Population genetics of the parasitic nematode *Strongyloides ratti*

Matthew C. Fisher

Submitted to the University of Edinburgh for the degree of Doctor of Philosophy

Institute of Cell, Animal and Population Biology
The University of Edinburgh
April, 1997

Abstract

This study investigates the genetic structure of natural populations of the nematode parasite of rats, *Strongyloides ratti*. *S. ratti* has an unusual life-cycle that may be wholly asexual, or include a free-living adult phase where conventional sexual reproduction occurs. The population genetic structure of *S. ratti* was investigated i) in relation to the partitioning of parasites within and between hosts, within and between different geographical locations and ii) in relation to the intensity of infection of different hosts. In this manner, the extent of population subdivision was characterised and the breeding structure measured.

A combination of molecular biological methods was used to isolate markers of genetic variation in *S. ratti*. Surveys of anonymous nuclear DNA loci resulted in several restriction fragment length polymorphisms. Attempts to use polymerase chain reaction amplified microsatellites were unsuccessful, and attempts to utilise mitochondrial DNA found levels of diversity that were not appropriate for use in this study.

S. ratti was sampled from rats from 11 sites in England and Scotland, there being 123 rats in total. 76 were infected giving an overall prevalence of infection of 62%. Small numbers of rats were also sampled from sites within Germany, 16% of which were infected. 1472 infective larvae were collected from these rats and the genotypes determined for each polymorphic locus. Analysis of variance and F-statistics were used to measure genetic variation at the following hierarchical levels; within parasites, within hosts, between hosts and between samples from spatially separated geographical areas. Mixed-genotype infections were common with 76% of rats containing two or more parasite genotypes. A large proportion of the total genetic diversity was found within single rats (96%). Rats with high-intensity infections tended to contain genetically more diverse parasite populations. A small amount (0 - 1.4%) of the total genetic diversity was attributable to variation between sampling sites, showing that limited population differentiation occurs. No isolation by distance was found between sites in Britain and no genetic differentiation was observed between sites in Britain and Germany.

The frequency of sexual reproduction was low within the parasite populations studied but appeared to be adequate to establish Hardy-Weinberg equilibria proportions of genotypes within most sample sites. However, some populations show deviations from Hardy-Weinberg proportions. In particular one population had a significant excess of heterozygotes. This is taken as evidence of limited allelic segregation as a consequence of the low levels of sexual reproduction within this parasite sample.

Acknowledgments

I would like to acknowledge with many thanks the generous help and support I have received from the following friends and colleagues during the course of the work described in this thesis:

My supervisors, Dr Mark Viney and Professor David Walliker, who without none of this would have been possible.

Mrs Yvonne Bissett and Mr Simon Harvey for sharing the load of culturing, counting and laboratory life. Miss Margaret McKinnon for invaluable help in teaching me the intricacies of SAS and Mr Steve Paterson for assistance in PCR and statistical matters. Professor Nick Barton, Dr Beata Nurnberger, Dr Lisa Ranford-Cartwright, Dr Andrew Read, Mr Alan Gemmill, Mrs Louise Groves, Dr Bruce Wedgewood Oppenheim and Dr Jane Carlton who all provided invaluable opinions and assistance.

Dr Alan McNicoll and Mrs Erica Gill from M.A.A.F. provided willing access to their captured rats. Dr Hans Pelz for sending me samples from his animals. Also, the many pest-control officers and farmers of Scotland and England who advised me on suitable sites and allowed access for trapping.

Professor Ali Anwar, Dr Tim Anderson, Dr Manuel Berdoy and Dr Joanne Webster for advice, access to field sites and laboratory equipment at Oxford.

All my friends and colleagues working at the I.C.A.P.B. who made my stay a pleasure.

Professor John Taylor for giving me lee-way to finish this thesis at Berkeley, California.

My family and friends for encouragement throughout.

And finally my friend Ayala Gill for proposing to, then marrying me in the midst of my writing.

I declare that with the assistance mentioned above the work presented and the opinions expressed herein are my own.

Contents

Chapter one

Introduction

1.1 Background

Few detailed studies exist on the extent and distribution of genetic diversity in species of parasitic nematodes. Fewer still systematically describe the population genetic structure of these parasites. This study considers the genetic structure of natural infections of *Strongyloides ratti*. *S. ratti* is an intestinal nematode parasite of rats. This parasite is diploid and is ideal for this type of study because (i) the basic genetics of its life cycle are understood, (ii) it is easily maintained in the laboratory and (iii) natural infections can be readily sampled. In the following chapters, I describe the use of molecular biological methods to characterise the genetic variation of *S. ratti*. I then perform a genetic analysis for a series of samples of *S. ratti* to determine the amount and distribution of genetic variation within parasites, within hosts, between hosts and between samples from spatially separated geographical areas.

There are several convincing reasons why we should be interested in the population genetics of nematode infections. First, incontrovertible evidence exists to show that nematode parasite populations are adapting in response to human control strategies. The control of disease by chemotherapeutic methods has resulted in the evolution of drug-resistant phenotypes within many nematode species. Resistance to all the major classes of anthelminthics has been documented within agriculturally important nematode species (Waller *et al.* 1996). In some areas, for example Paraguay, 75% of farms contain parasites that are resistant to benzimidazoles, levamisole and ivermectin. Benzimadizole resistance has been shown to be heritable, and thus the appearance of drug resistance reflects changes in the parasite gene pool (Prichard *et al.* 1980). A knowledge of the population genetic structure of the parasite will assist in determining how rapidly resistance alleles may arise and spread in natural populations. This is necessary to model control practices so as to minimise the potential for nematode populations to adapt in response.

Second, patterns of prevalence, intensity and rates of infection are well understood in medically important parasitic nematodes due to a wealth of cross-sectional and longitudinal surveys (Anderson & May 1985; Bundy 1988). It has been shown that the distribution of parasitic nematodes within natural populations is

overdispersed with a few hosts containing the majority of the parasites (Crofton 1971a). This distribution is thought to arise from heterogeneities in the distribution of infective stages, host immune responses and host behaviours that affect parasite acquisition (Anderson & May 1992). However, the role of parasite genotypes in the distribution is almost completely unknown. The evolution of gene-for-gene interactions have been shown to be commonplace between plants and their pathogens (Agrios 1988; Thompson & Burdon 1992). Similarly, coevolution between mammalian hosts and their parasites are thought to explain the diversity of alleles found at the major histocompatibility complex loci (Hughes & Queller 1993). Theoretical models, such as the Red Queen (Jaenike 1978; Lively & Apanius 1995), predict selection of parasites for those that are closely adapted to host genotype. If co-evolution between host and parasite occurs, then hosts with heavy infections may contain unique populations of parasites that have been selected to that particular host genotype. Consideration of the population genetic structure between hosts with heavy and light infections will show whether 'overdispersed' parasites comprise genetically unique subpopulations.

Third, a knowledge of the population genetic structure of a nematode parasite species is necessary when considering making a vaccine. The current vaccines produced by recombinant DNA methods are directed at specific subunits of antigens. For such vaccines to be effective, they must stimulate effective antiparasitic immunity against all naturally found epitopes of the antigens. If a parasite species is genetically highly substructured, or diverse, then there may be a range of polymorphic alleles for a given antigen (Targett 1995; Conway 1997). This diversity may translate into antigenic diversity as has been shown for a candidate vaccine for *Trichostrongylus colubriformis,* where substitutions naturally occur for 12/284 amino acids (Frenkel *et al.* 1989). Therefore, it is important to consider whether a candidate vaccine antigen will confer protection against all naturally found 'strains' of a parasite .

1.1.2 The biology of *Strongyloides spp.*

The genus *Strongyloides* is a member of the order Rhabditoidea. *Strongyloides stercoralis* was originally recognised in 1876 when unknown parasites were found in the faeces of French troops repatriated from Cochin China (Vietnam). A separate worm found in a soldier's gut after an autopsy was thought to be a closely related species (Grove 1989). These two "species" were in fact the parasitic and free-living stages of the same species. This was recognised by Grassi (Grassi, 1879a) and the genus *Strongyloides* was created to house *Strongyloides intestinalis*, subsequently renamed

Strongyloides stercoralis (Stiles and Hassal 1902). Since then, 52 species of *Strongyloides* have been described which infect a diversity of vertebrate host species, and which have a worldwide distribution in tropical and temperate latitudes (Speare 1989).

The members of the genus *Strongyloides* share a common life cycle that includes parasitic and free-living generations. Infection is achieved by infective third stage larvae (iL3s) penetrating the skin of a susceptible host. These larvae migrate through the body and eventually establish in the small intestine as adult parasites (Grove 1989). The route of migration has long been supposed to be along the Loos-Fuelleborn 'blood-lung' route (Abadie 1963). However, Tindall and Wilson (Tindall & Wilson 1988) demonstrated by compartmental analysis that at least 35% of larvae establishing in the small intestine migrated *via* the naso-frontal head region, and that this was an essential pathway for successful development to an adult parasite. Migration through the lung was found to occur in a low proportion of infecting larvae. This was however not essential for further development, in contrast to *Nippostrongylus brasiliensis*, for which migration by this route was a prerequisite for successful establishment (Tindall & Wilson 1988). This demonstrates that not all skin-penetrating nematodes have similar routes of migration.

The intestinal phase of the *Strongyloides* infection consists solely of adult female parasites embedded in the gut mucosa (Abadie 1963). These parasites reproduce parthenogenetically and the eggs are passed into the external environment in faeces, where further development occurs. *Strongyloides* species are unique amongst parasitic nematodes in that the free-living generation can undergo two types of development, termed homogonic and heterogonic (Figure 1.1). In homogonic development, larvae moult through first and second larval stages into infective third stage larvae (iL3s) which can infect a host directly. In heterogonic development, larvae develop into free living dioecious adults that mate and whose progeny develop into iL3s. Thus, the life cycle of *Strongyloides* can be completed with or without sexual reproduction. Although the cycle is common to all species of *Strongyloides*, only *S. planiceps* have been shown to be able to undergo multiple free living generations (Yamada *et al.* 1991).

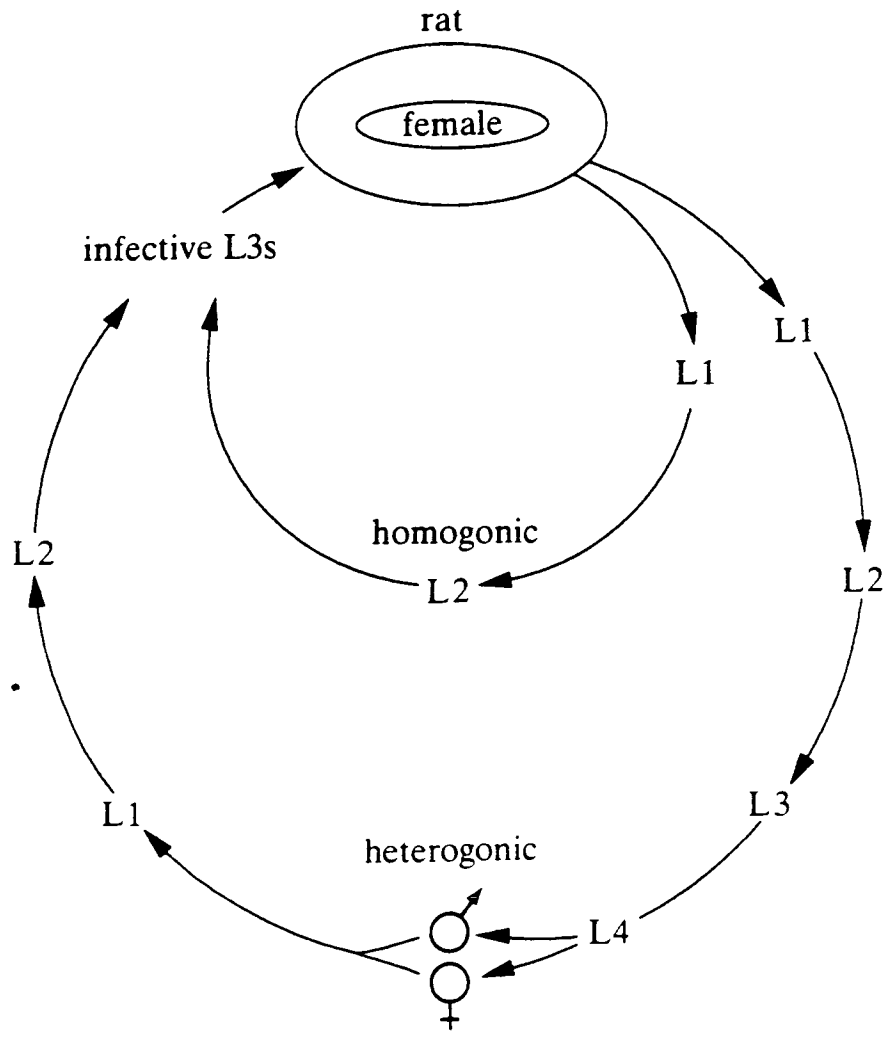


Figure 1.1. The life cycle of *Strongyloides ratti*. Larval stages are denoted by "L" (from Viney *et al.* 1993).

