Molecular markers, analysis and the population genetics of parasites

Clare Colleen Constantine

B.Sc (Hons.) UWA

Division of Veterinary and Biomedical Sciences

Murdoch University

Western Australia

This thesis is presented for the degree of Doctor of Philosophy of Murdoch University

2002

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

Clare Colleen Constantine

Abstract

In this study different molecular techniques are contrasted (RAPD's, allozyme, sequencing mtDNA, sequencing ribosomal spacers) and appropriate analytical methods (allelic and infinite-sites approaches; inbreeding and coalescent models) used for estimating population genetic parameters in parasites. A range of population genetic questions at different scales were chosen to emphasise the importance of tailoring techniques and analytical methods to the particular question being investigated.

The realisation that each question formulated has a particular scale means the appropriate technique and markers must be useful at that scale to attempt to answer the question. The useful scale of a technique depends several factors including the region of DNA examined, the density of sampling of the technique, and the mode of evolution of the markers. Each technique will produce a useful range of variability. Below the lower limit there is no variation, above the upper limit the variation is too high to produce useful comparisons.

Parasites are of interest for many reasons, primarily because they can cause disease and thus impact on their host's population dynamics. They are often closely associated with their hosts and may undergo co-evolution, as well as causing an ongoing immunological "arms race" with their hosts. The parasitic mode of live is found throughout nearly all taxonomic groupings and thus classical models of population genetics based on sexual, diploid vertebrates do not fit well with the entire diversity of parasite groups.

Genetic diversity within and among populations of *Echinococcus granulosus* was examined contrasting a RAPD dataset with an allozyme dataset. Two models of variation in *Echinococcus* have been proposed, those of Smyth and Rausch, and the expected genetic structure from each was compared to the observed genetic structure. The premise of Smyth’s model, predominant self-fertilisation, was supported, but the resultant pattern of genetic variation followed Rausch’s model.

RAPD data, being dominant, present challenges to analysis. An approach to overcome this dominance problem and allow standard allelic frequency analysis is described using

the selfing rate estimated from allozyme data. The RAPD data were also analysed using both band-sharing and nucleotide diversity approaches.

A population genetic study of *Ostertagia ostertagi* in the USA was extended to two different scales: within an Australian state and between the USA and Australian continents. Three alternative explanations for the observed discrepancy between genetic structure and differentiation in an important biological trait, hypobiosis, were explored. A number of programs and analyses were compared including coalescent geneflow estimates.

Variation among multiple copies of two spacer regions of rDNA was examined within individuals of *Ostertagia ostertagi*. Both the intergenic spacer and internal transcribed spacer 1 regions were found to include repeat regions, with different numbers of repeats creating length differences in clones from the same worm. Multi-copy genes present extra challenges in analysis to ensure that only homologous copies are being compared. Many studies fail to look for variation within populations or within individuals.

The two major conclusions from these examples are that:

1). The study of variation necessarily involves an implicit scale, and markers must be chosen that are appropriate to the question being explored.

2). Using several methods of analysis of genetic data allows contrasts to be made, and if different methods produce similar results gives much more confidence in the conclusions drawn. Incongruence in results leads to new questions and reexamination of the assumptions of each analysis.

Publications

Book chapter:

Thompson, R.C.A., Lymbery, A.J., Meloni, B.P., Morgan, U M, Binz, N, Constantine, C.C. and Hopkins, R.M. (1994). Molecular epidemiology of parasite infections. **In**: Biology of Parasitism (Edited by Ehrlich, R and Nieto, A). Edicio Trilce, Montevideo, Uruguay, pp 167-185.

Refereed journal articles:

Thompson, R.C.A., Constantine, C.C. and Morgan, U.M. (1998). Overview and significance of molecular methods: what role for molecular epidemiology? *Parasitology* **117**:S161-S175

Lymbery, A.J., Constantine, C.C. and Thompson, R.C.A. (1997). Self-fertilisation without genomic or population structuring in a parasitic tapeworm. *Evolution*. **51**(1):289-294.

Thompson, R.C.A., Lymbery, A.J. and Constantine, C.C. (1995). Variation in *Echinococcus*: Towards a taxonomic revision of the Genus. *Advances in Parasitology*. **35**: 145-176.

Constantine, C.C., Hobbs, R.P. and Lymbery, A.J. (1994). FORTRAN programs for analysing population structure from multilocus genotype data. *Journal of Heredity*. **85**: 336-337.

Conferences:

Constantine, C.C., M. Hankinson, A. J. Lymbery, and R. C. A. Thompson (1998). Use of RAPDs for analysing genetic variation at different scales in *Echinococcus. 73rd Annual Meeting of The American Society of Parasitologists, Kona, Hawaii, p81, August 1998*

Constantine, C.C., Blouin, M.S., Lymbery, A.J. and Thompson, R.C.A. (1997). Population genetic structure of *Ostertagia ostertagi* in Australia and the USA. *Genetics Society of Australia, Perth, p24, 1997*

Constantine, C.C., Blouin, M.S., Lymbery, A.J. and Thompson, R.C.A. (1996). Population genetic structure of *Ostertagia ostertagi* in Australia. *PASEAN , Bali, Indonesia, p 34, 1996*

Acknowledgments

Thank you to my supervisors, Professor R. C. Andy Thompson, who was always enthusiastic, and Dr. Alan J. Lymbery, who could always find a better way of writing something. Jennifer Walters helped in the fascinating task of picking worms out of abomasa, and provided accommodation. Margaret Hankinson meticulously produced the RAPD gels and scoring of bands for *Echinococcus granulosus*. Una Ryan, Richard Hopkins and Marion MacNish taught me a lot and were great fun to work with. To the members of the Parasitology group and the State Agricultural Biotechnology Center WA, thanks for creating a great atmosphere to research in. Special thanks to my family who supported me, even when it seemed I would never get finished.

Table of Contents

List of Tables

List of Figures

List of abbreviations

Fst, Fis, Fit See Table 3.5 and sections 1.5.3.3, 3.2.4.
Fst is a measure of genetic differentiation over subpopulations. The correlation between the probability that two randomly chosen gene copies picked from a subpopulation share an ancestor in the last generation relative to picked from the total population. Uses the partitioning of total genetic variation into variability within and between populations. Fst = (Fit-Fis)/(1-Fis) where t is total, i is individual and s is subdivision.

Gst See Table 3.6 and section 1.5.3.3.
A measure of genetic variation between populations relative to that within populations, a generalised version of Fst. Originally devised by Nei, with the following formula by Slatkin: Gst = 1/(4Nm[k/(k-1)]2 + 1)

Ho, Hs, Ht See Tables 3.2 and 3.5.
Ho = Average observed heterozygosity within groups. Ht = Total heterozygosity in the entire data set. Hs = Gene diversity within groups averaged over the entire data set. See Fst for general concept of subpopulation genetic differentiation.

ITS See Chapter 5 especially Figure 5.1.
Internal Transcribed Spacers are non-coding regions that are transcribed and later excised from the final gene products. There are two ribosomal ITS regions, one between 18S and 5.8S genes, and one between the 5.8S gene and 28S gene. These regions are more variable than the surrounding coding regions.

IGS See Chapter 5 especially Figure 5.1.
Inter genic spacers are non-coding regions beween genes that are not transcribed (although they often include functional motifs such as transcription initiation). Ribosomal IGS is often a large (>1kb) region which is highly variable.

Nm or NeM See Table 3.5 and section 1.5.3.4
A measure of gene flow. The product of effective population size (Ne) and effective migration rate (m). Often estimated using $N_em = 0.25(1-Fst)/Fst$ (Nei, 1987)

Nst See section 1.5.3.3.
A specific estimator equivalent to Fst (Lynch & Crease 1990).

RAPD See section 1.4.4.5 and Chapter 3 especially 3.1.4.
Random Amplified Polymorphic DNA is a molecular technique which uses short (10-16bp) single primers which bind wherever complementary sequence is found and when by chance two bind facing each other, close enough for PCR to amplify a fragment is produced which appears as a band when run on a gel. This technique can be applied to any DNA without any knowledge of its sequence. It is highly sensitive to changes in PCR conditions.

UPGMA See section 1.5.3.5.
Unweighted pair group method with arithmetic average to estimate genetic distance between groups. Produces a matrix of distance estimates that is then used by a clustering algorithm to produce a tree.

1 INTRODUCTION

1.1 The Importance of Scale

Inductive science is the study of variation to detect patterns from observations and then to discover the underlying processes which cause that pattern, that is develop hypotheses and test them. Implicit in all studies, but often not stated, is that variation is being examined at a specific temporal and spatial scale. There is no single correct scale at which to examine variation, and no description of the variability and predictability of the environment makes sense without reference to the particular range of scales that are relevant to the organisms or processes being examined (Levin, 1992). Different disciplines, such as ecology, population genetics and systematics are concerned with variation at different scales.

The problem of relating phenomena across scales is the central problem in biology and in all of science (Levin, 1992). The most challenging work is trying to integrate across discipline scales, for example trying to understand the relationship of micro- and macro-evolution. The key is understanding how information is transferred between scales (Levin, 1992). Patterns at one level may be the result of the sum of processes at lower levels, or may also be affected by new, higher level process, in which case they are said to exhibit emergent properties (Sole *et al.*, 1999).

Some lower level processes may not operate at higher levels. For example, fine scale structuring of relatedness due to limited seed dispersal or cloning in plant sub-populations may be irrelevant at higher scales if sufficient long distance dispersal occurs (for examples see Cloarec *et al.*, 1999; Williams and Waser, 1999 and Lasserre *et al.*, 1996). Similarly, over a longer time frame, short term structuring due to temporary gene flow barriers may be smoothed out by rare bursts of gene flow (for example recombination in *Trypanosoma brucei* is rare over a short time-scale but plays a notable role on an evolutionary scale: Tibayrenc and Ayala, 2000).

1.2 The Scale of Genetic Studies

Genetic studies are concerned with variation across a wide range of spatial and temporal scales from within individuals, among individuals within a population, among populations, to among species and higher level taxa. Different processes may operate at these different levels. Molecular genetics is principally concerned with the processes of development, mutation, aging and the transfer of genes from one generation to the next. Population genetics refers to the distribution of genetic variation among individuals within a population, and among populations within a species.

Species by most definitions are considered to evolve independently, that is they can change, bifurcate to produce new species, or go extinct but there is no genetic exchange between species. There is increasing evidence that this does not hold true for all organisms; in particular protozoans and bacteria have been found to undergo horizontal transfer of genes (Doolittle, 1999). Systematics is the study of relationships among species. Micro-evolution deals with changes seen within species, macro-evolution is speciation and higher level divergence and extinction of species. Macro-evolution is considered by some to involve more than the sum of micro-evolutionary processes (Sole *et al.*, 1999).

The ideal marker to examine genetic variation depends on the scale of the question being asked. Studies of change within an individual require a technique that can distinguish single base-pair changes. Studies to find variants causing certain traits need markers spread throughout the genome. Species diagnosis would ideally use a marker that is invariant within a species but distinguishes it from all other species. Higher level taxonomy requires markers that are present in all species but differ relative to the amount of time since they shared a common ancestor.

1.3 Genetic Studies of Parasites

Parasites have traditionally been considered to be specialised, even degenerate, and greatly constrained by their dependence on their host. Genetic studies of parasites have been mostly concerned with identification and diagnosis of species, with relatively little work below the species level, although that is rapidly changing. Price (1977; 1980) proposed a general model for the evolution of parasites that considered parasite species to be comprised of "small, relatively homozygous populations with little gene flow between populations, which results in many specialised races, rapid evolution and speciation without geographic isolation, and an abundance of sister species". These predictions have not been realised for most endoparasites studied so far (Nadler, 1990), where the genetic variability is similar to that observed in free-living species.

Genetic studies of parasites can be used to address a wide range of questions of both theoretical and practical importance (Thompson *et al.*, 1998). Parasitism is one of the most successful modes of life, as measured by the number of times it has evolved and the number of parasitic species (Poulin and Morand, 2000). Parasites are increasingly being recognised as agents of evolution, directly affecting their hosts, and having general ecosystem effects. Parasites have also been implicated in general evolutionary processes such as the predominance of sexual over asexual reproduction, despite a two-fold disadvantage (Hamilton *et al.*, 1990). There are a number of interesting aspects of genetic studies of parasites.

1.3.1 Parasites cause disease

Of major interest is the fact that parasites can cause diseases. By definition parasites are organisms that benefit from an obligatory (at least for one stage of their life cycle), close physical association with a single host, and the relationship causes the host to suffer. Parasitic diseases in humans and domesticated animals are a major source of mortality and morbidity (Larsen *et al.*, 1995; Walker and Fox-Rushby, 2000). It is not always straightforward to attribute sickness unequivocally to parasite infections. Some parasites may have little effect on their host unless the host is immuno-compromised or nutrient-restricted. Closely related parasite strains may vary greatly in their virulence, one strain causing no symptoms while a very similar one may cause significant disease.

Detecting and distinguishing closely related parasites is an essential first step to any control program. Knowledge of all possible hosts and whether different host species harbour distinct strains is critical to understanding the epidemiology of the parasite. Strain typing, using fingerprinting markers, can help to trace the source of outbreaks. Genetic markers can be developed as diagnostic tests or may be linked to traits of interest such as virulence or resistance (Morgan *et al.*, 1996; Costa *et al.*, 1997; Knight *et al.*, 1999).

Understanding the genetic structure, that is the amount and distribution of genetic variation within and between populations, of a parasite species is essential for effective control of parasitic disease. Resistance to current drugs used to treat parasitic diseases is developing faster than new drugs can be found (Sangster, 1999). Research is ongoing into alternatives to drenching, including: vaccines (Knox, 2000); developing nematode resistant sheep (Windon, 1996); and biological control of larval nematodes (Larsen, 1999). Nematodes have already demonstrated their ability to rapidly adapt, so it is quite likely that these approaches will also suffer from the rapid development of resistance.

Understanding patterns of genetic structure enables one to predict the consequences of local control programs on long-term population size and on the development of resistance to control agents. This approach has been used to develop guidelines to minimise the rate of development of resistance to drenches (Williams, 1997). Thus if alternative control approaches turn out to be feasible, similar guidelines should be established from the beginning to minimise the development of resistance.

Host-switching, where a parasite moves to a new host species, is of interest both as an evolutionary challenge and because it can cause the occurrence of new diseases especially since parasitic diseases are often more virulent in new hosts with which they have not co-evolved. There are also general evolutionary questions about the effects of parasites on their host populations, such as regulating population size or causing local extinction.

1.3.2 Host/parasite relationship

The fact that parasites spend most of their lives within or attached to a host means the processes affecting their hosts also affect parasites. For example, artificial movement of

domestic hosts causes gene flow in their parasites that tends to break up population genetic structures. Studies of several parasites of domestic hosts found high levels of gene flow and little structure across the entire range, in contrast to parasites of natural populations of wild hosts (Blouin *et al.*, 1995). In addition, spending their entire reproductive phase within a single host can cause inbreeding due to a small local breeding pool. The close link between some parasites and their hosts may provide two independent estimates of their shared phylogeny and gene flow between host/parasite populations. Parasites can also be used to provide information about host populations (Mosquera *et al.*, 2000).

1.3.3 Diversity of parasites

Parasites exhibit a huge range of life cycles and breeding systems even between closely related species (Poulin and Morand, 2000), this makes them ideal to test for general effects. However, they share common challenges such as evading immune responses, finding new hosts, penetrating or attaching to the host. To survive, a parasite must not redirect so many of the host resources that it dies before the parasite is ready to infect new hosts or the host population becomes extinct, but it must produce sufficient infectious offspring so that on average at least one new host is infected. The different methods by which parasites meet these challenges can reveal information about host immune systems, host population size dynamics, and epidemics.

1.4 Molecular markers for genetic studies

1.4.1 Molecular markers versus other markers

Markers are heritable variable traits that indicate the underlying genetic differences between individuals, populations, species or higher taxa. Traditional non-molecular markers include certain morphological traits, ecological traits, and fruit fly giant chromosomes. Non-molecular markers may have some severe limitations: their inheritance can only be proven by controlled breeding trials (some molecular markers also do not follow Mendelian inheritance); a trait in one group may not be present in other groups, making comparison between groups impossible; they may not reflect actual genetic variation due to phenotypic plasticity (same genotype producing different

phenotypes in different environments), or the same character state may be the result of analogous traits (similarity due to convergence rather than shared ancestry). Morphological and ecological traits should not be disregarded since they reflect the results of selection and there are examples of clear morphological and behavioural differentiation without genetic differentiation at neutral loci (Bensch *et al.*, 1999).

The key to the explosion of information in all areas of genetics in recent years, however, has been the development of molecular genetic markers utilising the polymerase chain reaction (PCR). Although molecular markers have been available for several decades, their use has grown very rapidly in the last few years. The variety of methods available, including many that will work without prior knowledge of the genome and on small amounts of DNA, means that there are now few groups of organisms for which no molecular data exists.

1.4.2 Markers and scale

Different molecular markers provide genetic information which is suited to only a certain range of questions at a particular scale (Waycott, 1998; Shaw *et al.*, 1999; Sole *et al.*, 1999). A very fine scale marker can detect and quantify differences between individuals but it becomes useless when applied to members of different species. For example Shaw *et al.*, (1999) used microsatellites because previous studies had found other markers to have uninformative levels of genetic variation in Atlantic herring. They found that microsatellites did detect structuring at a finer scale, but were less informative at larger scales of divergence. Similarly, a more conserved marker can be used to infer phylogenies of species but it may not detect any variation at all below the species level.

Different markers may therefore reveal different patterns simply because they detect variability at different scales (Raybould *et al.*, 1999). For example, many allozyme studies on marine fish have failed to detect any geographic structuring, although more rapidly evolving markers have in some cases revealed fine-scale structure (Neigel, 1994). Furthermore, it is likely that each evolutionary/genetic process may have somewhat different effects on each type of marker. For example, the effective population size for mitochondrial DNA (mtDNA) is 1/4 of that for nuclear markers,

therefore any processes that are population-size dependent will vary in their effect on mtDNA versus nuclear regions.

It is thus essential to choose a technique which is appropriate to the question being asked (Yakubu *et al.*, 1999). Studies using more than one molecular technique or region are thus much more powerful than studies that use a single technique, not only because they provide additional estimates of the pattern of variation but also because contradictions in multiple patterns can give insight into evolutionary processes.

1.4.3 Regions of DNA

Mutations may occur throughout the genome, however, the rate of incorporation of mutations differs greatly in different regions of the genome. Different regions of the genome perform different functions, for example mtDNA is involved with energy metabolism; ribosomal DNA (rDNA) with the translation and construction of proteins; chloroplast DNA (cpDNA) with photosynthesis. Stretches of DNA lying between genes and introns, regions which are cut out of precursor messenger RNA (mRNA), do not directly code for proteins. Although these non-coding regions are not translated, there is increasing evidence that secondary structure elements are conserved and these regions have significant functional roles. These differences affect the use and analysis of different markers (Table 1.1).

1.4.3.1 Coding regions versus non-coding regions

Coding regions are in general less variable than non-coding regions, presumably due to selection against changes that render the protein less functional than the original version. Insertions or deletions in coding regions may cause frame shift mutations that can completely change the rest of the protein amino acid sequence. Some nucleotide substitutions, however, do not cause changes in amino acids due to the redundancy of the genetic code; these are called silent or synonymous mutations.

1.4.3.2 Nuclear versus non-nuclear DNA

Nuclear DNA is by far the largest portion of an organisms total DNA, it is located in the nucleus and is usually organised into chromosomes. Ribosomal DNA (rDNA) genes are

a subset of nuclear DNA, but are quite distinct functionally and structurally. RNA-coding genes produce transfer RNAs (tRNA), ribosomal RNAs (rRNA) and others, such as small nuclear RNA, which make up the machinery of protein synthesis (Nei, 1987). rDNA is a multigene family, there are many copies in each genome. There appear to be some processes, for example concerted evolution, which affect multigene families, so that the different copies are not independent.

Mitochondrial DNA (mtDNA) is self-replicating DNA found in mitochondria. It is inherited predominantly maternally, although paternal and biparental inheritance occurs in some species (Quesada *et al.*, 1999; Jannotti-Passos *et al.*, 2001, Lunt and Hyman 1997). Generally, mtDNA can be considered as non-recombining and completely independent of nuclear DNA, although there are proven cases of mtDNA gene copies being incorporated into nuclear DNA (Perna and Kocher, 1996), and evidence of recombination in parasitic nematodes. The evolutionary pattern of variation of plant mtDNA differs greatly from animal mtDNA: gene order evolves much more rapidly, but nucleotide sequence substitution is up to 100-fold slower in plants than in animals. The size variation and gene order variability has greatly limited the usefulness of plant mtDNA for molecular systematic studies.

Chloroplast DNA (cpDNA) is transmitted maternally in most plants (biparentally in some and paternally in a few (Morgensen, 1988)), and varies in size from 120 to 217 kb. The rate of cpDNA evolution appears to be slow and it is often used for studies of higher level plant systematics (Avise, 1994; p 69).

Table 1.1 – Regions of DNA, their function, variability and limitations

Region	Function	Variability/Limitations
Nuclear coding	Genes translated into proteins	Low-medium
Non-coding	Not encoding proteins	High
rDNA	Produce RNA for protein synthesis machinery	Low-medium
mtDNA	Independent genome involved with energy production	High/Too high in plants to be useful for systematics
cpDNA	Independent genome involved with photosynthesis	Low/Plants only

1.4.4 Molecular techniques that produce markers

There is a range of marker types available and more are rapidly being developed. They all have different characteristics such as cost, ease of use and type of variation detected. There is no perfect marker that can be used to answer all questions. The choice of marker to use depends not only on considerations of cost and ease-of-use (Table 1.2), but most importantly on the questions for which answers are being sought (Yakubu *et al.*, 1999) (Table 1.3). Levels of variation for any marker are in large part based on the rate and mode of mutation relative to the effects of genetic drift (Scribner *et al.*, 1994). The different techniques in common use are discussed briefly below.

Table 1.2 – Characteristics of selected molecular techniques in terms of difficulty/cost, sampling, variability, prior information required and sensitivity.

Method	Difficulty/ Cost	Genome sampling	Variability	Prior information needed	Sensitivity
Hybridization & micro-complement fixation	Low	Complete /poor	Low	No	Low
Protein sequence	Medium	Good	Medium	Little	Low
Allozymes	Low	Good	High	None	Low
RFLP and restriction sites	Medium	High	Good	Little	Good (radioactive)
RAPD	Medium	High	High	None	High, but low specificity
Mini/micro satellites	High	Poor	Very high	Much	High
Sequence	High	Poor	Very high	Much	High

Table 1.3 – Scale of uses of selected molecular techniques (X – not useful; √ may be useful; √√ usually useful).

Method	Parentage	Genetic structure	Intraspecific phylogeny	Inter-specific diagnosis	Phylogeny
Hybridization & micro-complement fixation	X	X	X	√	√
Protein sequence	X	√	√	√√	√√
Allozymes	X	√√	X	√	√
RFLP and Restrict. sites	√√	√√	√ (sites only)	√	√(sites only)
RAPD	√√	√ (2)	X	√	X
Mini/micro satellites	√√	√	X	X	X
Sequence(1)	√	√	√√	√	√

(1) Note that sequence data are limited to relatively short stretches and usefulness is highly dependent on choosing a region with an appropriate level of variation.
(2) RAPD markers are dominant, which means heterozygotes cannot be detected, complicating analysis.

1.4.4.1 Hybridisation & immunological

Hybridisation estimates the similarity of two genomes by measuring how well they bind to each other, as indicated by higher melting point of the homo-duplex molecule. Immunological techniques measure similarity by degree of cross-reaction using micro-complement fixation, immunodiffusion, or precipitation. Although they examine only a small part of the genome, it has been shown that these immunological reactions are highly specific. These markers thus give only a single measure of distance/divergence (or similarity) between samples.

1.4.4.2 Allozymes

Allozymes are different forms of an enzyme produced by different alleles at a locus. They are detected by variations in electrophoretic migration of enzymes (stained for by using the enzyme function to complete a colour change reaction). This normally produces allelic data for which the heterozygotes can be detected (the markers are therefore co-dominant). Since these are functional enzymes, allozyme markers may be affected by selection (Ochando and Ayala, 1999). Not every nucleotide substitution

results in a change in amino acid sequence (silent or synonymous mutations) and further not all amino acid changes result in detectable change in electrophoretic migration of the enzyme. Thus allozyme analysis does not detect all genetic variation that is present at that locus.

1.4.4.3 Protein sequence

Protein sequences are obtained either directly by the order of amino acids in a mature protein, or indirectly from sequence data from mRNA or DNA. Protein sequences are more conserved than nucleotide sequences (due to redundancy in the genetic code) and phylogenies based on protein sequences can take into account functional similarities and substitution probabilities between amino acids, allowing more accurate similarity measures, in particular between distantly related species. Proteins are more subject to selection than non-coding regions.

1.4.4.4 RFLP and Restriction sites

RFLP analysis uses restriction enzymes to cut the DNA that is then separated by size using gel electrophoresis and the resulting fragments visualised. Presence and absence of each restriction site can be established by double digestions (using 2 enzymes at once) since for each particular band to be present both restriction sites at either end must be present. Statistical methods exist that take into account that a single site change may affect two or more bands (Nei and Li, 1979), however, since a number of assumptions are made it is best to derive actual site presence/absence data from banding patterns. Since bands are sorted only by size, errors can occur when two fragments are of very similar size but from different regions and Southern hybridisation is required to confirm the homology of bands.

1.4.4.5 RAPD

The technique of random amplification of polymorphic DNA (RAPD) is simply a polymerase chain reaction (PCR) amplification of genomic DNA by a single short (5-15bp) random oligonucleotide primer which produces complex patterns of anonymous polymorphic DNA fragments (Clark and Lanigan, 1994). The resultant DNA fragments

are separated on a gel and the bands scored as present/absent character states for each sample. There is no prior sequence information required but due to the low-stringency of the reaction, DNA must be purified, and while relatively little DNA is required the sensitivity is not as high as sequencing. There are a number of difficulties with RAPD which are discussed further in 3.1.4.

1.4.4.6 Microsatellites (or SSRs) and minisatellites

There are regions of DNA found throughout the genome that exhibit a particular form of variation in which short combinations of nucleotides re-occur sequentially to produce repeats that are highly variable in number between individuals. These are called VNTR (variable number of tandem repeat) loci. Microsatellites (also known as simple sequence repeats, SSRs) are tandem repeats of very short (1-5bp) motifs while minisatellites are segments of DNA composed of tandem repeats of short motifs (15-100bp); different mutational mechanisms are thought to produce each type of allelic variation (Scribner *et al.*, 1994). Minisatellite variation can be either revealed as presence/absence of multiple loci by using a probe containing motifs (multilocus fingerprinting); or a single polymorphic locus can be detected by using flanking primers to amplify a single locus (single-locus fingerprinting). Microsatellites are highly variable and usually codominant, and although difficult to find can easily generate lots of genetic data from small amount of sample (Queller *et al.*, 1993).

1.4.4.7 Sequencing

DNA sequence data are the ultimate in fine scale detection of variation because every substitution can be detected. However, sequencing is expensive and requires prior sequence information (to design appropriate primers). It also has a number of characteristics that make it quite different to the other forms of data. There are only four base pairs (character states) possible at each site, thus it is quite likely for a position to change back to an ancestral state (if the base pair changes were equally likely, which they aren't, reversion to an ancestral basepair would occur at 33% of the mutation rate). Each base pair is not independent of the ones adjacent to it, although most analyses treat them as if they were. The alignment can be critical to the analysis - gap weighting often affects the outcome more than the algorithm used.

1.5 Genetic analysis

Simply describing genetic variation is not directly useful. The data obtained from genetic markers must be analysed for a description of the pattern of genetic variation and to understand the processes responsible for the pattern. In order to make sense of the huge amount of variation it is necessary to construct models. Good models fit the observed data well and have predictive ability. Models are made by taking data from a sample of individual cases and reducing the number of variables until general relationships or rules can be formulated. Models are simplifications of the real world and thus contain assumptions. It is important to know what these assumptions are, because if they are greatly violated the analysis may be invalid and give misleading results. Many of the original population genetic models and methods of analysis were developed for vertebrates (diploid, mobile, separate sexes, generally outcrossing, no cloning) and may not be appropriate for all organisms.

1.5.1 Population genetic models and processes

The classic model is of an "ideal" population of diploid, strictly sexual individuals, infinite in size (no stochastic effects), with no selection, migration, mutation or overlapping generations, and mating equally likely between any two individuals (random mating) (Weir, 1996). For this ideal population, the frequency of each allele remains constant over time and genotype frequencies in one generation can be predicted from allelic frequencies in the previous generation (Hardy-Weinberg Equilibrium, HWE). Population genetic models attempt to describe the effects of one or a few processes that violate the assumptions of an ideal population. The spatial scale of analysis can affect the appropriate model, for example Planes *et al.* (1996) showed that the island model was appropriate at small scales but as scale increased an isolation-by-distance model became more appropriate.

1.5.1.1 Genetic drift

In a population of finite size, allelic frequencies will vary from generation to generation as a result of random sampling of gametes: this process is called genetic drift. Allele frequencies may drift to zero and that allele is then extinct from the population. This means that two populations founded from the same ancestral population, if completely isolated for a long period, may come by genetic drift alone to have different allelic frequencies and even have fixed allele differences. There must be a high degree of isolation or else gene flow will prevent the populations from drifting apart. In an expanding population, the probability of any drift occurring is greatly reduced as there is less likelihood of loss of any alleles. Conversely, a temporary dramatic decrease in population size, called a bottleneck, greatly increases the probability of allele loss. A related phenomenon called founder effect, occurs where a new population is started by a very few individuals and thus has few alleles. The resultant low diversity can have critical implications if variation is low enough to prevent adaptation to new challenges.

1.5.1.2 Mutation and mating system

All genetic variation first originates from mutations, it is the occurrence and subsequent persistence and spread of these mutations that creates the variation which allows evolution to occur. Another assumption made to simplify models is to ignore the occurrence of mutation. Breeding systems range from strict self-fertilisers through to random, strictly sexual mating, with some species able to vary their breeding system. Some species exhibit a range of asexual reproduction such as parthenogenesis, asexual multiplication during part of their life cycle or for some generations. Even among strictly sexual reproducing species assortative mating can occur, or if there is unrecognised sub-population structure within a sample this can lead to an apparent deviation from HWE even though within the sub-populations there is random mating. This is called the Wahlund effect and is revealed as an excess of homozygotes in a strictly sexual species. Care must be taken when collecting samples to try to ensure they are from the same population by collecting from the same locale at the same time.

1.5.1.3 Migration and gene flow

A migrant must disperse, breed and successfully pass on its genes in order to cause gene flow. However, care should be taken trying to use rates of dispersion estimated from direct observations to predict gene flow, since the direct observations cover only a small time frame and temporal rate variation and rare events may have a huge impact on gene flow. A common rule of thumb is that gene flow greater than 1 individual per generation is sufficient to prevent genetic differentiation of two populations through random drift. This is an oversimplification, as each gene may have different critical levels due to different levels of selection, and while the gene flow may be sufficient to prevent fixation of alternative alleles, it does not mean allelic frequencies will be the same in both populations (Allendorf and Phelps, 1981).

1.5.1.4 Selection

Most genetic models assume the genetic markers are selectively neutral. However, many of the markers used may be either subject to selection directly or linked to selected loci. It is ironic that while many studies use standard population genetic analyses that assume no selection, they also aspire to find markers that are linked to loci of functional interest. There is an ongoing debate about whether most genes, most of the time are neutral or under selection pressures.

1.5.2 Models of marker variation

Genetic markers provide useful information because they are variable within and/or among populations or species. In order to explain this variation in terms of population genetic processes, it's necessary to make certain assumptions about how the variation arose.

The infinite-alleles model was originally developed for allozyme electrophoresis. It assumes that each new version (allele) is equidistant from each other version (Tajima, 1996). Extensions exist which include different distances between alleles, for example for microsatellites the stepwise mutation model and for proteins the use of substitution frequencies to weight specific changes.
The infinite-sites model was originally developed for restriction sites. It assumes the genome is very large and each mutation occurs at a new site (thus ignores multiple

changes at a single site) (Tajima, 1996). Sequence data and restriction site analysis ideally use the infinite-site model. Although they can be analysed by the infinite-allele method this does not use all the information about how closely related each sequence is. These types of data can potentially reveal the phylogenetic history of the sequences. It cannot be used, however, when the mutation rate varies substantially among sites (Tajima, 1996).

These two models are related and parameters normally calculated by one method can be obtained from the other (Tajima, 1996). Even sequence variation, which at first glance seems the epitome of the infinite-sites model, shows allelic variation at a single site (furthermore the probability of change between the allelic states (four possible nucleotides) is not equal). The major advantage of genetic markers that provide genealogical information is that because the roles of genetic and demographic processes are so clearly separable, such markers may be used to estimate parameters governing the long-term demographic processes of interest to conservation and evolutionary biologists (Milligan *et al.*, 1994).

1.5.3 Statistics and methods of analysis

Many different methods have been proposed for analysing data from genetic markers. I have grouped these into 5 categories, generally increasing in scale.

1.5.3.1 Measures of diversity

Measures of diversity indicate the extent of genetic variation in a population. There is a large number of diversity measures applicable to different types of markers: percent of variable loci; average number of substitutions/site; heterozygosity, frequency of private alleles (Nei, 1987). To compare diversity statistics between populations or other species it is very important to ensure you do not bias your selection of markers i.e. do not discard invariable loci or pick loci which show the most differences between groups.

1.5.3.2 Distance measures

Genetic distance can be calculated between individuals or between populations. Numerous forms of distance measures exist for nearly all types of molecular data. Distance methods range from those without any genetic model such as simple Euclidean distance, or band-sharing (e.g. Jaccards distance) (Sneath and Sokal, 1973), through to methods incorporating complex theoritical and observational corrections specific to the type of data. Using the infinite-alleles model, Nei's genetic distance can be calculated between population pairs from allelic frequencies (Nei, 1978). Nucleotide divergence is an approach using the infinite-sites model which estimates the average number of substitutions per site from restriction site and RFLP data and is being developed for RAPD and AFLP data (Clark and Lanigan, 1994).

1.5.3.3 Hierarchical analysis

Hierarchical analysis finds the distribution of the total genetic variation between hierarchical levels. For example, how much variation is distributed within populations, among populations, among higher groupings. Fst analogs (Gst, Nst etc) all attempt to quantify the amount of the total variation that is due to differences between populations. Gst makes use of allelic frequencies while Nst takes into account the similarities between haplotypes (Pons and Petit, 1996). Analysis of molecular variance (AMOVA) uses Euclidean distance calculated between individuals to estimate proportion of variation within and between populations (Excoffier *et al.*, 1992).

1.5.3.4 Gene flow

Some individuals may leave the population and produce offspring in another population. The level of gene flow has significant consequences for the evolution of populations (and thus species) since high gene flow allows the spread of new alleles, reduces the genetic drifting apart of populations and may slow local adaptation, although there are circumstances where gene flow may enhance local adaptation. There are three indirect methods for estimating gene flow from observed (single timeframe)

genetic variation: estimation from Fst analogs; private alleles; and coalescent analyses (Slatkin and Maddison, 1989; Hudson *et al.*, 1992).

1.5.3.5 Clustering and phylogenies

Clustering involves grouping samples according to specific criteria. Many different approaches have been devised, the most common being similarity/distance. Genetic distances can be used to draw phenograms showing the similarity/distance between samples. There are numerous methods to produce phenograms, for example UPGMA and neighbour joining. An alternative approach to the distance measures is to use criteria such as parsimony or maximum likelihood to select a tree which best fits the data. These approaches require qualitative markers (character states), whereas other methods can use distance data as well as converting qualitative data to distance measures (Avise, 1994 p94). Parsimony is a method that works on character state data and attempts to draw a phylogenetic tree (bifurcating, no recombination) which minimises the number of state changes required to make the data fit the tree. For these trees, branch length is not proportional to genetic distance, although they can be scaled to show minimal estimates of character-state change.

Above the species level, and for non-recombining DNA regions, clustering produces an estimate of phylogeny, which is the order of bifurcations and additionally for some methods the length of the branches. If assumptions are made that the amount of inferred distance is proportional to time (molecular clock hypothesis) then timing of splits can be estimated. All available methods of inferring phylogenies have implicit assumptions - no single method works under all possible conditions (Nadler, 1990).

1.5.3.6 Coalescent analysis

A newer type of genetic analysis that avoids assumptions of equilibrium has been developed based on coalescent methods (Tavare, 1984; Hudson, 1990; Neigel, 1991; Crandall and Templeton, 1993; Neigel and Avise, 1993; Slatkin, 1993).The basic premise is that all genes sampled are derived from a single ancestor by bifurcation. In time, the extant copies of a gene will tend to descend from fewer and fewer ancestors,

and eventually only one (Tibayrenc and Ayala, 2000). Genetic markers that provide genealogical information allow the genetic and demographic processes to be separated, thus allowing estimates of parameters governing the long-term demographic processes (Milligan *et al.*, 1994). Coalescent analysis can be used across all scales and since it uses individual genes rather than estimates of group characteristics it avoids many of the assumptions imposed by other population genetic models.

1.5.4 Programs for genetic analysis

Table 1.4 contains a list of some of the large number of computer programs that are now available for the analysis of population genetic data. A more comprehensive listing can be found on the Phylip's Other Phylogeny Software page at http://evolution.genetics.washington.edu/phylip/software.html

Selecting the right program/s for a particular study requires consideration of the data type to be analysed, the available computer platforms, the type of analysis required and the sophistication of the analysis (e.g. does it provide confidence limits, how does it handle missing data). In general it may be best to use several programs to increase confidence in similar results and identify results which vary depending on the program used.

Table 1.4 - Selected computer programs for population genetic data analysis.

Name	Availability (http://)	Data type(s)	OS
Arlequin	acasun1.unige.ch/arlequin/	Allozyme RFLP microsat seq	Win, Mac
A program for population genetics analysis including estimation of gene frequencies, testing of linkage disequilibrium, and analysis of diversity between populations .			
Fstat	www.unil.ch/izea/softwares/fstat.html	co-dominant	Win 95, 98
Estimate and test gene diversity and fixation indices			
GDA	http://alleyn.eeb.uconn.edu/gda/	discrete	Win
Performs disequilibrium analyses and estimates population structure and inbreeding parameters for discrete genetic data			
Genetic Studio	www.GeneticStudio.com/	Allozyme, RAPD	Mac
This is a large program that analyzes population genetic data. Very user friendly, and easily expandable through plugins (R Dyer, unpublished).			
GenePop	wbiomed.curtin.edu.au/genepop/	Allozyme	Web based
Exact tests for HWE, pop differentiation, genotypic disequilibrium, converts file formats			
GeneStrut	wwwvet.murdoch.edu.au/vetschl/imgad/GenStrut.htm	Allozyme	Mac
Calculates genotypic and allelic frequencies, statistics for Hardy-Weinberg disequilibrium, diversity within and identities between populations, indices of structure (F-statistics) (Constantine *et al.*, 1994).			
PopGene	www.ualberta.ca/~fyeh/index.htm	Allozyme, RAPD	Win
It is a user-friendly computer freeware for the analysis of genetic variation among and within populations using co-dominant and dominant markers (Yeh *et al.*, 1997).			
RAPDistance	life.anu.edu.au/molecular/software/rapid.htm	RAPD	Win
A package for computing distance matrices for RAPD analyses. It has a comprehensive range of options for creating data files, editing and using application programs to analyse them (Armstrong *et al.*, 2000).			
TFPGA	www.public.asu.edu/~mmille8/	Allozyme, RAPD, AFLP	Win
Calculates genetic distances, descriptive statistics, and F-statistics, and performs tests for Hardy-Weinberg equilibrium, exact tests for genetic differentiation, Mantel tests, UPGMA cluster, and hierarchical data sets.			
DnaSP	www.bio.ub.es/~julio/DnaSP.html	Sequence	Win
Analysis of nucleotide polymorphism within and between populations, linkage disequilibrium, recombination, gene flow, gene conversion parameters, several tests of neutrality (Rozas and Rozas, 1999)			
Mega	www.megasoftware.net	Sequence	DOS
Parsimony, distance matrix and likelihood methods for molecular data (Kumar *et al.*, 1994)			
Phylip	evolution.genetics.washington.edu/phylip/phylip.html	Sequence protein distance etc	Win Mac
Package includes parsimony, distance matrix methods, maximum likelihood, and other methods on a variety of types of data, sequences, protein sequences, restriction sites, 0/1 discrete characters data, gene frequencies, continuous characters and distance matrices (Felsenstein, 1993)			
PAUP	www.lms.si.edu/PAUP/about.html	Sequence	Win Mac Unix
A software package for inference of evolutionary trees. It includes parsimony, distance matrix, invariants, maximum likelihood methods and many indices and statistical tests.			
Lamarc	evolution.genetics.washington.edu/lamarc.html	Sequence microsat allozyme	Win
LAMARC computes population parameters, such as population size, population growth rate and migration rates by using likelihoods for samples of data (sequences, microsatellites, and electrophoretic polymorphisms). It includes Coalesce, Migrate, Fluctuate and Recombine (Beerli and Felsenstein, 1999).			
Sites	heylab.rutgers.edu	Sequence	DOS
Analysis of multiple closely related DNA sequences to estimate population parameters			
SEND	Reference: (Nei and Jin, 1989)	Sequence, restriction site	Win
Standard errors for nucleotide divergence			
Microsat	human.stanford.edu/microsat/microsat.html	microsats	Mac
A program for calculating distances from microsatellite data.			
Restsite	www-genome.wi.mit.edu/~jmiller/restsite.htm	RFLP, restriction sites	Win
A package for computing distances between species based on restriction sites or restriction fragments			
MacClade	phylogeny.arizona.edu/macclade/macclade.htm	Any characters	Mac
MacClade is a computer program for phylogenetic analysis. Its analytical strength is in studies of character evolution. It has many tools for entering and editing data and phylogenies, and for producing tree diagrams and charts (Maddison and Maddison, 1992) .			

