M. (1992). A simple method for estimating and testing minimum-evolution trees. *Molecular Biology and Evolution*, **9**: 945-967.
- Rzhetsky, A. & Nei, M. (1993). Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Molecular Biology and Evolution*, **10**: 1073-1095.
- Rzhetsky, A. & Nei, M. (1994). METREE: a program package for inferring and testing minimum-evolution trees. *Computer Applications in Biosciences*, **10**: 409-412.
- Sarafis, K. & Isaac-Renton, J. (1993). Pulsed-field gel electrophoresis as a method of biotyping of *Giardia duodenalis*. *American Journal of Tropical Medicine and Hygiene*, **48**: 134-144.
- Sawyer, P.R., Brogden, R.N., Pinder, R.M., Speight, T.M. & Avery, G.S. (1976). Tinidazole: a review of its antiprotozoal activity and therapeutic efficacy. *Drugs*, **11**: 423-440.
- Schad, G.A. (1963). Niche diversification in a parasitic species flock. *Nature*, **198**: 404-406.
- Schad, G.A. (1965). Immunity, competition, and natural regulation of helminth populations. *American Naturalist*, **100**: 359-364.
- Schierwater, B. (1995). Arbitrarily amplified DNA in systematics and phylogenetics. *Electrophoresis*, **16**: 1643-1647.
- Schlegel, M., Elwood, H.J. & Sogin, M.L. (1991). Molecular evolution in hypotrichous ciliates: sequence of the small subunit ribosomal RNA genes from *Onychodromus quadricornutus* and *Oxytricha granulifera* (Oxytrichidae, Hypotrichida, Ciliophora). *Journal of Molecular Evolution*, **32**: 64-69.

- Schofield, P.J., Costello, M., Edwards, M.R. & O'Sullivan, W.J. (1990a). The arginine dihydrolase pathway is present in *Giardia intestinalis*. *International Journal for Parasitology*, **20**: 697-699
- Schofield, P.J., Edwards, M.R., Grossman, G. & Tutticci, E.A. (1995). Amino acid exchange activity of the alanine transporter of *Giardia intestinalis*. *Experimental Parasitology* **80**: 124-132.
- Schofield, P.J., Kranz, P. & Edwards, M.R. (1990b). Does *Giardia intestinalis* need glucose as an energy source? *International Journal for Parasitology*, **20**: 701-703
- Schupp, D.G., Reiner, D.S., Gillin, F.D. and Erlandsen, S.L. (1990) In vitro encystation of *Giardia*. Chapter 8, "*Giardiasis*" (E.A. Meyer, ed) 1990 Elsevier Science Publishers B.V.(Biomedical Division).
- Seiler, M., Eckert, J. & Wolff, K. (1983). *Giardia* and other intestinal parasites in dogs and cats in Switzerland. (*Giardia* und andere Darmparasiten bei Hund und Katze in der Schweiz.) *Schweizer Archiv fur Tierheilkunde*, **125**: 137-148.
- Selander, R.K., Beltran, P. & Smith N.H. (1991). Evolutionary Genetics of *Salmonella*. in *Evolution at the Molecular Level* (R.K. Selander, A.G. Clark & T.S. Whittan, eds.) Sinauer Associates inc. pp. 23-57.
- Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N., & Whittam, T.S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and Environmental Microbiology*, **51**: 873-84.
- Shaw, P.K., Brodsky, R.E., Lyman, D.O., Wood, B.T., Hibler, C.P., Healy, G.R., Macleod, K.I.E., Stahl, W. & Schultz, M.G. (1977). A communitywide outbreak of giardiasis with evidence of transmission by a municipal water supply. *Annals of Internal Medicine*, **87**: 426-432.
- Sheffield, H.G. & Bjorvat, B. (1977). Ultrastructure of the cyst of *Giardia lamblia* *American Journal of Tropical Medicine and Hygiene*, **26**: 23-30.