1.1.3 The biology of *Strongyloides ratti*

Two species of *Strongyloides* have been described from rats, *S. venezuelensis* Brumpt, 1934 from *R. norvegicus* in Venezuela and *S. ratti* Sandground, 1925 from *Rattus* sp. in Baltimore, USA. It is possible to differentiate these two species morphologically by their gut/ovary disposition and the shape of the stoma in *en face* view of the parasitic female (Little 1966a).

S. ratti shares the basic *Strongyloides* developmental plan in that a population of larvae isolated from a rat may develop by either the heterogonic or homogonic routes (Figure 1.1). Studies by Viney (Viney 1994) on the inheritance of genetic markers have shown that reproduction by the female parasite is by mitotic parthenogenesis and as a consequence the progeny are genetically identical to each other and to the parent. The mode of inheritance in free living adult worms was originally thought to be a form of meiotic parthenogenesis, and that inheritance was maternal only (Zaffagnini 1973; Triantaphyllou & Monocol 1977). That genetic exchange occured between free living males and females of *S. ratti* was demonstrated by Viney (Viney *et al.* 1993). In this study it was shown, using multilocus genetic markers, that segregation and recombination of alleles at several loci occured, and that these were consistent with normal sexual reproduction. Thus, the heterogonic cycle is the only phase in *S. ratti* life cycle in which genetic recombination occurs.

1.1.4 Development of *S. ratti*

The development of *S. ratti* is influenced by environmental and genetic components. The genetic component has been demonstrated by selection experiments showing that heterogonic and homogonic development are heritable traits (Viney 1996). This has been shown by selecting the larvae produced by an *S. ratti* line to develop by heterogonic or homogonic routes. These larvae were used to reinfect rats and the selection procedure repeated. Significant responses were observed to selection after 14 generations. Selection was successful in producing a line that was wholly homogonic; however there was a limit to selection for heterogonic development, suggesting a biological limit for this trait.

Work on the influence of the environment on the development of *Strongyloides spp.* has shown that extra-host factors, such as temperature, crowding and food availability are important (Schad 1989). A general conclusion is that an increase in heterogonic development occurs with "increasingly stressful environmental conditions" for the parasite. That the within host environment may also influence the

developmental route taken, has been demonstrated. Measurements show that the proportions of larvae developing heterogonically is high in certain isofemale lines of *S. ratti* (Viney *et al.* 1992). In these lines, infrapopulations (the parasites within a single host) of the parasite become steadily biased to development by the heterogonic route during the course of an infection (Viney 1996). At the start of the patent period, 75 - 90% of larvae develop homogonically into iL3s, however by the end of the infection, 75 - 85% of larvae develop heterogonically into free living adults. That the immune response is a major factor in determining the amount of heterogonic development has been confirmed by experimental manipulations of host immunity. In one experiment, three groups of rats were infected with *S. ratti*; immunocompromised, naive, and rats with previously acquired immunity to the parasite. Heterogonic development was least in the immunocompromised hosts, intermediate in naive rats, and greatest in rats with a previous history of exposure (Gemmil *et al.* 1997). Thus, it appears that larvae are more likely to develop into sexual adults as hosts develop immunity against parasitic females. Whether this is evidence that sexual reproduction is maintained in *S. ratti* as a mechanism for generating genetic diversity and thus avoidance of a genotype-specific host immune response is currently under debate.

1.1.5 The geographical and host distribution of *S. ratti*

S. ratti has been recovered from rats over a large proportion of the globe. Sites from which it has been recorded are listed in Table 1.1. The small size of the parasitic females (1.85 - 3.03 mm (Sandground 1925)) and the fact that they are embedded in the gut-wall means that faecal culture is usually necessary to diagnose infected rats. Neither species of *Strongyloides* have been reported from rats in Britain and Europe in studies designed to survey the parasitic fauna of *Rattus norvegicus* (Owen 1976; Webster & MacDonald 1995). However, the presence of *S. ratti* within Britain has been shown by Viney (Viney 1990), suggesting that the previous studies may have overlooked the infection.

Species	*Reference*	*Area*	*Prevalence of infection*
Rattus rattus	(Udonsi 1989)	Port Harcourt, Nigeria	14.7%
Rattus norvegicus	(Graham 1936)	New Jersey, USA	-
Sigmodon hispidus	(Boggs *et al.* 1991)	Oklahoma, USA	30.0%[a]
Rattus norvegicus	(Viney *et al.* 1992)	Britain	-
Rattus norvegicus	(Viney 1990)	Kagoshima, Japan	-
Bandicotaindica	(Sinniah 1979)	Malaysia	6.5%
Rattus tiomanicus,	"	"	73.9%
Rattus exulans	"	"	23.1%
Rattus norvegicus	"	"	8.6%
Rattus sabanus	"	"	16.7%
Rattus annandalei	"	"	35.9%
Rattus argentiventer	"	"	14.6%
Rattus rattus diardii	"	"	62.6%
Rattus rattus	(Hasegawa *et al.* 1994)	Lanyu, Taiwan	-
Rattus norvegicus	(Wertheim *et al.* 1970)	Tel-Aviv, Israel	35.7%[a]
Rattus argentiventer	(Hasegawa *et al.* 1992)	West Java, Indonesia	-

Table 1.1 Geographical distribution of rodent surveys that have recorded *S.ratti* infections. [a]species undifferentiated (*S. ratti* or *S. venezuelensis*)

In the absence of morphological markers, host fidelity of *Strongyloides* spp. has been used as a method of differentiating species. The genus generally shows high specificity of species to particular hosts. *S. ratti* is found within a range of rat species (Sinniah 1979) and can be maintained in inbred mice with some degree of success, depending on the strain of mouse used (Dawkins *et al.* 1980). Successful infections have also been established in the gerbil *Meriones unguiculatus* (Niamatali *et al.* 1992). However, the infection is refractory to rabbits (Sandground 1925) and man (Viney, pers. comm.). That natural *S. ratti* infections have not been reported in rodent species other than rats suggests that natural populations of the parasite are maintained solely within rat populations. However, it is not possible to discount the possibility that, if found within a non-rat host, the parasite may have been mis-identified as a separate species.

1.1.6 Epidemiology of *S. ratti* infections

The global prevalence of infection of *S. ratti* is 29% (standard deviation = 21%), calculated from 11 studies (Table 1.1). The infection is thus common in natural populations of rats.

The durations of *S. ratti* infections in the field is not known. Studies of *S. stercoralis* infections in man have shown that ex-far east prisoners of war retain life-long infections due to continuous cycles of autoinfection (Cook 1987). However, there is no evidence that this occurs in *S. ratti*. Chronic infections of *S. ratti* lasting from five months to a year have been reported in laboratory infections of *Meriones unguiculatus* (Niamatali *et al.* 1992). Post mortem examinations have found no intestinal development of iL3s, and thus autoinfection as a mechanism maintaining infection can probably be ruled out. Laboratory infections typically resolve in 4-6 weeks (Gemmil *et al.* 1997; Viney 1996). However infections in caged wild rats typically remain high over six months (Fisher 1997) (section 3.2.7b). This suggests that a difference in the strength of immunity between wild and laboratory rats exists. That the strength of the immune response is dependent on the size of parasite infections in laboratory rats has been demonstrated by Uchikawa (Uchikawa *et al.* 1989), where the rate of expulsion of worms was found to be dependent on the size of infective dose of iL3s. Moreover, chronic infections in rats lasting for 149 days (Graham (1940) can be established by infection with single iL3s. These studies suggest that naturally infected rats may have more persistent infections, due principally either to lower worm burdens, or to less effective immunity, than occurs in laboratory rats.

Transmission in the field is probably predominately by skin-penetration of rats by mature iL3s. However, transmission in the field may also occur by transmammary transmission between female rats and offspring. Kawanabe (Kawanabe *et al.* 1988) demonstrated that mothers transmitted infections to suckling offspring up to 36 hours post infection. This presumably occured by the offspring ingesting larvae migrating subcutaneously through the parent (Tindall & Wilson 1988). However, the relative importance of this process in the field is not known.

1.1.7 The epidemiology and distribution of nematode infections

The distribution of parasitic nematode infections share certain general characteristics. All helminth species examined have been found to have aggregated distributions in hosts, with a minority of hosts containing the majority of parasites (Crofton 1971b; Shaw & Dobson 1995). This aggregation is effectively described by

the negative binomial distribution and is characterised by variance to mean ratio greater than one ($s^2/x > 1$) (Anderson & May 1992). The overdispersion of helminths means that the majority of the parasite population is concentrated in a few hosts, while the remainder is thinly and patchily distributed throughout a large number of hosts.

Aggregated distributions are thought to arise as a consequence of heterogeneities within the host population (Grenfell *et al.* 1995). These heterogeneities may arise due to variation between hosts in one, or a combination of several, factors which include (i) the genetics of the host, (ii) the immunological status of the host and (iii) the rate of exposure of the host to the parasite (Anderson & May 1992).

(i) That resistance to infection exists, and is in part a genetically determined trait, has been shown in many breeding studies where small ruminants have been successfully selected for the criterion "ability to initiate and maintain anti-parasitic responses" (Woolaston & Baker 1996). The genetic component of the most often measured trait "reduction of faecal egg counts" (FEC) has been shown to be moderate with heritabilities of 0.2 - 0.3 (Gray *et al.* 1995). The bulk of the data indicates that the resistant phenotype is a polygenically determined trait and results from a combination of additive and non-additive genetic effects with environmental modifiers (Beh & Maddox 1996). A study of worm burdens in wild Soay sheep has found significant associations between alleles of the major histocompatibility complex and reduction in FECs (Paterson *et al.* 1997) suggesting that the genes involved in immune responses are important in determining susceptibility to infection. However, other studies on correlations between MHC alleles and resistance to infection in domestic animals have found no significant associations (Blattman *et al.* 1993).