1.6 Aims

The general aim of this thesis is to contrast different molecular techniques (RAPD's, allozyme, sequencing mtDNA, sequencing ribosomal spacers) and appropriate analytical methods (allelic and infinite-sites approaches; inbreeding and coalescent models) for estimating population genetic parameters in parasites. A range of population genetic questions at different scales were chosen to emphasise the importance of tailoring techniques and analytical methods to the particular question being investigated.

Chapter 2 describes the general materials and methods used. Chapter 3 contrasts the population genetic structure estimated from a RAPD dataset for *Echinococcus granulosus* in Australia with an allozyme dataset, and explores several methods of analysis of RAPD data, including a correction for dominance by using estimated selfing rate to calculate allelic frequencies. Chapter 4 extends a study of genetic variation in *Ostertagia ostertagi* in the USA to Australian populations and tests alternative hypotheses to explain the observation of little genetic differentiation despite differences in an important biological trait. Chapter 5 describes intra-individual variation in both the intergenic and internal transcribed spacer 1 of rDNA of *O. ostertagi*. Chapter 6 concludes with a general discussion.

2 General Materials and Methods

2.1 *Caenorhabditis elegans* culturing

Due to the low numbers of *Ostertagia ostertagi* worms recovered and the difficulties with extracting *O. ostertagi* DNA, *C. elegans* was grown to provide bulk DNA to test primers and serve as positive controls. The worms were obtained from Carol Behm, Australian National University. *E. coli* were cultured overnight on 2YT agar plates (16g Bactotryptone, 10 g Bacto yeast extract and 5g NaCl in 1 l water) or NGM plates (3g NaCl, 17g agar, 2.5 g peptone, 1ml (5mg/ml in ethanol) cholesterol and 975 ml water). *C. elegans* were washed off the old plates onto the fresh plates with 1ml M9 buffer (3g KH_2PO4, 6g Na_2HPO_4, 5g NaCl in 1 litre of water). Plates supported live worms for up to a month at 15˚C.

As a back up in case of culture contamination, worms were frozen by taking young plates of larvae, washing off with 1 ml M9 and adding an equal volume of freezing solution (5.85g NaCl, 6.8g KH_2PO_4, 300g gylcerol, 5.6ml 1M NaOH in 1 litre of water). 0.5ml aliquots of this mix were transferred to freezing vials and placed in styrofoam boxes at -70˚C (to freeze at approx. 1˚C/min). Worms frozen this way remained viable for several years.

2.2 DNA Extraction

2.2.1 *Ostertagia ostertagi* DNA extraction

A total of six different methods were tried for extracting *O. ostertagi* DNA, which proved to be quite difficult due to the very small amount of DNA in a single worm.

2.2.1.1 CTAB extraction

Worms were frozen using liquid nitrogen and 200µl of Lysis buffer I (50 mM Tris, 50mM EDTA, 100mM NaCl adjusted to pH 8.0 and autoclaved) added and ground using a mortar and pestle. 50µl of Proteinase K (20mg/ml) was added, mixed well and incubated at 56˚ C for two hours. 1ml of 2% CTAB (Hexadecyltrimethylammonium

bromide, Sigma H-5882) was added, causing a precipitate which was pelleted by spinning at 8,000 rpm for 5 mins. The pellet was resuspended in 250 µl of NE buffer (2.5M NaCl, 50mM EDTA adjusted to pH 7.5 and autoclaved). 250µl of TE (10mM Tris and 1mM EDTA adjusted to pH 7.5 and autoclaved) was added, then an equal volume of chloroform. After mixing, it was centrifuged at 14,000 rpm for 5 mins. The top aqueous layer was saved, avoiding the debris at interface. 1µl of 5M $MgCl_2$ and 1ml of ice-cold ethanol was added to precipitate DNA. After centrifugation, removal of the supernatant and washing in 70% ethanol, the pellet was dried under vacuum. The pellet was resuspended in 50 µl of TE.

2.2.1.2 Miniprep extraction

A second "miniprep" method was also tested. 0.2 ml of grinding solution (10mM Tris-HCl, 60mM NaCl, 10mM EDTA and 5% w/v sucrose) was added to the tissue, which was then crushed with a minipestle. 200µl of lysis solution (0.3M Tris-HCl, 1.1% w/v SDS, 0.1M EDTA, 5% w/v sucrose with 0.04ml of diethyl pyrocarbonate per 5ml of solution) was added and mixed by inversion. After placing on ice for 45 mins, it was centrifuged for 10 mins at 14,000 rpm and the supernatant transferred to a new tube. 1ml of ice-cold ethanol was added, well mixed and allowed to sit at room temperature for 30 mins. After centrifuging at 14,000 rpm for 10 mins, the supernatant was poured off and the pellet resuspended in 100µl of 0.5M ammonium acetate. 400µl of 100% ethanol was added and well mixed and the mix was kept at room temperature for 10 mins to allow the DNA to precipitate. After spinning at 14,000 rpm for 10 mins and removing supernatant, the pellet was vacuum dried. The DNA was resuspended in 100µl of TE or water.

2.2.1.3 Phenol-chloroform extraction

A third method taken from Sambrook *et al.* (1989) using phenol-chloroform extraction was also tested. This method followed the first method up to the proteinase K incubation. An equal volume (450µl) of phenol-chloroform was added and mixed until an emulsion formed. The mixture was centrifuged at 12,000 rpm for 15 secs and the aqueous phase transferred to a fresh tube. This step was repeated until no protein was visible at the interface. An equal volume of chloroform was added to re-extract. 25µl of 2M NaCl was added and mixed well. 1µl of 5M $MgCl_2$ (to a final concentration of

0.01M) and 950µl of ice-cold ethanol were added and the mixture kept on ice for an hour. The mixture was centrifuged at 0˚C and 12,000 rpm for 10 mins. The supernatant was removed and the tube half-filled with 70% ethanol. After centrifugation for 2 mins, the 70% ethanol wash was repeated and the pellet vacuum dried. The DNA was resuspended in TE.

These three methods were compared. The mini-prep outperformed the CTAB method while the phenol-chloroform method did not produce detectable DNA. However, despite a number of attempts, it was found that worms extracted by the miniprep method did not produce reproducible PCR results. Two other DNA extraction methods were then tested.

2.2.1.4 SDS extraction

Worms were incubated in 200µl of digest buffer (50mM Tris-HCl pH8.5, 1mM EDTA and 0.5% SDS with 200µg/ml proteinase K) at 55˚C for 3 hours. The proteinase K was deactivated at 95˚C for 10mins, cooled to 4˚C and centrifuged at 12,000g for 10 mins. Since SDS inhibits PCR, Tween20 had to be added to the PCR reaction mix. Trials showed that at 2% Tween20 only 2µl of the SDS extracted DNA mixture could be used before inhibition occurred (at 5% Tween20 up to 5µl of DNA mixture could be used). The trial also suggested grinding was not as effective as freezing.

2.2.1.5 Blouin's extraction

Blouin *et al.* (1992) successfully obtained PCR products from single *Ostertagia ostertagi* worms with this extraction method. This method was modified as follows: worms were freeze/thawed three times in 100µl of buffer (10mM TrisHCl pH8.3, 2.5mM $MgCl_2$, 50mM KCl, 0.1mg/ml gelatin, 0.45% NP40, 0.45% Tween20) and 10µl Proteinase K (200µg/ml). The mixture was incubated at 55˚C for one hour. The proteinase K was inactivated (56˚C for 10 mins).

Comparison of this method with the SDS method above showed this method was superior. Frozen worms were also found to be much better than worms preserved in 70% ethanol. All sequence data for individual worms were obtained from frozen worms extracted by Blouin's method.

2.2.1.6 Prep-A-Gene modified extraction

After the initial trial a "glass-milk" extraction method was tested which produced better PCR results, although enough worms for this study had already been extracted using Blouin's extraction. A single worm was placed in 80 µl tissue lysis buffer and freeze/thawed several times. Samples were incubated at 60˚C for one hour with 20µl proteinase K that was then inactivated by heating to 95˚C for 10 mins. 180µl of AL buffer (QIAGEN) was added. 10µl of Prep-A-Gene matrix (Biorad) was added and incubated at 72˚C for 10 mins. After centrifuging for 1 min and washing twice with 700µl AW wash, the pellet was vacuum dried. 50 µl of AE elution buffer was added and vortexed for 60 secs and then incubated at 72˚C for 10 mins. After centrifuging, the supernatant was transferred to a clean tube.

2.2.2 *Echinococcus granulosus* DNA extraction

DNA was prepared from 0.1ml packed, ethanol-preserved protoscoleces, washed with PBS and resuspended in 0.5ml extraction buffer by repeated freeze thawing in liquid nitrogen, 0.5mg Proteinase K and incubating at 56˚C for up to 16 hours. The lysate was extracted with an equal volume of phenol-chloroform-isoamylalcohol followed by an equal volume of chloroform. The DNA was precipitated by adding 1/10 volume of 2M sodium chloride and 2 volumes of cold absolute ethanol, and incubated for 2 hours at -20˚C. The DNA was pelleted, washed in 70% ethanol and dissolved in 500 µl TE buffer. RNA was removed by adding RNAase A (10 µg) and incubating at 37˚C for 30 minutes. The DNA was extracted with chloroform and precipitated by adding 1/10 volume of 3M sodium acetate and 2 volumes cold absolute ethanol, and left overnight at -20˚C. The DNA was pelleted by centrifugation at 12,000 rpm for 10 mins, washed in 70% ethanol and resuspended in 50-100µl of double distilled water. DNA concentration was estimated by electrophoresis through 0.8% agarose and comparison with DNA standard concentrations. For RAPD reactions, DNA was diluted to 10 ng/µl.

2.2.3 *C. elegans* DNA extraction

For DNA extraction, 3 plates of 17 day old *C. elegans* cultures were used. The worms were washed off the plates with 1ml of M9 buffer (see above) and allowed to settle at the bottom of the tube and excess liquid removed. After rinsing in 1 ml of M9, they were centrifuged down at 14,000g for 1 minute. 0.5ml extraction buffer (100mM NaCl,

100mM Tris-HCl (pH 8.5), 50 mM EDTA (pH 7.4), 1% SDS, 1% B-mercaptoethanol) was added with 2.5µl of 20mg/ml proteinase K. After incubating at 65˚C for 30 mins, the sample was extracted with 0.5 ml phenol, then 0.5 ml phenol/chloroform and finally 0.5 ml chloroform. The final aqueous layer was mixed with 1ml of cold ethanol to precipitate the DNA. After 10 mins, it was centrifuged down at 14,000g for 5 mins. The pellet was washed twice with 70% ethanol and once with 95% ethanol, before redissolving in 50 µl TE. A test to see how much DNA was lost at each step detected significant losses at each extraction step, especially the chloroform stage.

2.3 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out using 1.2% gels in TAE buffer (40mM Tris-HCl; 20 mM acetate; 2mM EDTA; pH adjusted to 7.9). Gels were stained using ethidium bromide, visualised under UV illumination and photographed using Polaroid Type 57 film or as a digital image with a red filter.

2.4 PCR

The basic PCR procedure was to add 2µl of DNA extraction to a PCR mixture containing 2µl (12.5pmoles) of each primer, 4µl of dNTPS, 4µl $MgCl_2$ (4mM), 2.5µl of 10x reaction buffer (Promega), 0.4µl Taq with dH_2O to a total volume of 25µl. The standard PCR was 1x[94˚C for 2 mins, 55˚C for 2 mins and 72˚C for 2 mins]; 30x[94˚C for 30 secs, 55˚C for 1 min and 72˚C for 2 mins]; 1x [94˚C for 30 secs, 55˚C for 1 min and 72˚C for 10 mins]. For long-range PCRs, an extension time of 6.5 mins was used.

2.4.1 Primer design

Primers were designed (Table 2.1) with the following guidelines where possible: CG clamp at the 3’ end; Tm should be over 50˚C (calculated roughly by Tm= 4 x # of C and G's + 2 x # of A and T's); no runs of any base >4 in a row; pair of primers should be matched to within 5˚C Tms; and the AT content similar to template. The program Amplify (William Engels, Genetics Department, University of Wisconsin), was used to assess primers and detect likely primer dimer or mispriming.

Table 2.1 List of primers designed for regions of *O. ostertagi*.

Name	Sequence	bp	Tm	Designed	Region
CeLSU	CTAAATCACTTGCATACGAC	20	56/61	*C.elegans*	rDNA SSU IGS
CeSSU	GCTCTAATGAGCCGTTCGC	19	60/66	*C.elegans*	rDNA LSU IGS
CeATleft	TATCTTAGCAACCCAAATGC	20	56/61	*C.elegans*	mtDNA
CeATright	CCCAAGACTTTTCTTATACG	20	56/61	*C.elegans*	mtDNA
LSU2end	ATACTGCGATCTGTTGAGAC	20	58/66	*C.elegans*	rDNA
SSUend	CTTGAACCGGGTAAAAGTCG	20	60/68	*C.elegans*	rDNA towards ITS
Ce5.8rev	ATCGATACTCGATGCAACCG	20	60/68	*C.elegans*	rDNA with SSUend
OoIGSrep	GTGAAATATGCATATTCAAA	20	50/58	*O.ostertagi*	IGS non-repeat
ATND5	TAGTGGATTATTTGGC	16	44/56	*O.ostertagi*	mtDNA 900bp AT
ATND6	CGACTTAACAACGAAATACC	20	56/61	*O.ostertagi*	mtDNA
LeftND4	ACACGGTTATACATCTACAC	20	54/57	*O.ostertagi*	mtDNA
RightCOI	AATCTAGTACCAACCATACC	20	56/59	*O.ostertagi*	mtDNA
CP17r	ATCACCACACCATGTGCAGC	20	62/62	*O.ostertagi*	Cysteine protease 1
CP16f	CAGCTGCGATGTCTGACAGG	20	64/64	*O.ostertagi*	Cysteine protease 1
CP13f	GGTTGAATATCTCCAAAAGAACC	23	64/59	*O.ostertagi*	Cysteine protease 1
CP18r	CTTCATCTGCGTGAAATCGG	20	60/60	*O.ostertagi*	Cysteine protease 1
Btub2f	CGGTATCCAGCCCGATGG	18	60/64	*O.ostertagi*	B tubulin
Btub3f	GAGGCAAATATGTCCCAC	19	58/60	*O.ostertagi*	B tubulin
Rtub6r	GAAGGCAGGTCGTGACTCC	19	62/64	*O.ostertagi*	B tubulin
Btub3r	CGTCCAAACACGTAATTATCTGG	23	66/61	*O.ostertagi*	B tubulin

Table 2.2 - Published primers used in this study.

Primer Name	Sequence	bps	Tm	Region (Reference)
Oo1Blouin	ATTTTACCAGCAAAAGAACAAGT	23	60/61	ND4 mtDNA (1)
Oo4Blouin	CAAAGTGATTCCAAGTCATTGGC	23	68/67	ND4 mtDNA (1)
Hp1Bluoin	CGACAAACCACCTTGAT	17	50/60	ND4 mtDNA (1)
Hp4Blouin	CAAAGTGATTCCAAGTCATTGGC	23	66/65	ND4 mtDNA (1)
LSUsrGasNC2	TTAGTTTCTTTTCCTCCGCT	20	56	rDNA ITS (2)
5.8sfGasNC1	ACGTCTGGTTCAGGGTTGTT	20	60	rDNA ITS1 (2)
$(CA)_8AC$	CACACACACACACACAAC	18	54/65/43	anchored SSR (3)
$(CA)_8GT$	CACACACACACACACAGT	18	54/65/43	anchored SSR (3)
$(CA)_8GC$	CACACACACACACACAGC	18	56/67/45	anchored SSR (3)
$(CA)_8AT$	CACACACACACACACAAT	18	52/62/41	anchored SSR (3)

References: (1) (Blouin *et al.*, 1995); (2) (Gasser *et al.*, 1993); (3) (Oliveira *et al.*, 1997)

2.5 Probing

2.5.1 Southern blotting, alkali method

The DNA to be probed was run on an agarose gel, which was then placed in distilled water. The gel was placed in the apparatus onto of a stack consisting of: blotting paper, nylon membrane, and plastic template which had been wet with 0.4M NaOH. The seal was checked and 0.4M NaOH added until the gutter was full. Vacuum to 5Hg was applied for 90 minutes. The membrane was rinsed in 2xSSC (0.03M Na_3 citrate; 0.3M NaCl; pH adjusted to 7.0) and put on blotting paper to dry. The membrane was then fixed using UV light.

2.5.2 Enhanced chemoluminescence (ECL) and Digoxigenin (DIG) labelling

These two commercial kits were used to try to localise the non-coding AT rich region in *O. ostertagi* mtDNA. Both enhanced chemoluminescence (ECL Amersham), which directly labels DNA or RNA with the enzyme horseradish peroxidase, and Digoxigenin (DIG Boeringer Mannheim, Biochmica) are non-radioactive labelling and detection systems. Neither of these methods was sensitive enough to detect the region given the small amount of sample, so radioactive labelling was then used.

2.5.3 Radioactive probing

The membrane was placed in a hybridisation bottle with 15 ml of hybridisation buffer (7% SDS 1%EDTA 1%SSPE) and placed in an oven for an hour at 42˚C. 50 ng of the probe was gamma-labelled by adding 1µl kinase buffer, 1µl kinase, 2.5µl H2O, 5µl dATP in a PCR tube, incubated at 37˚C for 30 mins and then 96˚C for 2 mins. The labelled probe was added in 6ml of hybridisation buffer to the bottle and hybridised overnight. The membrane was rinsed several times with a total of 100ml 20xSSC, and once in 2xSSC. The membrane was placed next to a film until the image developed.

2.6 Cloning

2.6.1 Preparation of competent cells

Cells were thawed and streaked on an agar plate overnight. Ten to twelve large, 2-3mm colonies were isolated with a plastic loop and placed into 250ml SOB in a 2l flask. The culture was grown to A600 of 0.6 at 18˚C with vigorous shaking (200-250rpm) and placed on ice for 10mins. This was centrifuged at 2500g (3000rpm Beckman J-6B) for 10mins at 4˚C. The pellet was resuspended in 80ml of ice cold TB, incubated in an ice bath for 10mins and then recentrifuged. The pellet was gently resuspended in 20ml TB, DMSO was added to 7% and then incubated in an ice bath for 10mins. The cells were aliquoted into 1-2ml freezing tubes and immersed in liquid N_2 to freeze.

2.1.1 Transformation

2µl of plasmid was added to 100µl of competent cells and placed on ice for 30 mins. The cells were heat shocked by incubating in a 42˚C water bath for 45 secs and replaced back on ice for 2 mins. 200µl Luria-Bertani broth (LB, 10g bacto-tryptone, 5g bacto-yeast extract and 10g NaCl dissolved in 950ml H20, pH 7.0) broth was added to the cells and incubated at 37˚C for an hour. The cells were plated out on LB plates with ampicillin which had been coated with 20µl of X-Gal at least an hour previously. Positive colonies were white (due to the insert disrupting the X-Gal enzyme that produces a blue colour).

2.1.2 Blunt end cloning

Blunt end cloning uses restriction enzymes which do not have any overhangs (e.g. EcoRV, SmaI, and Sfl). During ligation the RE is included so that any vector which self ligates (no insert) will be recut by the enzyme. Using a ratio of 100x insert to vector, forces the insert into the vector. Pfu DNA polymerase was used to blunt end the target fragment. During the ligation both the T4 DNA ligase and the restriction enzyme must act effectively, thus it is critical to find a buffer in which both enzymes will work well. Testing of T4 DNA ligase activity in different buffers by ligating HiLo marker, revealed

that NEB3 (New England Biolab) buffer was poor for T4 DNA ligation activity, NEB2 had some activity while NEB4 and OnePhorAll buffers were best.

2.1.3 Tvector cloning

The pGEM-T vector kit from Promega (U.S.A.) was used. Tvectors make use of the fact that Taq-polymerase adds a single deoxyadenosine to the 3' end of all duplex molecules during PCR. These A-overhangs are used to insert the PCR product into the T-vector which has 3' T-overhangs at the insertion site. Fresh PCR product was purified using a QiaQuick kit (this step was later left out). A 1:1 molar ratio of pGEM-T vector and PCR product was calculated (ng of insert = [ng vector x kb size of insert] /3kb (size of vector)).

The ligation was carried out at 15˚C for at least 3 hours and the reaction terminated by heating to 72˚C for 10mins. The reaction was cooled to room temperature and transferred to a new 1.5ml tube. 2µl of 0.5 M β-mercaptoethanol was added to 50µl of competent cells, the ligation reaction was added, mixed gently and placed on ice for 20 mins. The cells were then heat shocked for 30 secs at 42˚C and replaced on ice for 2 mins. 250µl of prewarmed SOC broth (Bacto tryphone 2%w/v; Bacto yeast extract 0.5%w/v; NaCl 10mM; KCl 2.5mM; $MgCl_2$ 10mM; $MgSO_4$ 10mM; glucose 20mM) was added and then incubated at 37˚C on a shaker (225 rpm) for one hour. 50µl and 200µl of the mixture were plated out on LB plates with ampicillin (50µg/ml), IPTG (40µl of 100mM) and X-Gal (40µl of 40mg/ml). After at least 18 hours at 37˚C the plates were placed at 4˚C for a few hours for colour development.

2.1.4 Insert screening by PCR

Half of each selected white colony was plated onto a LB (10g bacto-tryptone, 5g bacto-yeast extract and 10g NaCl dissolved in 950ml H20, pH 7.0) agar plate, the other half was removed with a pipette tip and added to 50µl of TE buffer containing 1% TritonX-100. The tubes were then incubated at 95˚C for 5 mins to lyse the cells, centrifuged for 2 mins to remove debris and supernatant and transferred to a clean tube. 5µl of this was then used in a PCR reaction using M13 forward and reverse primers. Samples with

correct size inserts or which did not produce a short fragment (negative - no insert) were re-screened with specific primers.

2.1.5 Small scale plasmid preps

Plasmid DNA was isolated using an alkaline lysis procedure. 5ml cultures in LB broth with antibiotic (ampicillin) were incubated on a shaker at 220 rpm, at 37˚C overnight. The cultures were placed on ice for 5 mins, then a 1.5ml eppendorf was filled and centrifuged down for 2 min at 14,000 rpm. This was repeated twice more so a total of 4.5ml of culture was centrifuged down. The pellet was resuspended in 200µl GTE buffer by pipetting up and down. 300 µl of freshly prepared 0.2N NaOH/1% SDS solution was added. The tube was mixed by inversion five times and put on ice for 5 mins. The solution was neutralised by adding 300µl 3M potassium acetate (pH4.8), mixed by inversion and placed on ice for 5mins. The solution was centrifuged for 10 mins and the supernatant transferred to a new tube. 10µl of RNAase (final concentration 20µg/ml) was added and incubated at 37˚C for 20 mins. The DNA was extracted with 400µl chloroform, the layers were mixed by hand for 30secs, then centrifuged for 1 min and the aqueous phase transferred to a clean tube. An equal volume (800µl) of isopropanol was added to precipitate the DNA, which was collected by centrifugation for 10 mins. The isopropanol was removed and the pellet washed in 500µl 70% alcohol. The pellet was dried under vacuum and resuspended in 50µl TE.

2.2 Sequencing

The PCR product was purified using a QIAQuick column (a modification of manufacturors instructions: DNA was resuspended in 50µl of warm dH_2O). The Taq DyeDeoxy™ terminator cycle sequencing kit (Applied BioSystems) was used. 8µl of the Dye-terminated premix was added to 1µl of primer (12.5 pmol), 200ng target DNA (for cloned DNA, only 50-100ng of PCR product required) and distilled water to a total of 20µl in a 200µl tube. Fragments were sequenced from both ends (separate reactions for each primer). For M13 primers (used for all cloned fragments), the sequencing reaction was: 96˚C for 2:20 [96˚C 10secs, 60˚C 4:05]x 30. The reaction was centrifuged and placed in new 600µl tube. 2µl of 3M sodium acetate and 50µl of 95% ethanol was

added to precipitate the DNA, the tube was mixed well and placed on ice for 15 mins. Spinning for 20mins at high speed pelleted the DNA. The ethanol was removed and the pellet washed with 250 µl of 70% ethanol, then recentrifuged for 5 mins, the ethanol removed and the pellet vacuum dried.

2.2.1 Sequence analyses

Sequences were manually corrected using SeqEd (Applied Biosystems) and aligned using Clustal V and X (Toby Gibson, Des Higgins, Julie Thompson, EMBL, Heidelberg, Germany). Phylogenetic inference was carried out using programs in the Phylip package (Felsenstein, 1993). MacClade (Maddison and Maddison, 1992) was used to trace characters onto trees generated by Phylip. Similar sequences were found by BLAST searching. ReadSeq (by D.G. Gilbert 1990 Biology Department Indiana University), was used to create matched sequence output (identical nucleotides below the guide sequence shown as dots, variant nucleotides shown).

3 Genetic diversity within and among populations: *Echinococcus granulosus* in Australia analysed using RAPD and allozyme data

3.1 Introduction

Species of the tapeworm *Echinococcus* (Cestoda: Taeniidae) are the causative agents of hydatid disease in humans and other mammals. Two hosts are required to complete their life cycle. The adult worm develops in the small intestine of a carnivorous definitive host, usually a dog or other canid (Thompson and Lymbery, 1990). After sexual reproduction, eggs are shed in the faeces and ingested by a herbivorous or omnivorous intermediate host. In Australia this host is usually sheep, macropods, pigs or humans. In the intermediate host, an egg develops into a cystic larva or metacestode, usually in the viscera. Within the cyst up to several thousand protoscoleces are produced by asexual multiplication, each of which if ingested by a definitive host may develop into an adult worm.

Taxonomy within the genus *Echinococcus* has been controversial for several decades (Thompson and Lymbery, 1988). Sixteen species have been described, however, currently there are only four accepted species (*E. granulosus, E voglei, E. multilocularis, E. oligarthus*), despite a number of informally designated strains within the species *E. granulosus* (Thompson and Lymbery, 1995). In Australia the "sheep" strain only occurs, which was almost certainly introduced with sheep soon after European settlement in the late 18th century. There are two different cycles of transmission in Australia, a domestic cycle principally between sheep and dogs, and a wildlife cycle involving mainly macropods and dingoes (Thompson and Kumaratilake, 1982).

There has been a proposal to revise the genus and recognise many strains with distinct genetic and biological traits as species (in several cases these groups have had species status in the past) (Thompson *et al.*, 1995). Even with this revision, there remains a major area of contention about the existence and importance of variability within *E. granulosus* strains.

3.1.1 Hypotheses on genetic variation and controversy

The extensive intra-specific variation of *Echinococcus* has been explained by two conflicting models, one attributable to J.D. Smyth and the other to R.L. Rausch (Thompson and Lymbery, 1988). Smyth's model assumes strict self-fertilisation, and with asexual reproduction in the cysts means a mutation can form a new strain. Even if low levels of cross-fertilisation occur, it will tend to occur between closely related individuals leading to an excess of homozygotes. Any genetic diversity within strains will be largely distributed between inbreeding family groups, with concentrations of genetically identical individuals in one area and spatial structuring of diverse family groups.

Rausch's model is based on the observation that strains are morphologically and biologically uniform over wide geographic ranges. Thus, he proposed extensive cross-fertilisation except where barriers such as different host species occur, with genetically diverse populations adapting to each host, resulting in different host strains. Within host strains, however, local differentiation is prevented and little spatial structuring would occur.

Many authors have reported the existence of genetic differences between strains of *E. granulosus* using Southern hybridisation of probes and RFLP (McManus and Smyth, 1979; McManus and Rishi, 1989). The fact that with these particular techniques they found no significant variability within strains led them to conclude there was no intra-strain variation. However, in most cases very small numbers of individuals within each strain were examined and there was no quantitation of variation, it was merely noted that profiles were similar within a strain and distinct between strains.

It was therefore assumed that *Echinococcus* was strictly self-fertilising (McManus and Smyth, 1986) and thus intraspecific variation could be explained by Smyth's model. When Lymbery and Thompson (1989) reported the existence of genetic variation between cysts within a single intermediate host, McManus (1990) suggested the observed allozyme differences were not real but due to artefacts.

A subsequent population genetic allozyme study of Australian isolates of *E. granulosus* by Lymbery *et al.* (1997) found that the parasite was predominantly self-fertilising. The high selfing rate was presumed to be due not only to actual self-fertilisation but also as a result of genetically identical protoscoleces from the same cyst being most likely to be adjacent when they have developed to adults in the gut of the definitive host, and thus any cross-fertilisation would be equivalent to self-fertilisation. Although the level of self-fertilisation was very high which supports Smyth's model, his prediction of spatial structuring of diverse family groups was not corroborated. Most of the variation found was within local populations, with very little attributed to different host species. This pattern fits the predictions from Rausch's model even though the level of cross-fertilisation was very low.

3.1.2 Recognition that disagreement is due to scale

The apparent contradiction between the allozyme study (Lymbery *et al.*, 1997), and other studies in the detection of within-strain variation is likely to be due to the scale at which various techniques reveal variation. The allozyme technique in this case can detect variation at a finer scale than RFLP of conserved regions or Southern hybridisation of those specific probes. Furthermore, the studies differed in their aims and therefore their sampling structure. Lymbery *et al.* (1997) aimed tp elucidate population genetic structure and therefore sampled a large number of individuals from each population, while the other studies were aimed at distinguishing strains/species, and therefore did not sample a large number of individuals from the same population.

3.1.3 Problems with the allozyme dataset

The allozyme study undertaken by Lymbery *et al.* (1997), was only just variable enough to be useful (7 variable loci found, but only one locus was highly variable). This meant that rare alleles had large effects on the estimation of variability. It also required relatively large amounts of material, and staining for some loci could only be done on adults grown from protoscoleces in dogs. Allozyme data are also restricted by a limited number of enzymes which can be stained and by the fact that since the technique examines active enzymes they may be affected by selection and thus may not be neutral markers, also not all amino acid changes are detected. Another technique was needed

that is more sensitive both in the amount of variation detected and the amount of material required for analysis. Ideally, it would also sample the genome more often and more evenly than the allozyme technique. This more sensitive technique could then test the genetic structure reported for the limited allozyme data.

3.1.4 Random Amplified Polymorphic DNA

Random Amplified Polymorphic DNA (RAPD) is a technique developed simultaneously by Williams *et al* 1991 and Welsh and McClelland (1990,1991). The technique, which is simply low stringency polymerase chain reaction (PCR) amplification of genomic DNA by a single short (5-15bp) oligonucleotide primer, produces complex patterns of anonymous polymorphic DNA fragments (Clark and Lanigan, 1994) . The resultant DNA fragments are separated on a gel and the bands scored as present/absent character states for each sample. The technique has numerous advantages, however the application and analysis of the resultant data presents some challenges.

RAPD requires no prior knowledge of the genome as the same primers can by used on any organism. Since the primers are random, the number of loci that can be examined is essentially unlimited. Since it is PCR-based, it is very sensitive and can be used when only small amounts of DNA are available. The fragments are separated by size on standard agarose gels with no need for radio-labelled probes. It detects a very high level of variation, nearly as high as sequencing but with much less expense. In many cases it reveals polymorphisms where other techniques such as allozymes have failed. It samples across the entire genome, not just a single locus that may be under selection and therefore meets the assumptions of most population genetic analyses.

There are, however, some disadvantages to the technique. RAPD is a low stringency technique, so any contaminating DNA will interfere with the banding pattern. Target DNA needs to be purified and negative controls always used. Sterilisation of buffers, tubes and tips is essential (Barral *et al.*, 1993). Reproducibility can be low, and the use of different types of Taq DNA polymerase and different thermocyclers may give different results (MacPherson *et al.*, 1993; Schierwater and Ender, 1993). Faint bands are difficult to score, especially when lane intensities vary and are better not utilised in

any evaluation. The amount of DNA used must be standardised and PCR inhibition can prevent bands being visible.

Homology of bands at the same position is assumed, although it is possible for two completely different DNA fragments to migrate at the same rate. This problem however, is less likely for closely related samples (e.g. within species), as the probability of two independent mutations producing a similar sized fragment is much less than the probability of them co-inheriting the same mutation. Homology can be tested by labelling DNA from a band and checking that it labels all co-migrating bands (Bachmann, 1994).

It is also possible that two different bands are created from the same region of DNA, meaning that two bands are alleles of the same locus and thus not independent. Proving independence of bands requires either breeding experiments to test inheritance or probing of every band with every other band to ensure they are not allelic forms. Additionally, the absence of a band may be due to a number of different mutations. Another approach to overcome this problem is to use only those bands present in >50% of individuals, therefore avoiding using two bands which are alleles of the same locus. Conversely, Lynch and Milligan (1994) suggest using only low frequency marker alleles. Non-Mendelian inheritance of RAPD bands has been detected although at a low level (Riedy *et al.*, 1992; Bucci and Menozzu, 1993; Levitan and Grosberg, 1993; Ayliffe *et al.*, 1994; Johnson *et al.*, 1994; Rothuizen and Van Wolferen, 1994). Non-Mendelian inheritance is of particular concern for parental assignment based on exclusion (Levitan and Grosberg, 1993).

To minimise these problems reactions should be chosen that:

1. Reveal polymorphisms;
2. Consistently produce strong bands;
3. Produce uniform bands between replicate PCRs ;
4. Are insensitive to DNA template concentrations (1-100 ng/µl) (Stewart and Excoffier, 1996).

Only fragments in the mid-molecular range should be scored, which minimises the chance that several different bands are produced from the same region by nested inverted repeats (Stewart and Excoffier, 1996).

Even when scored correctly, RAPD data present problems with analysis because of the dominant nature of the banding patterns. With RAPD data, it is not possible to distinguish individuals with two copies of DNA which produces a band from individuals with only one copy, as both show a band. The analysis of population structure with RAPD data is therefore hampered by the lack of complete genotypic information resulting from dominance (Lynch and Milligan, 1994).

Dominant data are less complete than that provided by co-dominant markers (e.g. isoenzymes), and thus more loci or more individuals need to be sampled. However, this requirement is completely offset by the huge number of primers available and high polymorphism of RAPD markers (Levitan and Grosberg, 1993; Lynch and Milligan, 1994). In strictly self-fertilising organisms, dominance is not a problem as no heterozygotes exist (except new mutations which become fixed or are lost in a few generations), thus band frequency is equal to frequency of the allele causing band presence. An independent estimate of selfing rate may also be used to obtain genotypic frequencies from RAPD data. RAPD data have been used before on *Echinococcus* but only to note whether different host samples showed different profiles (Siles-Lucas *et al.*, 1993).

3.1.5 Aims

In this chapter, genetic diversity is examined within and among populations of *Echinococcus granulosus* from different hosts and geographic regions in Australia, by comparing and contrasting RAPD and allozyme datasets. Analysis of RAPD data will be carried out using several methods including: estimation of allelic frequencies using different selfing rate; bandsharing; and nucleotide diversity.

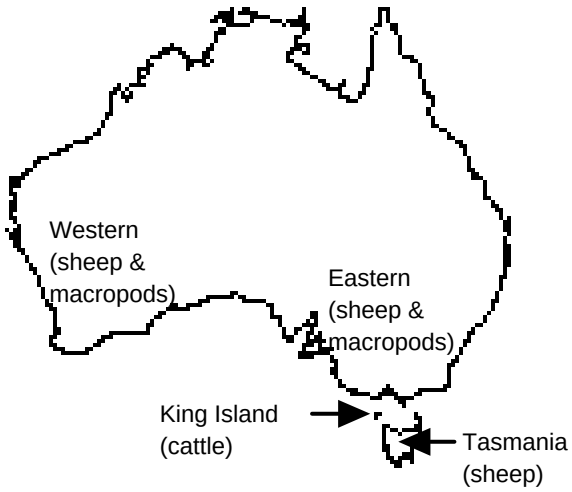
3.2 Materials and Methods

3.2.1 Datasets

Protoscoleces were collected from cysts from sheep, macropods (*Macropus giganteus*, *M. rufogriseus,* and *Wallabia bicolor* found only in eastern Australia and *M. fuliginouis* from Western Australia), and cattle across Australia as described in Lymbery *et al.* (1997). Each cyst was treated as a separate sample except for those shown in Appendix 1 with a final P representing pooled cysts from a single host. There were numerous cases of multiple cysts from the same host as shown in Appendix 1 by those sharing the same code except for the final letter.

Samples were divided into six populations (geographic/host origin): Eastern sheep; Eastern macropods; Western sheep; Western macropods; Tasmanian sheep; and King Island cattle (Figure 3.1). There were insufficient samples to include all three hosts from each area, due to the fact that Tasmanian macropods do not seem to be infected and cysts in cattle are rarely fertile on the mainland. The King Island samples (1988-1990) were from an outbreak found only in cattle on the island from which hydatid disease had been thought to have been eradicated in 1971 (Constantine *et al.*, 1991).

Figure 3.1 - Four regions of Australia from which *Echinococcus granulosus* samples were collected.



The allozyme data were collected by Lymbery *et al.* (1997) and five more bovine samples from King Island were added, analysed for the same enzyme loci. The RAPD data are described here for the first time. The number of samples from each population for each dataset is shown in Table 3.1. 24 samples were common to both datasets.

Table 3.1 - Sample sizes for *Echinococcus granulosus* from 6 Australian populations for allozyme and RAPD datasets (note that RAPDs had no missing data while allozymes had a high proportion of missing data, see Table 3.2 for average number scored). Each sample represents a pool of protoscolices from a cyst, for some hosts multiple cysts were collected, 24 samples were common to both datasets.

Population	Allozyme	RAPD
King Island bovine	5	4
WA sheep	26	4
WA macropods	28	9
Eastern sheep	50	13
Eastern macropods	14	8
Tasmanian sheep	39	5
	162	43

3.2.2 RAPD-PCR and Electrophoresis

RAPD-PCR reactions were carried out as described by Akopyanz *et al.* (1992). Reaction mixtures consisted of 2.5 µl 10X Tth Plus reaction buffer (Biotech Int), 5.0 µl 1.25 mM dNTPs, 2.0 µl primer (20 pmoles), 3.0 µl $MgCl_2$ (25 mM), and 1 unit Tth Plus DNA Polymerase (Biotech International) in a total volume of 24 µl. Reactions were layered with 50 µl of light mineral oil, with 1.0 µl (20 ng) of genomic DNA added above the oil and reactions mixed simultaneously by brief centrifugation.

Amplifications were performed in a thermal cycler (Omnigene) using the following program: four cycles of [94˚C, 5 min: 36˚C, 5 min: 72˚C, 5 min], thirty cycles of [94˚C, 1 min: 36˚C, 1 min: 72˚C, 2 min] and finally 72˚C, 10 min. Amplified products were resolved by electrophoresis on 2.0% Metaphor high resolution agarose (FMC) for 2.5 hours at 15 V/cm. DNA fragments were visualised by staining with ethidium bromide (1ug/ml) for 1 hour at room temperature and photographed prior to scoring of bands.

3.2.3 Primers

Forty arbitrary fingerprinting primers (Bresatec), twenty 10 mer and twenty 12 mer, were screened for profile production with genomic DNA from *E. granulosus* horse, pig and sheep strain and *E. multilocularis*, to identify primers capable of distinguishing isolates at species/strain level. In addition, two Australian sheep strain isolates demonstrating variability in allozyme studies were used to screen primers for their ability to detect sub-strain variation at the genomic level.

All forty primers produced profiles with each of the four DNA samples, 18 primers produced amplification patterns which distinguished all four samples from each other. Twenty seven primers detected polymorphisms in the two variable sheep strain isolates, these primers generally produced similar banding patterns with one or more unique bands in either of the isolates. From these primers, ten (5 x 10mer & 5 x 12 mer), were selected on the basis of producing the most clearly distinguishable polymorphic bands to reduce ambiguity in assigning identity among gels. The selected primers were: 10:01 5'-AAGCTGCGAG-3', 10:06 5'-TTCGAGCCAG-3', 10:14 5'-GTCCCGTGGT-3', 10:16 5'-CACCCGGATG-3', 10:19 5'-GGACGGCGTT-3', 12:03 5'-ACCTATGCCGAC-3', 12:04 5'-CAACCTTCGGAC-3', 12:06 5'-CAACCTTTGCGG, 12:07 5'-ACCCTGCTCATC-3' and 12:09 5'-ACCCACATCGGT-3'.

3.2.4 Analysis of Allozyme Data

Enzyme banding patterns were interpreted genetically, with mobility variants at the seven variable loci assumed to be the product of different alleles, as described in Lymbery and Thompson (1988). An infinite alleles model was used to estimate genetic variation within and among all populations. Genetic diversity within populations was described by three standard measures: P, the proportion of polymorphic loci, where a polymorphic locus has the frequency of the most common allele less than 95%; A, the mean number of alleles per loci; He, the total gene diversity or expected mean heterozygosity (Nei, 1978).