- Siddall, M.E. (1995). Phylogeny of adeleid blood parasites with a partial systematic revision of the haemogregarine complex. *Journal of Eukaryotic Microbiology*, **42**: 116-125.
- Siddall, M.E., Hong, H. & Desser, S.S. (1992). Phylogenetic Analysis of the Diplomonadida (Wenyon, 1926) Brugerolle, 1975: Evidence for Heterochrony in Protozoa and Against *Giardia lamblia* as a "Missing Link". *The Journal of Protozoology*, **39**: 361-367.
- Siddall, M.E., Martin, D.S., Bridge, D., Desser, S.S. & Cone, D.K. (1995) The demise of a phylum of protists: phylogeny of Myxozoa and other parasitic cnidaria. *Journal of Parasitology*, **81**: 961-967.
- Simms, E.L. (1996). The evolutionary genetics of plant-pathogen systems. Understanding the coevolution of hosts and parasites is key to understanding their ecology. *BioScience*, **46**: 136-145
- Simpson, G.G. (1951). The species concept. *Evolution*, **5**: 285-298.
- Smith, J.J., Scott-Craig, J.S., Leadbetter, J.R., Bush, G.L., Roberts, D.L. & Fulbright, D.W. (1994). Characterisation of random amplified polymorphic DNA (RAPD) products from *Xanthomonas campestris* and some comments on the use of RAPD products in phylogenetic analysis. *Molecular Phylogenetics and Evolution*, **3**: 135-145.
- Smithies, O. (1955). Zone electrophoresis in starch gels: Group variations in the serum proteins of normal individuals. *Biochemical Journal*, **61**: 629-641.
- Smothers, J.F., von Dohlen, C.D., Smith, L.H. Jr. & Spall, R.D. (1994). Molecular evidence that the myxozoan protists are metazoans. *Science*, **265**: 1719-1721.
- Sneath, P.H.A. & Sokal, R.R. (1973). *Numerical Taxonomy* W.H. Freeman & Co., San Francisco.
- Snider, D.P., Gordon, J., McDermott, M.R. & Underdown, B.J. (1985). Chronic *Giardia muris* infection in anti-IgM treated mice. *Journal of Immunology*, **134**: 4153-4162.

- Sogayar, M.I.L., Correa, F.M. & De, A. (1984). *Giardia* in dogs in Botucatu, Sao Paulo State, Brazil: a comparative study of canine and human species. *Revista de Ciencias Biomedicas*, **5**: 69-73.
- Sogayar, M.I.L. & Gregorio, E.A. (1989). *Giardia muris* and *Giardia duodenalis* groups: ultrastructural differences between the trophozoites. *Revista do Instituto de Medicina Tropical de Sao Paulo*, **31**: 242-247.
- Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A. & Peattie, D.A. (1988). Phylogenetic meaning of the kingdom concept. An unusual ribosomal RNA from *Giardia lamblia*. *Science*, **243**: 75-77.
- Sokal, R.R. & Sneath, P.H.A. (1963). *Principals of Numerical Taxonomy* W.H. Freeman & Co., San Francisco.
- Spratt, D.M. (1990). The role of helminths in the biological control of mammals. *International Journal for Parasitology*, **20**: 543-50.
- Stehr Green, J.K., Murray, G., Schantz, P.M. & Walquist, S.P. (1987). Intestinal parasites in pet store puppies in Atlanta. *American Journal of Public Health*, **77**: 345-346.
- Stranden, A.M., Eckert, J. & Kohler, P. (1990). Electrophoretic characterisation of *Giardia* isolated from humans, cattle, sheep, and a dog in Switzerland. *Journal of Parasitology*, **76**: 660-668
- Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology Evolution*, **13**: 964-969.
- Strimmer, K., Goldman, N. & von Haeseler, A. (1997). Bayesian probabilities and quartet puzzling. *Molecular Biology and Evolution*, **14**: 210-211.
- Sugano H., Fukase, T., Chinone, S. & Itagaki, H. (1989). Giardiasis in puppies from a breeder's kennel. *Journal of the Japan Veterinary Medical Association*, **42**: 68-71.