(ii) The immunological status of the host has been demonstrated as a cause of predisposition to infection, and thus as a mechanism in causing overdispersion. For instance, studies of *Heligmosomoides polygyrus* infections in mice have shown that overdispersed populations of parasites naturally establish themselves within laboratory populations of outbred mice. Work in which these mice were cured, then allowed to reinfect themselves, demonstrated a significant variation in predisposition to infection (Scott 1987). It was subsequently shown that greater parasite aggregation occurred within host populations that had acquired resistance to infection (Tanguay & Scott 1992). This shows first that there is heterogeneity between individual mice in ability to mount an immune response, and second that this heterogeneity causes overdispersion of the parasite population. The nature of the variation between individuals in their

ability to mount an immune response is unclear. A genetic component is clearly observed as discussed above in (i). The nutritional status of the host is also undoubtedly important. Low dietary protein content (2% as opposed to 8%) severely compromises the effectiveness of mice to clear *H. polygyrus* infections (Slater & Keymer 1986). Similarly, it has been shown that reducing the zinc content in rats' diet increases the effectiveness of establishment of *S. ratti* infections (Fenwick *et al.* 1990). It is unlikely that hosts are of similar nutritional status in natural settings and so this factor may be an important mechanism in generating overdispersion.

(iii) Heterogeneity between hosts in their exposure to the parasite may occur. This may be caused either by variation in host behaviour or by variation in the temporal or spatial distribution of infective stages of the parasite (Anderson & May 1992). Several studies have characterised the temporal dynamics of parasite prevalence and patterns ranging from long term stability (ie. Acanthocephalan parasites of fish (Kennedy 1987)) to cycles that vary seasonally (ie. helminths in field mice (Montgomery & Montgomery 1989) have been found. That these cycles may be governed by the density of the host distribution has been suggested by Hudson *et al.* in response to the observation that the population densities of *Trichostrongylus tenuis* and red grouse appear to be correlated (Hudson *et al.* 1992).

That the infective stages of helminths mirror the within-host distribution and are aggregated in the environment was demonstrated by Hominick *et al.* In this experiment, the application of damp pads on the ground in an area with endemic *Ancylostoma duodenale* infection showed a high degree of aggregation (Hominick *et al.* 1987). Studies of *Strongyloides stercoralis* suggest that in endemic communities such non-random distributions may result in transmission on a limited scale. Household clusterings of infection were observed in communities in Jamaica (Lindo *et al.* 1995) and Bangladesh (Conway *et al.* 1995) suggesting person to person transmission of infection within households. The prevalence of infection within these populations was low with 3.5% of the population positive by faecal egg counts. Furthermore, no reinfection by *S. stercoralis* was seen following mass chemotherapy aimed at controlling infection (Goulart *et al.* 1977). These data suggests that *S. stercoralis* populations are characterised by low transmission rates and limited geographical dispersion of transmission stages. Whether this is true for *S. ratti* is not known although the greater prevalence of infection in rats (Table 1.1) suggest that the transmission rate may be higher. Genetic evidence that limited transmission, and hence spatial clustering of parasites occurs has been shown by Anderson (Anderson *et al.*

1993), where humans were found to contain genotypically identical *Ascaris suum* more frequently than expected by chance. Unfortunately, no studies have examined whether limited spatial transmission or infection by (genetically) like parasites are important in generating overdispersed patterns of infection. This has been recognised in a recent symposium;

" A subsection of studies on predisposition should address the question; are heavily infected individuals becoming infected by their own macroparasites? This could be examined by comparing the variation in genetic relatedness between worms within, and between, individuals within a population."
(Hudson and Dobson, 1995).

1.1.8 The influence of nematode distribution on population genetic structure

The influence of non-random distributions of nematode infections on population genetic structure is unknown. It was stated by Price (Price 1980) that;

" A paradigm of parasite population structure is that parasitic organisms are characterised by small populations with high levels of inbreeding, low intrapopulation genetic variability and high levels of interpopulation differentiation due to genetic drift and founder effects "

It is apparent from the above quote that Price assumes that parasites experience limited transmission and strong isolation. He reasons that hosts may be perceived as spatially and temporally varying resources. The 'patchiness' of the host environment is a consequence of low probabilities of establishment and high probabilities of extinction of the parasitic infection. Price argued that as a consequence of host 'patchiness', the gene pool of the parasite species would be substructured into many small isolated inbreeding compartments (*ie.* subpopulations). These would be expected to differentiate as a consequence of the genetic processes operating in subpopulations.

1.1.9 The genetics of subdivided populations

Clearly, nematode parasites may exist as subdivided populations if transmission is limited and local. Furthermore, overdispersed populations such as those found in parasitic nematodes may be subdivided due to the limited opportunity

for outbreeding in the majority of infections. If population subdivision occurs, then the ways in which this will influence the population genetic structure are manifest.

Since studies by Dobzhansky and Wright in the '30s on the distribution of recessive lethal alleles within natural populations of *Drosophila pseudoobscura*, it has been apparent that genetic variation may be structured between natural populations. Wright (Wright 1931) held that genetic drift in the frequencies of alleles and heterozygosity would occur in isolated populations. This is due to stochastic effects, such as the genetic sampling that occurs during breeding. For neutral variation, random genetic drift may be an important mechanism in determining the amount of differentiation between populations. The rate of genetic drift depends on the numbers of individuals contributing progeny to the next generation. This was defined by Kimura and Crow (Kimura & Crow 1963) as the quantity of reproducing individuals in the parental generation that determine the degree of relatedness of the progeny, or the effective population size (N_e). N_e is determined by the amount of population subdivision within a species. Substructuring of a population into subpopulations reduces the size of *Ne* for each subpopulation, relative to that of the total population. If N_e is small, then the effect of sampling variation due to mating has an increased probability of being passed onto the next generation as a change in allele frequency (Hartl & Clarke 1989), and genetic drift occurs. For instance, stochastic models show that in a parasite population with N_e of 1000 and a generation time of 90 days, the average time to fixation of a neutral mutation is 1000 years. However, if N_e is reduced to 50, then fixation will be expected to occur within 50 years (Nadler 1995).

As well as population subdivision, the breeding system of the species will determine N_e. Species with biased sex ratios have reduced N_e relative to the total number of breeding adults. This is due to over-representation of alleles from the rare sex in each generation. The effect will be especially marked in species that show restricted genetic segregation as a consequence of facultative sexual reproduction, and may be expected in *Strongyloides ratti*.

N_e, and thus the amount of genetic differentiation between subpopulations that is attributable to the effect of drift, is dependent on gene flow. If gene flow between subpopulations is high then the relatedness between individuals within subpopulations will be low, and N_e will be large. Gene flow may be considered by under Wright's Island model in which populations exchange a certain number of individuals, m, each generation. If $N_e m > 1$ between populations then, at equilibrium, differentiation due to

drift will not occur. Conversely, for values of $N_em < 1$, the effects of gene flow are not strong enough to homogenise allele frequencies between populations and drift will occur (Slatkin 1987). Therefore, the genetic structure of nematode species will be shaped by the movement of individual parasites, or by groups of parasites, between subpopulations. This is in part a function of the transmission rate of the species, and will be determined by the same factors, such as the basic reproductive rate (Ro), host mobility and density, environmental resistance of infective stages and numbers of obligate hosts in the life cycle.

The amount of gene flow between subpopulations is important when considering the rate of dispersal of alleles. This is especially important when considering the effect on alleles that are under selection. For instance, the worldwide distribution of organophosphate resistance in the mosquito *Culex pipiens* has been shown to be derived from an allele that arose once (Raymond *et al.* 1991). The rate of spread of this rare resistance allele is determined by rates of gene flow, however the frequency of such alleles in subpopulations is, to a great extent, determined by selection. Within a closed nematode population (*ie.* there is no dispersion from adjacent populations) that is under constant selection pressure from a drug, resistance alleles increase in frequency with each successive parasite generation. The rate of increase in frequency of these resistance alleles depends on (i) the generation time of the parasite and (ii) the relative fitness of resistance and susceptible alleles (Crow 1986). The time for a significant accumulation of resistant individuals depends on (i) and (ii) as well as the initial frequency of the resistance allele in the population and the threshold at which resistance is recognised (Anderson & May 1992). However, parasite populations are not closed and gene flow from reservoirs of untreated parasites can occur. As shown in a theoretical paper by Comins (Comins 1977a), these individuals will prevent fixation of resistant mutations within the selected population if m is high enough. Conversely, gene flow may also act to spread resistance genotypes into populations that had not previously evolved specific responses against selective agents (Endler 1977). The introduction of a novel selected allele into a subpopulation will lead to an increase in its frequency at a rate that, while more rapid than that of a neutral mutation, is also dependent on the Ne of the subpopulation. That is, the expected fixation time of the resistance allele is greater in subpopulations with large Ne.

There is a growing body of empirical evidence to show that population subdivision is a potent force in evolution. Microevolution of phytophagous insects

such as leaf miners (*Stilbosis quadricustatella),* as a consequence of selection by the particular defence traits of individual host plants, has been cited as a mechanism by which sympatric yet genetically distinct races form (Mopper 1996). Experiments on several insect species show that individuals transplanted from natal to novel hosts suffer a competitive disadvantage *e.g.* (Karban 1989). Differentiation between populations occurs on small spatial scales (*ie.* between adjacent plants). This occurs even in species with relatively high migration rates. It is acknowledged that only a little gene flow is needed to homogenise populations (Slatkin 1987), and therefore strong selection by host traits seems to be the mechanism that maintains genetic differentiation within this system. This hypothesis is supported by the observation that intimacy of mode of feeding by the insect (and hence exposure to host defence mechanisms) and not dispersive ability is the major determinant of genetic isolation (Mopper 1996). There are obvious parallels between this system and endoparasitic species where the host defence mechanisms (*ie.* the immune system) are exquisitely sensitive to individual parasite traits.

1.1.10 Genetic variation in helminth populations

If parasitism predisposes species of a parasite to a subdivided population structure (Price 1980), then low overall genetic diversity relative to free-living species will result due to the effects of inbreeding and genetic drift (Hartl & Clarke 1989). There is strong evidence however that this is not the case for parasitic helminths.

Intraspecific genetic variation is widespread within natural populations. This conclusion is supported by studies on 1) the response of morphological and phenotypic characters to artificial selection (Fincham 1994) 2) the deleterious effects of inbreeding *e.g.* (Sved & Ayala 1970) and 3) surveys of whole-genomes for molecular variation *e.g.* (Lander & Botstein 1989). The genetic diversity of free living invertebrate species is as a rule higher than is observed for vertebrates, proportions of polymorphic loci ranging from 17.5% - 58.7% within the invertebrates compared to 14.5% - 33.6% in vertebrates (Selander 1976) .

Generally, helminth parasites exhibit little intraspecific morphological variation (Thompson & Lymbery 1990). Furthermore, if present, the use of variation in morphological characters as a method of characterising intraspecific variation is difficult. This is in part due to the effect of host-induced variation on parasite characters. For instance, factor analysis was used to discriminate between two strains of *Echinoccoccus granulosus* on the basis of rostellar morphology. Despite the fact that

these strains differ genetically and biologically, no difference was found. Rather, the morphological variation in rostellar characters correlated with the type of host most recently infected, showing that there is phenotypic plasticity for this character (Lymbery *et al.* 1990b). However, examples of genetically determined intraspecific morphological variation do exist. *Haemonchus contortus* characteristically has five vulval-morphs that can be maintained as inbred lines. The three main morphs respond to selection and segregate in crosses (LeJambre *et al.* 1972), showing that there is a genetic basis for the variation.

Many studies have surveyed parasitic helminths for variation in isoenzymes. Among 33 taxa of parasitic helminths surveyed for isoenzyme polymorphisms, 23% of loci were polymorphic (n = 11 - 38 loci) and of these loci, an average of 7% were heterozygous within individuals (Nadler 1990). The frequencies of polymorphic loci within these species are similar to those observed in free-living invertebrates where 39% of loci were polymorphic (Nevo 1978). Mean heterozygosity is also similar with levels of 10% observed in free-living invertebrate species (Nevo 1978).