Distances between populations were estimated using Nei's genetic distance, which were then clustered using the neighbour-joining procedure in Phylip. Genetic differentiation between populations was described by Wright's Fst values, calculated by the method of Weir and Cockerham (1984). Genestrut uses an extension of Nei's gene diversity analysis to a hierarchical population structure (Chakraborty, 1980). An equivalent analysis, AMOVA is carried out by Arlequin (Excoffier *et al.*, 1992). Gene flow (Nm) was estimated from Fst by the equation Nm = 0.25(1-Fst)/Fst (Nei, 1987).

3.2.5 Analysis of RAPD data

The first step was to score the gels by distinguishing different bands and scoring each individual as having each band present or absent. Difficulties include distinguishing fragments that run very close together and scoring faint bands that may not be visible in a sample due to less DNA rather than true absence. It is best to leave out bands that are difficult to resolve or are faint. There are computer programs available to analyse gels automatically, however, they are very sensitive to gel artefacts and changing settings can completely change the resultant distance matrices (Morgan *et al.*, 1995).

There have been a number of approaches to analysis of RAPD data. Three approaches will be discussed in detail: band sharing or distance measures; allelic (infinite-alleles); and nucleotide divergence (infinite-sites). Other potential methods of analysing RAPD data are possible, including methods that take into account both the position and relative brightness of bands.

3.2.5.1 Band sharing – distance and AMOVA approaches

Band sharing or simple distance methods involve calculating the number of bands shared and not shared and through various algorithms producing a distance matrix. This approach is not a genetic interpretation but a phenotypic approach based purely on similarity of banding patterns. Three algorithms of the 15 calculated by RAPDistance were chosen (numbers 2,5,15), one from each of the three main groups discussed in the RAPDistance documentation. One common algorithm used is Jaccard's, which has the characteristic that shared absence does not imply two individuals are more similar. This characteristic is beneficial for RAPD data, as many different mutations may be responsible for absence of a band. For example, different mutations at primer site, lack

of a priming site, insertions/deletions within the fragment making it too large or too small to be scored as the same locus, and secondary structure changes preventing primer from binding can all result in the absence of a band. The second algorithm was Pearson's Phi (algorithm 5), and the third was Excoffier's distance (algorithm 15).

The distance matrix can then be used to create a phenogram, by a number of different clustering techniques, such as UPGMA and neighbour joining (Huelsenbeck and Hillis, 1993). Alternatively if euclidean distance is calculated, it can be used in an analysis of molecular variance (AMOVA; Excoffier *et al.*, 1992). This procedure does not require any assumptions about breeding system. A correction for dominance has been suggested as an extension of the AMOVA procedure (Stewart and Excoffier, 1996).

3.2.5.2 Nucleotide diversity

Nucleotide diversity is an estimate of actual substitution rates (with assumptions). It is based on the infinite-sites model and can only be used for low divergences, as multiple changes at a single site cannot be detected. Clark and Lanigan (1994) discuss in detail the long list of assumptions and requirements to analyse RAPD data by calculating nucleotide diversity. The computer program RESTSITE uses this approach for fragment data and restriction site data (Nei and Miller, 1990). The two 10 mers and three 12 mers were analysed separately and a combined estimate of nucleotide diversity was calculated by RESTSITE.

3.2.5.2.1 Allelic with correction for dominance given inbreeding/selfing estimate

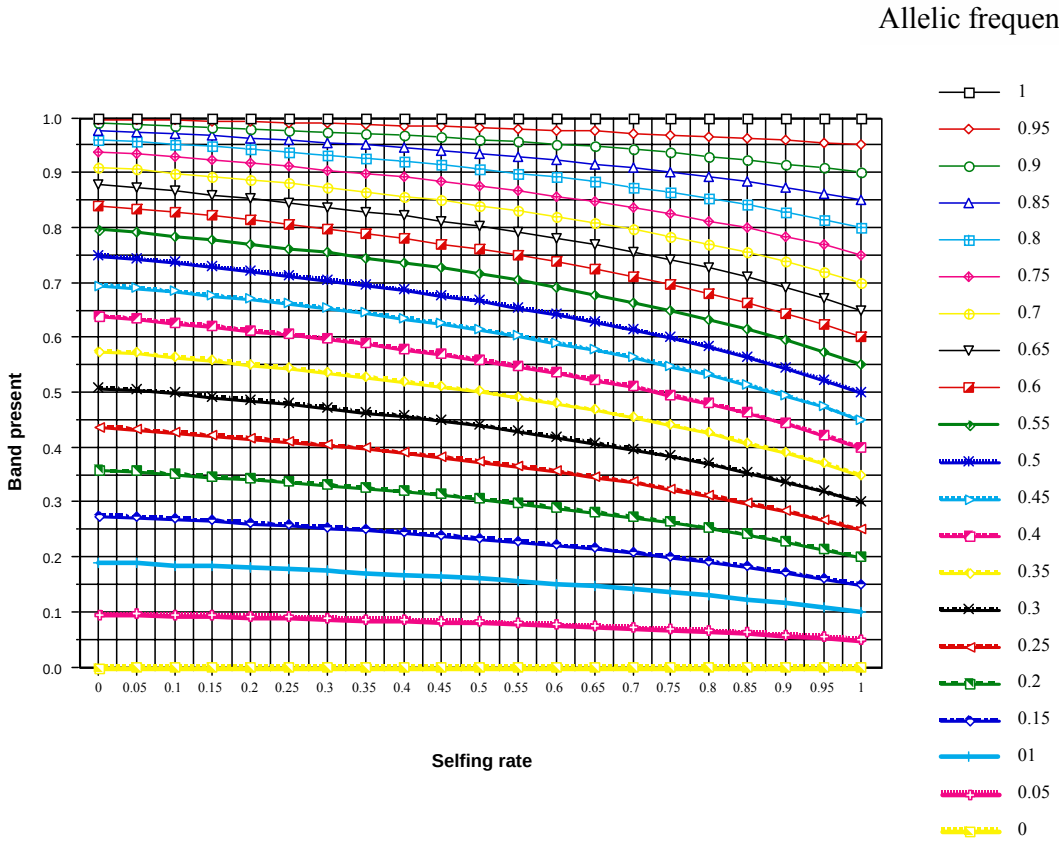
This is an infinite-alleles approach where each band position is treated as a separate locus, with individuals with two absent alleles not showing the band and individuals with the band being either heterozygous or have two present alleles. Calculation of allelic frequencies is straightforward if the organism is haploid or a strict self-fertiliser as band presence frequency is equal to present allele frequency. If the organism is cross fertilising, it can be assumed to be in Hardy Weinberg Equilibrium and allelic frequencies can be estimated from the HWE equation.

If there is an independent estimate (from co-dominate data) of selfing it is possible to calculate the actual allelic frequencies from the observed proportion of individuals with the band present. If P is the proportion of individuals with the band present; p is the frequency of the allele for band presence in a population and s is the selfing rate where s=1 means strict self-fertilisation (no heterozygotes); and s=0 for a population in HWE then:

$$p = P\left(1 + \left(\frac{(1-s)(1-P)}{1-s/2}\right)\right)$$

Figure 3.2 shows this relationship between frequency of band present allele and proportion of individuals with band present for selfing rates ranging from zero (Hardy-Weinberg Equilibrium) to one (self-fertilisation). The estimated allelic frequencies are then used to calculate genotypic frequencies and the standard population genetic parameters estimated as for allozyme data.

Figure 3.2 - Relationship between observed band frequency and allelic frequency for different selfing rates ranging from Hardy-Weinberg equilibrium (s=0) to strict self-fertilisation (s=1), assuming a 2-allele model (presence/absence).



3.3 Results

3.3.1 Genetic interpretation of banding patterns

The raw data, in terms of multilocus genotypes for each isolate in each population, are shown in Appendix 1 for RAPD data and Appendix 2 for allozyme data. The RAPD data were checked for monomorphic loci (all samples with bands present or all with bands absent), "singleton" loci (where only one isolate has the band uniquely present or absent - provide no useful clustering information), allelic bands (bands which never co-occur in any individual and one is always present; one example was found where one individual showed a unique absence of one band with two other nearby bands being uniquely present), and finally bands with identical patterns (one found).

The original data set from 10 primers with 147 bands was incomplete, and due to the difficulties of analysis with missing data the dataset was pruned to remove all missing data. This involved removal of both individuals (down to 43 samples) and band loci (140 variable bands down to 77) (Appendix 2). Allozymes are codominant, so genotypes were assigned as homozygotes or heterozygotes. Seven of the 23 loci were polymorphic over the entire dataset and heterozygotes were found only at 2 loci (Appendix 1). The allozyme data had a large proportion of missing loci due to small amounts of samples which did not allow all loci to be assessed for all isolates nor for loci to be repeated until they worked.

3.3.2 Genetic variation within populations

3.3.2.1 Allozyme data

There were few polymorphic loci with P values of ≤ 20% for all populations. Very few heterozygotes were observed (Ho), leading to an estimated Fis of 0.96, and a selfing rate of 0.989 (Lymbery *et al.*, 1997). The King Island population was the most variable by all measures despite its extremely low sample size (Table 3.2).

Table 3.2 - Genetic diversity within populations of *Echinococcus granulosus* in Australia, estimates from allozyme data: P - proportion of polymorphic loci, A - mean number of alleles per locus, Ho - observed heterozygosity, He - expected heterozygosity and N- average samples scored per locus for each population .

Pop	P (0.95)	A	Ho	He	N (scored)
King Island	20.00	1.200±0.396	0.025±0.001	0.087±0.042	3.70
Tasmanian Sheep	9.09	1.182±0.413	0.006±0.000	0.054±0.038	22.59
Eastern Sheep	8.70	1.174±0.241	0.001±0.000	0.029±0.019	24.57
Eastern Macropods	18.18	1.182±0.222	0.000±0.000	0.058±0.029	7.27
Western Sheep	9.09	1.136±0.280	0.000±0.000	0.037±0.026	13.68
Western Macropods	13.04	1.174±0.241	0.000±0.000	0.036±0.023	12.13

3.3.2.2 RAPD data

Since RAPD data are not codominant, individuals with a band may be either homozygous or heterozygous for band presence at that locus. If we wish to calculate genotypic and allelic frequencies from the raw data, then we must make an assumption about the selfing rate. Figure 3.2 shows how the proportion of individuals within a population with a band present at a RAPD locus, can be converted into allelic frequencies at different selfing rates from complete cross-fertilisation (or HWE, s=0) to complete self-fertilisation (s=1).

The effect on measures of genetic variation of 3 different assumptions were examined: s=0, s=1, s=0.989. These cases represent the extremes from complete cross-fertilisation (or HWE, s=0) to complete self-fertilisation (s=1) and the estimated rate of selfing from the allozyme dataset (s=0.989). Table 3.3 shows the two measures of population diversity calculated assuming strict selfing, selfing rate estimated from allozyme data, and Hardy-Weinberg Equilibrium.

Table 3.3 - Genetic diversity within populations of *E. granulosus* in Australia, measures from RAPD data: P - proportion of polymorphic loci, He - unbiased expected heterozygosity and N- sample size for each population of *E.granulosus* under three different selfing rate assumptions.

Population	N	S = 1		S = 0.989		S = 0	
		P (0.95)	He	P (0.95)	He	P (0.95)	He
King Island	4	40.8	0.1911	40.8	0.1906	40.8	0.1656
Tasmanian Sheep	5	52.6	0.2193	52.6	0.2205	52.6	0.2179
Western Sheep	9	30.3	0.1429	30.3	0.1424	30.3	0.1208
Western Macropods	4	34.2	0.1215	34.2	0.1225	34.2	0.1394
Eastern Sheep	13	72.4	0.2153	72.4	0.2170	64.5	0.2766
Eastern Macropods	8	61.8	0.2221	61.8	0.2229	61.8	0.2376

Expected heterozygosity estimates for each population for the RAPD dataset versus the allozyme dataset were not significantly correlated (Figure 3.3, r^2=0.108). King Island diversity was much higher from the allozyme dataset however, this was not seen in the RAPD dataset. Diversity measures are highly dependent on sample size and since sample sizes were low for some populations this may be expected to explain some of the variation, however there was no significant correlation between He and sample size for RAPD data and a negative correlation for allozyme data (Figure 3.4).

Figure 3.3 - Expected heterozygosity estimated from allozyme versus RAPD dataset for each population of *E. granulosus* in Australia, regression line shown. EA means eastern Australia, WA = Western Australia, Tas = Tasmania and King Is = King Island.

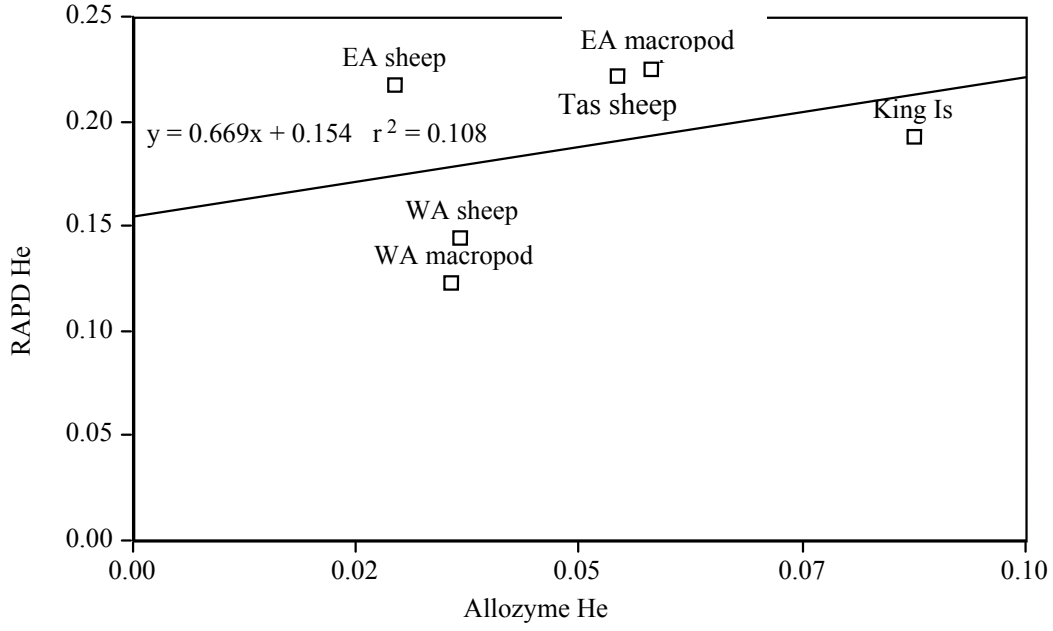
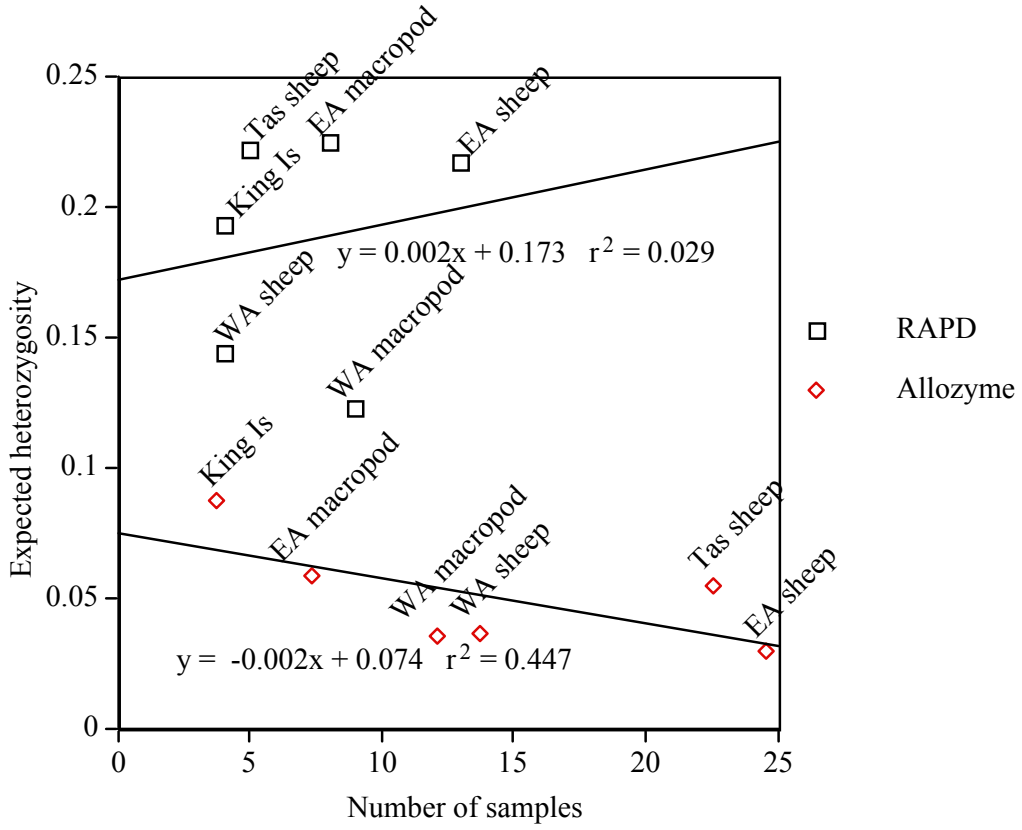


Figure 3.4 - Number of samples versus expected heterozygosity for populations of *E. granulosus* in Australia for both allozyme and RAPD data types, regression lines for each data type shown separately.



3.3.3 Genetic variation among populations

3.3.3.1 Allozyme data:

Genetic distances among populations were calculated by Nei's genetic distance (Nei, 1978). Figure 3.5 shows the similar estimated Nei's genetic distances for population pairs from allozyme data by 4 different programs. Two of the programs (GDA and PopGene) failed to correctly handle the situation of three loci not being scored for any individual in a population. Data had to be rerun on only 20 loci to avoid those two programs treating missing loci as fixed differences greatly inflating distance estimates (Figure 3.6). A phenogram of the populations clustered by neighbour joining of Nei's genetic distance is shown in Figure 3.7.

Figure 3.5 - Nei's genetic distances between populations of *E. granulosus* in Australia from allozyme data estimated by four computer programs. KB is King Island cattle, TS is Tasmanian sheep, ES is eastern Australian sheep, EM is eastern macropods, WS is Western Australian sheep, and WM is Western Australian macropods.

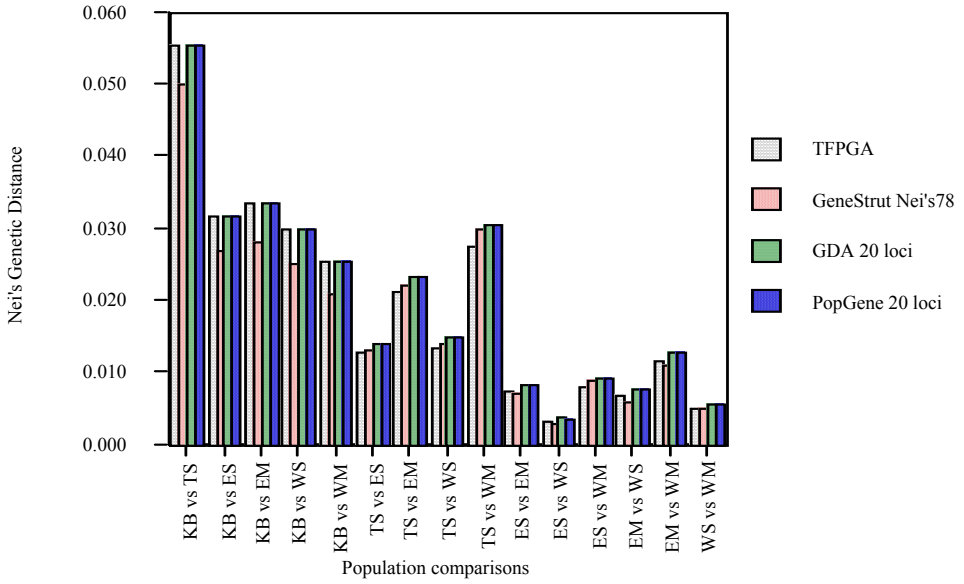


Figure 3.6 - Errors in calculating Nei's genetic distance between populations of *E. granulosus* in Australia due to missing loci by two programs, the line represents 1:1. KB is King Island cattle, TS is Tasmanian sheep, ES is eastern Australian sheep, EM is eastern macropods, WS is Western Australian sheep, and WM is Western Australian macropods.

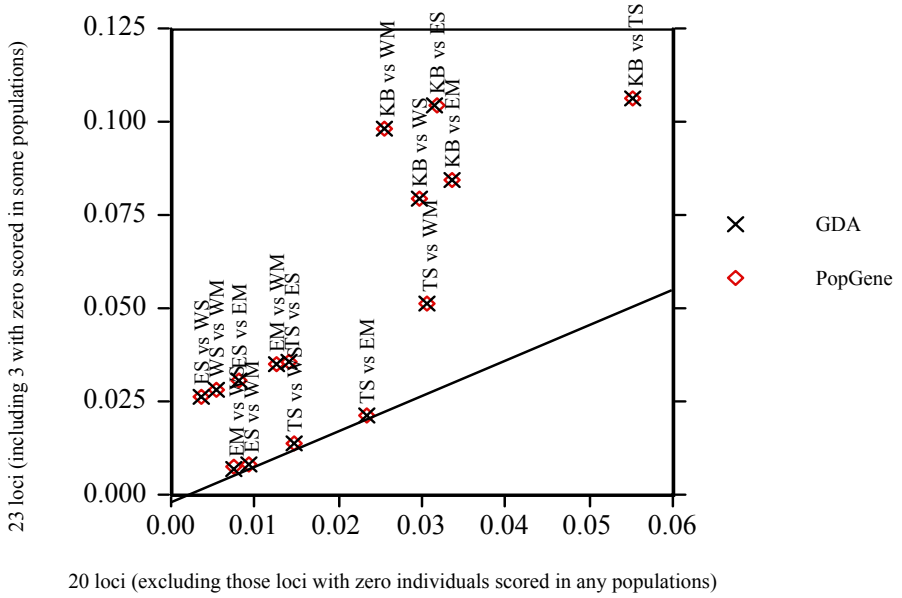
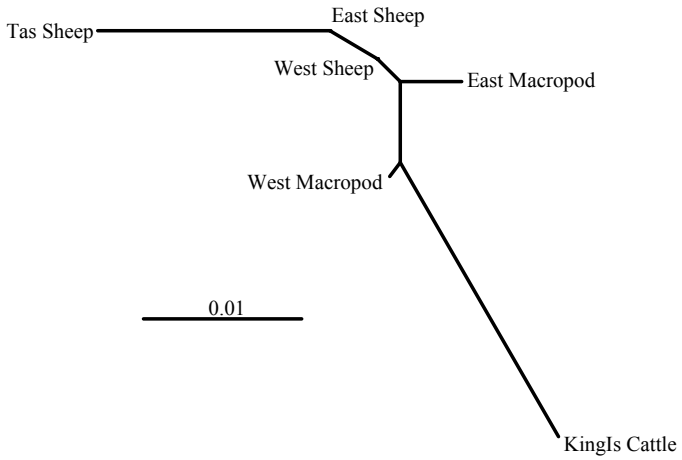


Figure 3.7 - Phenogram of populations of *E. granulosus* in Australia using neighbour joining of Nei's genetic distances from allozyme data.



3.3.3.2 RAPD data:

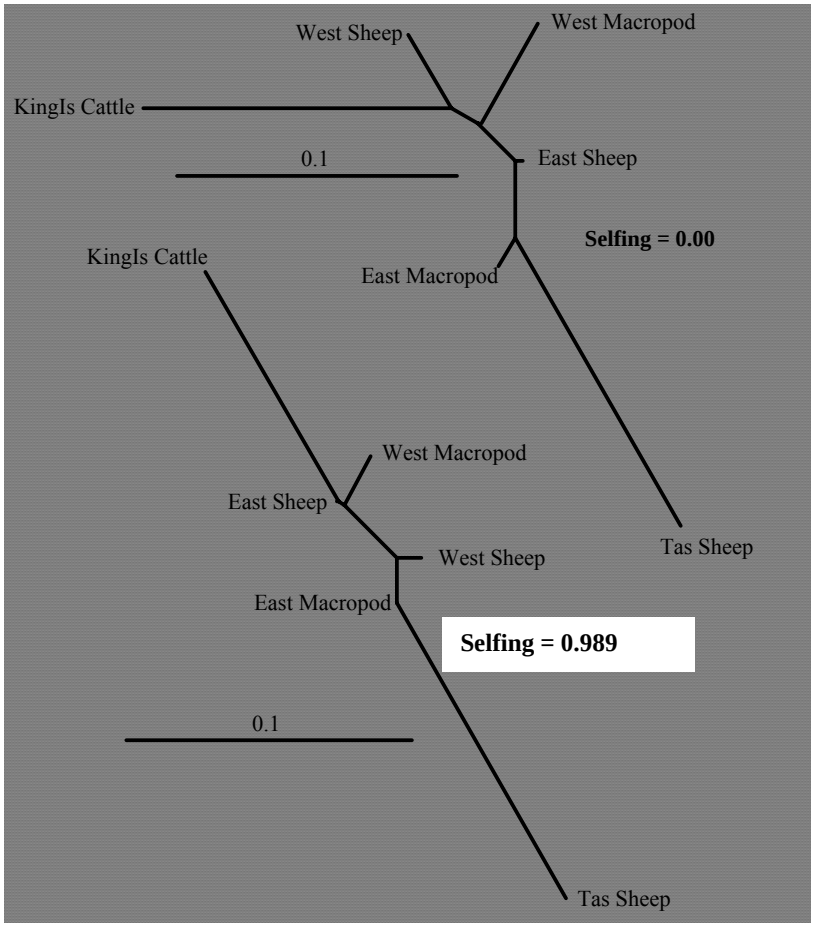
3.3.3.2.1 Allelic approach

Nei's genetic distance among populations was calculated under an infinite-alleles model, assuming selfing rates of s=0, s=0.989, s=1. The results are shown in Table 3.4. The genetic distances were higher when assuming HWE for all except the Tasmanian Sheep versus Eastern Sheep comparison. The distance values were clustered by neighbour joining (Figure 3.8), resulting in different phenograms for s=0 and s=1.

Table 3.4 - Nei's genetic distances between populations of *E. granulosus* in Australia estimated from RAPD data under 3 selfing rate assumptions.

	s=1	s=.989	s=0
KingIs Cattle vs Tasmanian Sheep	0.270	0.269	0.292
KingIs Cattle vs East Sheep	0.091	0.092	0.141
KingIs Cattle vs East Macropod	0.136	0.137	0.198
KingIs Cattle vs West Sheep	0.116	0.117	0.138
KingIs Cattle vs West Macropod	0.113	0.114	0.153
Tasmanian Sheep vs East Sheep	0.160	0.158	0.132
Tasmanian Sheep vs East Macropod	0.119	0.117	0.129
Tasmanian Sheep vs West Sheep	0.128	0.128	0.181
Tasmanian Sheep vs West Macropod	0.186	0.185	0.237
East Sheep vs East Macropod	0.041	0.041	0.059
East Sheep vs West Sheep	0.046	0.046	0.072
East Sheep vs West Macropod	0.020	0.020	0.054
East Macropod vs West Sheep	0.038	0.039	0.076
East Macropod vs West Macropod	0.047	0.048	0.085
West Sheep vs West Macropod	0.058	0.059	0.087

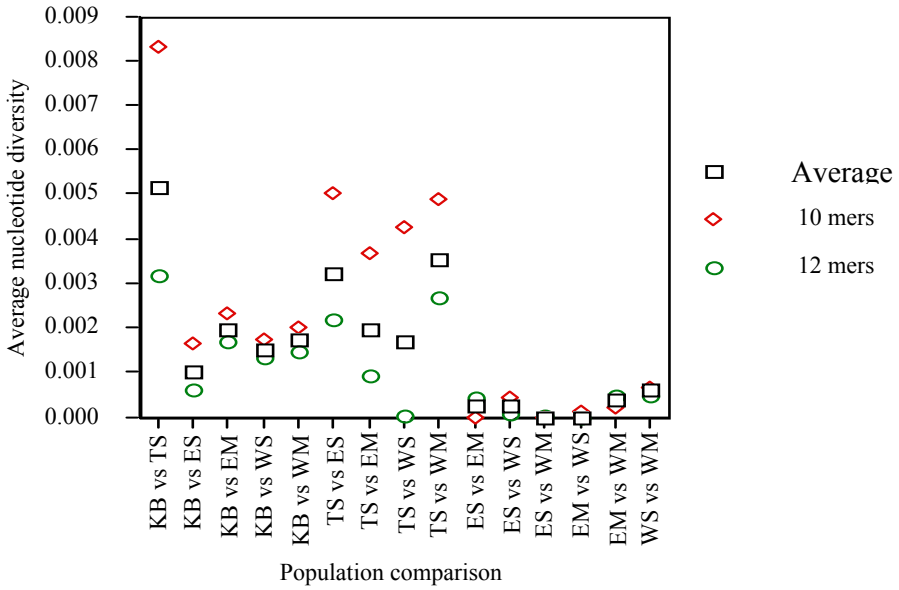
Figure 3.8 - Neighbour Joining phenogram of Nei's genetic distance for populations of *E. granulosus* in Australia, estimated from RAPD data assuming a) selfing=0 and b) selfing = 0.989 (s = 1.00 gave the same result).



3.3.3.2.2 Nucleotide diversity and band-sharing approaches

RAPD nucleotide diversity was calculated with the program RESTSITE which produces an estimate of the average rate of nucleotide substitution between pairs of populations. The estimates from the 10mer primers (2 primers, total 34 bands) were generally higher than those from 12mers (3 primers, total 43 bands) and there was significant variability between the primer sets (Figure 3.9).

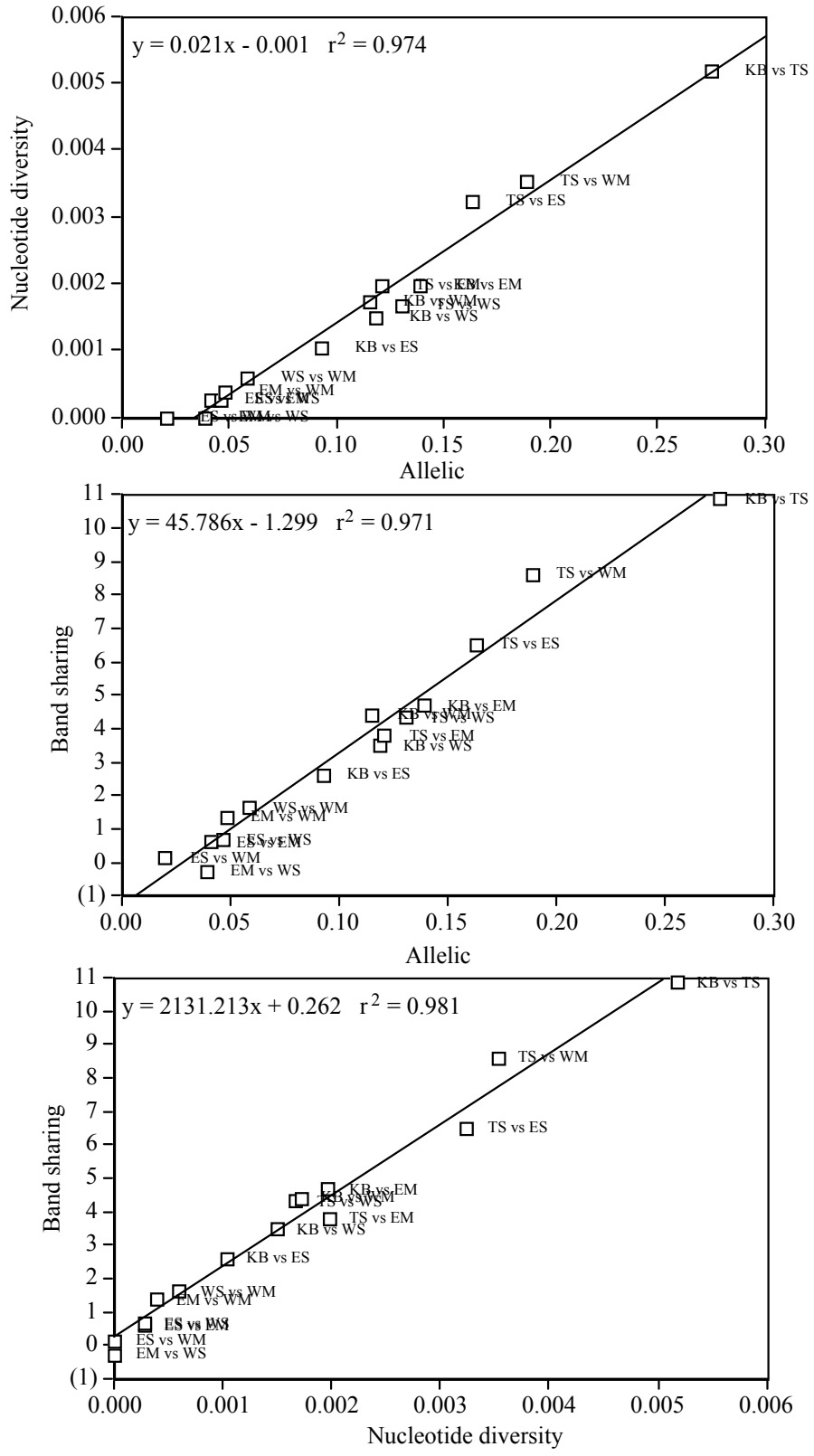
Figure 3.9 - Nucleotide diversity estimated between pairs of populations of *E. granulosus* in Australia from 10mer and 12mer primers and the average. KB is King Island cattle, TS is Tasmanian sheep, ES is eastern Australian sheep, EM is eastern macropods, WS is Western Australian sheep, and WM is Western Australian macropods.



Three different band sharing methods all from the program RAPDistance were used. They were highly correlated with each other (r^2 = 0.956 to 0.994, data not shown). Jaccards distance disregards shared absence of bands, however, it gave values very similar to Excoffier's distance.

These three different approaches to analysing RAPD data gave very similar population distance estimates as shown Figure 3.10, where one representative measure from each approach is shown.

Figure 3.10 - Correlation of three RAPD analysis approaches for estimating pairwise distances between populations of *E. granulosus* in Australia.

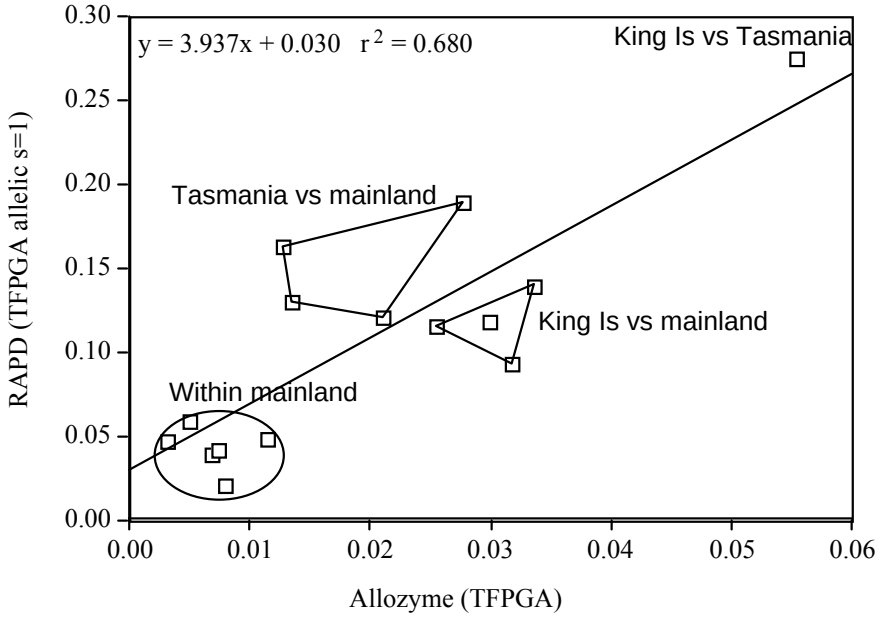


3.3.3.3 Contrasting RAPD and allozyme datasets for genetic variation among populations

Over both datasets and all methods generally there were three clusters of distances. The highest distance was between Tasmania and King Island, then all the comparisons between either Tasmania or King Is and mainland populations, and finally the smallest distances were seen within mainland comparisons. The distinctness of the King Island population from the other populations was reflected in finding that one RAPD band was present only in the four individuals from King Island, and similarly one allozyme allele (Np-2 C) was found only in 3/5 of the King Island individuals.

Allozyme distances however, showed that the Tasmania versus mainland comparisons (middle left polygram) were lower than King Island versus mainland (middle right triangle). The opposite was seen with the RAPD dataset, the King Island - mainland comparisons were slightly lower on average than the Tasmanian - mainland comparisons (Figure 3.11). Overall, the correlation between RAPD (allelic) and allozyme (allelic) population distance estimates was moderate (r^2 =0.68). This contrast was apparent across all comparisons of RAPD (three different types of analyses) and allozyme (analysis) (data not shown).

Figure 3.11 - Nei's genetic distances from allozyme data versus Nei's genetic distances from RAPD data for populations of *E. granulosus* in Australia.



3.3.4 Partitioning genetic variation within and among populations

Estimates of gene flow (Nm) between populations from Fst/Gst were lower for allozyme than RAPD datasets (Table 3.5, 3.6).

Table 3.5 - Allozyme locus diversity and gene flow estimates of *E. granulosus* in Australia. Ho is observed heterozygosity, Nm is geneflow.

Locus	Ho	Hs	Ht	Fis	Fit	Fst	Nm
Est	.067	.384	.422	.825	.841	.091	2.497
G6pd-2	.000	.106	.118	1.000	1.000	.105	2.131
Idh-1	.000	.144	.190	1.000	1.000	.240	0.792
Np-2	.042	.253	.366	.835	.886	.307	0.564
Pep	.000	.035	.034	1.000	1.000	-.021	undef
6pgd-1	.000	.079	.106	1.000	1.000	.251	0.746
6pgd-2	.000	.079	.106	1.000	1.000	.251	0.746
Av.	.016	.154	.192	.899	.919	.195	1.032

Table 3.6- RAPD diversity and gene flow estimates

	N	Ht	Hc	Hs	Gst	Gcs	Nm(Gst)	Nm(Gcs)
Mean 77 bands	43	0.238	0.170	0.216	0.092	0.212	4.942	1.857

Hierarchical analyses (Table 3.7) showed similar estimates of the proportion of genetic variation within and between populations for both datasets and both methods of analysis, the only exception to this was when HWE (s=0) was assumed for the RAPD data, which caused an increase in proportion of variation due to combining host populations on the mainland. Most variation was distributed within populations, with some among geographic areas (highest level in Table 3.7 which combines mainland, Tasmania and King Island), and very little variation among hosts within geographic areas (middle level).

Table 3.7 - Hierarchical analysis of genetic variation in *Echinococcus granulosus* in Australia, showing percent of total variation at three different levels. The lowest level represents the variation found among samples from the same host within the same area. The middle level represent the additional variation due to combining all hosts within a area, and the highest level represents the addition variation due to combining the three areas. Each column represent a separate analysis by different programs of one of the two datasets, for the last two columns the RAPD dataset was analysed assuming firstly no heterozygotes exist and secondly that the present and absence alleles are in Hardy Weinberg Equilibrium.

	PopGene Allozyme	GeneStrut Allozyme	Arlequin Allozyme	Arlequin RAPD	TFPGA RAPD haploid	TFPGA RAPD HWE
Combine Mainland/Tas/King	13.8%	16.0	12.1	12.3	12.3	13.4
Combine host in area	2.8	2.1	6.3	4.7	4.7	15.2
Within population	83.4	81.9	81.6	83.0	83.0	71.4

3.4 Discussion

3.4.1 Importance of breeding system

The breeding system of an organism has profound effects on its genetic structure. Critically however, it can also affect the analysis of certain types of genetic data. A number of different molecular techniques cannot detect heterozygotes and thus allelic frequencies can only be calculated if assumptions about breeding system are made. One aim of the study was to see what effect these assumptions could have on the analysis of data if they were incorrect.

Since *Echinococcus* has a sexual phase it has the potential to be in Hardy-Weinberg equilibrium (HWE), although studies have shown it to be mostly self-fertilising (Lymbery *et al.* 1997). When analysing a dominant data set, in which heterozygotes cannot be detected, it is possible to contrast the results from different assumptions about the breeding system. These assumptions can range from assuming the species is in HWE, through to assuming it is a strict self-fertiliser. Assuming the RAPD dataset was in HWE had major effects on the results (Table 3.7, contrast last two columns, and Figure 3.8).
For organisms in HWE, allelic frequencies can be estimated from RAPD data. For organisms in which the selfing rate is not 0 (HWE) or 1 (strict selfing) attempts to estimate allelic frequencies have not been made. If an independent measure of selfing rate has been made, allelic frequencies can be estimated for these intermediate cases. No current program will produce allelic frequencies from the proportion of individuals with a band present given a certain selfing rate, but it is easily calculated in a spreadsheet.

This procedure would allow researchers to see what effects lack of knowledge about true selfing rate may have on the resultant estimate of genetic structure. Bayesian statistics may help with intermediate cases between self-fertilising and HWE by incorporating the error of the estimate of selfing into an error for the estimate of allelic frequencies and subsequent parameters. It is interesting that so many authors emphasis dominance as a significant problem with RAPD data and yet RFLP and sequence data also rarely score heterozygotes.

The estimated selfing rate from allozyme data for *E. granulosus* was 0.989 which is very high, and heterozygotes are quite rare (8/162 individuals). Even this low level of outcrossing is however sufficient to have prevented drift causing divergence across the mainland between domestic/sylvatic cycles or between eastern/western populations. The gene flow barriers of water separating Tasmania and King Island and/or selection due to an eradication campaign on those islands has allowed some degree of divergence from mainland populations.