- Swabby, K.D., Hibler, C.P. & Wegrzyn, J.G. (1988) Infection of Mongolian gerbils (*Meriones unguiculatus*) with *Giardia* from human and animal sources. pp. 75-77 in: *Advances in Giardia research*. (P.M. Wallis and B.R. Hammond eds.) Calgary, Alberta, Canada.
- Swan, J.M., & Thompson, R.C.A. (1986). The prevalence of *Giardia* in dogs and cats in Perth, Western Australia. *Australian Veterinary Journal*, **63**: 110-112.
- Swofford D.L. (1993). *PAUP: Phylogenetic Analysis Using Parsimony*, Version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.
- Swofford, D.L., Olsen, G.J., Waddell, P.J. & Hillis, D.M. (1996). Phylogeny Inference. Chapter 11 in: *Molecular Systematics*. 2nd edition. (David M. Hillis, Craig Moritz and Barbara K. Mable eds) Sinauer Associates Inc Publishers, 1996.
- Sykes, T.J. & Fox, M.T. (1989). Patterns of infection with *Giardia* in dogs in London. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **83**: 239-240.
- Szymura, J.M. & Barton, N.H. (1986). Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina orientalis* and *B. variegata*, near Cracow in southern Poland. *Evolution*, **40**: 1141-1159.
- Tait, A. (1983). Sexual processes in the kinetoplastida. *Parasitology*, **86**: 29-57.
- Templeton, A.R. (1989). The meaning of species and speciation: a genetic perspective. pp 3-27 in: *Speciation and Its Consequences*. (D. Otte and J.A. Endler eds.) Sinauer Associates, Sunderland, MA.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**: 4673-4680.

- Thompson, R.C.A., Lymbery, A.J. and Meloni, B.P. (1990). Genetic variation in *Giardia* Kunstler, 1882: taxonomic and epidemiological significance. *Protozoological Abstracts*, **14**: 1-28.
- Thompson, R.C., Lymbery, A.J., Pearce, D.A., Finn, K.C., Reynoldson, J.A. & Meloni, B.P. (1996). *Giardia duodenalis*: exposure to metronidazole inhibits competitive interactions between isolates of the parasite *in vitro*. *Journal of Parasitology*, **82**: 679-683.
- Thompson, R.C.A. & Meloni, B.P. (1993). Molecular variation in *Giardia*. *Acta Tropica*, **53**: 167-184.
- Tibayrenc, M. Kjellberg, F. & Ayala, F.J. (1991). A clonal theory of parasitic protozoa: The population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proceedings of the National Academy of Science*, **87**: 2414-2418.
- Tibayrenc, M., Neubauer, K., Barnabe, C., Guerrini, F., Skarecky, D. & Ayala, F.J. (1993). Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proceedings of the National Academy of Sciences*, **90**: 1335-1339.
- Tonks, M.C., Brown, T.J. & Ionas, G. (1991). *Giardia* infection of cats and dogs in New Zealand. *New Zealand Veterinary Journal*, **39**: 33-34.
- Uji, A. Wallis, P.M. & Wenman, W.M. (1988). Comparison of *Giardia* Isolates by DNA-DNA Hybridisation. pp.165-167 in: *Advances in Giardia Research* (P.M. Wallis & B.R. Hammond eds.) University of Calgary Press, Calgary.
- Upcroft, J.A., Boreham, P.F.L. & Upcroft, P. (1989). Geographic variation in *Giardia* karyotypes. *International Journal of Parasitology*, **19**: 519-527.
- Upcroft, J.A., Boreham, P.F.L., Campbell, R.W., Shepherd, R.W. & Upcroft, P. (1995). Biological and genetic analysis of a longitudinal collection of *Giardia* samples derived from humans. *Acta Tropica*, **60**: 35-46.

- Upcroft, J.A., Chen, N. & Upcroft, P. (1996). Mapping variation in chromosome homologues of different *Giardia* strains. *Molecular and Biochemical Parasitology*, **76**: 135-143.
- Upcroft, J.A., Healey, A., Mitchell, R., Boreham, P.F. & Upcroft, P. (1990b). Antigen expression from the ribosomal DNA repeat unit of *Giardia intestinalis*. *Nucleic Acids Research*, **18**: 7077-7081.
- Upcroft, J.A. & Upcroft, P. (1994). Two distinct varieties of *Giardia* in a mixed infection from a single human patient. *Journal of Eukaryotic Microbiology*, **41**: 189-194.
- Upcroft, P.(1991). DNA fingerprinting of the human intestinal parasite *Giardia intestinalis* with hypervariable minisatellite sequences. *EXS*, **58**: 70-84.
- Upcroft, P., Mitchell, R. & Boreham, P.F. (1990a). DNA fingerprinting of the intestinal parasite *Giardia duodenalis* with the M13 phage genome. *International Journal of Parasitology*, **20**: 319-323.
- Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K. & Swings, J. (1996). Polyphasic Taxonomy, a Consensus Approach to Bacterial Sytematics. *Microbiological Reviews*, **60**: 407-438.
- Vickerman, K. (1990). Chapter 14d. Phylum Zoomastigina. Class Diplomonadida. in *Handbook of Protoctista*. (L. Margulis, J.O. Corliss, M. Melkonian and D.J. Chapman eds.) Jones and Bartlett Publishers, Boston.
- Vinayak, V.K., Kum Kum, Venkateswarlu, K., Khanna, R. and Mehta, S. (1987). Hypogammaglobulinaemia in Children with Persistent Giardiasis. *Journal of Tropical Paediatrics*, **33**: 140-142.
- Visvesvara, G.S., Dickerson, J.W. & Healy, G.R. (1988). Variable infectivity of human-derived *Giardia lamblia* cysts for Mongolian gerbils (*Meriones unguiculatus*). *Journal of Clinical Microbiology*, **26**: 837-841