Of the 33 parasitic helminths surveyed, 17 were Ascaridoid nematodes. Within this group, 37% of the loci were polymorphic and the mean frequency of heterozygous loci within individuals was 9.5% (n = 18 - 24 loci) (Bullini & Nascetti 1986). In this study, a significant difference was observed in both mean polymorphism and heterozygosity between the species with simple (one host) and complex (more than one host) life cycles. Frequencies of polymorphic loci of 24% and 64%, and heterozygosities of 4% and 15%, were observed for the species with simple and complex life cycles, respectively. It is possible that this difference results from natural selection maintaining genetic diversity within species that have adapted to more than one host. However, assessment of the heterogeneity of a parasites lifecycle is a subjective exercise, as even simple cycles involve a broad range of microenvironments presented within the bodies of single hosts (Read & Skorping 1995).

Measurements of the genetic diversity of samples of *Ascaris suum* collected from pigs and humans in Guatemala show levels of polymorphism at enzyme loci (38%) similar to that observed in other nematodes (Anderson *et al.* 1993). Similar frequencies were found in samples collected in the USA by Leslie (Leslie *et al.* 1982) (20%) and Nadler (Nadler *et al.* 1995) (46%). The genetic diversity of mitochondrial DNA (mtDNA) within Guatemalan populations was measured by restriction digests of PCR (polymerase chain reaction)-amplified mtDNA. Nucleotide diversity was 0.89%

and 1.57% for populations of parasites from pigs and humans respectively (Anderson *et al.* 1993). These values are higher than those usually found from the mtDNA of free-living taxa, which range from 0.05 - 0.51% (Lynch & Crease 1990). Parasites from pigs and humans within this area have been shown to be composed of almost completely genetically isolated populations, and the high mtDNA diversity may be explained by rare introgression events of haplotypes into the separate gene pools. However after correcting for the effects of the most extreme haplotypes, diversities were still high (0.36%).

A similar study within five species of trichostrongyle nematodes found mtDNA nucleotide diversities twice as high (Blouin *et al.* 1995). The mean value, 2.5%, was amongst the highest reported for any taxa. There are several mechanisms which might maintain the high mtDNA diversity observed within these populations, (i) an accelerated rate of nucleotide substitutions within nematode mtDNA, (ii) the sampled populations being mixtures of individuals from previously differentiated populations and (iii) large long term effective population sizes. While it is not clear which mechanism predominates within this system, the population sizes of Trichostrongyles are usually one to two orders of magnitude higher than those of *A. suum* (Nadler *et al.* 1995). Thus, the difference in mtDNA diversity between the two species may well be a consequence of the differences in population sizes. Genetic variation at nuclear DNA loci was measured for two β–tubulin genes of *Haemonchus contortus* by restriction surveys of PCR amplified loci (Beech *et al.* 1994). Nucleotide diversity was measured at 9.4% and 9.1% for each locus. This is much higher than that found in free-living taxa where values of 0.03 - 0.57% are characteristic (Lynch & Crease 1990). It is also higher than that measured in *H. contortus* mtDNA (Blouin *et al.* 1992). This result is unusual, the mtDNA having a mutation rate that is approximately 10 times greater than that observed in nuclear DNA (Thomas & Wilson 1991). Therefore, either these genes have an uncharacteristically high mutation rate or they are highly divergent due to extensive population subdivision.

In conclusion, helminth species show extensive genetic diversity at both nuclear and mtDNA loci. Moreover, there is no evidence to support the hypothesis that a parasitic mode of life results in reduced levels of genetic diversity relative to that found in free living organisms.

1.1.11 Population genetic structure of parasitic helminths

This section examines what evidence there is that the genetic variation seen in helminth species is structured into subpopulations. The population genetic structure of nematode parasites will depend to a large extent on the effective population size, N_e, of parasite subpopulations.

The effect of host movement on population genetic structure:

A major determinant of N_e, and thus of parasite population structure, will be the mobility of the host and the effect that this has on encountering and dispersing infective parasite stages. Studies on ascaridoid nematode parasites of pelagic marine mammals (*e.g.* seals) which use fish and invertebrates as intermediate hosts have shown low amounts of genetic structuring across distances of more than 5000 km (Paggi *et al.* 1991; Nascetti *et al.* 1993). The authors reported that 94-98 % of genetic variation was attributable to within sample site variation, lack of genetic differentiation between sites being consistent with the hypothesis that seal migration and dissemination of larvae in fish and current-born invertebrates causes high rates of gene flow across large geographic distances.

A comparative analysis of the distribution of mitochondrial diversity within five trichostrongylid species (Blouin 1995) was used to demonstrate the effect of host mobility on population genetic structure. These five species have direct life cycles with obligate sexual reproduction; however four are parasites of agricultural animals (*Ostertagia ostertagi* and *Haemonchus placei* in cattle, *H. contortus* and *Teladorsagia circumcincta* in sheep) and one a parasite of wild deer (*Mazamastrongylus odocoilei*). Nucleotide diversity was measured by sequencing a 600 bp fragment of mitochondrial DNA from 9-11 individuals of each species sampled from five sites across the USA. Genetic differentiation between the populations of parasites of agricultural animals was small, 96-99% of the total genetic diversity being distributed within sampling sites. These data are consistent with an earlier study (Blouin 1992) showing a similar lack of genetic differentiation between *O. ostertagi* populations as measured by total mitochondrial DNA diversity. The population structure of these parasite species is therefore that of a single, interbreeding population (panmixia). Populations of *M. odocoilei* were much more structured, 31% of the genetic variation being distributed between populations. Genetic distances between populations within this species were positively correlated with increases in geographical distance. This is consistent with the hypothesis that infrequent gene flow is occurring between *M. odocoeli* populations due to limited dispersal of deer relative to the large-scale movement of agricultural animals.

Therefore, it appears that host movement is a major determinant of nematode parasite population structure.

The effect of population size on population genetic structure:

Host movement is not the only factor determining population genetic structure. In contrasting *Ascaris suum* and *Ostertagia ostertagi*, it is evident that although both are parasites of livestock (*A. suum* of pigs and *O. ostertagi* of cattle) sharing similar life cycles, there are significant differences in the genetic structure of the species. Studies by Anderson *et al.*(1995) and Nadler *et al.* (1994) have shown significant population structure of *A. suum* infections of pigs in Venezuela and the USA. Genotypes were non-randomly distributed between hosts within and between geographic areas, and departures from Hardy Weinberg expectations in the USA samples provided evidence for non-randomly mating populations (Nadler *et al.* 1995). The main difference between the two species is in *Ne*. The average intensity of infection for *A. suum* in Venezuelan pigs was 4.8 (Anderson *et al.* 1993) compared to 10-100 000 *O. ostertagi* in cattle (Williams *et al.* 1983). Furthermore, the sex ratio in *A. suum* is usually biased, with two times the number of females compared to males, a factor that further reduces *Ne* in this species (Nadler *et al.* 1995). It was concluded that genetic drift may be the cause of population differentiation within this species.

1.1.12 Helminth population genetics and strains

Members of parasite species that vary in biological characters, such as host specificity, infectivity or morphology, are frequently classified as 'strains'. The existence of a parasite strain implies a degree of genetic isolation and thus divergence. However, patterns of genetic differentiation may, or may not, concur with patterns of biological variation. For instance, the cestode parasite *Echinococcus granulosus* has been traditionally recognised as comprising nine separate strains on the basis of differences in characters such as host specificity and development rate. Much effort has been expended on using genetic variation as a means of characterising these strains, but with equivocal results. Significant genetic differentiation has been measured between populations on mainland Australia and Tasmania, supporting the designation of these populations as strains on the basis of biological differences (Thompson & Lymbery 1996). These strains are not monomorphic 'clones', which suggests that cross-fertilisation is occuring among them. In Australia, the parasite has two separate transmission cycles, domestic, involving dogs, and sylvatic, involving dingoes. Despite differences in the onset of egg production, analysis of isoenzyme polymorphisms found no significant genetic differentiation between samples from the

two cycles arguing that the strains are not reproductively separate (Lymbery *et al.* 1990a). Furthermore, there is as much differentiation between the strains of *E. granulosus* as there is between *E. granulosus* and its conspecifics *E. multilocularis*, *E. vogeli* and *E. oligarthrus*. This illustrates the difficulty of defining species, let alone strains, on the basis of genetic criteria alone (Lymbery & Thompson 1996).

The techniques of population genetics have found useful application in elucidating the epidemiology of *Onchocerca volvulus*. Differences between the pathogenicities of rain forest and Savannah 'strains' of the parasite were observed in epidemics of sub-Saharan Africa (Anderson *et al.* 1974). A description of the taxonomy of the vector species *Simulium damnosum* showed that *S. damnosum* was a species complex, raising the possibility that variation in *O. volvulus* strains were due to adaptation to separate vector sibling species. The *O. volvulus* strains were assumed to be ecologically separate when the West African Onchocerciasis Control Project was being developed. To test this assumption, Flockhart *et al.* (Flockhart *et al.* 1986) used enzyme electrophoresis to determine the genetic distance between the forest and Savannah strains. A hierarchical analysis of variance found little inter-patient heterogeneity in infection and no significant differentiation in allele frequencies between villages within countries. Significant heterogeneity was found between separate countries, except for those that contained the forest forms of the parasite. These were genetically more similar to each other than to the Savannah isolates supporting the conclusion that there were ecologically separated strains of the parasite. The genetic distances measured were not great enough to group the strains as separate species suggesting (i) that there is some gene flow between the strains or (ii) that strain formation is a relatively recent event.

1.1.13 The effect of asexual reproduction on population genetic structure

The effects of asexuality on parasite genetic structure have rarely been investigated. The existence of asexual amplification stages in Digenean trematode parasites within intermediate hosts has the potential to superinfect hosts with like-genotypes due to clumping of infective stages. The platyhelminth *Fascioloides magna* parasitises white-tailed deer, hermaphroditic adults mating and releasing eggs from the liver. On hatching into miracidia and infecting a lymnaeid snail, asexual reproduction produces 300-600 infectious metacercaerial clones. Mulvey *et al.* (Mulvey *et al.* 1991b), in a study of *F. magna* in deer populations of South Carolina, found a highly aggregated distribution of parasites amongst hosts. Electrophoretic genotypes,

measured for both parasites and their hosts, showed that deer tend to be infected with populations of adult flukes of the same multilocus genotypes. The likelihood exists that clumping of identical metacercariae on deer browse increases the probability of deer ingesting parasites of like genotypes, thus structuring the populations so that most of the genetic variation is between, rather than within individual deer. Genetic substructuring of the *F. magna* and White-tailed deer populations between geographically separated sampling areas was negligible. However, the total study area was only 30 km across and the long life span of the parasite coupled with wide-ranging male deer would ensure high rates of gene flow between the parasite populations (Mulvey *et al.* 1991b). Lydeard *et al.* (Lydeard *et al.* 1989) sampled *F. magna* from widely separated localities in South Carolina and Tennessee. 18% of variation was attributable to between-State differentiation. This correlated with increasing geographical distance showing that the parasite subpopulations conform to a stepping-stone model of isolation by distance (Kimura 1964).

Several species of parasitic nematode, such as *S. ratti* and *Melidogyne* spp. have parthenogenetic modes of reproduction. Parthenogenesis may result in contrasting population genetic structures compared to equivalent dioecious species. Functionally, parthenogenetic systems can be divided into automictic and apomictic modes of reproduction. In the former, meiosis occurs during the formation of zygotes and a consequential doubling of chromosome number occurs at some point in the life cycle to reconstitute diploidy. In apomictic parthenogenesis, meiosis is absent and oocytes are produced by mitotic division. This parthenogenetic system predominates in taxa of cyclical and obligate parthenogens (Hughes 1989) and I will confine my consideration to the species utilising this mode of reproduction.