3.4.2 RAPD dataset compared to allozyme dataset

Overall, the RAPD dataset supported the finding from the allozyme data that most variation was found within populations. The premise of Smyth's model was supported (predominantly self-fertilisation), however, the resultant pattern of genetic variation followed Rausch's model. One possible explanation is that even a small level of cross-fertilisation can dramatically alter population genetic structure. In addition, the extensive artificial movement of domestic hosts may be assisting the low level of cross-fertilisation to maintain high within population diversity.

In terms of genetic distances the allozyme dataset showed that the King Island samples were on average more distant from the mainland than the Tasmanian samples (Figure 3.11). The RAPD dataset showed the distances were similar, if anything the King Island samples were more similar to the mainland samples. This is likely to be a result of the small sample size for the King Island population and the fact that a rare allozyme allele was found only in those few individuals, whereas the RAPD dataset had a much larger number of informative bands. It does not change, however, the overall conclusion that the King Island population does not appear to have originated recently from Tasmania or the mainland since it was twice as distant from either of these than distances seen among mainland populations.

There were insufficient mainland cattle samples to include due to their very low fertility (i.e. very few cysts develop protoscoleces). In contrast the King Island cattle cysts were very fertile. The difference in fertility of cysts also supports the observed genetic difference between King Island and the mainland. The few fertile mainland cattle

samples were very similar to mainland and not to King Island samples suggesting it is not a simple host effect.

3.4.3 Uses and limitations RAPD data

Despite drawbacks, the advantages of RAPD outweigh the limitations and RAPD has proved to be extremely useful for a wide range of applications (Barral *et al.*, 1993). RAPD data are useful for distinguishing strains and closely related species. They can assess relative variability of populations, and whether populations are similar enough to be treated as a single unit for conservation purposes. They can provide population genetic data with the constraints described above. They are also highly useful markers for genetic maps.

Diagnostic primers can be easily generated by detecting bands which are present only in the isolates of interest, and using the sequence of that band to design primers (Morgan *et al.*, 1996). There was one RAPD band which was present only in the four King Island samples which could be used in further studies to search for the origin of the outbreak on King Island. Further samples from other Bass Strait islands, Tasmania and New Zealand may detect other isolates with this band which may reveal the source of the King Island outbreak.

For simple applications, such as determining if two samples are the same or different, RAPD is ideal, as long as standard precautions are used. Standard precautions include running positive and negative standards on all gels, care when comparing across gels, running each sample at least twice, using many primers and testing for correlation between primers (which may indicate lack of independence). Problems arise when attempting to estimate absolute levels of difference and place phylogenetic interpretations on the data. It is important to check for bands that are allelic. Although it is usually impractical to Southern blot each pair of bands, there are simple checks that can rule out pairs of bands which are completely linked (always present together), or which never occur together in any individual (may be alleles).

RAPD data has been found to be of little use above species/sibling species level (Humbert and Cabaret, 1995; van de Zande and Bijlsma, 1995). This is due to the

differences above the species level being saturated, i.e. distance values are close to maximum and there is little detectable difference between interspecies distance and inter-genus distance. In some cases there may be no shared bands at all between OTUs, which makes an estimation of distance impossible. Thus RAPD data in general are not useful for constructing phylogenies. It is true for any molecular technique however, that there are upper and lower limits of usefulness, the lower level is set by detectable differences, and the upper by saturation.

In the next chapter, the usefulness of DNA sequencing for examining genetic variation within species is examined, not only as in this chapter at a between population level but also at a broader scale, between continents.

4 Genetic diversity within and among populations: *Ostertagia ostertagi* in Australia and the USA analysed using mtDNA sequences

4.1 Introduction

4.1.1 *Ostertagia ostertagi* and Ostertagiasis

Ostertagia ostertagi is a nematode parasite of cattle that lives in the abomasum and can cause significant production losses and mortality. *Ostertagia* is considered to be the most important parasite of cattle throughout the temperate zone (Anderson and Bremner, 1983). Regular drenching can control the parasite, however the emergence of resistance is threatening the long-term effectiveness of all of the existing classes of anthelmintics (Waller, 1999; Kohler, 2001). Generally, the development of anthelmintic resistance in cattle seems to be about 20 years behind that in sheep, probably due to less frequent treatment, however this lag may be decreasing and the use of sustained release devices may accelerate the appearance of resistance in bovines (Waller, 1994).

The lifecycle of the parasite begins with unembryonated eggs shed in faeces which after a day or two, hatch releasing first stage larvae and develop to L2 and then L3 larvae (Armour and Osbourne, 1982). This can take from a week in warm weather to several months in cool areas. Cattle, while grazing, ingest the L3 larvae which will burrow into the abomasal mucosa and develop to the fourth larval stage. L4 larvae eventually break out of the mucosa and return to the lumen to develop as adults. When they rupture the mucosal wall they create significant damage and a protein-losing enteropathy. In the abomasum, they become adults, reproduce sexually and release eggs from 17 days after infection.

Type I ostertagiasis is an acute or chronic abomastitis, which is characterised by profuse, watery diarrhoea, and hypoproteinaemia. Acutely affected animals can appear normal and die suddenly. Chronically infected animals are often emaciated and have a poor haircoat. Cattle do eventually develop an acquired immunity after a couple of years, which greatly reduces parasite loads (McKellar, 1993). Although in severe cases

it can cause death, most of the economic damage is due to lost production and the costs involved with drenching (Anderson and Bremner, 1983).

4.1.2 Hypobiosis

The L4 stage larvae, within the mucosa, do not always develop immediately but may undergo delayed development called hypobiosis and do not emerge until 4-5 months later (Gibbs, 1986). This hypobiotic trait allows the worms to avoid producing eggs during extreme climatic periods when the larvae would not survive on the field. This causes another type of Ostertagiasis (Type II) when the L4 larvae emerge from hypobiosis in a synchronous fashion. It generally affects a smaller percentage of the herd but it can be more severe than Type I.

There is a major division in this trait across the United States of America. In the north, *O. ostertagi* undergoes hypobiosis over winter, while in the south hypobiosis occurs over summer. This difference in the timing of hyposbiosis has been shown to be genetically based by reciprocal transplant experiments over several generations (Frank *et al.*, 1986; Frank *et al.*, 1988). There have been relatively few studies of hypobiosis in Australia and these have not found either the high levels nor the disjunction in timing seen in the USA (Smeal and Donald, 1981). One study found differences in the proportion of worms undergoing hypobiosis due to production type on the same farm (Smeal and Donald, 1982). Both dairy and beef production herds showed more inhibition (hypobiosis) in spring, but it was much more distinct in beef cattle (>50% inhibited versus 10% for dairy cattle). The authors suggested that this was due to the availability of susceptible hosts. In beef production herds, young susceptible animals occur only seasonally while in dairy production herds there is a constant supply of susceptible hosts.

4.1.3 Genetic structure in *Ostertagia ostertagi*

A study by Blouin *et al.*, (1992) sampled *Ostertagia ostertagi* from populations that were genetically differentiated for hypobiosis. Worms from Minnesota and Maine (in the north) exhibited hypobiosis over winter; Alabama, Louisiana and Tennessee (in the south) exhibited hypobiosis over summer. They carried out RFLP analysis of mtDNA (7 enzymes detecting 37 sites), and found very high within population diversity but little

geographic structuring across the USA. This finding of high diversity and low Fst values (Fst measures the degree of genetic differentiation of sub-populations) in mtDNA despite genetic differentiation in a key life history trait, presents a paradox.

Diversity within a population is determined by the effective number of alleles, which for mtDNA is equal to $2N_f\mu+1$, for a panmitic population in equilibrium, where N_f is the effective number of females and μ, the mutation rate. Thus diversity is proportional to both population size and mutation rate when migration and non-random mating are ignored. Fst includes the effects of both mutation and migration so low Fst values may be due to either high mutation or high gene flow. If migration rate (m) is much greater than mutation rate (μ) then Fst can be used to estimate gene flow using the formula $GST \approx 1/(4N_em+1)$ (Nei, 1987). It is also assumed that the populations have reached drift-migration equilibrium and the number of sub populations are high, so caution must be taken when estimating N_em from Fst values if the populations are not in equilibrium (Ruckelshaus, 1998).

If gene flow is greater than one individual per generation (i.e. $N_e\ m > 1$) this will be sufficient to prevent population differentiation (fixation of alternate alleles) due to genetic drift, assuming there is negligible selection. $N_e\ m>1$ is not however sufficient to maintain identical allelic frequencies (Allendorf and Phelps, 1981) and lack of structure at neutral loci does not prevent structure at selected loci. Low Fst values may be due to high migration (if mutation rate is much less than migration). Gene flow (when mutation is negligible) is the product of migration rate (m) and effective population size (N_e). Thus high gene flow could be due to high migration rate, or large effective population size (or moderate values of both).

There are thus three major alternative (not mutually exclusive) explanations for the observed high diversity and low Fst values for mtDNA in *O. ostertagi* despite differentiation in a key life history trait:

1. High mutation rate of mtDNA (High μ)
2. High migration (High m)
3. High effective population size (High N_e)

High mutation rate of mtDNA (High μ)

This possibility explains the contradiction as being due to different characteristics of mtDNA versus the rest of the genome. In many organisms, the rate of evolution of mtDNA has been estimated to be ten times faster than single copy nuclear DNA. High mutation rates in mtDNA would explain the observed high diversity and would invalidate the assumption used for gene flow calculations from Fst, that migration rate (m) is much greater than mutation rate (μ). By increasing the within population diversity, the Fst value (which is a ratio between the amount of variation between populations and the total variation) would decrease.

High migration (High m)

Since cattle are domesticated they have artificially high rates of movement, leading to high migration between populations. In order to maintain the distinct hypobiosis regions, all of the migrant individuals who end up in the wrong region must be heavily selected against. The genetic load on the populations on either side of the divide between summer and winter hypobiosis regions must be very high as the selection coefficient (s) must be greater than the migration rate (m) to prevent introgression (Lymbery, 1993).

High effective population size (High N_e)

Very high population sizes allow the maintenance of high diversity (since genetic drift is very low few alleles are ever lost). Similar to the high migration model, the low Fst values are due to high gene flow (N_em) however, since N_e is very high, actual migration rate (m) can be relatively lower, hence the actual number of worms dying due to selection for hypobiosis is much lower resulting in a more reasonable genetic load.

4.1.4 Aims

The aims of this study are to:

1 - Determine the genetic structure of *O. ostertagi* from cattle from different farms in Western Australia.

2 - Compare the genetic diversity of *O. ostertagi* in Australia with diversity of *O. ostertagi* in the USA.

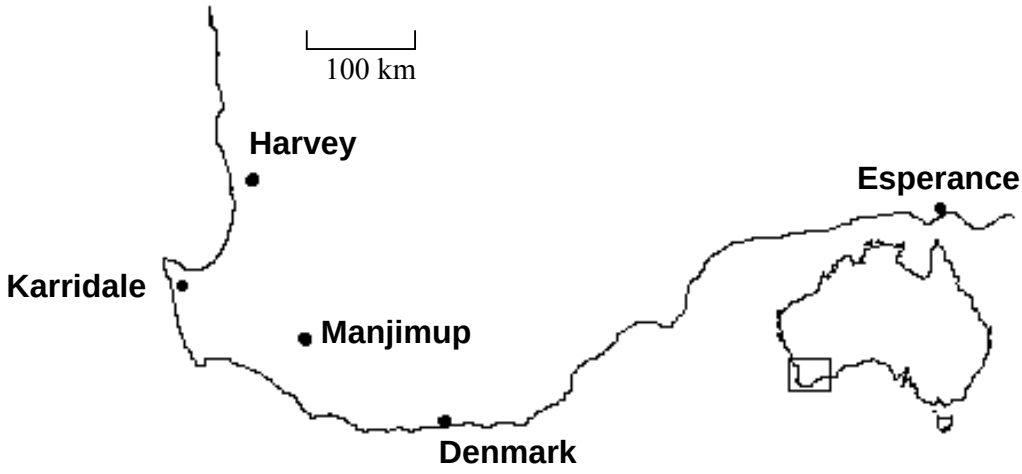
3 - Test alternative explanations for the observed discrepancy between genetic differentiation in mtDNA and in a key life history trait (hypobiosis).

4.2 Materials & Methods

4.2.1 Sampling Design

Twelve worms from each of five different farms across the south of Western Australia were analysed (Figure 4.1). A sample size of twelve was chosen as a compromise between detecting as much of the genetic variation present within a population as possible versus economic and time constraints. Abomasa from cattle were collected at an abattoir (Greens, Harvey) and corresponding tail tags noted. Abomasa from several cattle from the same herd were collected. The worms were collected by rinsing the abomasa in water and hand picking individuals. Due to the difficulties in finding worms the sampling was not complete, however numbers were indicative of relative abundance. The infection intensity showed a typical negative binomial distribution (Figures 4.2 and 4.3) (Smyth, 1994).

Figure 4.1 – Locations from tail tag traces of cattle sampled for *Ostertagia ostertagi* in Western Australia.



Initially the worms were stored in 70% alcohol, however this was found to interfere with PCR so later they were frozen in enough water to cover them. *Ostertagia* worms (Figure 4.2) were separated from the less numerous *Haemonchus* (Figure 4.3); both were then sexed and counted. These two genera were the only nematodes found in the abomasa in the sampled animals. The numbers of *Haemonchus placei* found were too low to permit population genetic study. Tail tags were traced to farmers who were contacted to gain information on whether the cattle were raised on the property (animals

recently bought from elsewhere were excluded). Information on anthelmintic treatment was also obtained.

Figure 4.2 – Number of *O. ostertagi* collected from cattle, sampled from 5 farms in Western Australia.

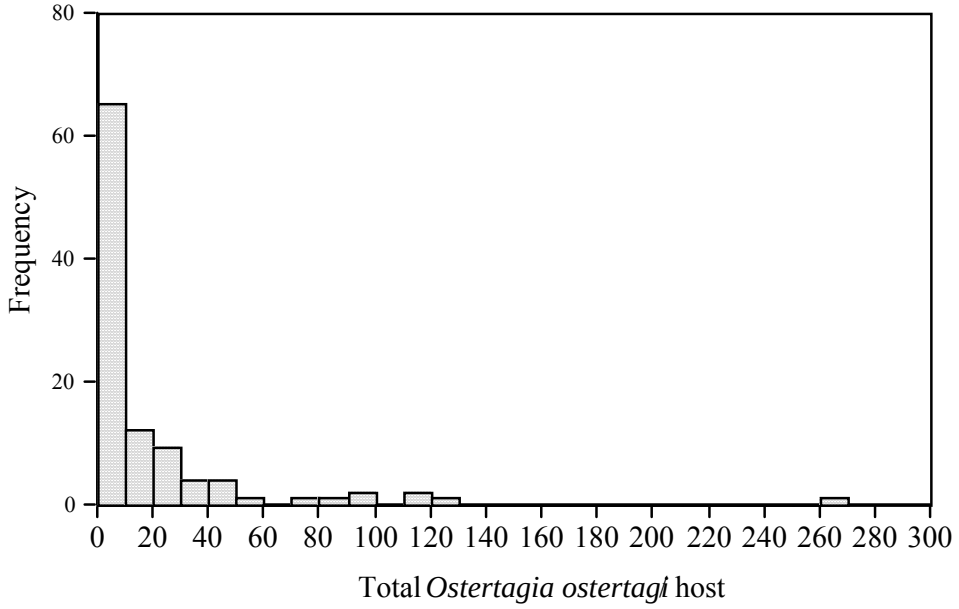
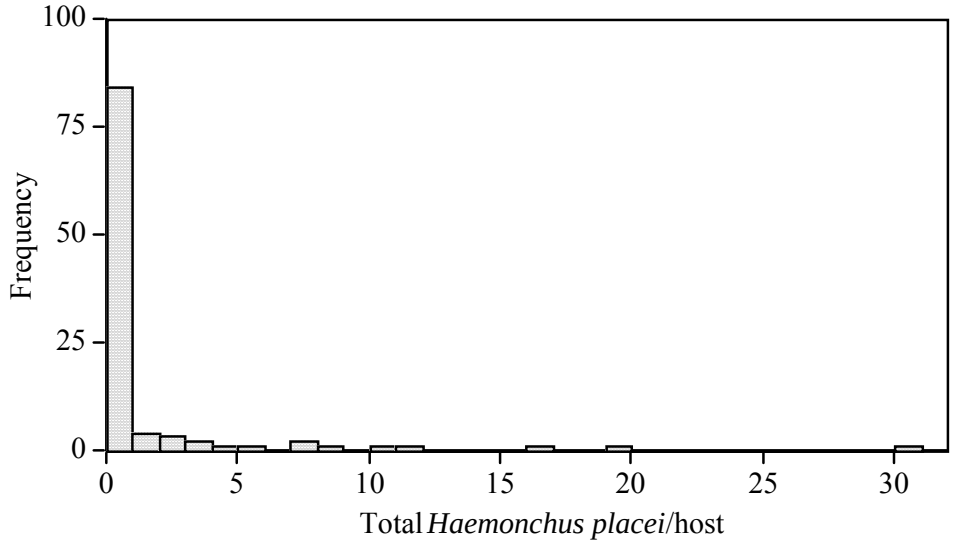


Figure 4.3 – Number of *Haemonchus placei* collected from cattle, sampled from 5 farms in Western Australia.



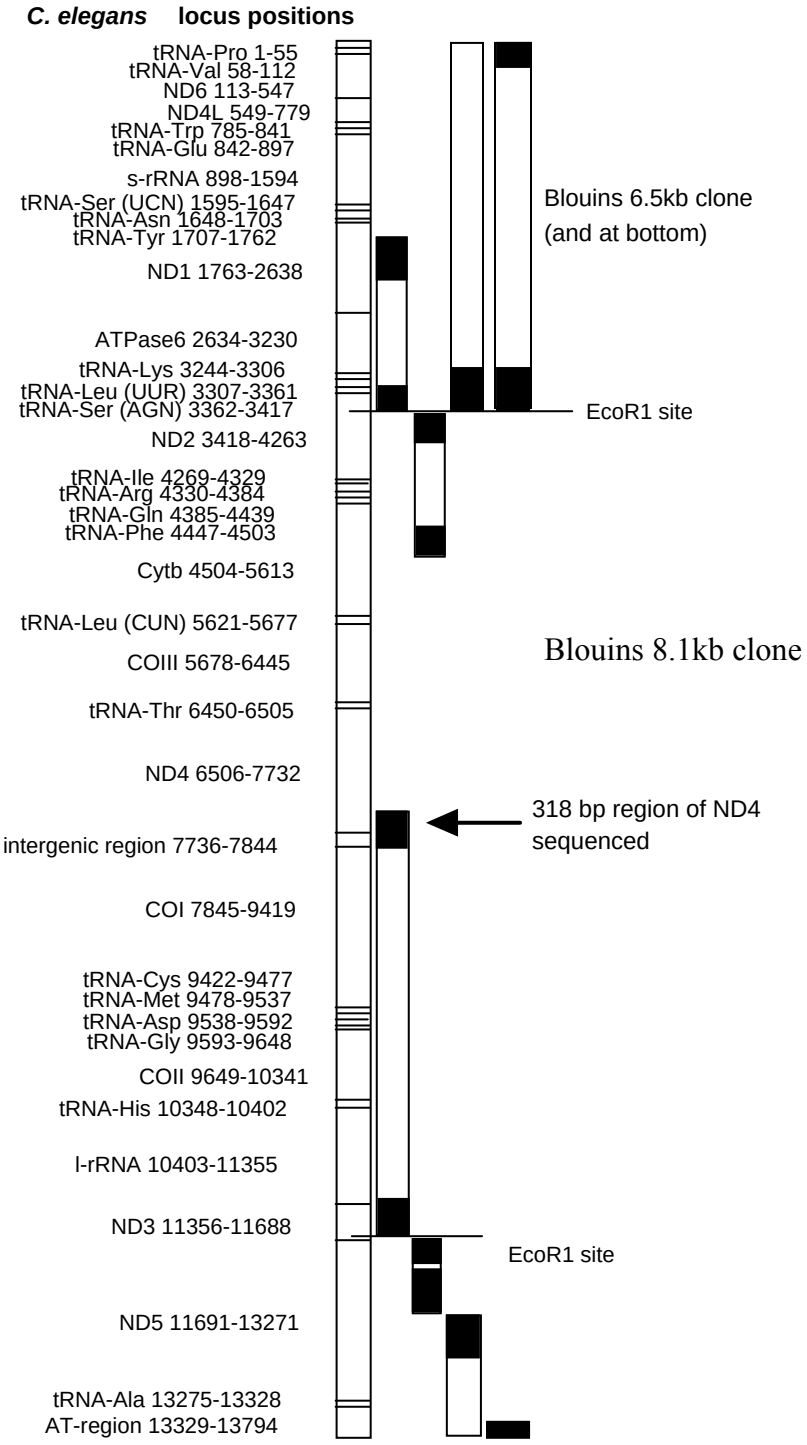
4.2.2 Choice of Genetic Marker

To extend the study of Blouin *et al.* (1992) further, it was decided to sequence a highly variable region within the mtDNA that would show more detail of the differences among individuals within a population. Two other nematodes (at the start of the study), *Ascaris suum* and *Caenorhabditis elegans* had had their entire mtDNA sequenced and provide information to target and design primers. They both have two non-coding regions which are likely to show the appropriate level of variation (Okimoto *et al.*, 1992). Only one, the "intergenic region" or long non-coding region (LNC) had a conserved position, i.e. the same in both species, lying between the COI and ND4 genes in both genomes.

Two clones (8.1kb and 6.5kb) were obtained from Dr Blouin, Department of Zoology, Oregon State University, which made up the entire *O. ostertagi* mtDNA. Since the primers designed from *C. elegans* for the LNC did not amplify *Ostertagia* DNA, the variable region had to be located within the clones. The region was located by labelling the region from *C. elegans* and probing a southern blot of *Ostertagia* mtDNA digested with various restriction enzymes. The labelling methods ECL and DIG (see 2.5.2 for descriptions) were tested but did not have the required sensitivity, so radioactive labelling was used.

Blunt-ended cloning produced 6 clones of various sizes covering most of the mtDNA and the ends were sequenced (Figure 4.4). The LNC region was detected in a 2kb clone, one end of which spans the intergenic region. At the stage of designing primers, Dame & Blouin reported that they had located the LNC region (Blouin *et al.*, 1995). They had discovered that the LNC was too variable so instead sequenced a region of the neighbouring ND4 gene which produced a useful level of variability for a population genetic study on *Ostertagia ostertagi* in the USA. In order to be able to contrast new data directly with their data, their primers were used for this study.

Figure 4.4 - Subclones of *O.ostertagi* mtDNA used to locate the non-coding region by comparison with *C. elegans* gene order, boxes represent subclones, solid regions were sequenced.



The ND4 partial sequence for each individual worm was obtained after PCR, QIAquick purification and sequencing. Each final sequence was a consensus between at least 2

sequence reactions that were aligned and trimmed using SeqEd. The results of the Australian sampling extends the study by Blouin *et al.* (1995) to include two additional scales, within state (between farms) and between continents, and also increases the sample size (Table 4.1).

Table 4.1 - Sample sizes of *Ostertagia ostertagi* from Australia and USA for mtDNA sequencing.

Continent	Population	Abbreviation	Number of worms
Australia	Harvey	Har	12
	Karridale	Kar	12
	Manjimup	Man	12
	Denmark	Den	12
	Esperance	Esp	12
USA	Minnesota	Mn	7
	Maine	Me	5
	Alabama	Al	3
	Louisiana	La	7
	Tennessee	Tn	6

Several other regions of DNA were examined in an attempt to rule out the possibility that the genetic structure observed was due to the characteristics of the mtDNA locus and not representative of the entire genome. Unfortunately, none of these attempts to obtain an alternative data set was successful. Due to the small amount of DNA from a single worm and the relatively simple extraction technique it was not possible to obtain reliable RAPD or anchored SSR banding patterns (Oliveira *et al.*, 1997) from single worm extracts. Both the intergenic spacer (IGS) and internal transcribed spacers (ITS1 & ITS2) of ribosomal DNA were amplified, cloned and sequenced, however, variation was detected even within clones from a single individual, making these markers unsuitable for population genetic analysis (see Chapter 5).

Primers for several single copy nuclear genes with large introns were tested on single *O. ostertagi* worms. Cysteine protease primers were designed from a published *O. ostertagi* sequence (Pratt *et al.*, 1992; accession number M88503), and B-tubulin primers were designed from sequences reported for *Haemonchus contortus*, *Trichostrongylus colubriformis* and *Teladorsagia* (*Ostertagia*) *circumcincta* (Elard *et al.*, 1979; accession Z69258). Failure to amplify these regions from single worms may be due in part to the small amounts of DNA, in contrast to both rDNA and mtDNA that have many copies present.

4.2.3 Analysis

mtDNA sequence data were analysed using the programs Arlequin, DnaSP, SITES, MEGA2, Phylip, and MacClade (see Introduction for program description and locations).

4.2.3.1 Nucleotide and amino acid diversity

Nucleotides were translated to amino acids, using the invertebrate mitochondrial code table. The number of variant basepairs in each of the three codon positions and the number of variant amino acids was counted by hand.

4.2.3.2 Within population diversity

Number of variant base pairs within each population was calculated but since this is strongly dependent on sample size, the average number of nucleotide differences (Tajima 1983, equation A3) within each population was also calculated.

4.2.3.3 Partitioning diversity within and among populations

Analyses of molecular variance (AMOVAs) were conducted on 1. Australian populations, 2. USA populations, 3. Australian and USA combined data and 4. USA summer and winter hypobiosis populations. AMOVA is equivalent to the analysis of Weir and Cockerham, (1984) in that it estimates the proportion of the total variation found within the low level grouping, and the proportion due to each higher level grouping. The AMOVA approach used in Arlequin (Excoffier *et al.*, 1992) takes into account the minimum number of mutations between molecular haplotypes. Simple structures of individuals within populations (as for the first two analyses) gives an Fst value which is identical to the weighted average F-statistic theta (Weir and Cockerham, 1984). In this study the third analysis had three levels: within population, between populations within each continent, and between continents. The significance is tested using a non-parametric permutation approach described in Excoffier *et al.* (1992), with at least 1000 permutations.

4.2.3.4 Gene flow among populations and isolation by distance

Gene flow between each pair of populations in Australia and in the USA was estimated using Nei's (1982) method. Slatkin's linerized Fst and resultant Nm values (from

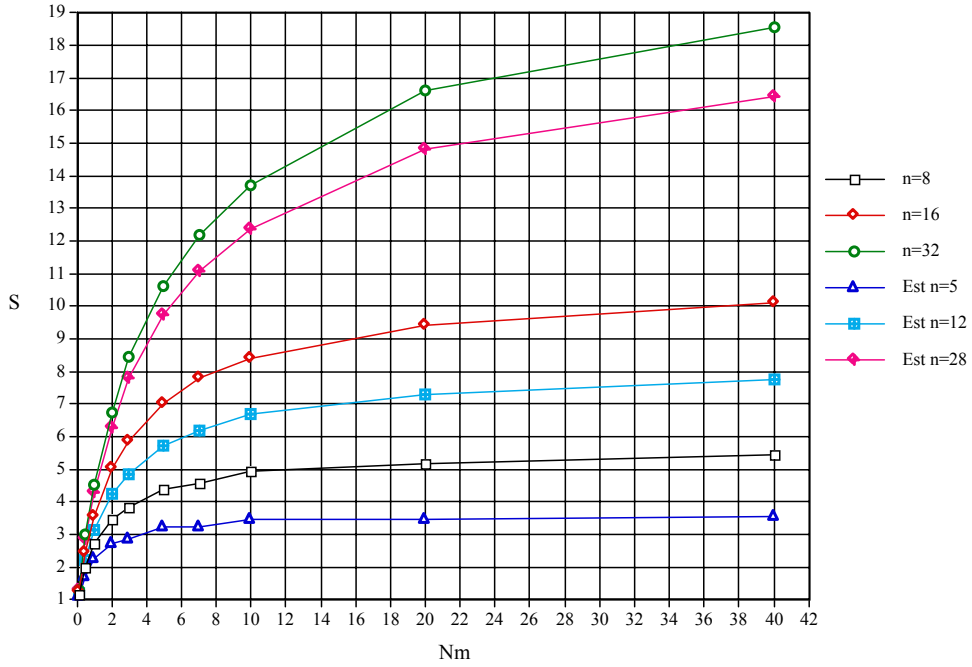
t/M=FST/(1-FST)) were calculated for each pair of populations. Slatkin (1993) recommended testing for isolation by distance by plotting M values against geographic distance, however since many Fst values in this study were zero the M values could not be calculated so instead Fst values were plotted against distance. M is inversely proportional to Fst so if isolation by distance exists you would expect to see a positive regression of Fst and distance.

4.2.3.5 Coalescent analysis

Traditional population genetic methods have concentrated on interpreting the distribution of genetic variation within and among populations. Coalescent theory looks at the phylogenies of different alleles and intra-specific phylogeography (the geographic distribution of alleles). Coalescent analysis is revolutionising the field of population genetics. Importantly, coalescent analysis does not have to assume populations are at equilibrium and can give information about historical changes in the rates of processes, i.e. can reveal information about the history of populations. Standard population genetic analysis assumes a state of equilibrium exists, which may not be justified for *O. ostertagi* due to extensive artificial movement of livestock hosts. The coalescent approach may give more accurate estimates of genetic parameters and it can also reveal extra information such as historic changes over time in the rate of gene flow.

Slatkin and Maddison (1989) presented a method for estimating the average level of gene flow among populations using a coalescent approach, assuming an island model. First a phylogeny of all the sequences is calculated. Population/deme is traced as a character onto the phylogeny. The minimum number of migration events (S) is the number of character state changes on the phylogeny (Slatkin and Maddison, 1989). This S value can be compared to values generated from 1000 random trees to test if it is significantly non-random (testing the null-hypothesis of complete panmixia). Using the method outlined in Slatkin and Maddison (1989), the S value is used to estimate N_em. Table 1 of Slatkin and Maddison (1989) shows the relationship between Nm and S for three different sample sizes (8, 16, 32) resulting from their simulations. The relationships for the sample sizes used in this study (5, 12, 28) are estimated by interpolation as shown in Figure 4.5.

Figure 4.5 - Gene flow from minimum number of migration events (S) as given in Slatkin and Maddison (1989) for sample sizes of n=8, 16 and 32, and as calculated by interpolation for my sample sizes of n= 5, 12 and 28.



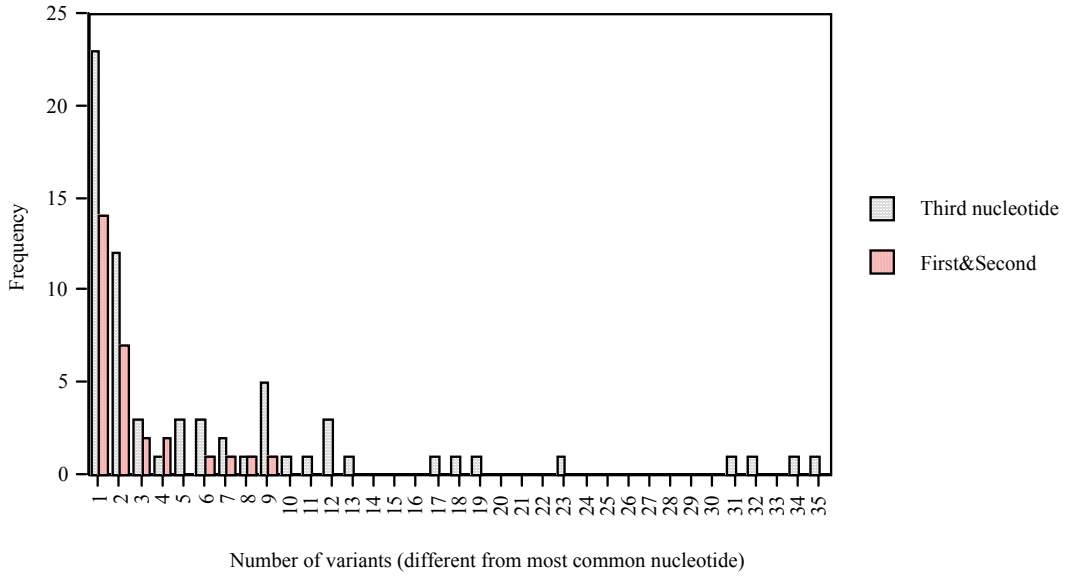
When more that two populations are being tested, the transformation suggested in Slatkin and Maddison (1989) is incorrect. They transform the multiple populations' case to the two-populations case by multiplying the number of migration events by 2/r, where r is the number of populations. This transformation should be to multiply by 1/(r-1), as this is the equation that correctly fits their simulated data displayed in their Figure 7 (Slatkin and Maddison, 1989).

4.3 Results

4.3.1 Nucleotide sequence and amino acid diversity

A 318bp region of ND4 was sequenced for 12 worms from each of 5 farms in Western Australia and compared to 28 sequences from 5 states of the USA. See Appendix 3 for full sequences, and Appendix 4 for corresponding amino acid sequences. Most nucleotide sequences were unique; only 3 pairs and a triplet of individuals were identical.

Figure 4.6 – Frequency of variant sequences by number of variant base pairs (different from the most common nucleotide) sequences for partial ND4 sequence for *Ostertagia ostertagi* (n=88).



This portion of the ND4 locus is highly variable, with 96/318 bp (30.2%) with more than one nucleotide. The variants occur mostly in degenerate third positions (Figure 4.6), 67/106 (63.2%), whereas only 29/212 (13.7%) of first and second position nucleotides were variable. Furthermore, more individuals showed variant 3rd position nucleotides; there were no variant first and second positions that had 10 or more variant individuals (out of 88). There was one base pair (bp 99) that showed a population difference in that all Australian (except one individual), and all Tennessee samples had A in that position, while the rest of the USA populations and one Australian isolate had G.

Overall, there were 23 variable amino acids out of 106 (21.7%). Fifteen of those amino acids were singletons (only one individual was different), five had 2 variant individuals and one site each of four, five and seven variant individuals. Most (63.6%) individuals showed no variant amino acids, 25% had one variant and 11.4% had two variant amino acids.

Sequencing errors are possible and will have increased the observed variability. However, the observation that variation is very non-random (highly biased towards third position). Also rarely more than one type of variant is observed at each site, suggests sequence errors are relatively low (since they should occur independently of coding significance).

4.3.2 Diversity within populations

Within population diversities were similar when comparing average number of nucleotide differences between individuals (Table 4.2). The Australian mean was 7.69 nucleotide differences and the USA mean was 8.43 nucleotide differences.

Table 4.2 Within population diversity, as measured by number of variant basepairs and average number of nucleotide differences, for five populations of *O. ostertagi* each from Australia and the USA

Continent	Population	Number	N. of variant bps	Av # nucleotide diffs
Australia	Harvey	12	37	7.18
	Karridale	12	35	8.08
	Manjimup	12	46	9.27
	Denmark	12	35	8.36
	Esperance	12	36	8.36
USA	Minnesota	7	26	10.19
	Maine	5	14	7.00
	Alabama	3	11	9.09
	Louisiana	7	27	9.52
	Tennessee	6	20	8.53

4.3.3 Partitioning diversity within and among populations

The distribution of variation in mtDNA within and among populations in Australia and the USA is shown in Tables 4.3 and 4.4. More variation was distributed among populations in the USA than in Australia (9.4% versus 3.8%), but in both countries the genetic structuring was significant.

Table 4.3 - Analysis of molecular variance for five Australia populations of *Ostertagia ostertagi*

Source of variation	d.f.	Sum of squares	Variance components	% of variation	Fst	Prob Fst ≥0
Among Populations	4	46.068	0.30206 Va	3.80	0.0379	0.04399
Within populations	59	451.588	7.65403 Vb	96.20		±0.00728
Total	63	497.656	7.95609			

Table 4.4 - Analysis of molecular variance for five USA populations of *Ostertagia ostertagi*

Source of variation	d.f.	Sum of squares	Variance components	% of variation	Fst	Prob Fst ≥0
Among Populations	4	27.833	0.46033 Va	9.42	0.0942	0.00098
Within populations	23	101.810	4.42650 Vb	90.58		±0.00098
Total	27	129.643	4.88683			

Combining the Australian and USA samples also revealed structuring with significant genetic variation due to differences both among populations (Fsc = 0.03877 $p \leq 0.0088$), and among continents (Fct=0.05828, $p \leq 0.01271$) (Table 4.5). None of the total variation in the USA could be ascribed to different regions where populations differed in hypobiosis timing (Table 4.6, Fct = -0.0078, p=0.628).

Table 4.5 – Hierarchical analysis of molecular variance for 5 USA and 5 Australian populations of *Ostertagia ostertagi*.

	df	SS	Variance	Percent	Prob Fxy≥0
Among continents	1	25.88	0.43449 Va	5.83%	0.0127
Among populations within continent	8	73.90	0.27218 Vb	3.65%	0.0088
Within populations	82	553.40	6.74875 Vc	90.52%	0.0059
Total	91	653.17	7.45542		

Table 4.6 – Analysis of molecular variance for 2 winter arresting populations and 3 summer arresting populations in the USA of *Ostertagia ostertagi*.

	df	SS	Variance	Percent	Prob Fxy≥0
Among hypobiosis regions	1	6.83	-0.03793	-0.78	0.6280
Within regions among populations	3	21.00	0.48398	9.93	0.0000
Within populations	23	101.81	4.42650	90.85	0.0010
Total	27	129.64	4.87255		

4.3.4 Gene flow among populations and isolation by distance

Gene flow, calculated by Nei's 1982 method, between Australian populations was higher (Table 4.7, mean Nm=12.3) than between USA populations (Table 4.8 mean Nm=3.0). This is not surprising since the Australian populations were all close and from one state, whereas the USA populations were from different, distant states. Intercontinental estimates of gene flow, between the Australian and USA, range from 5.15 to 9.09, depending on the method of estimation (Table 4.9).

Table 4.7 – Pairwise gene flow (Nm) estimates (calculated by Nei's 1982 method) between five Australian populations of *Ostertagia ostertagi*.

	Den	Esp	Har	Kar	Manjimu
N=	12	12	12	12	12
Denmark		11.5	10.1	12.4	11.2
Esperance			11.4	12.5	9.7
Harvey				12.3	17.4
Karridale					14.4

Table 4.8 – Pairwise gene flow (Nm) estimates (calculated by Nei's 1982 method) between five USA populations of *Ostertagia ostertagi*.

	Al	La	Me	Mn	Tennessee
N=	3	7	5	7	6
Alabama		3.6	1.8	3.5	2.4
Louisiana			3.5	4.2	2.7
Maine				3.2	2.0
Minnesota					3.3

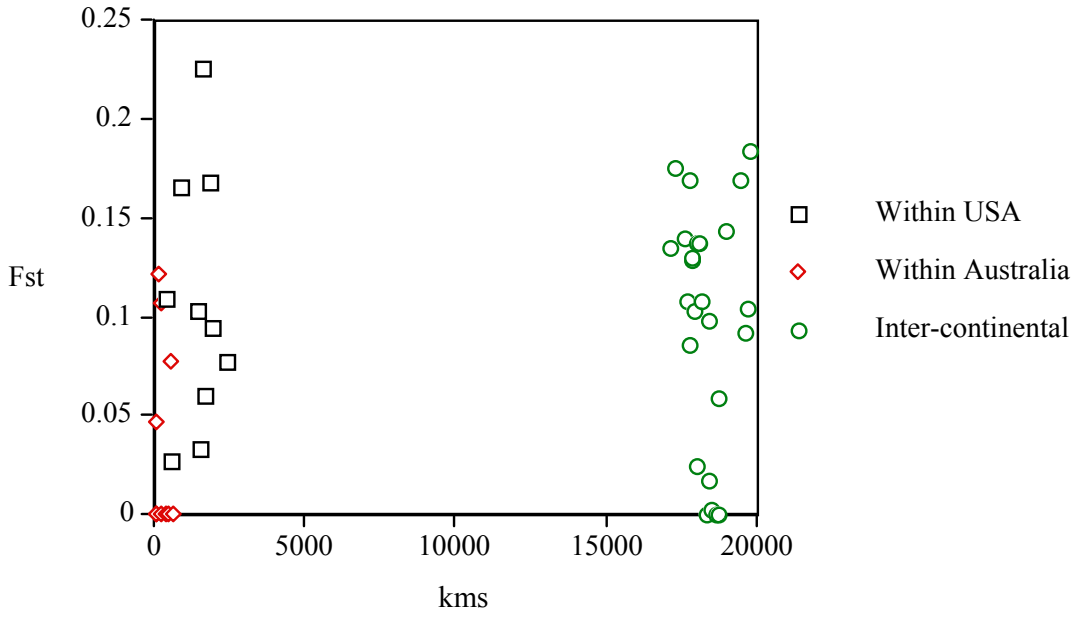
Table 4.9 - Intercontinental estimates of gene flow (Nm) for *O. ostertagi* between Australia and USA.