- Vitti, G.F., O'Sullivan, W.J. & Gero, A.M. (1987). The biosynthesis of uridine 5'-monophosphate in *Giardia lamblia*. *International Journal for Parasitology*, **17**: 805-812.
- Wallis, P.M., Buchanan-Mappin, J.M., Faubert, G.M. & Belosevic, M. (1984). Reservoirs of *Giardia* spp. in southwestern Alberta. *Journal of Wildlife Diseases*, **20**: 279-83
- Wallis, P.M., Erlandsen, S.L., Isaac-Renton, J.L., Olson, M.E., Robertson, W.J. & van Keulen, H. (1996). Prevalence of *Giardia* cysts and *Cryptosporidium* oocysts and characterization of *Giardia* spp. isolated from drinking water in Canada. *Applied and Environmental Microbiology*, **62**: 2789-2797.
- Wallis, P.M. & Wallis, H.M. (1986). Excystation and culturing of human and animal *Giardia* spp. by using gerbils and TYI-S-33 medium. *Applied and Environmental Microbiology*, **51**: 647-651.
- Weis, A.E. & Abrahamson, W.G. (1985). Potential selective pressures by parasitoids on a plant-herbivore interaction. *Ecology*, **66**: 1261-1269.
- Weiss, J.B., van Keulen, H. & Nash, T.E. (1992). Classification of subgroups of *Giardia lamblia* based upon ribosomal RNA gene sequence using the polymerase chain reaction. *Molecular and Biochemical Parasitology*, **54**: 73-86.
- Wenyon, C.M. (1926). Preface., in: *Protozoology: A manual for medical men, veterinarians and zoologists*. Bailliere, Tindall & Cox Publishers, London.
- Werman, S.D., Springer, M.S. & Britten, R.J. (1996). Nucleic Acids I: DNA-DNA Hybridization. Chapter 6 in: *Molecular Systematics*. 2nd edition. (David M. Hillis, Craig Moritz and Barbara K. Mable eds.) Sinauer Associates Inc Publishers.
- Whitehead, P. (1990). Systematics: an endangered species. *Systematic Zoology*, **39**: 179-184.
- Winsland, J.K.D., Nimmo, S., Butcher, P.D. & Farthing, M.J.G. (1989). Prevalence of *Giardia* in dogs and cats in the United Kingdom: survey of an Essex veterinary clinic. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **83**: 791-792.

- Wolff, K. & Eckert, J. (1979). *Giardia* infection of dogs and cats and its possible significance for man. (*Giardia* Befall bei Hund und Katze und dessen mogliche Bedeutung fur den Menschen.) *Berliner und Munchener Tierarztliche Wochenschrift*, **92**: 479-484.
- Woo, P.T. & Paterson, W.B. (1986). *Giardia lamblia* in children in day-care centres in Ontario, Canada, and susceptibility of animals to *G. lamblia*. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **80**: 56-59.
- Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling plans in regression analysis. *Annals of Statistics*, **14**: 1261-1295.
- Xiao, L. (1994). *Giardia* infection in farm animals. *Parasitology Today*, **10**: 436-438.
- Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Tree*, **11**: 367-372.
- Yee, J. & Dennis, P.P. (1993). The NADP-dependent glutamate dehydrogenase of *Giardia lamblia*: A study of function, gene structure and expression. *Systematic and Applied Microbiology*, **16**: 759-767.
- Zimmer, J.F. & Burrington, D.B. (1986). Comparison of four protocols for the treatment of canine giardiasis. *Journal of the American Animal Hospital Association*, **22**: 168-172.