For apomictic parthenogens, each reproducing female produces genetically identical clones of itself. These clones multiply as asexual lineages, sometimes known as 'races' or 'strains' (Hughes 1989). The patterns of genetic structure in populations of apomictic parthenogens are clear. In clonal reproduction, the entire genome is effectively linked. The higher the proportion of clonality (parthenogenesis) to genetic recombination (sexual reproduction), the stronger are the associations between alleles at different loci. Therefore, the repeated recovery of the same multilocus genotypes, especially over long distances, is strong evidence for clonal reproduction (Thompson & Lymbery 1990; Tibayrenc *et al.* 1990). In contrast, for sexually reproducing populations there are low linkage associations between alleles at separate loci and recovering the same multilocus genotype is a rare event. This method of detecting

clonal reproduction has found great use in characterising the mating systems of species where amounts of genetic recombination are rare or unknown. For instance, population genetic studies have shown that unicellular pathogens such as *Trypanosoma cruzi* and *Salmonella enterica* are almost exclusively clonal (Maynard-Smith *et al.* 1993), whereas the pathogenic fungus *Coccidioides immitis* has a completely recombining population structure, despite the lack of any known sexual phase in its life-cycle (Burt *et al.* 1996).

A general observation is that parthenogenetic species are, somewhat counter-intuitively, often genetically diverse. In parthenogens, there are several possible sources of diversity. First, within strains, levels of diversity will accumulate over time due to mutation (Muller 1964). Second, the species may have polyphyletic (multiple) origins from a sexual ancestor (Parker 1979), each origin giving rise to a new, genetically distinct, strain. For instance, high levels of genetic diversity are found in a species of parthenogenetic gecko, *Heteronotia binoei*, manifested both as high frequencies of heterozygous enzyme loci within clones, and genetic variation between clones. The clones were characterised by combinations of alleles found in two sexually reproducing diploid lizard populations; these loci were fixed for alternative alleles in the sexual lizard populations, and appeared as heterozygotes in the parthenogenetic species. This appeared to be consistent with the theory that the parthenogenetic species had arisen from multiple hybrid origins, each hybridisation event sampling the gene pools of the 'parental' species and creating new 'daughter' multilocus genotype (Moritz 1989). A contrasting situation is demonstrated by a species of trematode parasite, *Paragonimus westermani*. This species occurs as both dioecious diploid, and parthenogenetic triploid, forms. Although the diploid form has been found to be genetically variable, triploid parasites have been found to have fixed heterozygosities at five enzyme loci, and fixed homozygosities at four others. This suggests that the triploid form has arisen from a single hybrid event, and therefore has a monophyletic origin that reflects the genetic composition of the original mating event (Agatsuma & Habe 1985b).

The population structure of parthenogenetic nematodes has been little studied. *Melidogyne*, a genus of root knot nematodes, contains species that reproduce by obligate parthenogenesis (*M. arenaria, M. incognita, M. javanica*) and mictic reproduction (*M. hapla and M. naasi*). Studies by Dalmasso and Berge (Dalmasso & Berge 1983) on the isoenzyme profiles of the taxa at six polymorphic loci showed clear differences between each species. Restriction digestion of total genomic DNA

revealing restriction fragment length polymorphisms in repetitive DNA sequences differentiated strains within the parthenogenetic species of *Melidogyne* (Curran *et al.* 1986). Application of this technique revealed no intrapopulation genetic differentiation within species of sexually reproducing *Melidogyne* (Curran *et al.* 1985), therefore implicating a parthenogenetic mode of reproduction as a mechanism of maintaining independently evolving strains. Highly divergent groups of mtDNA haplotypes were found between each of the parthenogenetic species of *Melidogyne*, confirming the results of Dalmasso and Berge. Moreover, *M. arenaria* contained lineages that were as distinct from each other as they were from *M. javanica* (Hugall *et al.* 1994). This has caused confusion of the taxonomic status of these species and illustrates the difficulty of applying species concepts to asexual lineages. A modern interpretation of the species status of obligate parthenogens is that asexual lineages constitute separate evolutionary entities and therefore should be recognised as separate species (Frost & Wright 1989).

1.1.14 Facultative parthenogens

In *Strongyloides ratti* the presence of a dioecious generation will, to an extent, cancel out the effects of clonal reproduction. The longevity of genetically identical lineages will depend on the frequency of sexual reproduction within a population. Recombination will maintain genotype frequencies close to Hardy-Weinberg equilibrium if sexual reproduction is frequent. However, under conditions where sexual reproduction is rare, genotype frequencies will be distorted from random-mating expectations and persistent multilocus genotypes would be expected. The effects of infrequent sexual reproduction are well demonstrated by the cyclical parthenogen, *Daphnia pulex*.

The cladoceran *D. pulex* shows varying amounts of sexual reproduction. Temperate populations alternate between sexual and parthenogenetic reproduction (cyclic parthenogenesis) whereas Arctic populations reproduce exclusively by obligate parthenogenesis. Arctic populations of *Daphnia* have been shown to have rich genetic diversity, this existing as many genetically distinct clones. Weider *et al.* (Weider & Hobaek 1994) found 49 allozyme clones in a sample of 3357 *D. pulex* from the Svalbard archipelago. Over half of the sample consisted of two predominating closely related clones distributed over a range of 700km. It seems therefore that there is widespread dispersal of the species *via* factors such as waterbirds or wind. The distribution of genotypes was heterogeneous with only 1-2 clones found per pond. This may be explained by invoking rare founder effects as the main process by which

new populations are established. An alternative hypothesis was that the genetic heterogeneity was a consequence of competitive exclusion between competing genotypes. That fitness differences exist between clones has been suggested by the observation that different clones are associated with different biotic environments *i.e.* within the study clone G2 dominated high salinity environments while clone G1 was most often associated with low salinity pools. This is evidence for what is commonly known as the 'frozen niche hypothesis' (Vrijenhoek, 1979). Here, clonal genotypes are 'frozen' as specialists in the niche for which they are best adapted, which in this case is degree of pool salinity. The existence of multiple habitats may therefore be a mechanism that maintains genetic diversity in clonal populations (Weider & Hobaek 1994).

Studies on temperate cyclical parthenogenetic populations of *D. pulex* show a range of population structures, depending on the frequency of sexual reproduction. In small ponds that dry up, sexual reproduction is frequent with the result that genotypes are in Hardy-Weinberg equilibrium and clonal diversity is high, due to the mixing of parental genotypes. In permanent ponds, sexual reproduction is infrequent and genotypes deviate from expected Hardy-Weinberg proportions. The ponds also show a paucity of clones, as in the Arctic populations of *D. pulex.* That there is often an excess of heterozygotes within these stable populations (Hebert 1987b) suggests that heterozygosity may confer greater fitness on inbred lines of *D. pulex.* This process, known as heterosis, has been demonstrated by crossing clones of *D. magna* from England and Canada, the hybrid progeny showing greater heterozygosity as well as higher rates of population increase and resistance to changes in temperature and salinity (Hebert *et al.* 1982).

1.1.15 Intraspecific genetic variation in *Strongyloides spp.*

From the above discussion, several themes have emerged that are applicable to my study on *S. ratti.* First, parasitic nematodes show no evidence of genetic monomorphism relative to free-living species. Rather, levels of genetic diversity appear to be as high, or higher than expected. The mechanisms which maintain this diversity are subjects of debate. Second, the distribution of genetic variation is dependent to a large extent on host movement. Thus, the genetic structure of species of *Strongyloides* will be expected to depend to an extent on the type of host infected and its behaviour. Third, parthenogenesis will be expected to have an effect on population genetic structure by amplifying genotypic combinations when sexual reproduction is

infrequent and biasing genotypic frequencies from random mating expectations. This final section examines what is known about the genetic variation of *Strongyloides* spp.

That there is genetic variation within *Strongyloides* sp. was first indicated by Galliard (1938) who carried out artificial infections of *S. stercoralis* in dogs. It was shown that there were differences between Vietnamese, Indian and West African isolates of the parasite in the dynamics of infection. This was interpreted as being due to geographical races of *S. stercoralis* that varied in their virulence, and perhaps in their genetic composition. Viney and Ashford (Viney & Ashford 1990) attempted a classification of *Strongyloides* species using isoenzyme methods. Isoenzyme profiles were determined for 82 isolates of *Strongyloides*: 26 collected from Papua New Guinea (PNG) humans, 21 from PNG domestic animals and 14 from African non-human primates. The analysis showed that *Strongyloides* infecting humans in PNG was a sub-species of *Strongyloides fuelleborni,* previously thought to infect only primates in Africa, and that a separate species infected domestic animals. A significant amount of inter-isolate genetic heterogeneity in *S. fuelleborni* was also observed. The extent of the genetic heterogeneity found in parasites taken from a single host population was, in some cases, as great as the heterogeneity found between isolates that were from putatively separate populations. However, such heterogeneity was not found in a study on isolates of *S. ratti* (Viney *et al.* 1992). Seven isolates from America, Japan and the UK were cloned (populations derived from a single iL3) and examined for isoenzyme polymorphisms, but none was found at eight enzyme loci. This was despite measurements on the frequency of heterogonic development within the isolates, which showed extensive variation between lines. This variation was reproducible over a period of months, confirming that its origin was genetic.

1.2 Summary

The population genetic structure of parasitic nematodes is understudied. Only a few species have been examined. Within these studies, none have rigorously sampled from (i) many hosts and (ii) many sites. This has limited the extent to which hierarchical descriptions of population genetic structure may made. In still fewer studies have the effects of parasite asexuality on the amounts and distribution of genetic variation been considered. This thesis aims to characterise the population genetic structure of a parasitic nematode, *Strongyloides ratti,* by the use of a thorough sampling strategy and molecular genetic analysis. The life-cycle of this species will enable the effect of sexual reproduction on population genetic structure to be examined and in this way to bridge the gap between studies on asexual and obligately sexual nematode parasite species.

Chapter 2 describes the isolation of suitable genetic markers from the genome of *S. ratti* by the application of molecular genetic techniques.

Chapter 3 summarises the methods used to sample *S. ratti* from geographically separated populations of brown rats (*Rattus norvegicus*). The distribution of infection within one such population is described in detail.

Chapter 4 analyses the genetic structure of the sampled populations by a combination of statistical methods.

Chapter 5 discusses the results presented and summarises the contributions they make to understanding the biology of helminth parasites. Recommendations on the direction of future studies are made in light of the results presented herein.

Chapter two

Identification and characterisation of polymorphic loci in *Strongyloides ratti*

2.1 Introduction

This thesis describes the genetic structure of *Strongyloides ratti* by examining the distribution of polymorphic loci among *S. ratti* infections of rats. As was shown in the previous chapter a population genetic approach to describing the structure of macroparasite populations, by analysis of the distribution of genetic variation, has been used by several authors, notably Anderson (Anderson *et al.* 1995) Blouin (Blouin *et al.* 1995) and Nadler (Nadler *et al.* 1995). The technical developments that have made this possible are modern molecular genetic methods, specifically the polymerase chain reaction (PCR) and DNA sequencing. These techniques have allowed close scrutiny of genomes of organisms which has opened up an almost infinite pool of potentially informative genetic information. Exploitation of this genetic diversity has allowed previously intractable questions concerning the genetic structure of parasite populations to be addressed.