Method	Fst equivalents	Nm
Nei 1982	DeltaSt=0.0017 GammaSt= 0.0521	9.09
Lynch & Crease 1990 (with Jukes and Cantor correction)	Nst= 0.0885	5.15
Hudson, Slatkin & Maddison 1992	Fst= 0.0869	5.25

Table 4.10 - Slatkins linearized FSTs above diagonal; Nm values below diagonal (where t/M=FST/(1-FST)) for populations of *Ostertagia ostertagi* from Australia and the USA

	Har	Man	Kar	Den	Esp	Ten	Mai	Ala	Min	Lou
Harvey		0.046	0.121	0.107	0.078	0.059	0.104	0.018	0.129	0.137
Manjimup	10.82		0.000	0.000	0.000	0.000	0.092	0.000	0.086	0.103
Karridale	4.13	inf		0.000	0.000	0.000	0.184	0.098	0.130	0.137
Denmark	4.67	inf	inf		0.000	0.003	0.170	0.108	0.139	0.170
Esperance	6.44	inf	inf	inf		0.024	0.144	0.108	0.135	0.176
Tennessee	8.44	inf	inf	176.35	20.70		0.226	0.110	0.103	0.166
Maine	4.80	5.42	2.71	2.95	3.47	2.22		0.168	0.095	0.077
Alabama	28.56	inf	5.12	4.62	4.63	4.57	2.98		0.033	0.027
Minnesota	3.87	5.81	3.85	3.58	3.70	4.85	5.26	15.14		0.060
Louisiana	3.65	4.83	3.64	2.95	2.85	3.01	6.46	18.58	8.33	

Figure 4.7 –Geographic distance versus pairwise population Fst's for populations of *O. ostertagi* in Australia and USA.



Pairwise Fst and Nm values among populations from Australia and the USA are shown in Table 4.10. There was no linear correlation between pairwise estimates of inter-population Fst's and inter-population geographic distances (Figure 4.7). Fst values between Australia and each of the USA populations fell within the range of those between the states in the USA. The low sample sizes for the USA populations may have contributed to the higher between population Fst estimates in the USA.

4.3.5 Coalescent analysis of gene flow

A maximum likelihood neighbour-joining tree (Figure 4.8) was drawn for all individuals from Australia (60) and the USA (28). In order to carry out Slatkin & Maddison's 1989 coalescent estimate of geneflow the comparisons had to have the same number of worms for each population, so five different groupings of randomly selected subsets of the total data set were analysed (Table 4.11). Figure 4.9 shows a pruned version of the tree in Figure 4.8, with geographic origin of each of 28 samples from Australia and 28 individuals from the USA traced on, to illustrate how the minimum number of migration steps was calculated (last comparsion in Table 4.11).

Table 4.11 thus shows the minimum number of migration steps for different combinations of populations, and the resultant estimate of gene flow (M). The migration rate between Australian farms, and also between all the populations at once was too

high to estimate. The geneflow estimate between the four USA state populations was the lowest (3.5). Grouping all the summer USA hypobiosis versus the winter samples gave a high geneflow estimate (15.5). Grouping all Australian samples versus all USA samples grouped gave a geneflow estimate of 7 (tree shown in Figure 4.9).

Figure 4.8 – Maximum likelihood and neighbour joining tree of 318 bp of ND4 sequence from 28 USA and 60 Australian *Ostertagia ostertagi*.

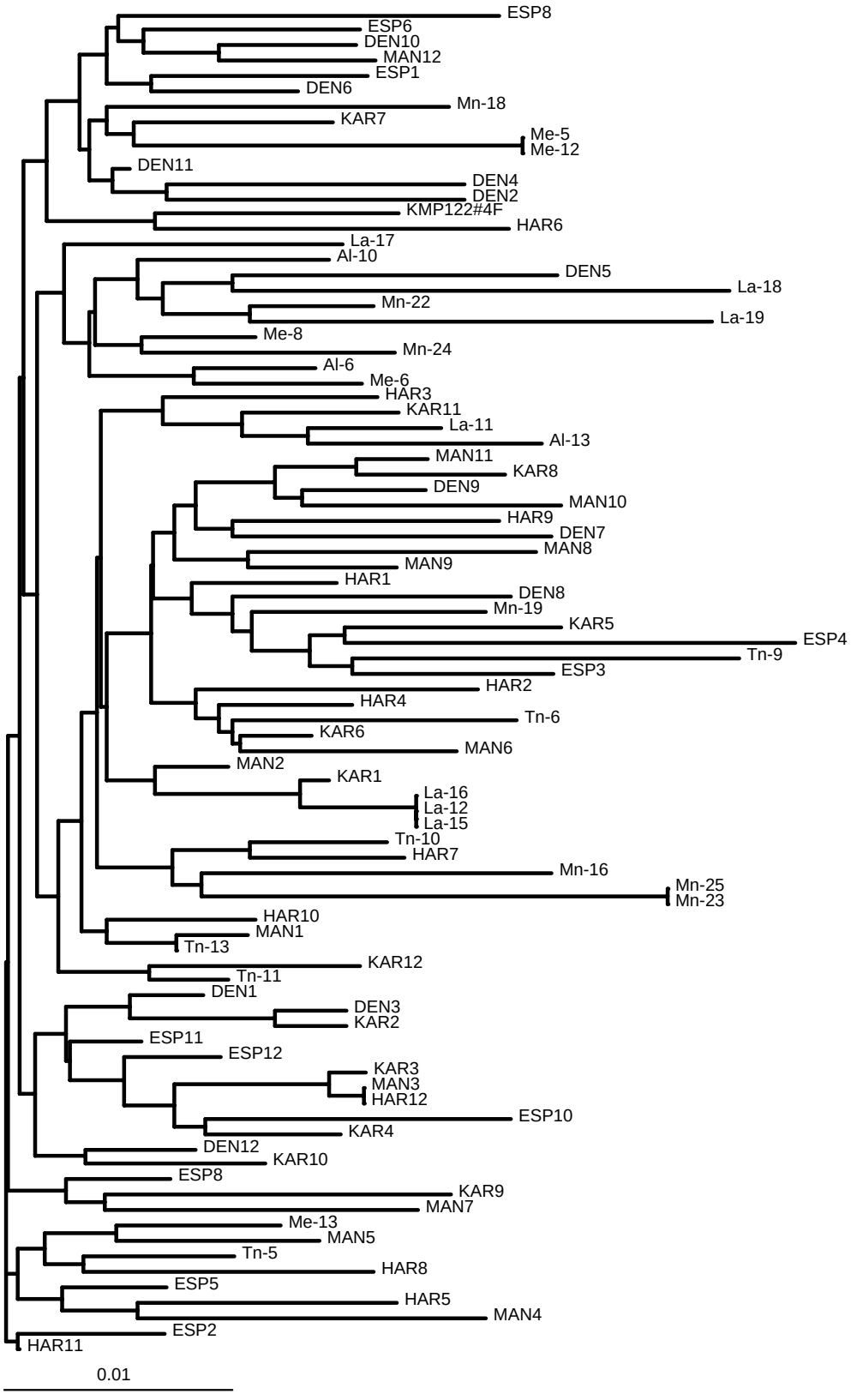


Figure 4.9 – Location traced onto maximum likelihood neighbour joining tree for 28 samples each from Australia and USA of *O. ostertagi*.

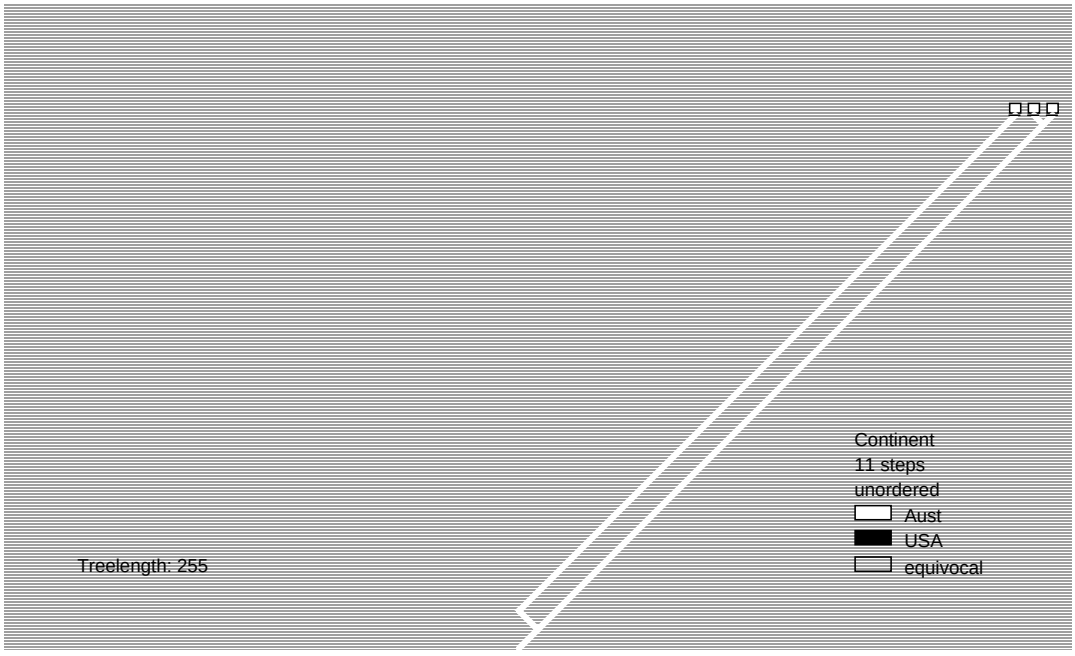


Table 4.11 – Gene flow estimated from phylogenies of *Ostertagia ostertagi* from different populations using the modified Slatkin & Maddison (1989) method.

Comparison	Populations/ demes	N each population	Steps	S/(r-1)	M
Australian farms	5	12	38	9.50	Inf
USA states	4	5	9	3.00	3.5
All populations (Australian and USA)	9	5	31	3.88	Inf
USA winter vs summer hypobiosis	2	10	6	6.00	15.5
Australian vs USA	2	28	11	11.00	7

4.4 Discussion

Diversity within the sequenced portion of the ND4 mtDNA locus of *Ostertagia ostertagi* is very high with most haplotypes being unique. Although within population diversity is very high, there was also a significant proportion of variation attributable to between population variation. Gene flow between the sampled populations was very high, although not panmitic. The genetic structure in Australia was similar to that seen in the USA. Although there was higher between population structure in the USA, probably due to each USA population being from distant states, while the Australian herds sampled were all from one state.

The existence of a disjunction between northern and southern USA in an important trait, hypobiosis, which exists despite the low Fst values, presents a paradox. In the introduction three main explanations for the observed discrepancy between genetic variation and variation in hypobiosis were presented:

1. High mutation rate of mtDNA (High μ)
2. High migration rate (High m)
3. High effective population size (High N_e)

Each of these alternatives would be expected to have different consequences for the observed pattern of genetic diversity.

4.4.1 High mutation rate of mtDNA (High μ)

High mutation rate (for either that particular locus or for mtDNA in nematodes) would increase within population diversity and thus proportionally decrease Fst values. A high mutation rate would also result in most haplotypes being unique. This is supported by the data: few samples are identical (only 3 pairs and a triplet out of 88, see Figure 4.8) and most of those are from the same population.

However, if high mutation rate is responsible for the observed lack of genetic diversity among populations, the underlying structure should be observable by using a coalescent analysis, since each unique haplotype is not treated as if it is completely different from all other haplotypes (as occurs with allelic analysis). The relationships between the

haplotypes should still reflect geographic history even though they are independently accumulating mutations. Coalescent analysis should therefore result in lower gene flow estimates since coalescent analysis can detect real levels of gene flow through the noise of the high mutation rate. There was, however, little structure detected by looking at parsimony (or maximum likelihood, or distance) clustering of the samples. The gene flow estimates between populations within Australia, within the USA or between continents were very similar using either an Fst or a coalescent approach. Thus, the present analysis does not support the theory that a high mutation rate in mtDNA is responsible for the discrepancy between variation in hypobiosis, and genetic variation in part of the ND4 locus.

4.4.2 High migration rate (High m)

High current migration rates would result in high within population diversity, as many migrants bring in different haplotypes. However, there should also be many cases of the same haplotype being found in different populations, reflecting recent migration events. This was not observed as there was only a single case of the same haplotype occurring in different populations. Coalescent and standard gene flow estimates would be expected to be similar, fitting the observations. There would also be a distinct increase in Fst values (and hence markedly lower Nm values) between continents, as there are few adult animals exchanged between countries, but an isolation by distance effect within continents would be expected. There was little support for either isolation by distance within continents or an increase in Fst due to continental isolation (Figure 4.7, Table 4.9).

High migration would also have to be countered by high selection pressures to maintain the observed disjunction in hypobiosis. In populations near the disjunction, this would result in a very high genetic load as all of the migrants with the "wrong" hypobiosis timing for that climate must be quickly selected out. Lack of genetic differentiation at neutral loci does not however rule out differentiation in other characters (Bensch *et al.*, 1999). A study by Peterson and Denno (1997) is an example of selection maintaining differentiation in a trait despite high gene flow. The hypobiosis trait may not be as simple as suggested. For example, there may be an environmental component which would decrease the required selection, and hence genetic load. The existence of

different levels of hypobiosis on the same farm in Australia between dairy and beef herds shows that differentiation can occur at a local scale.

High migration rate is thus not well supported due to (1) only one case of the same haplotype in neighbouring populations; (2) no large increase in Fst values for intercontinental comparisons and (3) the high genetic load required to maintain the observed differentiation in the hypobiosis trait.

4.4.3 High effective population size (High N_e)

The final alternative explains the observed N_em values as being mainly due to high effective population sizes with proportionally fewer actual numbers of migrants. High effective population sizes would result in high within population diversity with many unique haplotypes expected, as was observed. Gene flow estimates using the standard or coalescent methods would be similar and show less effect of isolation by distance. Since the number of actual migrants is low the genetic load required to maintain the differentiation in hypobiosis would be lower. This alternative thus seems to best fit the observations. Quantitative estimation of nematode effective population size (Ne) is very difficult (Nadler, 1995; Anderson *et al.*, 1998). Blouin et al (1992) made an estimate using an average nucleotide substitution rate of 0.02 substitutions between random pairs of individuals. Using the formula $N_f=1/4N_e$ and assuming nematode mtDNA evolves at the same rate as vertebrate mtDNA (0.01 substitutions per lineage per million years) then the average time to common ancestry between pairs of random individuals is one million years. With one to two generations per year the long term Ne was estimated as four to eight million individuals.

4.4.4 Recent radiation/colonisation

The population genetics of parasites of domestic hosts is greatly affected by the artificial movement of livestock and is likely to be very different from that of parasites of wild hosts. In particular, the high rate of artificial movement and the relatively short period of domestication means that domestic animals and their parasites are unlikely to have reached equilibrium between drift and migration, therefore many of the assumptions of standard population genetic models are not met. Since domestic

livestock has only recently colonised both the USA and Australia (2-300 years) from what is likely to be a basically common stock (European), the populations have not had sufficient time to reach a drift/migration equilibrium.

The observed high gene flow (or high S values or low Fst values) may result not from current gene flow but due to the fact that populations are similar since they have been recently derived from a common stock. The inability of both standard and Slatkin & Maddison methods to distinguish between current gene flow and historical association is discussed further in Slatkin & Maddison (1989). The low bootstrap values and star-like (all branches radiating from the centre) phylogeny suggest low phylogenetic signal in the dataset. This would also explain why the intercontinental Fst values were no larger than the within USA comparisons. Effectively, each population is a sample from the original stock with insufficient time for lack of continental gene flow to cause divergence.

The difference shown by Blouin *et al.* al (1992), between the nematodes of a wild host, which show a high degree of structuring, versus the domestic nematodes, which show little structure, may not be due to the high rate of artificial host movement but to the lack of time to reach equilibrium given that domestic hosts have only recently (a few hundred of years) radiated, whereas wildlife hosts have maintained similar ranges for presumably thousands of years. Most nematode studies have been on parasites of either humans and domestic livestock, or commensals (Anderson *et al.*, 1998) which have probably been subject to recent range expansion, and have not had time to reach equilibrium.

4.4.5 Conclusions

This study has extended a study of USA *O. ostertagi* to two further levels of scale, farms within a state and intercontinental. It also analysed the combined data by a variety of software programs using different methods and a coalescent approach (Slatkin and Maddison, 1989). Three possible explanations of the observed paradox between a fixed difference in biological trait, hypobiosis, across the USA and lack of genetic differentiation are examined in detail and predictions of their resultant effects on genetic structure compared with the measured genetic structure. The high effective population size alternative is found to be best supported. Since this estimate of genetic structure is based on only a single mtDNA locus ideally other loci should be examined to ensure the structure is consistent with nuclear loci. Since a highly variable region is required to detect intra-population variation necessary for a population genetic analysis the spacer regions (ITS and IGS) of ribosomal DNA were examined as described in the next chapter.

5 Genetic diversity within and among individuals: Ribosomal spacer regions, ITS1 and IGS, of *Ostertagia ostertagi* both show a repeat region with intra-individual variation in repeat number

5.1 Introduction

In the previous chapter, variation at a mtDNA locus was examined within and among populations of *Ostertagia ostertagi* in Australia. To determine if the genetic structure shown by mtDNA was representative of the entire genome, a nuclear region was sought that should show variation at the same scale (that is, variable within populations) to allow a population genetic analysis.

5.1.1 Ribosomal DNA

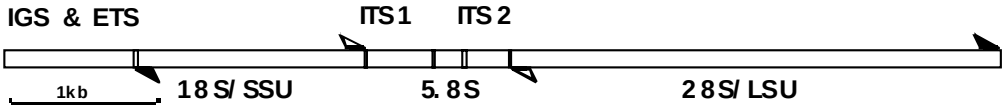
Ribosomal DNA (rDNA) in eukaryotes typically is present in several hundred tandemly repeated multiple copies and the different genes and regions show different degrees of conservation (Hillis and Dixon, 1991). Multiple copies make it easy to amplify rDNA even from small amounts of sample.

The different genes vary in their level of conservation from the 28S (or LSU) and 18S (SSU) genes which have highly conserved regions and are ideal for designing "universal" primers, to the ITS regions which are usually moderately variable, and the IGS region which is highly variable. This allows the ribosomal region to be used across most phylogenetic scales. The conserved regions are used for 'tree of life' phylogenies and the highly variable regions (internal transcribed spacers, ITS1 & ITS2 and the intergenic spacer, IGS) for the differentiation of closely related species (detection of cryptic species) or for detecting variation within and among populations of the same species.

In *Ostertagia ostertagi* the entire rDNA unit is 7.5kb (Figure 5.1). It is directly repeated and represents 0.9% of the total genomic DNA (Dame *et al.*, 1991). The rDNA of *O. ostertagi* is similar to that of *Caenorhabditis elegans* in the size of the coding regions, unit length, uniformity of the repeat units, the percentage of the genome the units

represent and the fact that rDNA is the only simple satellite sequence in the genome (Files and Hirsh, 1981; Dame *et al.* 1991).

Figure 5.1 - A rDNA unit of *Ostertagia ostertagi* showing coding regions (shaded) and spacer regions, arrows represent primers used, open arrows for ITS region and closed for IGS region.



5.1.2 Ribosomal spacers

The region that separates the rDNA units is called the intergenic spacer or IGS, previously called the non-transcribed spacer or NTS. It evolves the most rapidly and is responsible for most of the observed length variations among rDNA units. It often contains subrepeating elements that serve as enhancers of transcription (Hillis and Dixon, 1991) and differences in the number and sequence of these repeats account for most of the length variability of rDNA units (Baldridge and Fallon, 1992). Other features found in the IGS sequence include RNA polymerase I promoters, spacer promoters and transcriptional terminators (Baldridge and Fallon, 1992). Transcription is initiated within the IGS and that part of the IGS region which is transcribed is called the external transcribed spacer (ETS).

The internal transcribed spacers lie between the 18S and 5.8S (ITS1) and between the 5.8S and 28S (ITS2) coding regions. The entire rDNA unit is transcribed by RNA polymerase I as a single 45S precursor molecule (Schlotterer *et al.*, 1994). The transcribed spacers contain signals for processing the rRNA transcript (Hillis and Dixon, 1991). The secondary structure of the spacers is important for processing of the rRNAs and in *Drosophila* approximately 40% of the spacer sequence is not free to diverge (Schlotterer *et al.*, 1994).

Previous studies suggested that for population genetic analyses of *Ostertagia*, only the more variable regions (ITS and IGS) were likely to be useful (Hillis and Dixon, 1991; Hoste *et al.*, 1995). Primers were therefore designed for both the ITS and IGS regions.

Using these primers, substantial variation was detected at a finer scale than anticipated: within individual worms. This complicated population genetic analyses, but led to interesting questions about the extent and source of intra-individual variation.

5.1.3 Aim

The aim of this chapter is to investigate the extent and source of intra-individual length variation observed in both IGS and ITS regions of *Ostertagia ostertagi*, by examining the underlying sequence variation in clones. The significance of this variation is discussed with respect to use of these spacer regions for population genetic studies and for species diagnosis, and the implications of concerted evolution failing to homogenise these multi-copy regions.

5.2 Materials & Methods

5.2.1 DNA extraction

DNA was extracted by a variation of the method of Blouin *et al.* (1995), by placing individual worms in tubes with 100µl extract buffer (10mM TrisHCl pH8.3; 2.5mM $MgCl_2$; 50mM KCl; 0.1mg/ml gelatin; 0.45%NP40; 0.45% Tween20). The tubes were freeze/thawed three times in liquid nitrogen, and 10µl proteinase K 10mg/ml was added. The tubes were incubated for 1 hour at 55˚ C, and then proteinase K was heat inactivated by heating to 95˚ for 10 mins.

5.2.2 PCR Amplification

The IGS region was amplified using primers designed from the large subunit (3' end, ATACTGCGATCTGTTGAGAC) to the small subunit (5' end, GCTCTAATGAGCCGTTCGC) sequences for other nematodes. Similarly the SSU ITS region primer was designed from database sequences (SSU 5' end, CTTGAACCGGGTAAAAGTCG), and a LSU primer designed by Gasser *et al.* (1994), primer NC2, TTAGTTTCTTTTCCTCCGCT was used. PCR was carried out with 4mM $MgCl_2$, 200µM dNTP, 2.5µl reaction buffer, Tth 0.1µl, 12.5pmoles of each primer and 2µl of DNA (see extraction above) with water to a total of 25µl. The thermocycler program was: 94˚ 2mins [94˚ 30secs 50˚ 1 min 72˚ 2 mins]x50 72˚ 7mins. When visualised, both these PCR reactions from single worms produced distinct multiple (3-5) bands approximately 100bp different in size.

5.2.3 Cloning of multi-sized PCR product

Purification using the QIAquick (QIAGEN) kit was carried out for IGS only, prior to cloning using the TA cloning kit (Invitrogen). Screening of the white colonies was done by taking half the colony with a pipette tip and adding 50µl TE with 1% TritonX, boiling for 5-10 mins, then spinning down and transferring the supernatant. 5µl was used in PCR with M13 primers. All colonies with no PCR band and those with bands of the expected size were retested with specific primers. Five to nine positive colonies for each region were examined, from each of three worms for the IGS and a single worm

for ITS. Further worms were not sequenced since the within individual variability observed meant that it would be necessary to clone and sequence numerous clones from each of the sixty worms sampled, which was not economically feasible and may not have produced useable population genetic data. Sequencing of selected clones was carried out using Applied BioSystems Taq DyeDeoxy™ terminator kit.

5.2.4 Analysis

Sequences were aligned using Clustal X. Since the fragments were cloned prior to sequencing, any PCR errors are incorporated in the clone, so only variation which is present in two or more clones is discussed, although all variation was included in Phylip analyses. PCR and cloning both preferentially produce smaller fragments, thus the actual ratio of different size spacers in each individual may be different to that observed after cloning. Maximum likelihood and parsimony analyses were used to determine phylogenetic relationships among clones. MacClade was used to map the characters (repeat size, repeat position, worm origin) on the phylogenies. Trees shown are maximum likelihood with neighbour joining, the bootstrap values (from 100 bootstraps jumbled 10 times) that were greater than 50% for a clade are shown, and >50% bootstrap values from parsimony are given in brackets.

5.3 Results

5.3.1 IGS

From three different worms from Denmark, Western Australia (see location map, Figure 4.1), there was a total of 23 positive clones with between 1-4 copies of a 104bp simple repeat region. From both the observed density of the multiple bands produced by direct PCR and from the clones isolated, the proportions of copies with different number of repeats seemed to vary between worms (Table 5.1, band densities not shown).

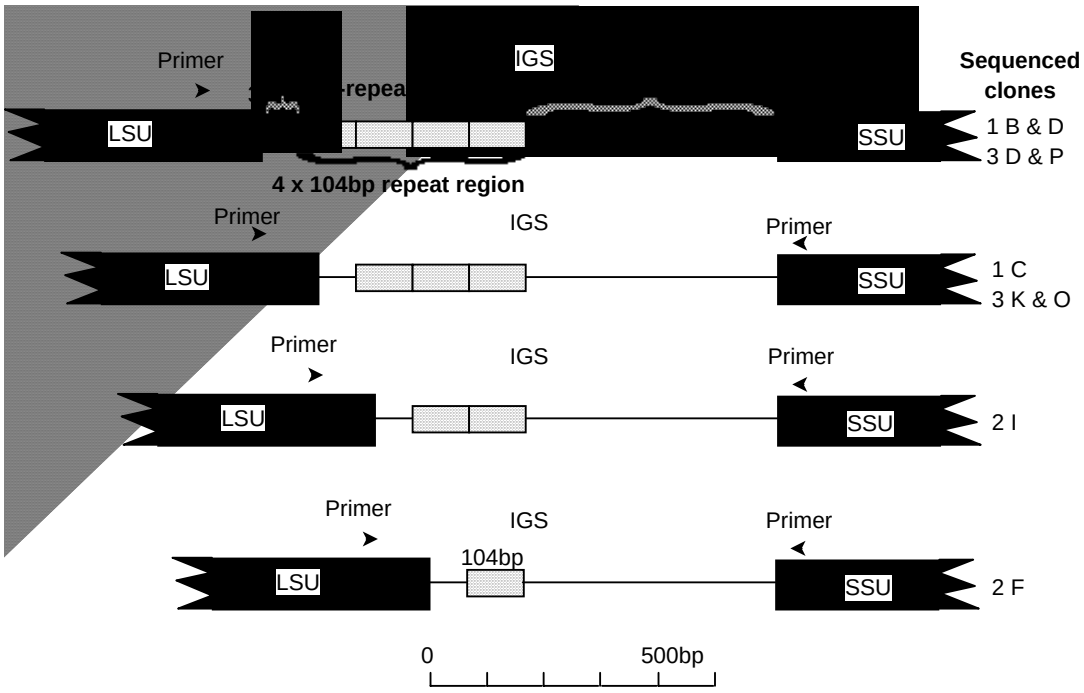
Table 5.1 - Numbers of 104bp repeats in cloned IGS fragments from three *Ostertagia ostertagi* worms, number of clones sequenced in brackets.

Worm	Clones with 4 repeats	Clones with 3 repeats	Clones with 2 repeats	Clones with single copy	Total clones
1	7 (2)	2 (1)	0	0	9 (3)
2	3	0	5 (1)	1 (1)	9 (2)
3	2 (2)	3 (2)	0	0	5 (4)
Total	12 (4)	5 (3)	5 (1)	1 (1)	23 (9)

5.3.1.1 IGS sequence variation

Sequencing of nine clones revealed some variation both outside the repeat region and within (Appendix 5 shows the complete sequence for each clone). There are several pairs/triplet of bases which appear to be linked, that is if one has changed so has the other, both within and outside the repeat region (shown by circles in Appendix 5). The difference in length due to different numbers of repeats is shown diagrammatically in Figure 5.2.

Figure 5.2 - Diagram of IGS repeat units in different clones of *O. ostertagi*. Under the right heading "Sequenced clones", the numeral represents the worm (1-3) and letters represent the clones.



5.3.1.2 IGS phylogenies

Phylogeny of the IGS repeat regions (Figure 5.3) showed approximate grouping by repeat position (final number), with no evidence that repeats within an individual (start with same number), were more similar than between individuals. This is shown clearly in Figure 5.4, in which repeat position (first to fourth) is traced onto sequence phylogeny. It is also apparent from Appendix 5, which shows that all first copies show an A in the second position, while all others have G. Furthermore, there are several other changes which occur at different frequencies in different copy positions. Phylogeny of non-repeat regions showed no sign of clones within worms clustering together, though there was some suggestion that clones with the same number of repeats tend to cluster together (Figure 5.5).

Figure 5.3 - Phylogeny of IGS repeat regions only from 3 different worms (1-3), clone letter (B,C,D,F,I,K,O,P), and repeat position (1-4). Arrows indicate the four repeats found in clone B of worm 1.

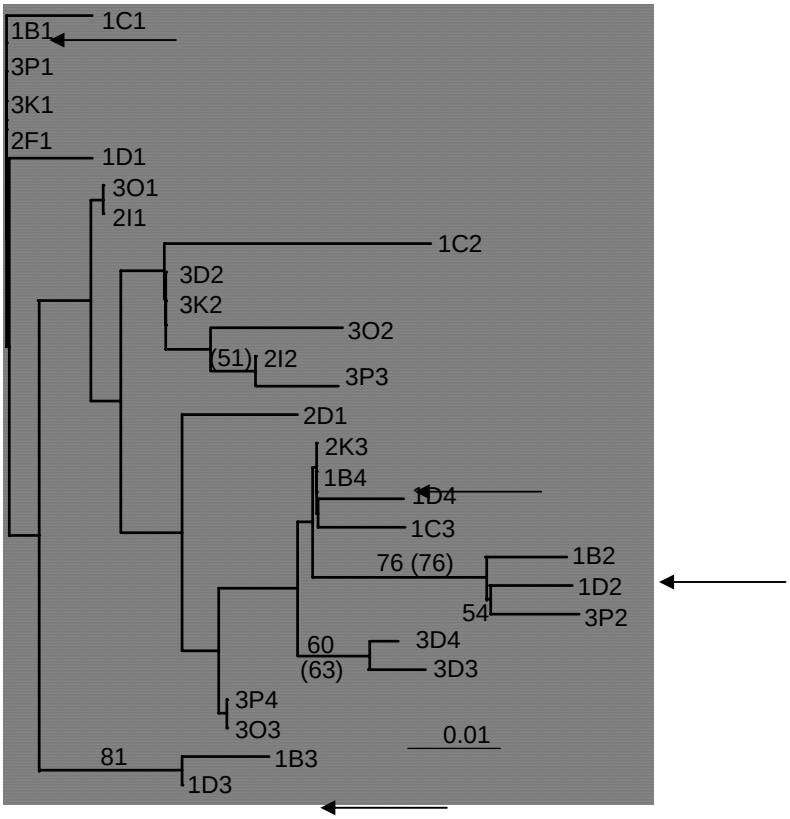


Figure 5.4 - Phylogeny of IGS repeat regions of *O. ostertagi* with repeat position traced, first number is the worm, the letter is the clone, final number is position of repeat, which is also shown by shading.

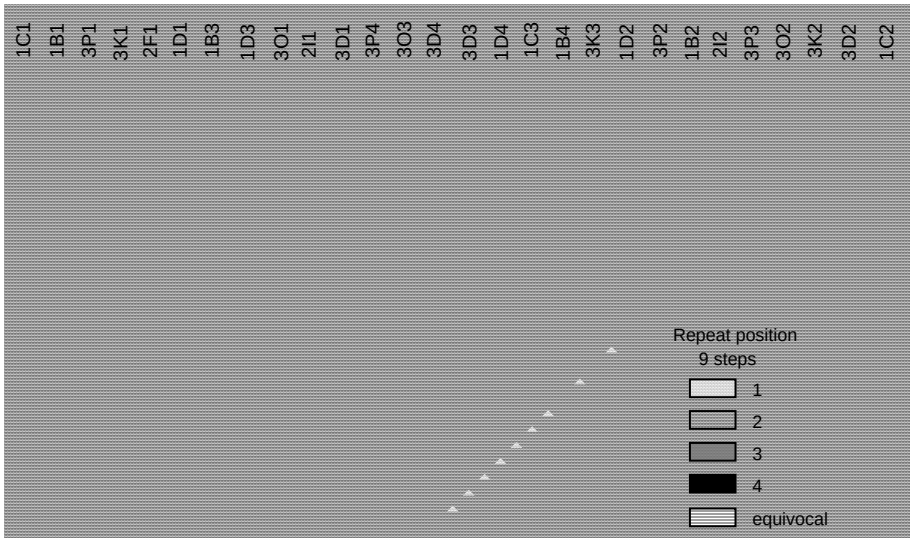
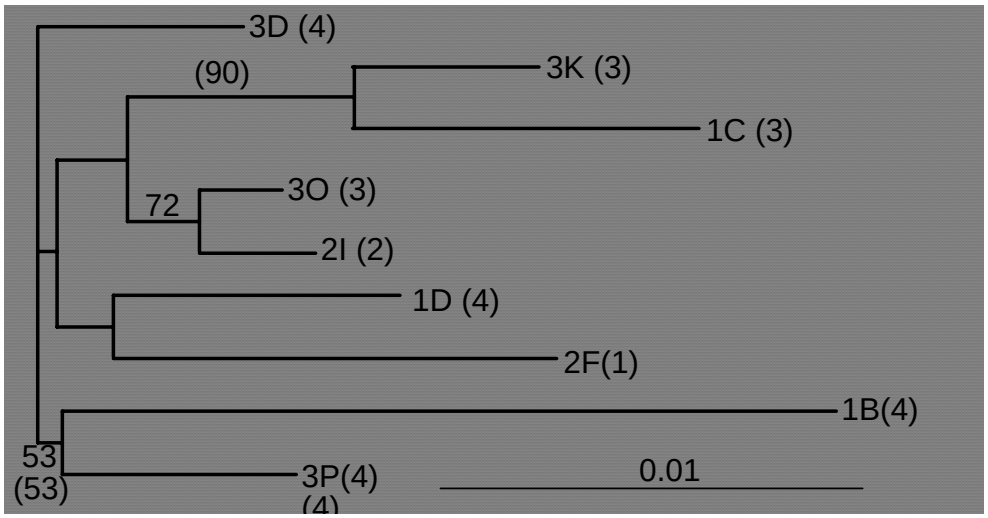


Figure 5.5 - Phylogeny of non-repeat IGS regions of *O. ostertagi* cloned from three worms (number of repeats shown in brackets), using Kimura's distance and neighbour joining. Numbers on branches are bootstrap values above 50 obtained from consensus trees for maximum likelihood, and parsimony (in brackets).



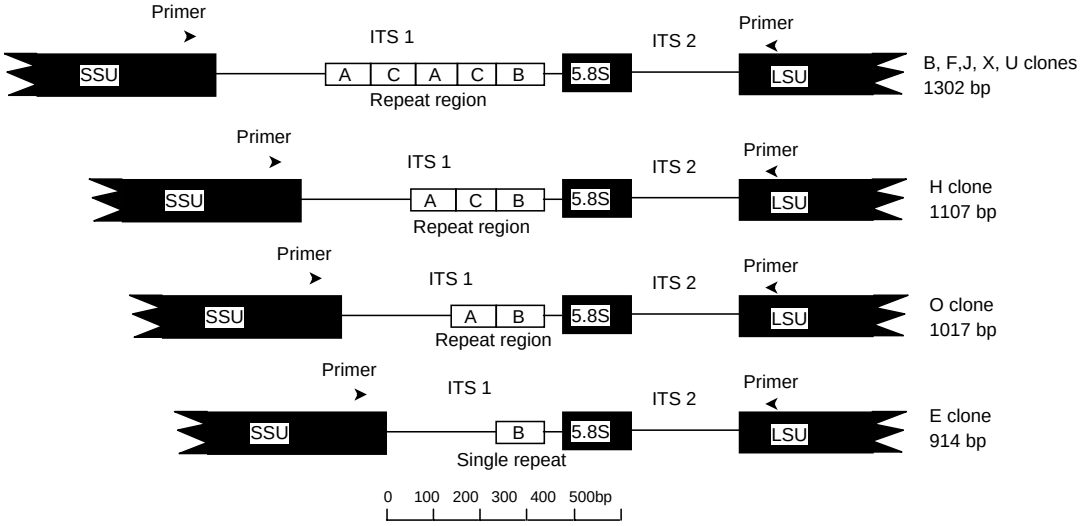
5.3.2 ITS

Of the 8 ITS fragments cloned from a single worm from Karridale (see map Figure 4.1), 5 clones had 5 repeats, and 1 clone each had 3, 2 and 1 repeat(s) (Figure 5.6). The X clone was not fully sequenced although the total length was consistent with having 5 copies.

5.3.2.1 ITS sequence variation

The repeat versions were grouped according to size where A=113bp, B=108bp and C=91bp. Appendix 6 shows the full sequence for each clone and Figure 5.6 shows a diagrammatic representation of the variants.

Figure 5.6 - Diagram of ITS1 repeat units in different clones of a single *O. ostertagi*. Letters on the right represent different clones.



5.3.2.2 ITS Phylogenies

Phylogenetic analysis of the ITS repeats shows that repeats are clustered mainly by length which is not surprising as difference in length is treated as multiple gap characters (Figure 5.7). However, the same analysis using only the first 91 bp (which is common to all repeats, Figure 5.8) produces a very similar phylogeny. Tracing position and repeat type onto the phylogeny (Figure 5.9), reveals that type C's still all group together, with the addition of O2. This O2 repeat is identical to X4 (a C type), both lack a C at position 84 which all other C types show. Since the O clone has only 2 copies of the repeat region it may be the result of a deletion which combined a C type with the end of a B type. The A and B types were intermixed in the phylogeny.

Figure 5.7 - Phylogeny using Kimura distance and neighbour joining of ITS repeats from clones of a single *O. ostertagi* worm, first letter indicates clone, number is the repeat position, and final letter is the repeat type. Numbers on branches are bootstrap values above 50 obtained from consensus trees for maximum likelihood, and parsimony (in brackets).

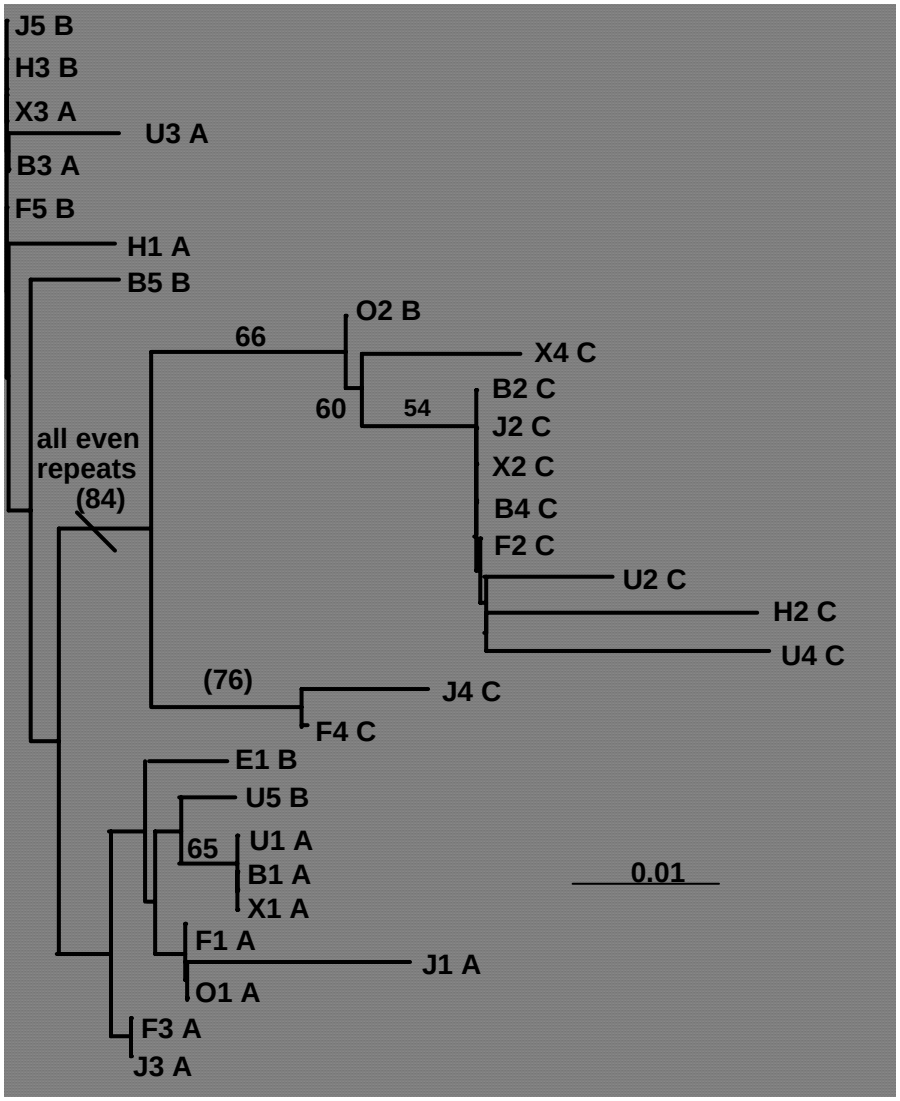


Figure 5.8 - Phylogeny of ITS repeats only first 91 basepairs, from clones of a single *O. ostertagi* worm, first letter indicates clone, number is the repeat position, and final letter is the repeat type. Numbers on branches are bootstrap values above 50 obtained from consensus trees for maximum likelihood, and parsimony (in brackets).