Many methods have been developed to characterise molecular genetic variation (Avise 1994), however, there is a great deal of difference in the quality and type of information that is obtained. The genetic markers used in this study had to satisfy strong criteria to be appropriate for use. They had to be (1) polymorphic within British *S. ratti* samples, (2) usable on a fraction of the DNA extracted from a single parasite, (3) rapidly and reproducibly scorable, (4) undergo normal Mendelian segregation and (5) be selectively neutral. This chapter will review the main methods now available for characterising genetic variation. It will then go on to describe which methods were selected and exploited to identify polymorphic genetic loci in *S. ratti* .

2.1.1 Molecular genetic variation

There are two genetic sources that can be used to analyse variation within metazoan eukaryotes. These are (1) mitochondrial DNA sequences and (2) nuclear DNA sequences.

2.1.2 Mitochondrial DNA

Mitochondrial DNA (mtDNA) exists as multiple cytoplasmic copies in cells. In nematodes the mitochondrial genome is a circular molecule of about 13-14 kb in length (13 794 bp in *C. elegans*, 14 284 in *A. suum* (Okimoto *et al.* 1992). mtDNA is often considered to provide ideal genetic markers for the following reasons;

(1) MtDNA has been shown to be principally maternally inherited, and this has been illustrated in the nematode species *A. suum* (Anderson 1994). As the molecules do not normally undergo recombination each genotype exists as a matrinlinearly inherited clone or haplotype. Severe bottlenecks occur during oogenesis resulting in populations of the molecule being principally of one haplotype within an individual (Takahata 1985). Mitochondrial genes are thus more sensitive than nuclear genes as markers for investigating population structure for the following reasons. Factors important in determining the amount of structure between two populations are the effective population sizes (*Ne*) and the rate of migration between the populations (*m*). Although *m* is the same for nuclear and mitochondrial DNA, *Ne* for nuclear genes is 4 times that of mtDNA within sexually reproducing organisms. This is because at each nuclear locus both parents potentially contribute two alleles of each nuclear gene to the next generation rather than the female parent contributing a single maternal mitochondrial haplotype. The resulting difference in effective population sizes means that mtDNA haplotype frequencies will homogenise between geographically separated populations less rapidly than nuclear genes (Anderson *et al.* 1995).

(2) It has been shown that mutations accumulate 5-10 times faster in mtDNA relative to nuclear DNA (Hillis & Moritz 1990). This is because mtDNA polymerase has no proof-reading ability. As a consequence, extensive intraspecific variation of mtDNA may build up. mtDNA therefore tends to be more informative in providing population genetic markers than is nuclear DNA, when variation per nucleotide analysed is considered (Thomaz *et al.* 1996; Blouin *et al.* 1995).

(3) mtDNA is generally considered to be selectively neutral. This point is based on two arguments. First, relatively high mutation rates of the molecule caused by non-synonymous substitutions are evidence for relaxed constraints on the proteins coded within. Second, there is accordance with rates of evolution predicted by neutral theory (Kimura, 1983). However, that rates of nucleotide divergence vary between different mtDNA genes suggest that accordance with strict neutral expectations may not be strictly valid (Williams *et al.* 1995).

(4) Ease of characterisation of mtDNA variation. MtDNA haplotypes differ in DNA sequence. These sequence differences are a consequence of cumulative point mutations that accrue within individual mtDNA molecules. This variation can be assayed by restriction enzyme digestion where the whole, or part, of the molecule is subjected to digestions by a panel of restriction endonucleases. The gain or loss of enzyme recognition sites, resulting in changes in the profiles of the enzyme-restricted DNA between individuals provides the mtDNA haplotype, and a raw estimate of sequence divergence. Considerable sequence divergence has been found in mtDNA in surveys of nematode parasite populations (Anderson *et al.* 1995; Blouin *et al.* 1995).

Although mtDNA is theoretically capable of providing excellent markers, there are some drawbacks associated with its use. For example, it has become apparent with the advent of PCR amplification techniques that mtDNA sequences can undergo insertion into the nuclear genome, forming pseudogenes. This has been shown to occur within the genomes of old world monkeys and hominoids (Collura & Stewart 1995). These nuclear sequences are amplified together with the mtDNA sequences confounding population genetic analyses. It has not been determined to what extent these insertion sequences exist within and between the genomes of different organisms, as no thorough surveys have been undertaken.

2.1.3 Nuclear DNA variation

Genetic variation within the nuclear genome exists as discrete differences in the form of insertions/deletions of DNA sequences or single base mutations; this is known as *single locus* genetic variation. Physically separated sequences can be classified together on the basis of similarity and the genetic variation within them analysed together; this is known as *multilocus* genetic variation.

2.1.4 Multilocus genetic variation

Variable numbers of tandem repeats (VNTRs):

Jeffreys *et al.* (Jeffreys *et al.* 1985) described a class of loci characterised by sequences consisting of a core motif 10-15 bp long arranged in tandem arrays. These arrays were found to occur at multiple sites throughout the genome. These 'minisatellites' show variable numbers of tandem repeats (VNTRs) between individuals due to unequal crossing over at each locus during meiosis (Jarman & Wells 1989). DNA 'fingerprinting' involves probing Southern blots with the VNTR motif resulting in a highly variable banding profile which distinguishes most individuals

from each another. Problems with this technique stem mainly from the need to have large quantities of DNA, making the genetic analysis of small parasites impossible. However, the approach of Viney *et al.* (Viney *et al.* 1993) has shown how single *S. ratti* parasites can be cloned and sufficient DNA extracted from the offspring to perform a genetic analysis using a VNTR probe. A further problem intrinsic in the use of such multilocus probes in population genetics is that although inheritance of alleles at each locus is Mendelian, allelic identity cannot be established between bands. This makes it very difficult to establish the amounts of heterozygosity within an individual. For these reasons multilocus VNTRs were not considered appropriate for use in this study.

Randomly amplified polymorphic DNA-PCR (RAPD-PCR):

RAPD-PCR is a powerful and fast method for detecting polymorphisms in species that have relatively uncharacterised genomes. This technique amplifies fragments of genomic DNA using single PCR primers of arbitrary sequence. Such primers (usually 10 base pairs in length) are used at low annealing temperatures in PCR reactions, the low stringency of the PCR resulting in the primer annealing to several sites within the genome. If primers anneal to sequences that are relatively close to each other and in the correct orientation then amplification of discrete bands will result. Often sufficient sites exist within a genome that simultaneous amplification of a number of loci occurs, resulting in a multi-band profile, or RAPD pattern (Welsh & McClelland 1990; Williams *et al.* 1990).

The majority of amplified fragments will be expected to be similar among individuals but for some, polymorphisms occur. These are due to sequence variation in primer sites preventing primer annealing, or insertions/deletions between primer sites resulting in length variations of the amplified fragment. This variation is detected as size differences between the pattern of amplified bands from separate individuals.

RAPD analysis is performed by comparing the number of unique and shared bands between individuals, and is therefore a phenotypic rather than a genotypic description of the organism. An important limitation of the technique is that an allelic variant of a given sequence resulting from mutation within the primer site will result in the disappearance of a fragment only if it is in the homozygous state. Therefore, heterozygotes ("band, no band") appear the same as homozygotes ("band,band"). A second limitation of the technique is that it is technically demanding. Due to the low annealing temperature of the PCRs slight fluctuations in the initial quantities and

qualities of template DNA can result in large differences to the final RAPD pattern. This results in low experimental reproducibility (Meunier & Grimont 1993) (Kernodle *et al.* 1993). Due to these considerations, RAPD-PCR was not considered a suitable technique for finding genetic markers in this study.

2.1.5 Single locus molecular variation

Protein electrophoresis:

Characterising polymorphisms of protein-coding loci by variations in the electrophoretic mobility of their protein products has long been used as a technique for detecting genetic variation for use in population genetic studies (Selander 1976). However, the levels of polymorphism observed are generally low due to the selective constraints on mutations within protein-coding loci and the redundancy of the genetic code. A study by Viney *et al.* (Viney *et al.* 1992) on seven isolates of *S. ratti* from different geographical regions found no genetic variation at eight enzyme loci. Following these results, and the fact that large quantities of parasite material are required to characterise isoenzymes (Viney & Ashford 1990), this technique was not considered further in the context of this work.

PCR based methods for detecting small genetic polymorphisms:

The development of PCR techniques has opened up a number of methods for detecting DNA sequence variation. The finest resolution is provided by the direct sequencing of PCR products. However, in practice this may be too slow to detect mutations within large DNA fragments. In response to this, several methods have been developed for rapidly scanning sections of the genome for variation (McPherson *et al.* 1995). These methods rely on PCR products being generated by either i) designing primers to fragments fully or partially sequenced from a genomic or cDNA library or ii) specifically isolating known polymorphic sequences (such as microsatellites, see below) by the use of specific hybridisation probes, and designing PCR primers to these sequences once found. Once a fragment can be amplified by PCR, it may be surveyed for sequence variation;

PCR Restriction Fragment Length Polymorphisms (RFLPs): This method relies on digesting amplified PCR products with a restriction endonuclease. These enzymes cut at specific sequences of bases. If such a sequence is present within a PCR product then digestion of the fragment occurs, resulting in a specific pattern of bands following electrophoresis (Maniatis *et al.* 1982). Analysis of the number and lengths of bands resulting from digestion by several enzymes enables a restriction map to be made of the

PCR fragment, identifying the position of restriction sites. Sequence differences between individuals may result in the gain or loss of restriction sites. These polymorphisms are easily scored by comparing the banding patterns from each individual and heterozygotes can be distinguished from homozygotes.

Large numbers of individuals can be rapidly surveyed for genetic variation by this technique once the initial effort in identifying polymorphic restriction sites has been accomplished. As these loci are randomly selected from the genome they are often non-coding and therefore likely to be selectively neutral (Karl *et al.* 1992). This technique has been used to provide clear and useful genetic markers for the nematodes *Haemonchus contortus* (Beech *et al.* 1994) and *Ascaris suum* (Anderson 1995). In *S. ratti*, the existence of an *Mnl* I RFLP within the Actin IV gene (Viney 1994) shows that this method can be used to isolate polymorphic genetic markers in this organism.

Microsatellites: Microsatellites, or simple sequence repeats, are analogous to VNTRs in that they consist of families of sequence motifs repeated in arrays and dispersed throughout the genome. The difference between the two classes of markers lies in the size and number of repeat sequence motifs. Microsatellites occur as 1-6 bp nucleotide motifs, tandemly repeated 5-100 times and flanked by unique sequences, whereas VNTRs tend to be much longer due to the larger size of the repeat motif and the greater number of times that it is repeated. Both types of marker can be hypervariable for the number of repeat motifs in the array.

The genomes of all phyla examined, including unicellular eukaryotes, prokaryotes and organelle DNA (Field & Wills 1996) have been shown to contain microsatellites. However, the numbers of each type of microsatellite present varies dramatically from one organism to another. Field *et al.* (1996) showed that more AT rich microsatellites occur in organisms with reduced genome complexity. Most microsatellites are thought to be non-coding and therefore selectively neutral (Stallings *et al.* 1991) although roles for microsatellites in gene regulation have been shown in prokaryotes (Moxon *et al.* 1994). A number of human diseases have been shown to be a consequence of increases in the number of triplet repeats with particular microsatellite loci, for example fragile-X syndrome (Hirst *et al.* 1994), spinocerebellar ataxia (Orr *et al.* 1993) and Huntington's chorea (The Huntington's disease collaborative research group, 1993).