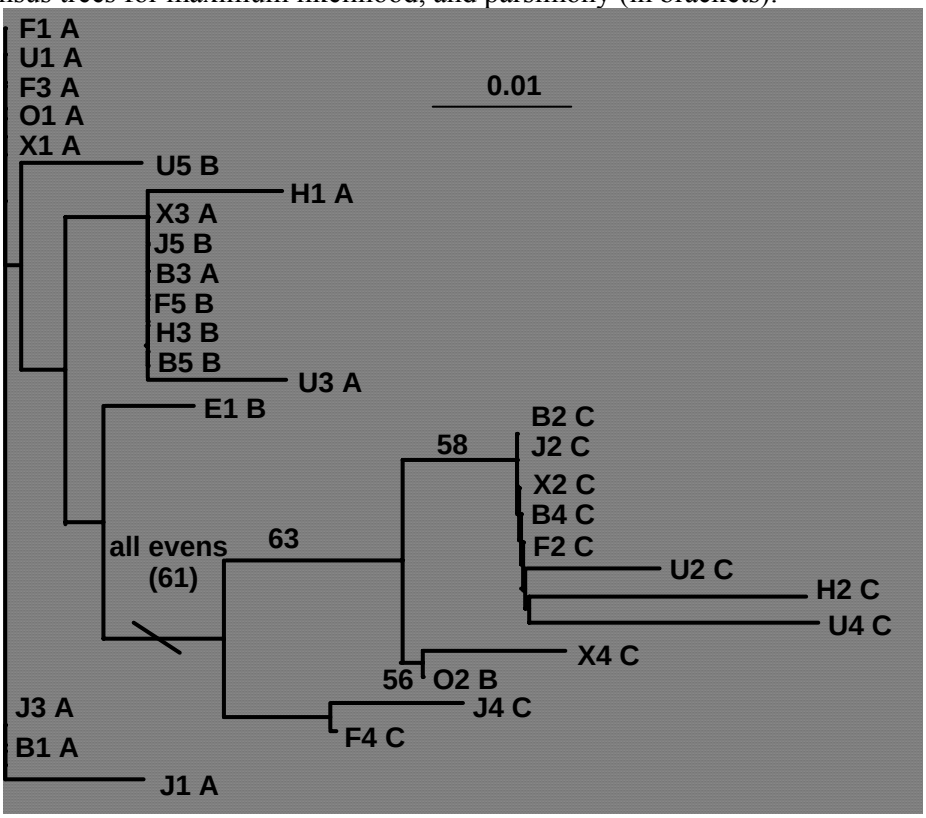


Figure 5.9 - Phylogeny of ITS repeat regions only the first 91 basepairs with position traced, from clones of a single *O. ostertagi* worm, lower letter indicates clone, number is the repeat position, and top letter is the repeat type.

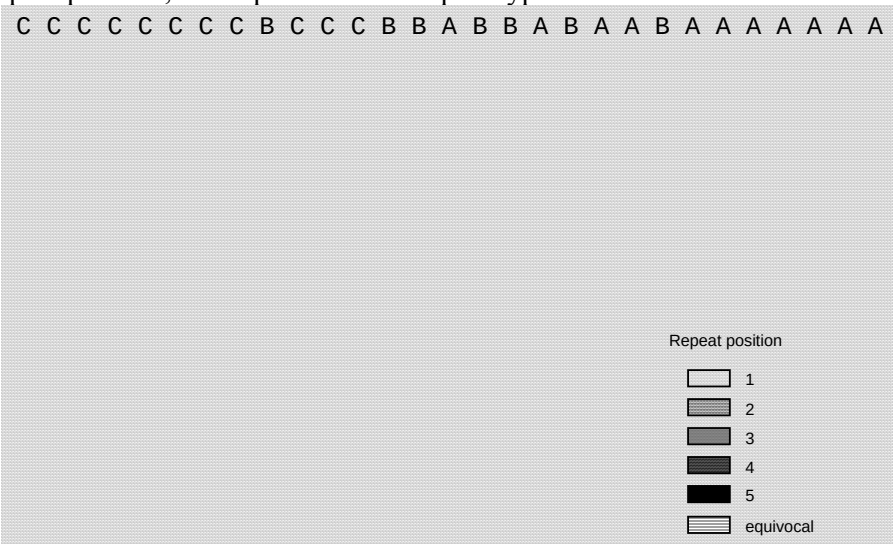
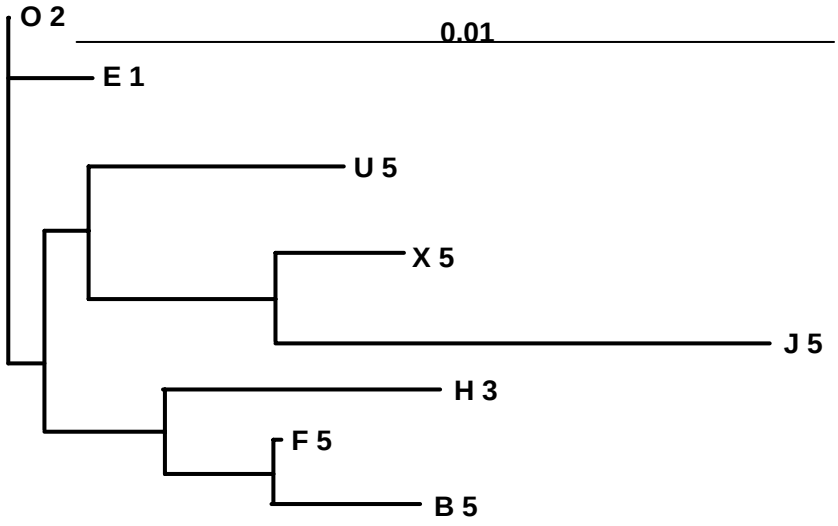


Figure 5.10 - Phylogeny of ITS non-repeat regions (sequence before and after repeat region) of clones from a single *O. ostertagi* . Letter is clone and number is the number of repeats found in that clone.



5.3.2.3 Comparison with ITS1 from other species

All of the clones had a terminal B type and alignment with published ITS1 sequences of other *Ostertagia* spp. and closely related nematodes *Teladorsagia circumcinta*, *Haemonchus contortus* and *H. placei* (all from Zarlenga *et al.*, 1998b) revealed that all other *Ostertagia* and species from related genera have a single copy of this region which is similar to the B type (Figure 5.11).

Figure 5.11 - ITS1 region sequence for various trichostrongyle nematodes (first is a consensus from this study; others are from Zarlenga *et al* 1998b AF044927-34)- the * marks an insert for *O. ostertagi and O. lyrata* and the [] mark the region which is repeated in the insert.

```
O.ostertagi  TCGAAACCA-AaACACATGGTTCCTTTGATTATGAGAAACCAACATGCAAGaTTTTTCACGACTTTGTCG
O.ostertagi  ..........A.....................................W.................
O.lyrata     ..........A.....................................A.................
O.leptospi   .........C.A...A.............C.............C..T..AA....-...........
O.mossi      .........CTA.................C.................TT.-......G...........
O.dikmansi   .........CTA.................C.................TT.-......G...-.......
T.circumsi   ..........A.................C.C............-..TTCA.....-...........
H.contortus  ........-T.A.....A...........C.C............-..T.TG.....--...........
H.placei     ........-T.A.....A...........C.C............-..T.TG.....--...........

O.ostertagi  TGAAACGTTGGGAGTATCACCCCCGTTAAAGCTCTATATT--TGAGGTGTCTATGTATGACATGAGTCGT
O.ostertagi  ............................................................................
O.lyrata     ............................................................................
O.leptospi   .....T.....................G..........ATG.....................................
O.mossi      ..........A.............................A.....................................
O.dikmansi   ..........A.............................A.....................................
T.circumsi   .....-.....................G..........G.....................................
H.contortus  .A...-.................................-.ACA.....................C...
H.placei     .A...-.................................-.ACM.....................C...

O.ostertagi  TCATGAGTGGCGACTGTGATTGTTCATGCAAAGTTCCCATTCCAATGGTTGGTTATGGTTGAGCTTGAGA
O.ostertagi  ............................................................................
O.lyrata     ............................................................................
O.leptospi   ..G.....................................YATTG..........A..........
O.mossi      ..G.........G.............................T..CATTG..A.................
O.dikmansi   ..G.........G.............................T..CATTG..A.................
T.circumsi   ..G.........................G.......T..CA.CT.TA.....G............T...
H.contortus  ..GA........G..............G.......T..CA---------T.G...............
H.placei     ..GA........G..............G.......T..CA----------..G..A............

O.ostertagi  CTTAATGAGTATTGCTATAATACCGCCTCATTGCA*[TTTATAATGGTGGTTATGTACATAC----------
O.ostertagi  ....................................*[.............Y.....................
O.lyrata     ....................................*[.................................
O.leptospi   ......................T.........A.T--[.....................C.T..........
O.mossi      ......................T.........A.T--[.....................C.T..........
O.dikmansi   ......................T.........A.T--[.....................C.T..........
T.circumsi   ......................T.........A.T--[.....................................
H.contortus  ......A..............T......CC.T--[..AT............A....GA..CAAATTACTT
H.placei     ......A..............T.....GCC.T--[..AT............A....GA..CAAATTACTT

O.ostertagi  -----GTGTA-GTATGTaCGGTACCTGGcTGT--CAGGAAACCTTAATGATCTCGCCTAGCTTTTTATGT
O.ostertagi  ................A..........T.....................................
O.lyrata     ................A..........T.....................................
O.leptospi   ...GGC....CA.G...A..........T............T..........CG....G..C...C....
O.mossi      ...GGC....CA.G...A..........C.....................CG..T.G...A.......
O.dikmansi   ...GGC....CA.G...A..........C.....................CG..T.G...A.......
T.circumsi   .....CGCAYG..G...A..........T.............................C..C.....
H.contortus  CTTGAAGTATGTGG...A.T.....C.AT.A.AT.G..G............A...G...ACA..C..T.
H.placei     CTTGAAGTATGTGG...A.T.....C.AT.A.AT.G..G............A...G...ACA..C..T.

O.ostertagi  TT]AAAGTTTGCAGAaAT--GTGTACACAGAAA--TGTGTCACAATT---GAC
O.ostertagi  ..]............A.................................
O.lyrata     ..]............A.............................C........
O.leptospi   ..]............A........T....TT..TT............ATGT..
O.mossi      ..]............A........T....TTC.TT............ATGT..
O.dikmansi   ..]............A........T....TTC.TT............ATGT..
T.circumsi   ..]...T........A.C......T....TT--TT............ATC..T
H.contortus  ..].C..........AC.TA....T....TTC.TT..........A.ATC..T
H.placei     ..].C..........AC.TA....T....TTC.TT..........A.ATC..T
```

5.3.2.4 ITS2 variation and comparison with other species

There was little variation seen in the ITS2 region except for a linked pair of sites where half the clones had C and A and the other half had T and G (Appendix 6). A published

sequence for *O. ostertagi* correlates exactly with the first variant (Stevenson *et al.*, 1996). Figure 5.12 shows ITS2 from several closely related species.

Figure 5.12 - ITS2 sequence of *Ostertagia ostertagi* from this study and previously reported sequences for *O. ostertagi* X86027, *O. leptospicularis* X86025, *Teladorsagia circumcincta* X86026 (Stevenson *et al.*, 1996)[1]); *O. arctica* AJ250657, *O. gruehneri* AJ400716, *Marshallagia marshalli* AJ400715 (Dallas *et al.*, 2000)[2]; and *O.trifurcata* AJ251124 (Chilton *et al.*, 2001)[3].

```
O.ostertagi    aatgaaactactacagtgtggctagttyataacactgtttgtcgaatggt
O.ostertagi[1] ........................c.....................
O.artica[2]    .......................acac.....................
O.gruehneri[2] .......................acac.....................
O.leptospic.[1] ......................acac.....................
O.trifurcat.[3] ......t...............acat.....................
M.marshalli[2] ........a.............acat.....................
T.circumci.[1] .......................acat.....................c

O.ostertagi    atttattactttattgtgataattcccattccagttcaagaatatgaaat
O.ostertagi[1] ................................................
O.artica[2]    .....c....a.................t...........-.ac.c..
O.gruehneri[2] .....c....a.................t...........-.ac.c..
O.leptospic.[1] .....c....g.................t...........-.ac.c..
O.trifurcat.[3] .....c....c.................t............ac.c..
M.marshalli[2] .....c....c.........c......t............ac.c..
T.circumci.[1] .....ym...k.................t............ac.t..

O.ostertagi    gcaacatgacgttaaca---ttgt----------taacgttactgaatga
O.ostertagi[1] ................................................
O.artica[2]    ................tga.............................
O.gruehneri[2] ........r........tga.............................
O.leptospic.[1] .......a........taa...................g........
O.trifurcat.[3] .......a........-.ga-...................c........
M.marshalli[2] ................-.ga-...................c........
T.circumci.[1] ..............-...-.gacg..attaccgtcg.......c........

O.ostertagi    tactgaatatat--taccactatttgaatgtactcaatgaatatgagatt
O.ostertagi[1] ................................................
O.artica[2]    ..-.........at..t..............................
O.gruehneri[2] ..-.........at..t..............................
O.leptospic.[1] ..-.........at..t..............................
O.trifurcat.[3] ..-.....g......g.............................g..
M.marshalli[2] ..-.....g.........g...................g.........c
T.circumci.[1] ..-....cgcg.at.g.t..............................

O.ostertagi    gattcaaatagggacatgtatggtattgtatattcratgtatcatttatat
O.ostertagi[1] ...................................a............
O.artica[2]    ........................a.a...g...a..........g...
O.gruehneri[2] ........................a.a...g...a..........g...
O.leptospic.[1] ........................a.....g...a..........ga.a
O.trifurcat.[3] .....g.....a.............c....cg...a..........g...
M.marshalli[2] ....t......a.............c.....g...a..a........g...
T.circumci.[1] ......g....a.............a.a...g...a..........g...
```

5.4 Discussion

5.4.1 IGS

In many organisms the IGS region is very long and complex with multiple imperfect repeat regions (e.g. *Xenopus* 1.8-5.5kD with 3 repeat regions, Wellauer *et al.* 1976; *Aedes* ~6.3kb one major repeat region with 17 x~205bp repeats, Baldridge and Fallon 1992; *Daphnia pulex* 4.8kb 2 repeat regions, Crease 1993; *Avena* 3.89kb with 5 repeat regions, Polanco and Delavega 1994).

The IGS structure found in this study for *Ostertagia ostertagi* was shorter and simpler compared to most other organisms, with a total length from ~600 to 900 bp and a single imperfect repeat region showing 1-4 copies of 104bp repeats. Another nematode, *Meloidogyne arenaria* showed similar variability with the 2kb IGS showing deletions due to a reduction in the number of 129bp tandem repeats (9, 10, 12, 14 repeats) (Vahidi and Honda, 1991). These repeats were perfect except for a single base pair change in repeat number 4, similar to the *O.ostertagi* first repeat showing a fixed second base pair difference from all other repeats.

O. ostertagi IGS repeats showed a tendency to cluster by position (Figure 5.3). *Daphnia pulex* shows even more position specific clustering of repeats and a similar finding of one dominant class of repeat size (18/21 clones) with a few with variant number of repeats (3/21) (Crease, 1995). One study on an amphibian reported even less variation in IGS (which was 1.5kb long) with only a maximum of 120bp length differences observed (Morgan and Middleton, 1992).

5.4.2 ITS

Unlike the IGS, intra-individual variation in size and the presence of repeat regions in ITS has more rarely been reported. *O. ostertagi* has a repeat region within the ITS1 which shows length variation within an individual worm. Similar repeat regions are seen in some trematodes (van Herwerden *et al.*, 1999) including *Schistosoma* (Kane and Rollinson, 1994; van Herwerden *et al.* 1998). The repeats within the ITS region were not perfect, showing both length and sequence variation, much of which could be

attributed to repeat position (for example Figure 5.8 shows all the even position repeats clustered together and all but one were C type length). There is also a low level of sequence variation outside of the repeat region of ITS1 and in the ITS2 between clones from the same worm. In nematodes, multiple different sequences of ITS are frequently found within individual worms (Anderson *et al.*, 1998).

5.4.3 PCR-induced deletions may be responsible for length variation

PCR-induced deletion of repeat regions has been reported (Fenton *et al.*, 1998). This is explained as template movement forward to a similar region during PCR resulting in a deletion between the repeat regions. Since the *O. ostertagi* spacer regions were first amplified by PCR prior to cloning and have repeat regions in both IGS and ITS1, similar template movement may account for the cloned shorter variants which appear to have fewer numbers of repeats. However, each clone was unique outside the repeat regions and none were the result of a simple deletion event from another clone. If PCR-induced deletions are common, many studies reporting variation in length may need to be re-examined.

5.4.4 Implications for rDNA as a diagnostic tool

A number of studies have used rDNA sequences, particularly ITS1 and ITS2, as markers to distinguish between nematode species. A study by Zarlenga *et al.* (1998a), examined the ITS1 of 8 species including 5 species of *Ostertagia, Teladorsagia circumcincta*, *Haemonchus placei* and *H. contortus*. They reported that *Ostertagia ostertagi* and *O. lyrata* had ITS regions that were approximately 400bp larger than the other *Ostertagia* and *Haemonchus* spp. They developed a semi-quantitative method for determining the percentage of *O. ostertagi* eggs within a mixed population. They described the observed variation as *O. ostertagi* and *O. lyrata* having a 408bp insert consisting of 2x204bp repeats. All species (of *Ostertagia* and *Haemonchus*) showed a B type region, which in *O. ostertagi* and *O. lyrata* has undergone duplication to create up to 5 imperfect copies of the region. Similarly, Newton *et al.* (1998) distinguished numerous nematode species by RFLP of a fragment containing both ITS regions and the 5.8S gene. They also found that all species produced a fragment of ~ 870bp in length except *O. ostertagi* which was ~1250bp.

The present study, however, suggests that although 5 copies of the region in *O. ostertagi* is the most common (12/23 clones) it is possible to get other length versions from PCR of the ITS1 region of *O. ostertagi*, including a rare (1/23 clones) version with a single copy making it of similar length to the other *Ostertagi* and *Haemonchus* spp. Zarlenga *et al.* (1998a) did observe some weaker bands but did not sequence them. This may lead to an underestimate of true proportion of *O. ostertagi* when using the semi-quantitative method and potentially classifying a pure *O. ostertagi* sample as containing a proportion of other species.

5.4.4.1 Compensatory mutations in spacer regions to maintain secondary structure

Secondary structure is maintained despite continuous evolution of the primary sequence by compensatory mutations between paired nucleotides (Hillis and Dixon, 1991). The present study revealed numerous examples of compensatory mutations in both the IGS and ITS regions (shown by circles Figures 5.1 and 5.5). An analysis in several species of *Drosophila* estimated that 40% of the ITS was not free to diverge due to structural constraints (Schlotterer *et al.*, 1994). The observed tendency for repeats in the same position to be more similar is also likely due to structural constraints which opposes unequal crossing-over resulting in homogenization of repeats.

5.4.5 Implications for concerted evolution

Concerted evolution is the non-independent evolution of repetitive DNA sequences, resulting in a sequence similarity of repeating units that is greater within than among species. This intra-specific homogenisation of sequences is said to take place via the poorly understood mechanism of molecular drive (Elder and Turner, 1995). Concerted evolution appears to operate variably in different species, with some showing complete homogenization and others maintaining intra-individual variants for long periods of time even through speciation events (Kane and Rollinson, 1994; van Herwerden *et al.*, 2000). There are several possible mechanisms which may affect the operation of molecular drive.

Firstly, it seems likely that if rDNA exists on different clusters located on different chromosomes, molecular drive is less efficient (Arnheim *et al.*, 1982; Vogel and

DeSalle, 1994), however heterogeneity is observed even in species with only one rDNA locus (Kane and Rollinson, 1994; van Herwerden *et al.*, 1998; van Herwerden *et al.*, 1999). Secondly, a site-specific transposable element (R4) has been reported from *Ascaris* which inserts in some rDNA sequences and may inhibit concerted evolution by reducing homology and pairing between adjacent repeats (Anderson *et al.*, 1998).

Thirdly, it is possible that rare selective sweeps occur when a rDNA variant arises which is positively selected, and rapidly replaces all other existing neutral variants. Finally, hybridization may lead to mixing of rDNA types (Sang *et al.*, 1995). In *Plasmodium* there are stage specific rDNA's, with the different copies being expressed at different stages and evolving independently, however *Plasmodium* lacks some of the standard mechanisms of rDNA sequence correction (Rogers *et al.*, 1995).

Results from this study show that concerted evolution is not homogenising the spacer regions, as both sequence variation and variation in repeat numbers are maintained. Each unit is not however accumulating mutations randomly; there are correlated base changes and repeat-specific changes that reveal that secondary structure is constraining variation. Neutral variants are being incorporated at a rate faster than any homogenising mechanisms can remove them.

5.4.6 Use of multi-copy regions

The presence of intra-individual and intra-species variation in different rDNA copies can complicate the use of these regions for phylogeny or diagnostics, although it may still be possible to get useful information (Vogel and DeSalle, 1994). Since it is apparent that the rate of homogenization varies greatly across different species groups (Schlotterer *et al.*, 1994), it is important to look for intra-individual and intra-population variation when using multi-copy regions.

Sequencing of only one or a few clones, or direct sequencing of bulked individuals may not reveal all the existing variation. A study on Scots pine found that most of the variation in spacer regions was within populations (86% including intra-individual, half of the variants were found only in single individuals) and only 14% between populations (Karvonen and Savolainen, 1993). While this may constrain the use of such

multi-copy regions in population genetic studies, it seems that for distinguishing species the species-specific features of these regions may still be useful (Novak *et al.*, 1993).

5.4.7 Conclusion

This study reveals that *O. ostertagi* shows intra-individual variation in repeat lengths within both the IGS and ITS1 spacers when amplified from single worms, which has not been previously reported despite studies utilising the ITS1 region length of *O. ostertagi* for diagnostic purposes. At present, it cannot be excluded that this length variation is due to PCR-induced deletions, and not representative of the actual genome however in either case it affects the use of these regions for diagnostic or population genetic studies.

There is also a significant level of sequence variation in both regions apart from repeat length. This sequence variation seems to be constrained by secondary structure, as there were paired nucleotide changes and repeat position specific variants. Concerted evolution is not completely homogenising these highly variable spacer regions within individuals. This has implications for the use of multi-copy genes to ensure that intra-individual variation is accounted for prior to using the region for higher level comparisons. It makes this region unsuitable for population genetic studies as it would require excessive amounts of cloning and sequencing to characterise intra-individual variation before inter-individual variation could be examined.

6 General Discussion

Parasites are a group defined by ecological and not phylogenetic criteria. Essentially a parasite is an organism that is intimately associated with, and metabolically dependent upon another living organism (Lincoln *et al.*, 1982). In the past, parasites have often been considered as more constrained and even degenerate compared to free-living organisms due to their compulsory association with their hosts. However, the range of groups which has evolved parasitic lifestyles reveals that they are very diverse in characteristics such as breeding system. The common challenges of a parasitic lifestyle, such as avoidance of host immune responses and the adverse effects of the environment on infectious stages, combined with the diversity of groups that are parasitic, mean they provide opportunities for detecting general evolutionary effects. These evolutionary effects may be apparent at different scales of measurement, from within populations to among species and higher taxa.

In this thesis, genetic variation has been examined at different scales of measurement in two parasite species, *Echinococcus granulosus* and *Ostertagia ostertagi*. Different molecular techniques (allozyme electrophoresis, RAPDs, sequencing of partial mtDNA, sequencing of rDNA spacers) and different analytical methods (allelic and infinite site models, inbreeding and coalescent methods) have been examined to calculate a range of population genetic parameters.

In Chapter 3 results from two molecular markers, allozymes and RAPDs, were compared to assess the genetic diversity within and among populations of *Echinococcus granulosus* in Australia. Allozyme analysis of *E. granulosus* in Australia had previously been found to be only just variable enough to be useful and so a more variable technique (RAPD) was used to confirm and expand the description of the genetic structure.

The dominant nature of RAPD markers has led to many researchers dismissing it as a technique, especially in situations where the breeding system is unknown. It is shown that it is possible to analyse RAPD data assuming alternative breeding systems (fromstrict self-fertilisation to Hardy Weinberg equilibrium) to assess the effect of

different assumptions on the estimated genetic structure. The technique becomes especially powerful when the breeding system can be determined independently and estimated selfing rates used to determine genotype frequencies from dominant patterns.

In addition to this allelic approach to the analysis of RAPD data, two other approaches to the analysis of RAPD data were examined; band sharing distance calculations and average nucleotide diversity estimations. All three of these approaches gave similar estimates for genetic distances between pairs of populations (r^2 = 0.971 to 0.981). RAPD markers were found to be more numerous and more variable than allozyme loci, which increased the confidence in the estimates of differentiation between populations of *E. granulosus* from Tasmania, King Island and the mainland.

Samples of *E. granulosus* from other areas of the world, particularly New Zealand and other Australian islands, are needed to determine if the King Island samples are likely to represent a relic population or a reintroduction. Reintroduction is unlikely to have occurred recently from Tasmania, since the King Island samples were most different from the Tasmania population.

Chapter 4 extended a study of mtDNA variation in *O. ostertagi* in the USA to samples from Australia and allowed the investigation of different methods of analysis including coalescent analysis from the within-farm scale to between-continents scale. The genetic structure of *O. ostertagi* in Australia was similar to that observed in the USA; high within population diversity and low Fst values between populations. Gene flow was higher between the Australian populations which were all from a single state, whereas the USA populations were from different states.

A structured assessment is made of three alternatives to explain the paradox of lack of genetic differentiation despite a fixed difference in a trait of biological significance (hypobiosis or delayed development within the host). The most likely alternative is that the effective population sizes are very large, reducing drift and allowing the migration rate to be low, thus reducing the genetic load of strong selection causing the observed differentiation of hypobiosis timing from north and south USA populations. The low level of drift is revealed firstly in the lack of geographic structuring in the ND4 gene-tree and also in the low Fst values between Australian and USA populations. The

relatively recent radiation during colonisation of the host and parasite populations means there has been insufficient time to reach an equilibrium between drift and migration. These populations thus violate the assumptions of classical population genetic models. Furthermore, since population structure is ideally estimated from nearly neutral markers, it is not surprising to find incongruence with markers likely to be under significant selection.

Further extensions of coalescent approaches (Kingman, 2000) are allowing precise testing of classical assumptions, for example population growth (Slatkin and Hudson, 1991; Innan and Stephan, 2000) and recombination (Hudson, 1983; Schierup and Hein, 2000). As coalescent analysis becomes more advanced it should be possible to separate the confounding effects causing similarity due to current gene flow versus recent ancestral descent. Studies on other parasites of domestic hosts and humans will reveal whether the high artificial host movement has minimised differentiation in other parasites or whether *O. ostertagi* has little structure due to its high effective population size and/or recent radiation.

A second marker, particularly a nuclear one, is required to confirm that the population genetic structure revealed from the partial ND4 locus for *O. ostertagi* is representative of the entire genome. Several techniques were tried to attempt to find a second marker including RAPDs, anchored simple sequence repeat (aSSR) banding patterns (Oliveira *et al.*, 1997), PCR of introns in cysteine protease and B-tubulin genes, sequencing of the intergenic spacer (IGS) and internal transcribed spacers (ITS1 & ITS2) of ribosomal DNA as described in Chapter 5. The first three techniques were not reproducible or did not amplify using the single worm DNA extracts and the last one showed intra-individual variation.

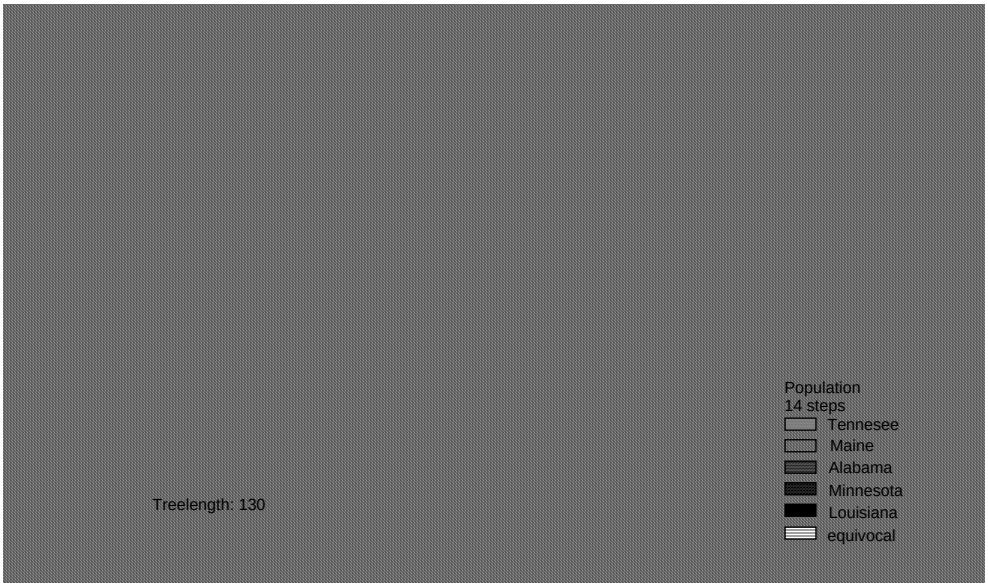
It is only possible to make a very general comparison of the population genetic structure of *E. granulosus* in Australia and *O. ostertagi* from Australia and the USA as sampling and molecular techniques were different. *O. ostertagi* was not sampled across Australia, as was the case with *E. granulosus*. However, there is no reason to suspect there would be any more structure within Australia than across the USA, thus comparing *O. ostertagi* across the USA with *E. granulosus* across Australia suggests higher gene flow estimates in *O. ostertagi*. There was somewhat more structure seen in *E. granulosus* as

evidenced by the less random nature of the phenogram for individuals, with more clustering by population (Figure 6.1a vs Figure 6.1b).

Both of these parasites presumably were introduced at similar times during European colonisation. The extra structure seen in *E. granulosus* may be due to numerous factors including this parasite's predominant self-fertilisation (the effective rate of which is greatly increased by asexual amplification in the cyst leading to genetically identical worms most likely to be proximal), possible lower effective population size, utilisation of wildlife cycle (less artificial host movement), selection (perhaps increased by the eradication campaign in Tasmania) or founder effects.

Figure 6.1a - Neighbour Joining clustering of Jaccards distances from RAPD data for 43 individual *E. granulosus* from 6 populations across Australia (KB=King Island bovine, ES = eastern sheep, EM = eastern macropod, TS = Tasmanian sheep, WS = western sheep, WM = western macropod).

Figure 6.1b - Neighbour Joining clustering of maximum likelihood distance from partial ND4 mtDNA sequence for 28 *Ostertagia ostertagi* individuals from 5 USA states (Al = Alabama, La = Louisiana, Me = Maine, Mn = Minnesota, Tn = Tennessee).



Variation within populations is the basis of evolution but is often overlooked by studies examining higher scales. Studies which include only one or a few isolates or which bulk many individuals to get a single representative sample will not detect possible variation within the group or even within a single individual. Chapter 5 describes the cloning and sequencing of different length spacer regions, both the intergenic and inter transcribed 1 spacer regions of rDNA, found within individual *Ostertagia ostertagi* worms. The length differences were found to be due to different numbers of repeats in both spacer regions.

This is an example of intra-individual variation that precludes the use of those regions for higher scale studies, due to the time and expense of fully characterising all the variation. Any multi-copy DNA needs to be used carefully, as variation between copies, due to incomplete molecular drive, may lead to comparisons of non-homologous regions. Variation within a population (either of individuals or of gene copies) means that as populations diverge the gene-tree at first shows polyphyly, then paraphyly and finally reciprocal monophyly (Avise, 1994). Only at this final stage does the divergence fit the bifurcation assumptions implicit in estimating phylogeny.

The use of the size of the ITS1 region as a semi-quantitative species diagnostic when there is variation within a species, including a variant that is similar to that found in other species, may lead to false results. Further studies on the precise mechanisms responsible for molecular drive and more information on how many multi-copy regions actually exhibit homogenisation, will reveal the relative importance of intra-individual variation in multi-copy regions.

A number of general conclusions are apparent from these studies:

I. The study of variation necessarily involves an implicit scale. Genetic variation can be studied across a huge range of scales and the molecular techniques must be chosen to ensure they are appropriate for the scale. The question being explored also affects the choice of technique. For diagnosis of a parasite infection a simple presence/absence test is all that is needed, whereas a technique to carry out an epidemiological study requires markers which are variable within a population, but do not change over time for an individual strain. All markers are limited in the scales at which they are useful. At the bottom limit the marker shows no variation, while at the top limit maximum variation is reached, which may be constrained by processes such as codon bias or secondary structure constraints (Blouin *et al.*, 1998; McDonnell *et al.*, 2000), all samples maybe equally different (or the marker may not be present) and thus no conclusions can be drawn about the relationships between samples. It is important to look at variation at the scale below the one of interest, to ensure sampling is adequate and that samples are representative of the entire group.

II. The use of several methods of analysis on the same genetic data allows comparisons to be made, and if the different methods produce similar results gives more confidence in the conclusions drawn. It is critical to understand the assumptions used by each method of analysis, as all analyses have implicit assumptions that if violated can invalidate the conclusions reached. Data should be explored using all available methods of analysis, not just a single method as the data may not be appropriate for that method, which would be revealled by discordant results from other analyses. Similarly, wherever possible, genetic data should be combined with ecological, morphological and historical data. Coalescent and other new methods of analysis currently being developed have the potential to avoid many of the assumptions of traditional equilibrium analysis which clearly do not apply well to parasites of domestic hosts and humans, which have undergone high rates of artificial movement.

Appendix 1 - Allozyme data set for *Echinococcus granulosus* in Australia; 6 populations from region/host: King Island cattle, WA sheep, WA macropods, Eastern Australia sheep, EA macropods, and Tasmanian sheep. Diploid genotypes with common allele designated A, other alleles B and C, ?? = not scored. 23 loci in order: Est G6pd-2 Idh-1 Np-2 Pep 6pgd-1 6pgd-2 Ada Ak G6pd-1 Gpi Gdh Got Hk-1 Hk-2 Idh-2 Mdh Me Mpi Np-1 Pgm Sod-1 Sod-2.

```
King Island bovine (5)
B291188#1 AA AA ?? AC AA AA AA AA AA ?? AA ?? AA ?? ?? AA AA AA AA ?? AA AA AA
B291188#2 ?? BB AA CC ?? ?? AA AA AA ?? ?? ?? AA ?? ?? AA AA ?? AA ?? ?? ?? ??
B160589   AA ?? ?? ?? AA AA AA AA AA ?? ?? ?? AA ?? ?? AA AA ?? AA ?? AA ?? ??
B160590   AA AA ?? CC AA ?? ?? AA AA ?? AA AA AA ?? AA AA AA AA AA ?? AA AA AA
B231189   AC AA ?? AA AA BB AA AA AA ?? AA AA AA ?? AA AA AA AA AA ?? AA AA AA
WA sheep (26)
S1602881c ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S18028812 ?? ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ?? AA AA ?? ?? ?? AA ?? ??
S180387   ?? ?? ?? ?? AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA AA AA ?? ?? ?? ?? ??
S1602881b AA AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S1602881d AA AA ?? AA ?? AA ?? ?? ?? AA AA ?? ?? ?? ?? AA AA AA AA ?? AA AA AA
S1602881e AA AA ?? AA ?? AA ?? ?? ?? AA AA ?? ?? ?? ?? AA AA AA AA ?? AA AA AA
S1602881f AA AA ?? AA ?? AA ?? ?? ?? AA AA ?? ?? ?? ?? AA AA AA AA ?? AA AA AA
S1602881g AA AA ?? AA ?? AA ?? ?? ?? AA AA ?? ?? ?? ?? AA AA AA AA ?? AA AA AA
S1602881P AA AA ?? AA AA AA ?? AA ?? AA ?? ?? AA AA AA AA ?? ?? AA ?? AA AA ??
S1602882a AA AA ?? AA AA AA ?? AA AA ?? AA ?? AA ?? ?? AA AA ?? AA ?? AA ?? ??
S1602882c AA ?? ?? AA AA ?? ?? AA AA ?? AA ?? AA ?? ?? AA AA AA AA ?? AA AA AA
S1602882d AA ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? AA ?? ?? AA AA ?? ?? ?? AA ?? ??
S3101902? AA ?? ?? ?? AA ?? ?? AA AA ?? ?? ?? AA ?? AA AA AA AA AA ?? ?? ?? ??
S3101903A AA ?? ?? AA AA ?? ?? ?? ?? ?? AA AA AA ?? ?? AA AA AA AA ?? AA AA AA
S3101904B AA AA ?? BB AA AA ?? ?? ?? ?? AA AA AA ?? ?? AA AA AA AA ?? AA AA AA
S3101905B AA ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? AA AA AA AA ?? AA AA AA
S080889   AA AA ?? ?? AA ?? ?? ?? ?? ?? ?? AA AA ?? ?? AA AA ?? ?? ?? AA AA AA
S250987   AA AA AA BB AA AA ?? AA AA AA AA ?? AA ?? AA AA AA AA AA ?? AA AA AA
S1602882b BB ?? ?? AA AA ?? ?? AA AA ?? AA ?? AA ?? ?? AA AA ?? AA ?? AA ?? ??
S18028811 BB ?? ?? AA AA ?? ?? AA AA ?? AA ?? AA ?? ?? AA AA ?? AA ?? AA ?? ??
S3101901B BB AA ?? ?? AA AA ?? ?? ?? ?? AA ?? AA ?? AA ?? AA AA ?? ?? AA AA AA
S3101901A BB AA ?? ?? AA AA AA ?? ?? ?? AA ?? AA ?? AA AA AA AA ?? ?? AA AA AA
S3101905A BB AA ?? AA AA AA ?? ?? ?? ?? AA AA AA ?? ?? AA AA AA AA ?? AA AA AA
S3101905C BB ?? ?? BB ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? AA AA AA AA ?? AA AA AA
S070286   BB AA AA ?? AA AA AA ?? AA AA AA AA AA ?? ?? AA AA AA AA ?? AA AA AA
S3101904A CC AA ?? AA AA AA AA ?? ?? ?? AA AA AA ?? AA AA AA AA AA ?? AA AA AA
WA macropods (28)
M250288d  ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? ?? ??
M1909881d ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ??
M030388   ?? AA ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? AA AA ??
M1202886a AA ?? ?? AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ??
M1202886b AA ?? ?? BB ?? AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA AA ?? ?? ??
M1202886c AA ?? ?? AA AA AA ?? AA AA ?? AA ?? AA ?? ?? AA AA ?? AA ?? AA ?? ??
M1202886d AA ?? ?? AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA AA ?? ?? ??
M1202886P AA AA ?? BB AA AA ?? AA AA AA ?? ?? AA AA AA AA ?? AA AA AA AA AA AA
M1202886P AA AA ?? AA AA AA ?? AA AA AA ?? ?? AA AA AA AA ?? AA AA AA AA AA AA
M250288a  AA AA ?? AA AA AA ?? AA AA ?? ?? ?? AA ?? AA AA AA AA AA ?? AA AA AA
M250288b  AA AA ?? AA AA AA AA AA ?? ?? AA ?? AA ?? ?? AA AA ?? AA ?? AA ?? ??
M1909881a AA AA ?? BB AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
M1909881b AA AA ?? BB AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA AA ?? ?? ??
M1909881c AA ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? AA ?? AA ?? AA ?? ??
M1909881P AA AA ?? BB AA AA ?? AA AA ?? AA ?? AA ?? AA AA AA AA AA AA AA AA AA
M1909882a AA AA ?? BB AA AA ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
M1909882b AA AA ?? AA AA AA ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
M1909882c AA ?? ?? ?? AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
M1909882d AA ?? ?? ?? AA ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? AA ?? AA ?? AA AA AA
M1909882P AA AA ?? BB AA AA ?? AA AA ?? AA ?? AA ?? AA AA AA ?? AA ?? AA AA AA
M1909882P AA AA ?? AA AA AA ?? AA AA ?? AA ?? AA ?? AA AA AA ?? AA ?? AA AA AA
M080786   AA AA ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? AA AA ?? AA ?? AA AA AA
M080986   AA AA AA AA AA AA ?? AA AA ?? AA ?? AA ?? AA AA AA AA ?? ?? AA AA AA
M130488   AA ?? ?? AA ?? AA ?? AA AA ?? ?? ?? AA ?? ?? AA AA ?? AA ?? AA AA AA
```