Microsatellites are of great value as genetic markers due to their high mutation rates relative to the rest of the genome. This increased mutation rate is thought to occur as a consequence of an elongation by slippage mechanism (Weber 1990), a hypothesis that is supported by *in vitro* experiments (Schlotterer 1992) and observations on the distributions of mutations within natural populations (Weber & Wong 1993). The process results in the addition or subtraction of one or two repeat motifs to the array by repair enzymes. As a consequence, a diversity of length variants for each microsatellite locus accumulates within the species. Due to the relatively short length of microsatellites and their unique flanking sequences, alleles at a particular locus are assayed by amplifying the locus from individuals by PCR and electrophoresing the products on polyacrylamide gels capable of resolving molecules varying by single bases. The products are visualized by incorporation of radioactively or fluorescently labeled nucleotides or primers during the PCR reaction. Alleles at each locus are scored by size variation, this being due to variation in the number of microsatellite repeat units present. The microsatellite alleles are unambiguous and co-dominant, heterozygotes being observed in diploid genomes as two amplification products separated by a multiple of the length of the repeat motif. Reproducibility is generally high and the technique uses small quantities of DNA making it ideal for use in studies on organisms with limited amounts of DNA available.

If microsatellites from a given organism have not previously been sequenced and deposited in genomic DNA databases then they have to be isolated. This is done by screening genomic libraries for microsatellite sequences, which is a time consuming task. One approach that can be used is to use PCR-primers to homologous microsatellite-containing loci from putatively closely related species (within which microsatellites have already been identified). Amplification from homologous loci has been shown to work between species of whale (Schlotterer *et al.* 1989) and fish (Ciro & Hewitt 1996) despite divergence from common ancestry of up to 470 million years.

Due to our interest in considering the microgeographic population structure of *S. ratti*, this project required markers that had high levels of intraspecific variation and would show fine levels of genetic substructuring. mtDNA is an ideal marker for this type of study as it has shown by a variety of studies that i) ample genetic variation exists within nematode populations and ii) this can be surveyed using PCR based techniques. I therefore aimed to survey *S. ratti* mtDNA for genetic variation using RFLP and sequencing methods. However, to be able to describe levels of inbreeding within *S. ratti* samples by observations on the amount of heterozygosity it was also

necessary to include nuclear markers. To this end I attempted to isolate microsatellites and PCR-RFLPs from the genome of *S. ratti*. The possibility of using homologous microsatellite loci isolated by the *C. elegans* genome project with *S. ratti* was also explored. The remainder or this chapter describes the isolation and characterisation of loci from the mtDNA and nuclear genome of *S. ratti*.

2.2 Materials and methods

2.2.1 *S. ratti* lines used

To avoid ambiguity, several terms are used specifically when describing *S. ratti* :

a) Isolate- a group of parasites collected from a wild-caught animal and kept as cryopreserved material.

b) Line- A population of parasites derived from an isolate and maintained by serial passage *in vivo*.

c) Isofemale line- A population of parasites derived from a single parasitic female and maintained by serial passage (Viney 1994).

d) Clone- The term 'clone', despite having been used in the literature (Viney & Ashford 1990), is misleading with respect to *S. ratti* as it implies genetic homogeneity. It has been shown that *S. ratti* 'cloned lines' undergo genetic recombination (Viney 1994) if they undergo the heterogonic life cycle. Such lines should therefore be referred to as isofemale lines.

Isolates are presumed to be genetically heterogeneous and to contain mixtures of wild-type parasites. Lines will preserve some of this genetic heterogeneity. However, genetic variation will be expected to be reduced within lines as a consequence of drift when small sub-populations are used to passage the parasite for laboratory maintenance. Isofemale lines, despite being derived from a single individual still contain genetic variation due to the diploid nature of the parasite and the occurrence of genetic recombination following mating between individual males and females in culture.

The lines used in this study are shown in Table 2.1.

Line	*Notes*
Isofemale line 5:	Isofemale line made from an isolate collected in Philadelphia in 1960 (Viney *et al.* 1992)
Line 5 Heterogonic:	Line derived from isofemale line 5 and selected for heterogonic development (Viney 1996)
Isofemale line 132:	Isofemale line made from rat collected in Kashogima, Japan (Viney *et al.* 1992) (Viney 1996)
Line 132 Heterogonic:	Line derived from isofemale line 132 and selected for heterogonic development (Viney 1996).
Isofemale line 54:	Isofemale line made from rat collected in Kashogima, Japan (Viney *et al.* 1992)
Isofemale line 68:	Isofemale line made from a genetic cross between isofemale lines 5 and 54 (Viney *et al.* 1993)
Line 43:	Made from isolate 43, derived from parasites collected from a dead rat collected in Leith (Scotland) by M. Viney, 2.6.89.
Line 29:	Made from isolate 29, derived from a captive rat collected in Wiltshire (England) by M.A.A.F. 2.6.89.
Line 32:	Made from isolate 32, derived from rat-faeces recovered in Craigmillar (Edinburgh, Scotland) by M. Viney, 10.11.89
Line B352	Made from isolate B352, derived from a rat collected in Berks.hire (England) by M.A.A.F. 5.5.95
Line B362	Made from isolate B362, derived from a rat collected in Berks.hire (England) by M.A.A.F. 5.5.95
G2924	Made from isolate G2924, derived from a dead rat collected in Germany by H. Pelz, March '95
G2932	Made from isolate G2924, derived from a dead rat collected in Germany by H. Pelz, March '95

Table 2.1 Origins of *S. ratti* used in the search for genetic markers.

2.2.2 Cryopreservation of *S. ratti*

Infective third stage larvae (iL3s, Figure 1.1) of *S. ratti* were stored in liquid nitrogen. 20 000 iL3s were mixed with 1 ml of cryopreservation fluid (10% w/v

Dextran (Sigma), 10% v/v dimethyl sulphoxide) and agitated at room temperature for 1 hour. The larvae were placed in the vapour phase of liquid nitrogen until completely frozen, then transferred to liquid nitrogen containers for long term storage.

Cryopreserved iL3s were recovered by adding 1 ml of RPMI medium (Sigma) and incubated at 37 °C until thawed. This solution was made up to 10 ml with RPMI and the parasites sedimented by centrifugation. Re-suspension and centrifugation were repeated three times, followed by a final re-suspension in 200 μl of RPMI (Nolan *et al.* 1988). The iL3s were subsequently used to infect a rat (section 2.2.3)

2.2.3 *In vivo* maintenance of S. ratti

Parasites were maintained by serial passage in 3-4 week old Wistar rats (B&K Universal, Hull). Rats were inoculated subcutaneously with 500 iL3s suspended in 200 μl RPMI. Such infections became patent on day 5 post infection (pi) and substantial numbers of *S. ratti* larvae are passed in faeces for the following 2-3 weeks. The numbers of *S. ratti* in the faeces declined over this period and the infection generally resolved 4-5 weeks pi. iL3s collected over the course of the infection were used to re-infect rats every 4 weeks in order to maintain the line.

2.2.4 Culture of *S. ratti*

Faeces were collected from infected rats which had been kept overnight in cages with grid floors. This allowed faecal pellets to drop onto dampened paper liners from which faecal collections were made. 3-4 faecal pellets were rinsed with distilled water and placed on a watch glass. This was dampened and placed in a covered glass petri dish with a small amount of distilled water in the bottom. The petri dishes were incubated at 19 °C for 3 days and the cultures examined using a stereo-microscope for the presence of *S. ratti*. Heterogonic adult males and females were generally found around the rim of the watch glass in the water surrounding the faecal pellets. iL3s were found in the water in the base of the petri-dishes due to their migration away from the faecal material. These iL3s were removed and suspended in 10 ml of distilled water and washed twice by sedimentation with centrifugation, followed by re-suspension in 10 ml distilled water. These iL3s were either stored at -20°C for further analysis or used to infect rats.

2.2.5 Isofemale lines

Isofemale lines were initiated and maintained by M. Viney using the method described in Viney *et al.* (Viney *et al.* 1992).

2.2.6 Extraction of Genomic DNA

100 000 iL3s were physically disrupted by snap-freezing in liquid nitrogen and grinding in polypropylene micro-homogenisers (Biomedix), repeated three times. The homogenate was added to an equal volume of TNESST (10 mM Tris/HCl pH 7.4, 60 mM NaCl, 10 mM EDTA pH 8.0, 0.15 mM spermidine, 0.15 mM spermine, 0.5% v/v Triton X-100). SDS (Sigma), sodium deoxycholate (Sigma), Proteinase K (Boehringer Mannheim), Collagenase (Sigma) and Trypsin (Sigma) were added to final concentrations of 0.3% w/v, 0.15% w/v and 0.1 ng/ml respectively. This solution was incubated for four hours at 37 °C with gentle agitation and then successively extracted with equal volumes of phenol, 25:24:1 phenol/ chloroform/ isoamyl alcohol and 24:1 chloroform/isoamyl alcohol. An equal volume of 5M LiCl was added to the resulting aqueous phase, held at -20 °C for 30 minutes and then centrifuged at 10 000 *g*. The supernatant was removed to a fresh tube and 2.5 volumes of ethanol added. This was held at -20 °C for 12 hours to precipitate the DNA and then centrifuged at 10 000 *g* for 30 minutes to pellet the DNA. After removal of the supernatant, the pellet was washed twice in 70% v/v ethanol, vacuum dried and resuspended in TE (10mM Tris HCL, 1mM EDTA, pH 7.6). An aliquot of the DNA preparation was electrophoresed on a 2 % w/v agarose gel and visualized to acetain the quality of the DNA. The quantity of DNA present was quantified spectrophotometerically (Genequant, Pharmacia).

2.2.7 Extraction of DNA from a single-parasite

Single *S. ratti* larvae or free-living adults were transferred from faecal culture to 2 ml of distilled water to remove adhering debris. The worms were transferred into an microfuge tube in 5 μl of distilled water, together with 0.5 μl 10 X PCR buffer (Dynazyme) and 0.5 μl Proteinase K (20 mg ml^{-1})(Boehringer Mannheim), then overlaid with a drop of mineral oil (Sigma). After centrifugation at 10 000 *g* for 3 minutes the sample was held at -70 °C for 30 minutes, 55 °C for 90 minutes and 95 °C for 30 minutes in a thermocycler (Biometra). The DNA preparations were stored at -20 °C.

2.2.8 Isolation of microsatellite loci

2.2.8a Construction of an *S. ratti* size-selected genomic library

1 μg of genomic DNA from line 43 (Table 2.1) was digested to completion with 10 units of *Alu* I and *Hae* III and separated on a 1.5% w/v low melting point agarose gel. The size fraction between 300-500 bp was excised and purified with a DNA cleanup column (Promega). The size selected DNA was blunt-end ligated into the vector pGEM 4Z (Promega). The vector had previously been prepared by digestion with *Hind*II, then dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs). Vector: Insert molar ratios of 1:1, 1:3 and 3:1 were ligated overnight at 4 °C with 1 unit of T4 DNA ligase (Promega). Control reactions were set up to monitor the efficiency of the ligation reaction of vector+ligase, vector only and insert only. Competent JM109 cells were then transformed with the ligation reactions following the manufacturers' protocol (Promega technical bulletin No. 095). To monitor the efficiency of the transformation, 0.1ng of plasmid was used as a control transformation of the competent cells. The transformed cells were subsequently plated on LB plates containing ampicillin (Sigma), IPTG (Sigma) and X-Gal (Sigma) at final concentrations of 100 μg/ml, 0.5 mM and 40 μg/ml respectively. The plates were then grown up overnight at 37 °C and blue/white selection used to distinguish between recombinant and non-recombinant colonies. To confirm that the size selection procedure had worked, 20 white colonies were picked at random from the plates, suspended in 5 μl distilled water and used as templates in a PCR reaction with M13 forward and -21 primers (Promega). All colonies were confirmed as having inserts between 300 and 500 bp in length.