Appendix 1 continued - Allozyme data set for *Echinococcus granulosus* in Australia

```
M290886   AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA ?? AA AA AA
M020986   BB AA AA ?? BB AA AA ?? AA AA AA AA AA AA AA AA AA AA AA ?? AA AA AA
M250288c  CC ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
M250288e  CC ?? ?? ?? AA AA ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
Eastern sheep (50)
S060488e  ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? ?? ??
S1908881  ?? ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? AA ?? ?? AA AA ?? ?? ?? AA ?? ??
S1908882  ?? ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? AA ?? ?? AA AA ?? AA ?? AA ?? ??
S150188a  ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? AA ?? AA ?? ?? ?? ??
S190986a  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986b  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986c  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986d  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986e  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986f  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986g  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986h  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986i  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986j  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986k  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986l  AA AA ?? AA AA AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ??
S190986P  AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA ?? AA AA AA
S0503871P AA AA AA AA AA AA ?? AA AA ?? AA AA AA ?? AA AA AA AA AA ?? AA AA AA
S0503871d AA AA AA AA AA AA ?? AA AA ?? AA AA AA ?? AA AA AA AA AA ?? AA AA AA
S140587P  AA AA BB AA AA AA ?? AA AA AA AA ?? AA ?? AA AA AA AA AA ?? AA AA AA
S1405871  AA AA BB ?? AA AA ?? AA ?? AA ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? ?? ??
S1405873  AA ?? ?? AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ??
S150188b  AA AA ?? AA AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S150188d  AA AA ?? AA AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S150188e  AA AA ?? AA AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S100588b  AA ?? ?? AA AA ?? ?? AA ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S060488a  AA AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S060488b  AA AA AA AA AA ?? ?? ?? ?? AA AA ?? AA ?? AA AA AA AA AA AA ?? AA AA
S060488c  AA ?? AA AA ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? AA ?? ?? AA AA ?? ?? ??
S060488d  AA ?? AA ?? AA ?? ?? AA ?? ?? AA ?? AA ?? ?? AA AA ?? AA AA AA ?? ??
S1908884  AA ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S1908885  AA ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? ?? ?? ?? AA AA ?? AA ?? AA AA AA
S2011863  AA ?? ?? ?? AA ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? AA AA ?? ?? AA AA AA
S2011861  AA ?? ?? ?? AA ?? ?? ?? ?? ?? AA AA ?? ?? ?? ?? AA ?? AA ?? AA AA AA
S080187P  AA AA AA AA AA AA AA AA AA AA AA AA AA ?? AA AA AA AA AA ?? AA AA AA
S080187P  AA AA AA AA BB AA AA AA AA AA AA AA AA ?? AA AA AA AA AA ?? AA AA AA
S250587   AA AA AA AA AA AA ?? AA AA AA AA AA AA AA AA AA AA AA AA AA ?? AA AA AA
S250986   AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA ?? AA AA AA
S2011862a AA ?? ?? ?? AA AA ?? AA AA AA AA ?? AA ?? AA AA AA AA AA ?? AA ?? AA
S100588c  AC ?? ?? AA AA ?? ?? AA ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S0503871a BB AA AA AA AA AA ?? AA AA ?? AA AA AA ?? AA AA AA AA AA ?? AA AA AA
S0503871b BB AA AA AA AA AA ?? AA AA ?? AA AA AA ?? AA AA AA AA AA ?? AA AA AA
S0503871c BB AA AA AA AA AA ?? AA AA ?? AA AA AA ?? AA AA AA AA AA ?? AA AA AA
S0503871e BB AA AA AA AA AA ?? AA AA ?? AA AA AA ?? AA AA AA AA AA ?? AA AA AA
S1908883  BB ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? AA ?? ?? AA AA ?? AA ?? AA ?? ??
S060587   BB AA AA AA AA AA AA AA AA AA AA AA AA ?? AA AA AA AA ?? ?? AA AA AA
S060587   BB AA AA AA BB AA AA AA AA AA AA AA AA ?? AA AA AA AA ?? ?? AA AA AA
S150188c  CC AA ?? AA AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
S150188f  CC AA ?? AA AA ?? AA ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA AA ?? ?? ??
S100588a  CC ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? AA ??
Eastern macropods (14)
M1601892  ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? ?? ?? ?? ?? ??
M2309871  AA BB AA AA AA AA ?? AA AA AA AA AA AA AA AA AA AA AA ?? ?? AA AA ??
M2309872a AA AA ?? AA AA AA ?? ?? ?? ?? ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
M2309872b AA AA ?? AA AA AA BB ?? ?? AA ?? ?? ?? ?? ?? AA ?? ?? AA ?? ?? ?? ??
M2309872P AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA ?? AA AA ??
M231087a  AA ?? ?? ?? AA ?? ?? AA ?? ?? ?? ?? AA ?? ?? AA ?? ?? AA ?? AA ?? ??
M231087b  AA ?? ?? AA AA ?? ?? AA AA ?? AA ?? AA ?? ?? AA AA ?? AA ?? AA ?? ??
M231087P  AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA ?? AA AA ??
```

Appendix 1 continued - Allozyme data set for *Echinococcus granulosus* in Australia

M1601893	AA	??	??	??	??	??	??	??	??	??	??	??	??	??	??	AA	??	??	AA	??	??	??	??
M040789	AA	AA	??	??	AA	AA	BB	??	??	??	AA	??	AA	??	AA	AA	AA	AA	??	??	AA	AA	AA
M091286	AA	AA	??	BB	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	??	AA	AA	AA
M231087d	BB	??	??	AA	AA	??	??	AA	AA	??	AA	??	AA	??	??	AA	AA	??	AA	??	AA	??	??
M1601891	BB	??	??	AA	??	AA	??	??	??	??	??	??	??	AA	??	??	AA	??	??	AA	??	??	??	??
M050589	BB	AA	??	AA	AA	AA	AA	??	??	??	??	??	??	AA	??	??	AA	??	??	AA	??	??	??	??
Tasmanian sheep (39)																								
S090490A1	??	??	??	AA	??	??	??	AA	AA	??	??	??	??	??	??	AA	AA	??	AA	??	AA	??	??
S0602903D	AA	??	??	AA	AA	??	??	AA	??	??	AA	AA	AA	??	??	AA	AA	AA	AA	??	AA	AA	AA
S090490A6	AA	??	??	AA	??	??	??	AA	AA	??	??	??	AA	??	??	AA	AA	??	AA	??	AA	??	??
S090490B3	AA	??	??	??	??	??	??	AA	??	??	AA	??	??	??	??	AA	AA	AA	AA	??	AA	AA	AA
S090490B4	AA	??	CC	??	AA	??	??	AA	AA	??	AA	??	AA	??	??	AA	AA	??	AA	??	??	??	??
S090490B5	AA	AA	AA	AA	AA	??	??	AA	AA	??	AA	??	AA	??	AA	AA	AA	AA	AA	??	??	AA	AA
S09049C10	AA	AA	??	AA	AA	??	??	AA	AA	??	AA	??	??	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S090490C2	AA	AA	CC	AA	AA	??	??	AA	AA	??	AA	??	AA	??	AA	AA	AA	??	AA	??	??	??	??
S090490C4	AA	AA	??	AA	AA	??	??	AA	AA	??	AA	??	??	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S090490C6	AA	AA	??	AA	AA	??	??	AA	AA	??	AA	??	??	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S090490C8	AA	AA	??	AA	AA	??	??	AA	AA	??	AA	??	??	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S090490C9	AA	AA	??	AA	AA	??	??	AA	AA	??	AA	??	??	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S3003881	AA	??	??	AA	??	??	??	AA	AA	??	??	??	AA	??	??	AA	AA	AA	AA	??	AA	??	??
S3003882	AA	AA	??	??	AA	??	??	??	??	??	AA	AA	AA	??	??	AA	AA	AA	AA	??	AA	AA	AA
S06029090	AA	??	??	??	AA	??	??	??	??	??	??	AA	AA	??	??	AA	AA	??	??	??	AA	AA	AA
S0090490A2	AB	??	??	AA	??	??	??	AA	AA	??	AA	??	AA	??	??	AA	AA	AA	AA	??	AA	AA	AA
S0090490A3	AB	AA	??	AA	AA	??	??	AA	AA	??	AA	??	AA	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S0090490A4	AB	AA	??	AA	??	??	??	AA	AA	??	AA	??	AA	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S0090490A5	AC	AA	??	AA	AA	??	??	AA	AA	??	AA	??	AA	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S0090490B2	AC	??	??	??	AA	??	??	AA	??	??	AA	??	??	??	??	AA	AA	AA	AA	??	AA	AA	AA
S0602903A	BB	AA	??	??	AA	??	??	AA	AA	??	??	??	AA	??	??	AA	AA	AA	AA	??	??	??	??
S0602903B	BB	??	??	??	AA	??	??	AA	AA	??	??	??	AA	??	??	AA	AA	AA	AA	??	??	??	??
S0602903C	BB	AA	??	AA	AA	AA	AA	??	??	??	AA	AA	AA	??	??	AA	AA	AA	AA	??	AA	AA	AA
S0090490B1	BB	??	??	??	??	??	??	??	AA	??	??	??	AA	??	??	AA	AA	??	AA	??	AA	??	??
S0090490B6	BB	??	??	AA	??	??	??	AA	AA	??	??	??	AA	??	??	AA	AA	AA	AA	??	AA	??	??
S0090490C1	BB	??	??	AA	??	??	??	AA	AA	??	??	??	AA	??	??	AA	AA	AA	AA	??	AA	??	??
S09049C11	BB	??	??	AA	AA	??	??	AA	AA	??	AA	??	AA	??	??	AA	AA	AA	AA	??	AA	AA	AA
S0090490C3	BB	AA	??	AA	AA	??	??	AA	AA	??	AA	??	??	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S0090490C5	BB	AA	??	AA	AA	??	??	AA	AA	??	AA	??	??	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S0090490C7	BB	AA	??	AA	AA	??	??	AA	AA	??	AA	??	??	??	AA	AA	AA	AA	AA	??	AA	AA	AA
S031189b	BB	AA	??	AA	AA	??	??	AA	AA	??	??	??	AA	??	??	AA	AA	??	AA	??	??	??	??
S031189c	BB	AA	??	??	AA	AA	??	??	??	??	AA	??	AA	??	AA	AA	AA	AA	??	??	AA	AA	AA
S031189d	BB	??	AA	??	AA	??	??	??	??	??	AA	??	??	??	??	AA	AA	AA	AA	??	AA	AA	AA
S0602903A	BB	??	??	AA	AA	??	??	AA	AA	??	??	??	AA	??	AA	??	AA	??	AA	??	??	??	??
S0602903B	BB	AA	??	AA	AA	??	??	AA	AA	??	AA	??	AA	??	AA	AA	AA	AA	AA	??	??	??	??
S0602903C	BB	AA	??	AA	AA	AA	AA	??	??	??	AA	AA	??	??	??	AA	AA	AA	AA	??	AA	AA	AA
S0602903D	BB	??	??	??	AA	??	??	AA	??	??	AA	AA	AA	??	??	AA	AA	AA	AA	??	AA	AA	AA
S240986	BB	AA	AA	??	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	??	AA	AA	AA
S280486	BB	AA	BB	??	AA	AA	??	??	??	??	AA	??	??	??	AA	AA	AA	AA	AA	??	AA	AA	AA

Appendix 2 - RAPD data for *Echinococcus granulosus* in Australia. 43 Individuals (one per line) scored as 1 for band presence or zero for absence for 77 bands. Six populations with sample size in brackets: King Island bovine(4), Tasmanian sheep(5), Eastern sheep(13), Eastern macropods(8), WA sheep(4), WA macropods(9) Primer names (number of bands): 10:01(18), 10:19(16), 12:03(11), 12:06(15), 12:07(17).

```
King Island Bovine
10101111111111011110001111111011001111010101111011100010010010111011011110100
11111110101001000110001111111011001110010101100011100110111110101110010001001
11111110101001000110001111111011001110011111001011001111111111101110010101001
11111110101001000110001111111011001110011111111111101111100111011100010001001
Tasmanian Sheep
00010010110000100000000000100001100001001010110001100000000000000001001001000100
11001110100001000100000011100001100001011010110001100000000001111010011000000 1
11001110100001000100010110000001100011000110100001100000010000010011000000000
11001110100001000100010111000011000010111101101111000010000011101110110000001
11001110100001000100010111001011001110111101101111100010110011101110110001001
Eastern Sheep
00000000000000000000000011100001100011011100100001101000000000010010010000000 0
01001010010000100000000111100000101111011111110101101001000000011110111010000
01011100000000000100001111111011001110111101111111101010111011101110010001001
01011110100000000100001111111011001110111101101011101010110011101110110001001
01111110100001000100001111111011000110111011100001100011111110101110000001011
11011110100001000100001111111011001110111111100011100010110011101110110001001
11101110100001000110011111111011001110111011101111100010111111101110111001001
11111110100001000110001111111011011110110111111111110001111111101110000001011
11111110100001000100001111111011001110110111101111100111111111101110000001001
11111110100001000100001111111011001110111111101111100110111011101110110001001
11111110100001000100001111111011001000110111100101110110111111101110000001011
11111110100001001110001111111011001110111111111111110001011011110111001000 1001
11111110100001001100001111111011001110111011101111100111101111101110010001011
Eastern Macropod
01001100000000100010000011110001100011011010110101110001000001110100010100010 1
11001100000000000100001111111011001010110101101111100010111011101110100101101
11001110100001000100000110100011000010101101001011000000000001111010010000001 1
11001110100001000101000111110100101110111111101011101010000011110101100001001
11011110100000000110001111111011001110110011001011110110111011101110110101001
11111110100000000100001111111011001110110111100111100010111111101110010101011
11111110100001000100001111111011001110110101101111100010000011101110010101001
11111110100001000100001111111011001110110000101011100010111111101110100001001
Western Sheep
11001110100001000100001111111011000010110101101011100010110011101110010001001
11001110100001000100001111111011000010110101101011100010100011101110000000000
11011110100001000110011111111011001101111111101011000010000011101110111000001
11101110100001000110001111111011001110110011101111100110111111101110100001001
Western Macropod
01111110100001000100001111111011001110111111101111100111111111101110100001001
01111110100001000100001111111011001110111111101111100110111111101110100001001
01111110100001000100001111111011001110111111100001100010111111101110010001001
01111110100001000100001111111011001110111111101111100110111111101110100001001
11001100100001000100010111111011001110111111101011000010000011110101100001001
11011110000001000100000111111011001110111111101111000010000011101110110101001
11011110100001000110010111111011001110111111101011000010000011101110010101001
11111110100001000100001111111011001110111011011000101111111011101000 01001
111111101000010001000011111110110011101110111011111001111111110111010
0001001
```

Appendix 3 - 318 bp of ND4 locus for 88 Australian and 28 USA *O. ostertagi*.

```
HAR1  GGAAAATTAAGTGGCCTAATGATAATGTTGGCTCACGGTTATACGTCAACGTTAATATTCTATATAATTGGGGAATTTTA
HAR2  ............................A.............................................................
HAR3  ..........................................................................G.....
HAR4  .......................C......................................................................
HAR5  ..G.............A.....................................................................
HAR6  ..................A.............T.....................T.....................................
HAR7  ..................................T..........G.....................................
HAR8  ............T.....A.....................G.....................................
HAR9  A.......................C.....................................................
HAR10 ..........................................................................
HAR11 ..................A.....................................................................
HAR12 ..........................................................................
MAN1  .........................C.....................................................
MAN2  ..........................................................................
MAN3  ..........................................................................
MAN4  ..........C........A.....................................T.....................
MAN5  ..................A.....................................................................
MAN6  ..........C.....................................................................
MAN7  ..............................C.....................................................
MAN8  ..G.....................................................................
MAN9  ..G..........T.....................................................................
MAN10 ..................A.....................................................................
MAN11 ..................A.....................................T.....................
MAN12 ..................A.....................................G.....T.....................
KAR1  ......C.....................................................................
KAR2  ......C.......T.....................................................................
KAR3  ..........................................................................
KAR4  ..........................................................................
KAR5  ..................................C.....................................................
KAR6  ..............................C.....................................................
KAR7  ..........C.....................................................................
KAR8  ..........C........A.....................................T.....................
KAR9  ..T.............A..........C.....................................A........
KAR10 ..........................................................G.....................
KAR11 ....................................C.....................................................
KAR12 ............TT...A..........C.....C.....................................................
DEN1  ..........................................................................
DEN2  ............T.....................................................................
DEN3  ......C.......T.....................................................................
DEN4  ..............T.................T.....................................................
DEN5  ............T.....................................................................
DEN6  ..................A.....................................................................
DEN7  ....................................C..........A.....................................
DEN8  ..................A.............T.....................................................
DEN9  ..................A.....................................................................
DEN10 ................G..A.....................................T.....................
DEN11 ..........................................................................
DEN12 ..................A.....................................................................
ESP1  ..........................................................................
ESP2  ..................A.....................C.....................................................
ESP3  ..........................................................................
ESP4 ...........................................................................
ESP5  ..................A.....................................................................
ESP6  ................G.............C........C.....................................
ESP7  ..................A.....................................................C.....G.....
ESP8  ..................A.....................................................................
ESP9  ................................T.....................................................
ESP10 ..........................................................................G.....
ESP11 ..........................................................................
ESP12 ..G.....................................................................
 Tn-5 ............T.....A.....................................................................
 Tn-6 ............T.....................................G.....................................
 Tn-9 ............T.....A.............C.....................................................
Tn-10 ..G.............................T.....................................................
Tn-11 ..................A.............T..C.....................................................
Tn-13 ..........................................................................
 Me-5 ....................................................A.......C.....................
 Me-6 ..................A.....................................................................
 Me-8 ..........................................................................
Me-12 ....................................................A.......C.....................
Me-13 .A................A.....................................................................
```

Appendix 3 continued - 318 bp ND4 locus for 88 Australian and 28 USA *O.ostertagi*

```
 Al-6 ...................A.....................................................
Al-10 ...............T.....................................T...................
Al-13 ...............T.....................C.....................................
Mn-16 ..............................C..T.....................................
Mn-18 ...........................................................T...................
Mn-19 ...................................C.....................................
Mn-22 ...............TT....................................................
Mn-23 A..............................T..C.....................................
Mn-24 ..G.................A.............T.....................................
Mn-25 A..............................T..C.....................................
La-11 ...................................C.....................................
La-12 ......C.....................................................................
La-15 ......C.....................................................................
La-16 ......C.....................................................................
La-17 ....................A.....................................................
La-18 ..T...........TT....A.................................G.............A........
La-19 ..............TT................C.....................................

HAR1  TCACGCCAGATCAAGTCGAATAATTTATTTTTTAAATAGTTTTTTAAGCTCTAGAATATTATTTGGTATTTTATTTTCTT
HAR2  ...T.....................................G.....................................
HAR3  ...T.....................................G..T..................................
HAR4  ...T.....................................G.....................................
HAR5  ...T..T..................................C..G...........G.....................
HAR6  ...T.....................................G........G..G.....................
HAR7  ...T.....................................G.....................................
HAR8  ...T..T..................................G.....................................
HAR9  ...T.....................................G.....................................
HAR10 ...T.....................................G.....................................
HAR11 ...T.....................................G...........G.....................
HAR12 ...T.....................C...............G...........G.....................
MAN1  ...T.....................................G.....................................
MAN2  ...T.....................................G.....................................
MAN3  ...T.........................C...........G...........G.....................
MAN4  ...T..T..................................G...........G.....................
MAN5  ...T.....................................G...........G.....................
MAN6  ...T.....................................G.A.....................................
MAN7  ..........................................................G.....C..............
MAN8  C..T.....................................G.....................................
MAN9  ...T.....................................G.....................................
MAN10 ...T.................G...................G.....................................
MAN11 .........................................G.....................................
MAN12 ...T.....................................G.....................................
KAR1  ...T.....................................G.....................................
KAR2  C..TA....................................G...........G.....................
KAR3  ...T.........................C...........G...........G.....................
KAR4  ...T.....................................G...........G.....................
KAR5  ...T.....................................T.....................................
KAR6  ...T.....................................G.....................................
KAR7  ...T.....................................G...........G..G..................
KAR8  ...TA....................................G.....................................
KAR9  ...T.....................................G...........G.....................
KAR10 ...T.....................................G...........G.....................C
KAR11 ...T........G.............................G..T..................................
KAR12 ...T.....................................G.....................................
DEN1  C..T..T..................................G...........G.....................
DEN2  ...T..T..................................G...........G.....................
DEN3  C..T.....................................G...........G.....................
DEN4  ...T.....................................G...........G.....................
DEN5  C..T.T............G.....................................................
DEN6  ...T.....................................G...........G...........C..........
DEN7  ...T.....................................G.....................................
DEN8  ...T.....................................................................
DEN9  ...T.....................................G.....................................
DEN10 ...T.....................................................................
DEN11 ...T.....................................G...........G.....................
DEN12 ...T.....................................G...........G.....................
ESP1  ...T.....................C...............G.....................................
ESP2  ...T.....................................G...........G.....................
ESP3  ...T.....................................A.....................................
ESP4  ...T.............................G..............T..C..............G...........
ESP5  ...T.....................................G..T........G.....................
```

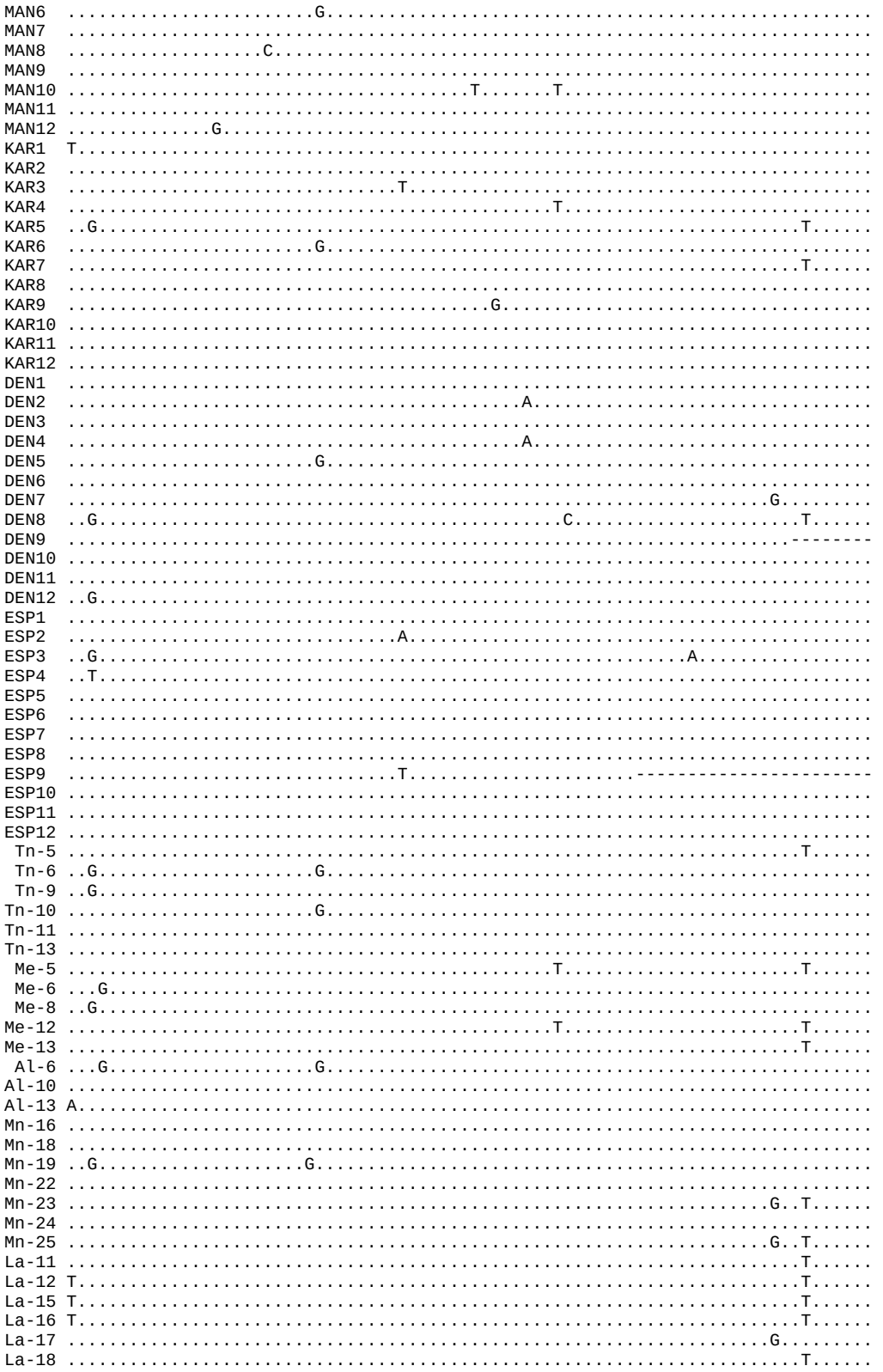
Appendix 3 continued - 318 bp ND4 locus for 88 Australian and 28 USA *O.ostertagi*

```
ESP6  ...T.....................................G.............................
ESP7  ......T..................................G.............................
ESP8  ...T.....................................G..........G..G...............
ESP9  ...T.T...................................G........G..G.................
ESP10 C..T..T..................................G.............................
ESP11 ...T.....................................G..........G..................
ESP12 ...T.....................................G..........G..................
 Tn-5 ...T.T...................................G..........G..................
 Tn-6 ...T.....................................G.............................
 Tn-9 .........................................................................
Tn-10 ...T.....................................G.............................
Tn-11 ...T.....................................G.............................
Tn-13 ...T.....................................G.............................
 Me-5 ...T.............G.....................G...........C..................
 Me-6 ...T.............G................G....G.............................
 Me-8 ...T.............G.....................G.............................
Me-12 ...T.............G.....................G...........C..................
Me-13 ...T.............G.....................G.............................
 Al-6 ...T.............G.....................G.............................
Al-10 C..T.............G.....................G.............................
Al-13 ...T.............G.....................G..T..........................
Mn-16 ...T.............G.....................G..T.....T..G..................
Mn-18 ...T.............G.....................G..........G..................
Mn-19 ...T.............G.....................G.............................
Mn-22 ...T.............G.....................G.............................
Mn-23 ...T.............G.....................G...........A..................
Mn-24 C..T.............G.....................G.............................
Mn-25 ...T.............G.....................G...........A..................
La-11 ...T.............G.....................G..T..........................
La-12 ...T.............G.....................G.............................
La-15 ...T.............G.....................G.............................
La-16 ...T.............G.....................G.............................
La-17 ...T..........A..G.....................G..........G..................
La-18 ...T.TT..........G.....................G.............................
La-19 ...T.............G..G...................G.................C............

HAR1  TGGTATTTTTATCCAACAGCGGTGTACCCCCTTCATTGTCTTTTTTATCGGAATTTATTATTATTGTTAATAGATTTATG
HAR2  ...............T.....................................C.................
HAR3  .................T...................................C.................
HAR4  .........................................................................
HAR5  .................A...................................C.................
HAR6  .................T...................................C.................
HAR7  ............................C..G.....................C.................
HAR8  .................T..........C.........................C.................
HAR9  ......................C.....................................................
HAR10 .................T...................................C.................
HAR11 .................T...................................C.................
HAR12 .................T.......T.............................C.................
MAN1  .................T..........A..........................C.................
MAN2  .........................................................C.................
MAN3  .................T.......T.............................C.................
MAN4  .A.......................................................C.................
MAN5  .................T..C.....................................C.................
MAN6  ......................G..................................C.................
MAN7  ..............T..T..........C.........................C.................
MAN8  ............................C.........................C.................
MAN9  ...................................................A.....................
MAN10 ...........T.............C.....................................................
MAN11 ............T.....................................................................
MAN12 .................T.....................................................................
KAR1  .................T.............G.....................C.................
KAR2  .................T...................................C.................
KAR3  .................T.......T.............................C.................
KAR4  ........................T.....................................................
KAR5  ............................C.....................................................
KAR6  .........................................................C.................
KAR7  .................T.....................................................................
KAR8  ...........T.....................................................................
KAR9  .................T.....G.....C.....................C.................
KAR10 .................T.........................A.........C.................
KAR11 .................T.........................A.........C.................
```

Appendix 3 continued - 318 bp ND4 locus for 88 Australian and 28 USA *O.ostertagi*

```
KAR12 ..................T.........................................C.................
DEN1  ..................T.........................................C.................
DEN2  ..................T.....G..G.................................................
DEN3  ..................T....C.....................................C.................
DEN4  ..................T...........................A.............G.................
DEN5  .............T.....................................................-C.............
DEN6  C.................T.....................................................C.....
DEN7  .............T...........G.....................................................
DEN8  ..........................................................C.....................
DEN9  .............T..T..............A.....................................................
DEN10 ..................T.....................................................C.....
DEN11 ..................T.....................................................C.....
DEN12 ..................T...........................A...........C.....................
ESP1  C.................T.....................................................C.......
ESP2  ..................T.........................................C.....................
ESP3  ..........................T.....................................C.....................
ESP4  ..................T........T..C.................A.....................C.....
ESP5  ..........................................................C.....................
ESP6  ..................T.....................................C.....................
ESP7  ..................T...........A.....................C.....................
ESP8  ..................T...........C.....................C.....................
ESP9  ..................T.....................................C.....................
ESP10 ..................T........T.....G.....................C.....................
ESP11 ..................T..............G.....................C.....................
ESP12 ..................T........T.....C.....................C.....................
 Tn-5 ..................T.....................................C.....................
 Tn-6 ..........................................................C.....C.............
 Tn-9 ......................C.....T.....................A.....C.....................
Tn-10 ..............................C.....C.....................C.....................
Tn-11 ..................T.....C.....................C.....................
Tn-13 ..................T...........A.....C.....................
 Me-5 ..................T.....C.....................C.....................
 Me-6 ..................T.....C.....................A...........C.....................
 Me-8 ................T..T.....C.....................C.....................
Me-12 ..................T.....C.....................C.....................
Me-13 ..................T.....C.....................C.....................
 Al-6 ..........................C.....C.....................C.....................
Al-10 ..................T.....C.....................C.....................
Al-13 ..........................T.....C.....................C.....................
Mn-16 ..................T...........C.....................A...........C.....................
Mn-18 ................T..T...........C.....C.....................
Mn-19 ..........C.....C.....................C.....................
Mn-22 ..........C.....A.....................
Mn-23 ..................T...........C..G.....C.....................
Mn-24 ................T..T.....C.....................C.....................
Mn-25 ..................T...........C..G.....C.....................
La-11 ..........C.....A...........C.....................
La-12 ..................T.....G.....C.....................
La-15 ..................T.....G.....C.....................
La-16 ..................T.....G.....C.....................
La-17 ..................T.....G.....C.....................
La-18 ..........C.....C.....................
La-19 ..................T...........A.....C.....C.....

HAR1  CTAAGAAAAATATTATTTTTTATAATTTTTGTGTATTTTATAGTGGCCTTTTATTATTCTTTGTTTTTAATCACTTGT
HAR2  .....G.....----------
HAR3  ..........A.....T......
HAR4  ..........G.....
HAR5  ..T.....--------
HAR6  T..........A.....----------
HAR7  .....
HAR8  ..........----------
HAR9  .....G.....
HAR10 .....C.....A.....
HAR11 .....
HAR12 .....A.....
MAN1  .....
MAN2  T.....
MAN3  .....A.....
MAN4  .....G.....T......
MAN5  .....G.....T......
```

Appendix 3 continued - 318 bp ND4 locus for 88 Australian and 28 USA *O.ostertagi*



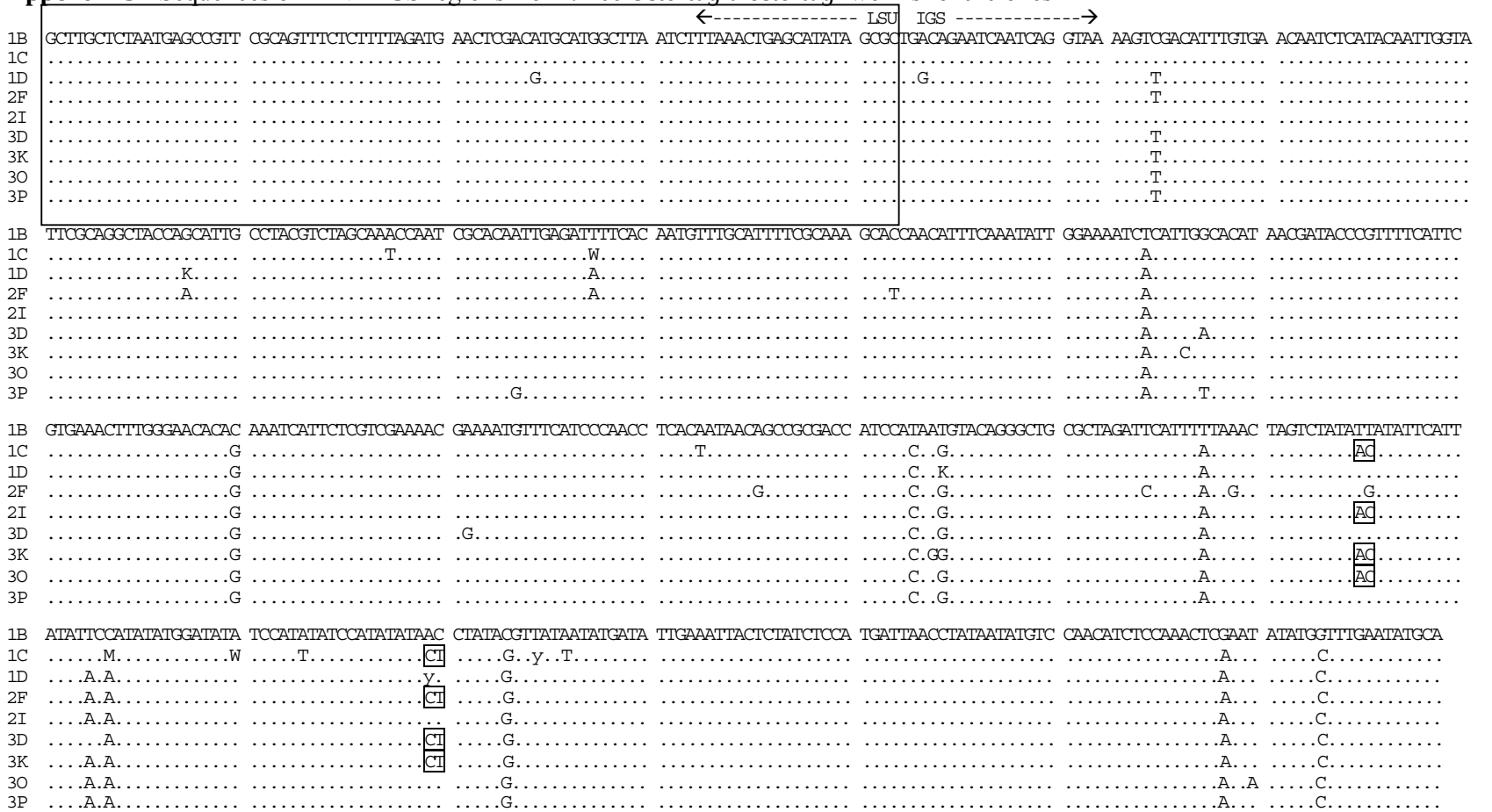
Appendix 4 - ND4 *Ostertagia ostertagi* 88 worms 318 nucleotides translated to 106 amino acid sequence

```
HAR1  GKLSGLMMML AHGYTSTLMF YMIGEFYHAS SSRMIYFLNS FLSSSMLFGI LFSLVFLSNS GVPPSLSFLS EFIIIVNSFM LSKMLFFMIF VYFMVAFYYS LFLITC
HAR2  .......... .......... .......... .......... .......... .......... .......... .......... ......???? ?????????? ??????
HAR3  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..L....... ......
HAR4  ........I. .......... .......... .......... .......... .......... .......... .......... ........V. .......... ......
HAR5  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ...???
HAR6  .......... .......... .......... .......... .......... .......... .......... .......... .........L ..???????? ??????
HAR7  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
HAR8  .......... .......... .......... .......... .......... .......... .......... .......... ......???? ?????????? ??????
HAR9  S......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
HAR10 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
HAR11 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
HAR12 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
MAN1  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
MAN2  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
MAN3  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
MAN4  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
MAN5  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
MAN6  .......... .......... .......... ..........N..... .......... .......... .......... ........V. .......... ......
MAN7  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
MAN8  .......... .......... .......... .......... .......... .......... .......... ...T...... ......S... .......... ......
MAN9  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
MAN10 .......... .......... .......... .......... .......... .......... .......... .......... .......... ...L...... ......
MAN11 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
MAN12 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
KAR1  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
KAR2  .......... .......... ........T. .......... .......... .......... .......... .......... .......... .......... ......
KAR3  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
KAR4  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
KAR5  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
KAR6  .......... .......... .......... .......... .......... .......... .......... .......... ........V. .......... ......
KAR7  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
KAR8  .......... .......... ........T. .......... .......... .......... .......... .......... .......... .......... ......
KAR9  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
KAR10 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
KAR11 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
KAR12 .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
DEN1  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
DEN2  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
DEN3  .......... .......... .......... .......... .......... .......... .A........ .......... .......... .......... ......
DEN4  .......... .......... .......... .......... .......... .......... .......... ....V..... .......... .......... ......
DEN5  .......... .......... ........V. .......... .......... .......... .......... ...?...... ........V. .......... ......
DEN6  .......... .......... .......... .......... .......... ...S...... .......... .......... .......... .......... ......
DEN7  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ......
DEN8  .......... .......... .......... .......... .......... .......... .......... .......... .......... ......L... ......
```

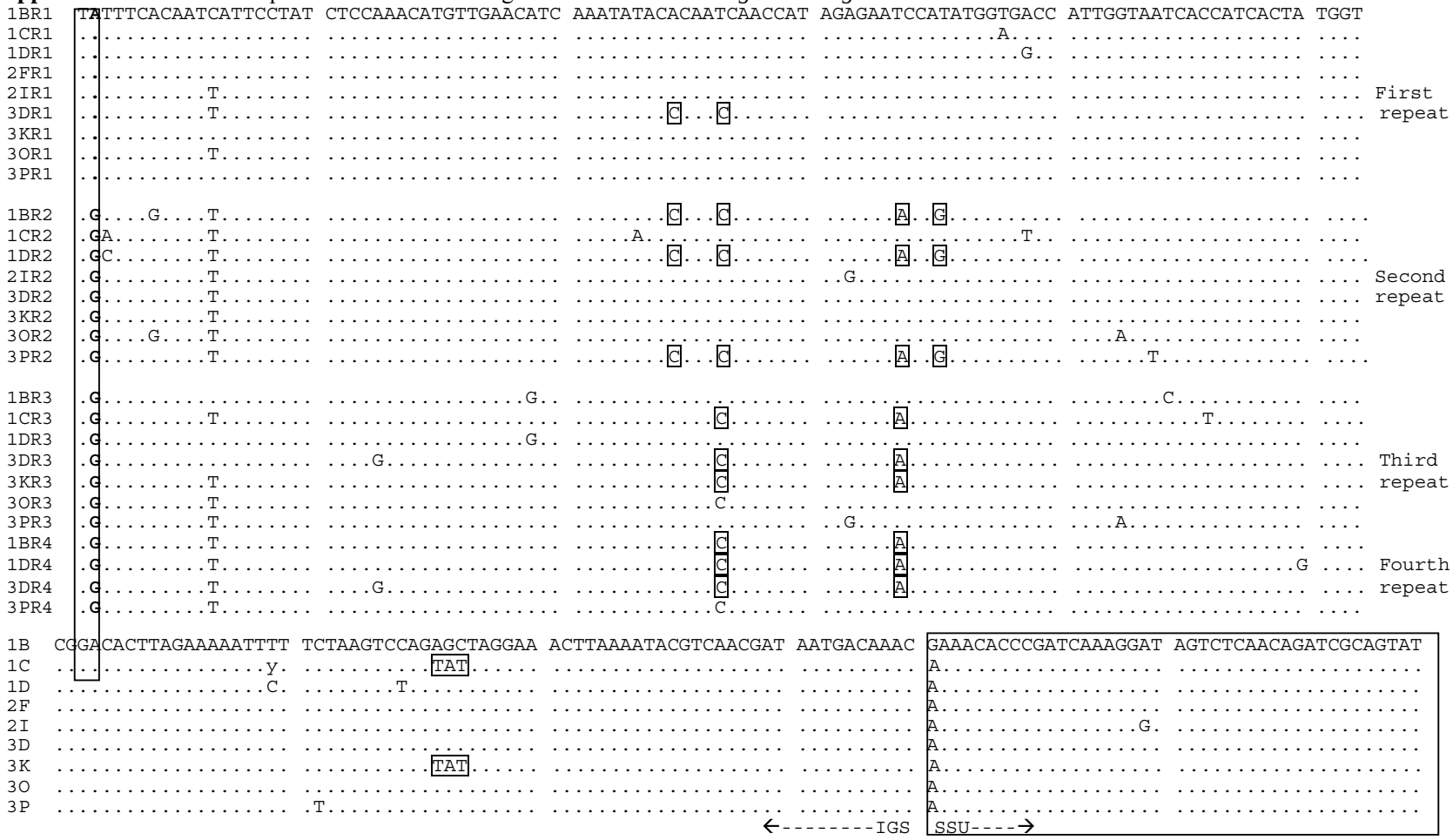
Appendix 4 continued - ND4 *Ostertagia ostertagi* 88 worms 318 nucleotides translated to 106 amino acid sequence

```
DEN9  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ...???
DEN10 ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
DEN11 ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
DEN12 ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
ESP1  ..........  ..........  ..........  ..........  ..........  ...S......  ..........  .......T..  ..........  ..........  ......
ESP2  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
ESP3  ..........  ..........  ..........  ..........  ..N.......  ..........  ..........  ..........  ..........  ..........  M.....
ESP4  ..........  ..........  ..........  ..........  .........V  ..........  ..........  ..........  ..........  ..........  ......
ESP5  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
ESP6  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
ESP7  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
ESP8  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
ESP9  ..........  ..........  ........V.  ..........  ..........  ..........  ..........  ..........  ..........  ........?? ??????
ESP10 ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
ESP11 ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
ESP12 ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Tn5   ..........  ..........  ........V.  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Tn6   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ........V.  ..........  ......
Tn9   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Tn10  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ........V.  ..........  ......
Tn11  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Tn13  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Me5   ..........  .......S..  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Me6   ..........  ..........  ..........  .........G  ..........  ..........  ..........  ..........  .G........  ..........  ......
Me8   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Me12  ..........  .......S..  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Me13  E.........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Al6   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  .G......V.  ..........  ......
Al10  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Al13  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  M.........  ..........  ......
Mn16  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Mn18  ..........  ..........  ..........  ..........  ..........  ..........  ..........  .S........  ..........  ..........  ......
Mn19  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Mn22  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Mn23  S.........  ..........  ..........  ..........  ......M...  ..........  ..........  ..........  ..........  ..........  ......
Mn24  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
Mn25  S.........  ..........  ..........  ..........  ......M...  ..........  ..........  ..........  ..........  ..........  ......
La11  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
La12  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
La15  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
La16  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
La17  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
La18  ..........  ..........  ........V.  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
La19  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ......
```

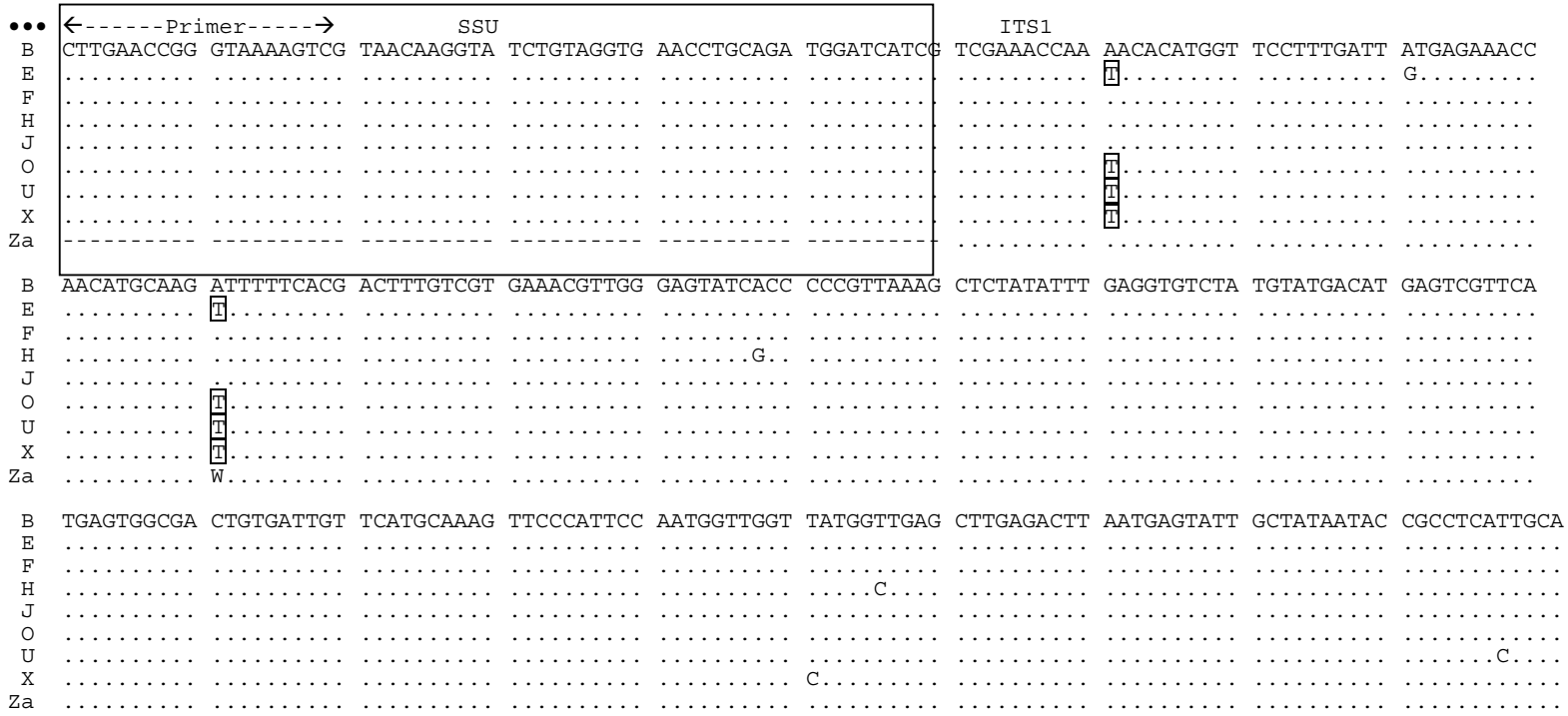
Appendix 5 - Sequences of rDNA IGS regions from three *Ostertagia ostertagi* worms for 9 clones



Appendix 5 continued - Sequences of rDNA IGS regions from three *Ostertagia ostertagi* worms for 9 clones



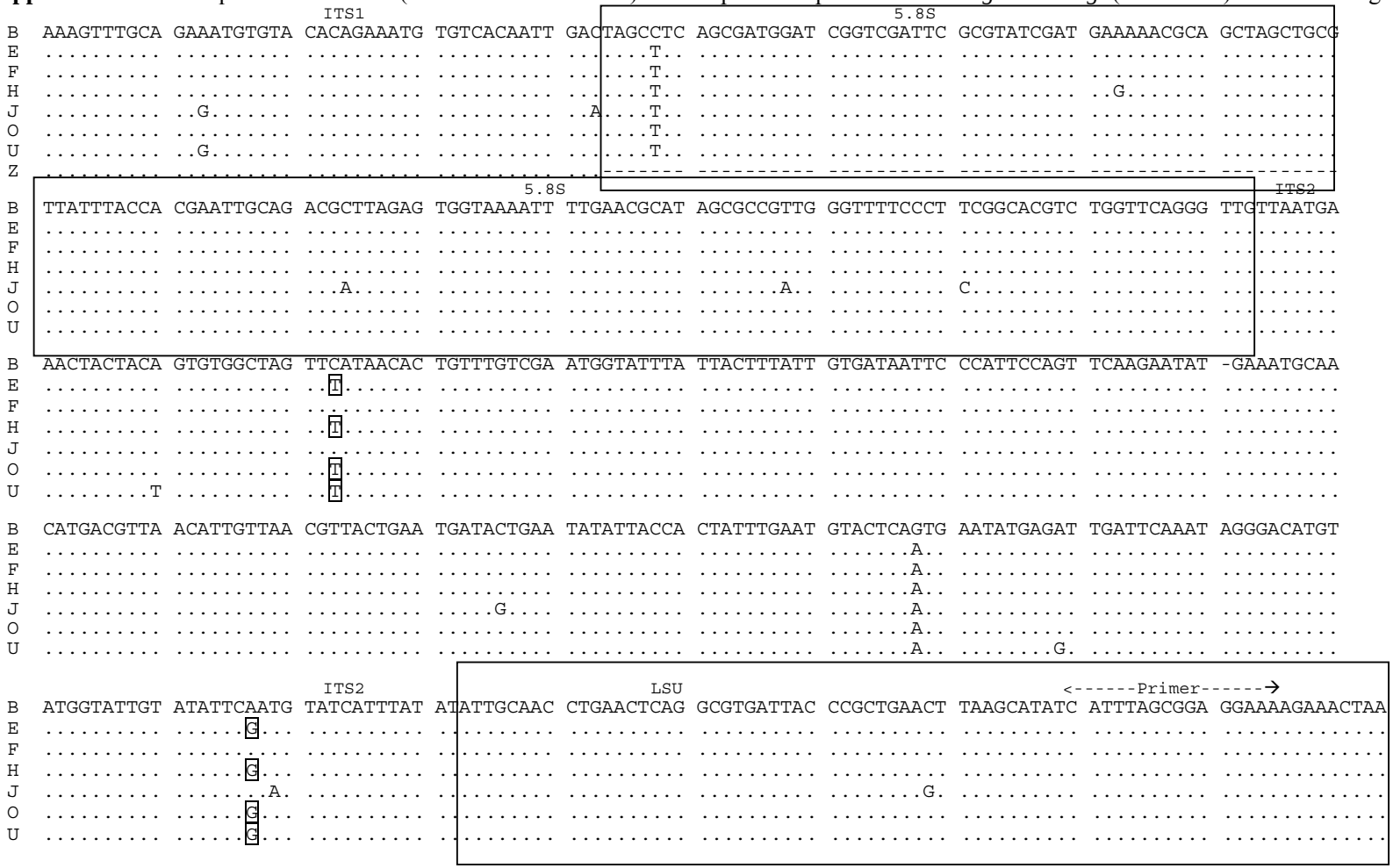
Appendix 6 - Sequence of 8 clones (accession numbers AF304560-7) and a previously reported sequence of *Ostertagia ostertagi* (AF044927) for the ITS regions



Appendix 6 continued - Sequence of 8 clones (accession numbers AF304560-7) and a previously reported sequence of *Ostertagia ostertagi* (AF044927) for the ITS regions. Each clone has 1 to 4 copies of this region so actual sequence order is e.g. B (page 123) B1 B2 B3 B4 (all page 124) B (page 125)

```
Repeats
B1  TTTATAATGG TGGTTATGTA CATACGTGTA GTATGTACGG TACCTGGTTG TCAGGAAACC TTAATGATCT CGCCTAGACG CCATTATAAA ACACAACTTT TTATGTTTGATTA
X1  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..............
U1  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..............
Z1  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..............
F1  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ............C.
J1  .......... .......... .......... .......... .......... .......... .......... .......... G......... ........G. ............C.
O1  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ............C.
F3  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ........A..C.
J3  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ........A..C.
B3  .......... .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ........A..C.
H1  C......... .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ........A..C.
U3  .......... .......... ...T...... .......... .......C.. .......... .......... .......... .......... .......... ........A..C.
X3  .......... .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ........A..C.
Z3  .......... .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ........A..C.
B5  .......... .......... .......... .......... .......C.. .......... .......... .......... .......... ..G....... ........
F5  .......... .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ........
H3  .......... .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ........
J5  .......... .......... .......... .......... .......C.. .......... .......... .......... .......... .......... ........
Z5  .......... .......... .......... .......... .......C.. ..T....... .......... .......... .......... .......... ........
E1  .......... .......... .......... .......... .......... .......... .......... .......... ...C...... .......... ........
U5  .......... .......... .......... ......G... .......... .......... .......... .......... .......... .......... ........
O2  ......G... .......... .......... ......G... .......A.. .......... .......... .......... .......... .......... ........
Z4  ......G... .......... .......... ......G... .......A.. .......... .......... .......... ...C...... .
B2  ......G... .......... .......... ......G... .......A.. .......... .......... .......... ...C...... .
B4  ......G... .......... .......... ......G... .......A.. .......... .......... .......... ...C...... .
F2  ......G... .......... .......... ......G... .......A.. .......... .......... .......... ...C...... .
H2  ......G... .......... .......... ......G... .......A.A ........T. .......... .......... ...C...... .
J2  ......G... .......... .......... ......G... .......A.. .......... .......... .......... ...C...... .
U2  ......G... .......... ........C. ......G... .......A.. .......... .......... .......... ...C...... .
U4  ......G... .......... .......... ......G... .......A.. .......... .......... ........GC ...C...... .
X2  ......G... .......... .......... ......G... .......A.. .......... .......... .......... ...C...... .
X4  ......G... .......... .....C.... ......G... .......A.. .......... .......... .......... .......... .
Z2  ......G... .......... .......... ......G... .......A.. .......... .......... .......... ...C...... .
F4  ......G... .......... .......... .......... .......C.. .......... .......... .......... ...C...... .
J4  C.....G... .......... .......... .......... .......C.. .......... .......... .......... ...C...... .
```

Appendix 6 cont. - Sequence of 8 clones (accession AF304560-7) and a reported sequence of *Ostertagia ostertagi* (AF044927) for the ITS regions



References

Akopyanz, N., N. O. Bukanov, T. U. Westblom, S. Kresovich and D. E. Berg (1992). DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. *Nucleic Acids Research* **20**(19): 5137-42.

Allendorf, F. W. and S. R. Phelps (1981). Use of allelic frequencies to describe population structure. *Canadian Journal of Fisheries and Aquatic Sciences* **38**: 1507-1514.

Anderson, N. and K. C. Bremner (1983). Life cycle and pathogenesis of helminths of major economic importance. The Epidemiology and Control of Gastrointestinal Parasites of Cattle in Australia. N. Anderson and P. J. Waller, Commonwealth Scientific and Industrial Research Organization**:** 23-34.

Anderson, T. J., M. S. Blouin and R. N. Beech (1998). Population biology of parasitic nematodes: applications of genetic markers. *Advances in Parasitology* **41**: 219-83.

Armour, J. and C. D. Ogbourne (1982). Bovine Ostertagiasis: a Review and Annotated Bibliography. Commonwealth Agricultural Bureaux.

Armstrong, J., A. Gibbs, R. Peakall and G. Weiller (2000). RAPDistance Package. Canberra, Australian National University, Canberra.

Arnheim, N., D. Treco, B. Taylor and E. M. Eicher (1982). Distribution of ribosomal gene length variants among mouse chromosomes. *Proceedings of the National Academy of Sciences of the United States* **79**: 4677-4680.

Avise, J. C. (1994). Molecular Markers, Natural History and Evolution. New York, London, Chapman & Hall.

Ayliffe, M. A., G. J. Lawrence, J. G. Ellis and A. J. Pryor (1994). Heteroduplex molecules formed between alleles sequences cause nonparental RAPD bands. *Nucleic Acids Research* **22**(9): 1632-1636.

Bachmann, K. (1994). Molecular markers in plant ecology. *New Phytologist* **126**: 403-418.

Baldridge, G. D. and A. M. Fallon (1992). Primary structure of the ribosomal DNA intergenic spacer from the mosquito, *Aedes albopictus*. *DNA and Cell Biology* **11**(1): 51-59.

Barral, V., P. This, D. Imbert-Establet, C. Combes and M. Delseny (1993). Genetic variability and evolution of the *Schistosoma* genome analysed by using random amplified polymorphic DNA markers. *Molecular and Biochemical Parasitology* **59**: 211-222.

Beerli, P. and J. Felsenstein (1999). Maximum likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**(2): 763-773.

Bensch, S., T. Anderson and S. Akesson (1999). Morphological and molecular variation across a migratory divide in willow warblers, *Phylloscopus trochilus*. *Evolution* **53**(6): 1925-1935.

Blouin, M. S., J. B. Dame, C. A. Tarrant and C. H. Courtney (1992). Unusual population genetics of a parasitic nematode: mtDNA variation within and among populations. *Evolution* **46**: 470-476.

Blouin, M. S., C. A. Yowell, C. H. Courtney and J. B. Dame (1995). Host movement and the genetic structure of populations of parasitic nematodes. *Genetics* **141**(3): 1007-1014.

Blouin, M. S., C. A. Yowell, C. H. Courtney and J. B. Dame (1998). Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. *Molecular Biology and Evolution* **15**(12): 1719-27.

Bucci, G. and P. Menozzu (1993). Segregation analysis of random amplified polymorphic DNA (RAPD) markers in *Picea abies* Karst. *Molecular Ecology* **2**: 227-232.

Chakraborty, R. (1980). Gene-diversity analysis in nested, subdivided populations. *Genetics* **96**: 721-726.

Chilton, N. B., L. A. Newton, I. Beveridge and R. B. Gasser (2001). Evolutionary relationships of trichostrongyloid nematodes (Strongylida) inferred from ribosomal DNA sequence data. *Molecular Phylogenetics and Evolution* **19**(3): 367-386.

Clark, A. G. and C. M. S. Lanigan (1994). Prospects for estimating nucleotide divergence with RAPDs. *Molecular Biology and Evolution* **10**(5): 1096-1111.

Cloarec, A., C. Rivault and M. L. Cariou (1999). Genetic population structure of the German cockroach, *Blattella germanica*: Absence of geographical variation. *Entomologia Experimentalis et Applicata.* **92**(3): 311-319.

Constantine, C. C., R. P. Hobbs and A. J. Lymbery (1994). Fortran programs for analyzing population-structure from multilocus genotypic data. *Journal of Heredity* **85**(4): 336-337.

Constantine, C. C., A. J. Lymbery, R. C. A. Thompson and D. L. Obendorf (1991). The origin of a new focus of infection with *Echinococcus granulosus* in Tasmania. *International Journal of Parasitology* **21**(8): 959-961.

Costa, J. M., M. L. Darde, B. Assouline, M. Vidaud and S. Bretagne (1997). Microsatellite in the beta-tubulin gene of *Toxoplasma gondii* as a new genetic marker for use in direct screening of amniotic fluid. *Journal of Clinical Microbiology* **35**(10): 2542-5.

Crandall, K. A. and A. R. Templeton (1993). Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134**: 959-969.

Crease, T. J. (1993). Sequence of the intergenic spacer between the 28S and 18S rRNA encoding genes of the crustacean, *Daphnia pulex*. *Gene* **134**(2): 245-249.

Crease, T. J. (1995). Ribosomal DNA evolution at the population level - nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. *Genetics* **141**(4): 1327-1337.

Dallas, J. F., R. J. Irvine and O. Halvorsen (2000). DNA evidence that *Ostertagia gruehneri* and *Ostertagia arctica* (Nematoda: Ostertagiinae) in reindeer from Norway and Svalbard are conspecific. *International Journal of Parasitology* **30**(5): 655-658.

Dame, J. B., C. A. Yowell, C. H. Courtney and W. G. Lindgren (1991). Cloning and characterization of the ribosomal RNA gene repeat from *Ostertagia ostertagi*. *Molecular and Biochemical Parasitology* **45**: 275-280.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* **284**: 2124-2128.

Elard, L., A. M. Comes and J. F. Humbert (1979). Sequences of beta-tubulin cDNA from benzimidazole-susceptible and -resistant strains of *Teladorsagia circumcincta*, a nematode parasite of small ruminants. *Molecular and Biochemical Parasitology* **79**(2): 249-253.

Elder, J. F. and B. J. Turner (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Quarterly Review of Biology* **70**(3): 297-320.

Excoffier, L., P. E. Smouse and J. M. Quattro (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package). Seattle, Distributed by the author. Department of Genetics, University of Washington, Seattle.

Fenton, B., G. Malloch and F. Germa (1998). A study of variation in rDNA ITS regions shows that two haplotypes coexist within a single aphid genome. *Genome* **41**(3): 337-345.

Files, J. G. and D. Hirsh (1981). Ribosomal DNA of *Caenorhabditis elegans*. *Journal of Molecular Biology* **149**(2): 223-40.

Frank, G. R., R. P. Herd, K. S. Marbury and J. C. Williams (1986). Effects of transfer of *Ostertagia ostertagi* between northern and southern U.S.A. on the pattem and frequency of hypobiosis. *International Journal of Parasitology* **16**: 391-398.

Frank, G. R., R. P. Herd, K. S. Marbury, J. C. Williams and E. R. Willis (1988). Additional investigations on hypobiosis of *Ostertagia ostertagi* after transfer between northern and southern U.S.A. *International Journal of Parasitology* **18**: 171-177.

Gasser, R. B., N. B. Chilton, H. Hoste and I. Beveridge (1993). Rapid sequencing of rDNA from single worms and eggs of parasitic helminths. *Nucleic Acids Research* **21**(10): 2525-2526.

Gasser, R. B., N. B. Chilton, H. Hoste and L. A. Stevenson (1994). Species identification of trichostrongyle nematodes by PCR-linked RFLP. *International Journal for Parasitology* **24**(2): 291-293.

Gibbs, H. C. (1986). Hypobiosis in parasitic nematodes - An update. *Advances in Parasitology* **25**: 129-174.

Hamilton, W. D., R. Axelrod and R. Tanese (1990). Sexual reproduction as an adaption to resist parasites (a review). *Proceedings of the National Academy of Science USA* **87**(9): 3566-73.

Hillis, D. M. and M. T. Dixon (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly review of Biology* **66**(4): 411-453.

Hoste, H., N. B. Chilton, R. B. Gasser and I. Beveridge (1995). Differences in the second internal transcribed spacer (ribosomal dna) between five species of *Trichostrongylus* (Nematoda, Trichoshrongylidae). *International Journal for Parasitology* **25**(1): 75-80.

Hudson, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology* **23**(2): 183-201.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7:** 1-44.

Hudson, R. R., M. Slatkin and W. P. Maddison (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**(2): 583-9.

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

Humbert, J. F. and J. Cabaret (1995). Use of random amplified polymorphic dna for identification of ruminant trichostrongylid nematodes. *Parasitology Research* **81**(1): 1-5.

Innan, H. and W. Stephan (2000). The coalescent in an exponentially growing metapopulation and its application to *Arabidopsis thaliana*. *Genetics* **155**(4): 2015-9.

Jannotti-Passos, L. K., C. P. Souza, J. C. Parra and A. J. Simpson (2001). Biparental mitochondrial inheritance in the parasitic trematode *Schistosoma mansoni*. *Journal of Parasitology* **87**(1): 79-82.

Johnson, S. L., C. N. Midson, E. W. Ballinger and J. H. Postlethwait (1994). Identification of RAPD primers that reveal extensive polymorphisms between laboratory strains of zebrafish. *Genomics* **19**: 152-156.

Kane, R. A. and D. Rollinson (1994). Repetitive sequences in the ribosomal DNA internal transcribed spacer of *Schistosoma haematobium*, *Schistosoma intercalatum* and *Schistosoma mattheei*. *Molecular and Biochemical Parasitology* **63**(1): 153-156.

Karvonen, P. and O. Savolainen (1993). Variation and inheritance of ribosomal DNA in *Pinus sylvestris* L (Scots Pine). *Heredity* **71**(Part 6): 614-622.

Kingman, J. F. (2000). Origins of the coalescent. 1974-1982. *Genetics* **156**(4): 1461-3.

Knight, M., A. N. Miller, C. N. Patterson, C. G. Rowe, G. Michaels, D. Carr, C. S. Richards and F. A. Lewis (1999). The identification of markers segregating with resistence to *Schistosoma mansoni* infection in the snail *Biomphalaria glabrata*. *Proceedings of the National Academy of Sciences of the United States of America* **96**(4): 1510-5.

Knox, D. P. (2000). Development of vaccines against gastrointestinal nematodes. *Parasitology* **120**(Suppl): S43-61.

Kohler, P. (2001). The biochemical basis of anthelmintic action and resistance. *International Journal of Parasitology* **31**(4): 336-45.

Kumar, S., K. Tamura and M. Nei (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *CABIOS* **10**(2): 189-191.

Larsen, J. W., A. L. Vizard and N. Anderson (1995). Production losses in Merino ewes and financial penalties caused by trichostrongylid infections during winter and spring. *Australian Veterinary Journal* **72**(2): 58-63.

Larsen, M. (1999). Biological control of helminths. *International Journal of Parasitology* **29**(1): 139-46.

Lasserre, F., F. Gigault, J. P. Gauthier, J. P. Henry, M. Sandmeier and R. Rivoal (1996). Genetic variation in natural populations of the cyst nematode (*Heterodera aveanae* Woll.) submitted to resistant and susceptible cultivars. *Theoretical and Applied Genetics* **93**(1-2): 1-8.

Levin, S. A. (1992). The problems of pattern and scale in ecology. *Ecology* **73**(6): 1943-1967.

Levitan, D. R. and R. K. Grosberg (1993). The analysis of paternity and maternity in the marine hydrozoan *Hydractinia symbiolongicarpus* using randomly amplified polymorphic DNA (RAPD) markers. *Molecular Ecology* **2**: 315-326.

Lincoln, R. J., G. A. Boxshall and P. F. Clark (1982). A Dictionary of Ecology, Evolution and Systematics. Cambridge, Cambridge University Press.

Lunt D. H. and B.C. Hyman (1997). Animal mitochondrial DNA recombination. *Nature* 15;387(6630):247.

Lymbery, A. J. (1993). Migration, selection and population size in *Ostertagia ostertagi*. *Parasitology Today* **9**(2): 37-38.

Lymbery, A. J., C. C. Constantine and R. C. A. Thompson (1997). Self-fertilization without genomic or population structuring in a parasitic tapeworm. *Evolution* **51**(1): 289-294.

Lymbery, A. J. and R. C. A. Thompson (1988). Electrophoretic analysis of genetic variation in *Echinococcus granulosus* from domestic hosts in Australia. *International Journal for Parasitology* **18**(6): 803-811.

Lymbery, A. J. and R. C. A. Thompson (1989). Genetic differences between cysts of *Echinococcus granulosus* from the same host. *International Journal for Parasitology* **19**(8): 961-964.

Lynch, M. and T. J. Crease (1990). The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution* **7**: 377-394.

Lynch, M. and B. G. Milligan (1994). Analysis of population genetic structure with RAPD markers. *Molecular Ecology* **3**(2): 91-99.

MacPherson, J. M., P. E. Eckstein, G. J. Scoles and A. A. Gajadhar (1993). Variability of the random amplified polymorphic DNA assay among thermal cyclers, and effects of primer and DNA concentration. *Molecular and Cellular Probes* **7**: 293-299.

Maddison, W. P. and D. R. Maddison (1992). MacClade: Analysis of Phylogeny and Character Evolution. Version 3. Sunderland, Massachusetts, Sinauer Associates.

McDonnell, A., S. Love, A. Tait, J. R. Lichtenfels and J. B. Matthews (2000). Phylogenetic analysis of partial mitochondrial cytochrome oxidase c subunit I and large ribosomal RNA sequences and nuclear internal transcribed spacer I sequences from species of Cyathostominae and Strongylinae (Nematoda, Order Strongylida), parasites of the horse. *Parasitology* **121**(Pt 6): 649-59.

McKellar, Q. A. (1993). Interactions of *Ostertagia* species with their bovine and ovine hosts. *International Journal for Parasitology* **23**(4): 451-462.

McManus, D. and J. Smyth (1979). Isoelectric focusing of some enzymes from *Echinococcus granulosus* (horse and sheep strains) and *E. multiocularis*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **73**(3): 259-265.

McManus, D. P. (1990). Points in question. Genetic diversity in *Echinococcus granulosus*. *International Journal for Parasitology* **20**(6): 723.

McManus, D. P. and A. K. Rishi (1989). Genetic heterogeneity within *Echinococcus granulosus*: isolates from different hosts and geographical areas characterized with DNA probes. *Parasitology* **99**(Pt 1): 17-29.

McManus, D. P. and J. D. Smyth (1986). Hydatidosis: Changing concepts in epidemiology and speciation. *Parasitology Today* **2**(6): 163-168.
Milligan, B. G., J. Leebens-Mack and A. E. Strand (1994). Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology* **3**: 423-435.

Morgan, G. T. and K. M. Middleton (1992). Conservation of intergenic spacer length in ribosomal DNA of the tailed frog, *Ascaphus truei*. *Gene* **110**(2): 219-223.

Morgan, U. M., C. C. Constantine, P. O'Donoghue, B. P. Meloni, P. A. O'Brien and R. C. A. Thompson (1995). Molecular characterisation o*f Cryptosporidi*um isolates from humans and other animals using RAPD (Random Amplified Polymorphic DNA) Analysis. *American Journal of Tropical Medicine and Hygiene* **52**: 559-564.

Morgan, U. M., P. A. O'Brien and R. C. A. Thompson (1996). The development of diagnostic PCR primers for *Cryptosporidium* using RAPD-PCR. *Molecular and Biochemical Parasitology*. **77**: 103-108.

Morgensen, H. L. (1988). The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany* **83**(3): 338-404.

Mosquera, J., M. Gomez-Gesteira and V. Perez-Villar (2000). Using parasites as biological tags of fish populations: a dynamic model. *Bulletin of Mathematical Biology* **62**(1): 87-99.

Nadler, S. (1990). Molecular approaches to studying helminth population genetics and phylogeny. *International Journal for Parasitology* **20**(1): 11-29.

Nadler, S. A. (1995). Microevolution and the genetic structure of parasite populations. *Journal of Parasitology* **81**(3): 395-403.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590.

Nei, M. (1982). Evolution of human races at the gene level. Human Genetics Part A: The Unfolding Genome. B. Bohhe-Tamir, P. Cohen and R. N. Goodman. New York, Alan R Liss**:** 167-181.

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

Nei, M. and L. Jin (1989). Variances of the average numbers of nucleotide substitutions within and between populations. *Molecular Biology and Evolution* **6**(3): 290-300.

Nei, M. and W.-H. Li (1979). Mathematical models for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Science* **76**(10): 5269-5273.

Nei, M. and J. C. Miller (1990). A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics* **125**: 873-879.

Neigel, J. E., et al (1991). Estimation of single generation migration distances from geographic variation in animal mitochondrial DNA. *Evolution* **45**: 423-432.

Neigel, J. E. (1994). Analysis of rapidly evolving molecules and DNA sequence variants - alternative approaches for detecting genetic structure in marine populations. *California Cooperative Oceanic Fisheries Investigations Reports* **35**: 82-89.

Neigel, J. E. and J. C. Avise (1993). Application of a random walk model to geographic distributions of animal mitochondrial DNA variation. *Genetics* **135**: 1209-1220.

Newton, L. A., N. B. Chilton, I. Beveridge, H. Hoste, P. Nansen and R. B. Gasser (1998). Genetic markers for strongylid nematodes of livestock defined by PCR-based restriction analysis of spacer rDNA. *Acta Tropica* **69**(1): 1-15.

Novak, E. M., M. P. de Mello, H. B. Gomes, I. Galindo, P. Guevara, J. L. Ramirez and J. F. da Silveira (1993). Repetitive sequences in the ribosomal intergenic spacer of *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* **60**(2): 273-280.

Ochando, M. D. and F. J. Ayala (1999). Fitness of wild-caught *Drosophila melanogaster* females: allozyme variants of GPDH, ADH, PGM, and EST. *Genetica* **105**(1): 7-18.

Okimoto, R., J. L. Macfarlane, D. O. Clary and D. R. Wolstenholme (1992). The mitochondrial genome of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* **130**: 471-498.

Oliveira, R. P., A. M. Macedo, E. Chiari and S. D. J. Pena (1997). An alternative approach to evaluating the intraspecific genetic variability of parasites. *Parasitology Today* **13**(5): 196-200.

Perna, N. T. and T. D. Kocher (1996). Mitochondrial DNA: molecular fossils in the nucleus. *Current Biology* **6**(2): 128-9.

Peterson, M. A. and R. F. Denno (1997). The influence of intraspecific variation in dispersal strategies on the genetic structure of planthopper populations. *Evolution* **51**(4): 1189-1206.

Planes, S., R. Galzin and F. Bonhomme (1996). A genetic metapopulation model for reef fishes in oceanic islands - the case of the Surgeonfish, *Acanthurus triostegus*. *Journal of Evolutionary Biology* **9**(1): 103-117.

Polanco, C. and M. P. Delavega (1994). The structure of the rDNA intergenic spacer of *Avena sativa* L.: - a comparative study. *Plant Molecular Biology* **25**(4): 751-756.

Pons, O. and R. J. Petit (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**: 1237-1245.

Poulin, R. and S. Morand (2000). The diversity of parasites. *Quarterly Review of Biology* **75**(3): 277-93.
Pratt, D., R. J. Boisvenue and G. N. Cox (1992). Isolation of putative cysteine protease genes of *Ostertagia ostertagi*. *Molecular and Biochemical Parasitology* **56**: 39-48.

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

Price, P. W. (1980). Evolutionary Biology of Parasites. New Jersey, Princeton University Press.

Queller, D. C., J. E. Strassmann and C. R. Hughes (1993). Microsatellites and kinship. *Trends in Ecology and Evolution* **8**(8): 285-288.

Quesada, H., R. Wenne and D. O. Skibinski (1999). Interspecies transfer of female mitochondrial DNA is coupled with role-reversals and departure from neutrality in the mussel *Mytilus trossulus*. *Molecular Biology and Evolution* **16**(5): 655-65.

Raybould, A. F., R. J. Mogg, R. T. Clarke, C. J. Gliddon and A. J. Gray (1999). Variation and population structure at microsatellite and isozyme loci in wild cabbage (*Brassica oleracea* L.) in Dorset (UK). *Genetic Resources and Crop Evolution.* **46**(4): 351-360.

Riedy, M. F., W. J. Hamilton and C. F. Aquadro (1992). Excess non-parental bands in offspring from known pedigrees assayed using RAPD PCR. *Nucleic Acids Research* **20**(4): 918.

Rogers, M. J., G. A. Mcconkey, J. Li and T. F. Mccutchan (1995). The ribosomal DNA loci in *Plasmodium falciparum* accumulate mutations independently. *Journal of Molecular Biology* **254**(5): 881-891.

Rothuizen, J. and M. Van Wolferen (1994). Randomly amplified DNA polymorphisms in dogs are reproducible and display Mendelian transmission. *Animal Genetics* **25**: 13-18.

Rozas, J. and R. Rozas (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174-175.

Ruckelshaus, M. H. (1998). Spatial scale of genetic structure and an indirect estimate of gene flow in eelgrass, *Zostera marina*. *Evolution* **52**(2): 330-343.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular Cloning. A Laboratory Manual, Cold Spring Harbour Laboratory Press, New York.

Sang, T., D. J. Crawford and T. F. Stuessy (1995). Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA - implications for biogeography and concerted evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**(15): 6813-6817.

Sangster, N. C. (1999). Anthelminthic resistance: past, present and future. *International Journal of Parasitology* **29**(1): 115-24.
Schierup, M. H. and J. Hein (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**(2): 879-91.

Schierwater, B. and A. Ender (1993). Different thermostable DNA polymerases may amplify different RAPD products. *Nucleic Acids Research* **21**(19): 4647-4648.

Schlotterer, C., M.-T. Hauser, A. von Haesler and D. Tautz (1994). Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. *Molecular Biology and Evolution* **11**(3): 513-522.

Scribner, K. T., J. W. Arntzen and T. Burke (1994). Comparative analysis of intra- and interpopulation genetic diversity in *Bufo bufo*, using allozyme, single-locus microsatellite, minisatellite, and multilocus minisatellite data. *Molecular Biology & Evolution* **11**(5): 737-748.

Shaw, P. W., C. Turan, J. M. Wright, M. O'Connell and G. R. Carvalho (1999). Microsatellite DNA analysis of population structure in Atlantic herring (*Clupea harengus*), with direct comparison to allozyme and mtDNA RFLP analyses. *Heredity.* **83**(Part 4): 490-499.

Siles-Lucas, M., C. Cuesta-Bandera and M. Cesar-Benito (1993). Random amplified polymorphic DNA technique for speciation studies of *Echinococcus granulosus*. *Parasitology Research* **79**: 343-345.

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**(1): 264-279.

Slatkin, M. and R. R. Hudson (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**(2): 555-62.

Slatkin, M. and W. P. Maddison (1989). A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**: 603-613.

Smeal, M. G. and A. D. Donald (1981). Effects on inhibition of development of the transfer of *Ostertagia ostertagi* between geographical regions of Australia. *Parasitology* **82**: 389-399.

Smeal, M. G. and A. D. Donald (1982). Inhibited development of *Ostertagia ostertagi* in relation to production systems for cattle. *Parasitology* **85**: 21-25.

Smyth, J. D. (1994). Introduction to Animal Parasitology. Cambridge, Cambridge University Press.

Sneath, P. H. A. and R. R. Sokal (1973). Numerical Taxonomy. San Francisco, W. H. Freeman and company.

Sole, R. V., S. C. Manrubia, M. Benton, S. Kauffman and P. Bak (1999). Criticality and scaling in evolutionary ecology. *Trends in Ecology & Evolution* **14**(4): 156-160.

Stevenson, L. A., R. B. Gasser and N. B. Chilton (1996). The ITS-2 rDNA of *Teladorsagia circumcincta*, *T. trifurcata* and *T. davtiani* (Nematoda, Trichostrongylidae) indicates that these taxa are one species. *International Journal for Parasitology* **26**(10): 1123-1126.

Stewart, C. N. and L. Excoffier (1996). Assessing population genetic structure and varability with RAPD data: Application to *Vaccinium macrocarpon* (American Cranberry). *Journal of Evolutionary Biology* **9**: 153-171.

Tajima, F. (1993). Unbiased estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* **10**(3): 677-688.

Tajima, F. (1996). Infinite-allele model and infinite-site model in population genetics. *Journal of Genetics* **75**(1): 27-31.

Tavare, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoretical Poplation Biology* **26**: 119-164.

Thompson, R. C. A., C. C. Constantine and U. M. Morgan (1998). Overview and significance of molecular methods: what role for molecular epidemiology? *Parasitology* **117**: S161-S175.

Thompson, R. C. A. and L. M. Kumaratilake (1982). Intraspecific variation in *Echinococcus granulosus*: the Australian situation and perspectives for the future. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **76**(1): 13-15.

Thompson, R. C. A. and A. J. Lymbery (1988). The nature, extent and significance of variation within the genus *Echinococcus*. *Advances in Parasitology* **27**: 210-258.

Thompson, R. C. A. and A. J. Lymbery (1990). *Echinococcus*: Biology and strain variation. *International Journal of Parasitology* **20**: 457-470.

Thompson, R. C. A. and A. J. Lymbery (1995). *Echinococcus* and Hydatid Disease, CAB International.

Thompson, R. C. A., A. J. Lymbery and C. C. Constantine (1995). Variation in *Echinococcus* - Towards a taxonomic revision of the genus. *Advances in Parasitology*. **35:** 145-176.

Tibayrenc, M. and F. J. Ayala (2000). Molecular epidemiology and evolutionary genetics of pathogenic microorganisms: analysis and interpretation of data. Molecular Epidemiology of Infectious Diseases. R. C. A. Thompson. London, Arnold.

Vahidi, H. and B. M. Honda (1991). Repeats and subrepeats in the intergenic spacer of rDNA from the nematode *Meloidogyne arenaria*. *Molecular and General Genetics* **227**(2): 334-336.

van de Zande, L. and R. Bijlsma (1995). Limitations of the RAPD technique in phylogeny reconstruction in *Drosophila*. *Journal of Evolutionary Biology* **8**(5): 645-656.

van Herwerden, L., D. Blair and T. Agatsuma (1998). Intra- and inter-specific variation in nuclear ribosomal internal transcribed spacer 1 of the *Schistosoma japonicum* species complex. *Parasitology* **116**(Part 4): 311-317.

van Herwerden, L., D. Blair and T. Agatsuma (1999). Intra- and interindividual variation in ITS1 of *Paragonimus westermani* (Trematoda : Digenea) and related species: Implications for phylogenetic studies. *Molecular Phylogenetics & Evolution.* **12**(1): 67-73.

van Herwerden, L., D. Blair and T. Agatsuma (2000). Multiple lineages of the mitochondrial gene NADH dehydrogenase subunit 1 (ND1) in parasitic helminths: implications for molecular evolutionary studies of facultatively anaerobic eukaryotes. *Journal of Molecular Evolution* **51**(4): 339-52.

Vogel, A. P. and R. DeSalle (1994). Evolution and phylogenetic information content of the ITS-1 region in the Tiger beetle *Cicindela dorsalis*. *Molecular Biology and Evolution* **11**(3): 393-405.

Walker, D. and J. Fox-Rushby (2000). Economic evaluation of parasitic diseases: a critique of the internal and external validity of published studies. *Tropical Medicine and Internal Health* **5**(4): 237-49.

Waller, P. J. (1994). The development of anthelmintic resistance in ruminant livestock. *Acta Tropica* **56**: 233-243.

Waller, P. J. (1999). International approaches to the concept of integrated control of nematode parasites of livestock. *International Journal of Parasitology* **29**(1): 155-64; discussion 183-4.

Waycott, M. (1998). Genetic variation, its assessment and implications to the conservation of seagrasses. *Molecular Ecology* **7**(7): 793-800.

Weir, B. S. (1996). Intraspecific differentiation. Molecular Systematics. D. M. Hillis, Moritz, C., Mable, B.K. Sinderland, Sinauer**:** 385-405.

Weir, B. S. and C. C. Cockerham (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**(6): 1358-1370.

Wellauer, P. K., I. B. Dawid, D. D. Brown and R. H. Reeder (1976). The molecular basis for length heterogeneity in ribosomal DNA from *Xenopus laevis*. *Journal of Molecular Biology* **105**: 461-486.

Welsh, J. and M. McClelland (1990). "Fingerprinting genomes using PCR with arbitrary primers." *Nucleic Acids Research* **18**(24): 7213-8.

Welsh, J. and M. McClelland (1991). "Genomic fingerprinting using arbitrarily primed PCR and a matrix of pairwise combinations of primers." *Nucleic Acids Research* **19**(19): 5275-9.

Williams, C. F. and N. M. Waser (1999). Spatial genetic structure of *Delphinium nuttallianum* populations: inferences about gene flow. *Heredity* **83**(Part 5): 541-550.

Williams, J. C. (1997). Anthelmintic treatment strategies: current status and future. *Veterinary Parasitology* **72**(3-4): 461-77.

Williams, J. G. K., A. R. Kubelik, K.J. Livak, J. A. Rafalski, and S. V. Tingey (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6531-6535.

Windon, R. G. (1996). Genetic control of resistance to helminths in sheep. *Veterinary Immunological Immunopathology* **54**(1-4): 245-54.

Yakubu, D. E., F. J. R. Abadi and T. H. Pennington (1999). Molecular typing methods for *Neisseria meningitidis*. *Journal of Medical Microbiology* **48**(12): 1055-1064.

Yeh, F. C., R.-C. Yang, T. B. J. Boyle, Z.-H. Ye and J. X. Mao (1997). POPGENE, the user-friendly shareware for population genetic analysis, Molecular Biology and Biotechnology Centre, University of Alberta, Canada.

Zarlenga, D. S., L. C. Gasbarre, P. Boyd, E. Leighton and J. R. Lichtenfels (1998a). Identification and semi-quantitation of *Ostertagia ostertagi* eggs by enzymatic amplification of ITS-1 sequences. *Veterinary Parasitology* **77**(4): 245-257.

Zarlenga, D. S., E. P. Hoberg, F. Stringfellow and J. R. Lichtenfels (1998b). Comparisons of two polymorphic species of *Ostertagia* and phylogenetic relationships within the Ostertagiinae (Nematoda, Trichostrongyloidea) inferred from ribosomal DNA repeat and mitochondrial DNA sequences. *Journal of Parasitology* **84**(4): 806-812.