2.2.8b Identification of recombinant colonies containing microsatellites

Ten agar plates were grown with an average density of 500 colonies per plate. Hybond-N (Amersham) nylon membranes were laid on the plates and pin-pricks made through both membrane and agar so as to allow subsequent re-orientation of the membrane to the plate. Replica filters were also lifted as a control for false positives. The membranes were incubated at 37 °C on LB Ampicillin plates until the colonies were 0.5 - 1 mm in diameter and then denatured by resting, colony side up, on Whatman 3MM paper soaked in denaturing solution (Appendix 1) for 7 minutes. The membranes were neutralised in neutralising solution (Appendix 1) and washed in 2 x SSC (Appendix 1), then air-dried and fixed by baking at 80 °C for 2 hrs. After the colony lifts, plates were incubated at 37 °C to re-grow the colonies in order that positive colonies could later be identified and picked. The plates were stored at 4 °C.

Microsatellite probes were prepared by end-labeling 1 μg of either of the following oligonucleotides; $(AC)_{15}$, $(GA)_{15}$ or $(CAG)_{15}$ (Applied Biosystems). The probes were end-labeled with γ-^{32}P dATP (5000 Ci mmol^{-1}, Amersham) using T4 Polynucleotide Kinase (Promega) following the manufacturers' protocol. Non-incorporated nucleotides were removed from the probes using a chromaspin-10 column (Clontech) and the labeled oligonucleotide suspended in 50 μl of TE.

Colony lift membranes were pre-hybridised for 4 hrs. in pre-hybridisation solution (appendix 1) at 60 °C. The labelled oligonucleotide was added to the pre-hybridisation solution to a final concentration of 10 ng ml^{-1} and hybridised overnight. Membranes were washed twice in pre-heated 2X SSPE, 1% SDS for 30 mins at 60 °C, and exposed to Kodak XAR film in autoradiography cassettes with intensifying screens at -70 °C.

Autoradiographs of master and replica filters were aligned to detect true from false positives. Putative positive colonies were picked from the agar plates by keying the plate to the autoradiograph and picking the colony into 100 μl of LB medium. This was then streaked onto a fresh LB Ampicillin plate for secondary screening using the same method as above. Colonies identified as positive from the secondary screen were resuspended into 100 μl of LB Ampicillin medium. 5 μl of this suspension was used to inoculate 8 mls of LB-amp medium and grown at 37 °C in a shaking incubator. 0.85 ml of this suspension was mixed with 0.15 ml of glycerol and stored at -70 °C to provide a permanent source of the positive colony.

2.2.8c Sequencing of plasmid inserts

Plasmid DNA from the positive colonies was isolated from 4 mls of LB Ampicillin medium using a Wizard mini-prep (Promega) and eluted in 50 μl TE. 30 μl of double-stranded DNA was denatured by incubation with 5 μl 4N NaOH, 20 μl 1 mM EDTA for 30 minutes at 37 °C. The plasmid DNA was precipitated by addition of 30 μl of 3M sodium acetate and 200 μl ethanol, held at -70 °C for 30 minutes and centrifuged at 10 000 *g* for 15 minutes. The pellet was then washed with 70% ethanol, vacuum dried and resuspended in 13 μl TE.

Each plasmid was then manually sequenced using the Sequenase version 2.0 sequencing kit (United States Biochemical Corp.) using the dideoxy chain termination method (Sanger *et al.* 1977). 6 μl aliquots of each plasmid were sequenced with the

M13 -21 and M13 reverse primers following the manufacturers' instructions. Sequencing reactions were denatured at 95 °C for 3 minutes and electrophoresed on 6% denaturing polyacrylamide gels made using a Sequi-Gen Nucleic Acid Sequencing Cell (BioRad). Gels were dried onto Whatman 3MM paper and exposed to Kodak Bio-Max film overnight. The sequence obtained was electronically stored and analysed using the University of Wisconsin Genetics Computer Group DNA sequence analysis software (GCG, version 8).

2.2.9 PCR amplification of microsatellite loci

To amplify the microsatellite loci isolated from *S. ratti* two types of PCR were used, (1) non-radioactive and (2) radioactive reactions. Both types of PCR were performed initially using standard conditions, and the reaction conditions varied subsequently to optimise the reaction.

2.2.9a Standard non-radioactive PCR

Non-radioactive PCR reactions were performed in a 50 μl volume in 0.5 ml microfuge tubes. All reactions were set up on ice. Deoxynucleotide triphosphates (dNTPs)(Boehringer Mannheim) were used at a final concentration of 75 μM, and primers at a final concentration of 100 nM. Template DNA was either 1 μl of 1/100 dilution of a genomic DNA preparation or 1 μl of a single worm preparation. PCR buffer (Cambio) was used at 1X concentration, *Taq* DNA (Boehringer Mannheim) at a final concentration of 0.5 units ml^{-1} and $MgCl_2$ at a final concentration of 1.5 mM. The reaction mixture was overlaid with mineral oil (Sigma), centrifuged briefly to settle the contents and cycled on a TRIO-thermoblock (Biometra) for 40 cycles of 95 °C for 1 min, 50 °C for 1 min. and 70 °C for 1 min. 10 μl of the PCR reactions were electrophoresed on a 2% w/v agarose gel containing ethidium bromide (0.5 $\mu g\ ml^{-1}$).

2.2.9b Standard radioactive PCR

Radioactive PCRs were performed as above, except that they were performed in a volume of 10 μl with dATP used at 1/10 normal concentration, supplemented with 1 μCi of α-^{32}P dATP (3000 Ci $mmol^{-1}$, Amersham). The PCR reactions were electrophoresed on a 6 % denaturing sequencing gel, transferred to 3MM paper, dried and exposed to XAR film (Kodak) overnight in autoradiography cassettes with intensifying screens at -70 °C.

To optimise amplification of both types of basic PCR reaction, conditions were varied as follows;

1) Primer design

Primers were designed to amplify target sequences containing microsatellites with several criteria in mind: (i) the DNA fragment to be amplified would be 100-200 bp in length to allow resolution on a sequencing gel; (ii) the primers would have low self and between-primer complementarity to minimise primer-dimerisation during the PCR reaction; (iii) the annealing temperatures (Ta) of the primers would be in the range 50-65 °C and similar; (iv) the primers would have a GC content of about 50 % and (v) the primers would have a GC rich 3' 'sticky' end, terminating in a G or C base. Design was aided by the use of the OLIGO v3.3 computer program (Rychlik & Rhoades 1989). Once designed, oligonucleotides were synthesised by Oswell DNA service (Edinburgh University), aliquoted into 50 μl volumes and stored at -20 °C.

2) Annealing temperatures of the PCR

Annealing temperatures were increased in 2 °C increments from an initial annealing temperature of 45 °C to the calculated Ta of the primer pair. Ta for oligonucleotides up to 20 bases long is calculated approximately by the equation

$$Ta = 4(G + C) + 2(A + T)$$

(Don *et al.* 1991):

3) Primer concentration

Primers were used at 100 nM or 500 nM final concentrations

4) $MgCl_2$ concentration

$MgCl_2$ concentration was increased in 0.5 μM increments between 0.5 - 4.5 μM, final concentrations.

5) dNTP concentration

dNTP concentration was either 75 or 150 μM

6) Temperature cycle

Four types of cycle were used for each locus;

Cycle A: 93 °C, 3 min.; 40 X (93 °C 1 min., Ta 1 min., 72 °C 1 min.)

Cycle B: Touchdown PCR. 40 X (93 °C 1 min., Ta reduced in 2 °C increments every second cycle from 65 °C to the estimated primer Ta, 72 °C 1 min.)

Cycle C: 93 °C 3 min.; 10 X (93 °C 1 min., T_a - 5°C 1 min., 72 °C 1 min.) 30 x (93 °C 1 min., T_a 1 min., 72 °C 1 min.)

Cycle H: Hot start (polymerase added to reaction at first 93 °C denaturation) + cycle A.

7) Polymerase type

Three types of *Taq* polymerase were used;

1) *Taq* polymerase (Boehringer Mannheim)

2)*Taq* expand high fidelity (Boehringer Mannheim); a mix of *Pwo* and *Taq* polymerases.

3) Dynazyme (Flowgen)

8) PCR additives

Denaturation reagents can be used to increase the effective stringency of the annealing step (Don *et al.* 1991). Those included in the PCR reactions were dimethyl sulphoxide (DMSO) at 1 % w/v, single-stranded DNA binding proteins (1 % v/v) (Perfect match, Promega), bovine serum albumin (BSA) (0.2 % w/v) and 7 deaza-GTP (0.75 mM) (Boehringer Mannheim) and used as detailed in table 2.4.

In all PCR reactions, amplification of the appropriate cloned microsatellite was included as a positive control, and *S. ratti* DNA (both genomic and single-worm preparations) were tested for quality by amplification using primers recognising a non-microsatellite locus. A negative control reaction containing no parasite DNA was included.

2.2.9c Nested PCR reactions

Several loci were amplified by nested PCR reactions in order to increase the sensitivity of the reaction. This technique relies on a two-step PCR reaction.

Step 1: A standard non-radioactive PCR reaction was performed as described in section 2.2.9a except that the products of the reaction were not electrophoresed.

Step 2: A second PCR reaction was performed as in Step 1 with three differences. i) A pair of 'nested' primers were used that had been designed to sequences internal to the primers used in step 1, ii) 1 μl of the step 1 PCR reaction was used as a DNA template and iii) the reaction was thermocycled 35 times instead of 40. The annealing temperatures of the reactions were varied over the range described in section 2.2.9b to optimise the reactions. The step 2 reactions were either radioactive or non-radioactive

as described in sections 2.2.9a/b. Products from the reactions were subsequently electrophoresed on either 2% w/v agarose gel or 6% denaturing acrylamide gels.

2.2.9d PCR reconstitution experiments

The presence of genomic DNA contaminants that might specifically inhibit the amplification of microsatellite loci was tested by performing reconstitution experiments. Here, low concentrations of the cloned locus (1 or 10 ng) was used as a template in non-radioactive PCR reactions (described in section 2.2.9a) with increasing concentrations of *S. ratti* genomic DNA (0, 1, 10, 100 ng/PCR reaction). These reactions were electrophoresed on 2% w/v agarose gels and the relative intensity of the PCR products compared visually. The expectation was that if genomic DNA contained a PCR inhibitor, then amplification of the cloned locus would be reduced with increasing concentrations of *S. ratti* genomic DNA.

2.2.10 Southern blot analysis

3 μg of *S. ratti* genomic DNA (isofemale line 5 Heterogonic, Table 2.1) was digested overnight with *Eco* RI or *Rsa* I (Boehringer Mannheim) in the appropriate 1X reaction buffer. An aliquot of digested DNA was electrophoresed to ensure that digestion had occurred. 0.8 and 1.6 μg aliquots of digested DNA were separated on a 2% w/v agarose gel, the gel depurinated in 0.25 HCL for 10 minutes, rinsed in distilled water and capillary blotted with Hybond-N+ (Amersham) following the manufacturers' protocol. Following blotting, the membrane was rinsed in 2 X SSC (Appendix 1) and pre-hybridised at 60 °C as described previously (section 2.2.8b).

Probes were 1 μg of a microsatellite locus, PCR-amplified from the cloned insert and labelled with α^{32}P dATP (3000 Ci mmol^{-1}) using a random prime DNA labeling kit (Boehringer Mannheim) following the manufacturers' instructions. Unincorporated nucleotides were removed using a Chromaspin 30 column (Clontech) following the manufacturers' instructions. 0.1 mg of sheared Salmon sperm DNA was added to the probe, the volume made up to 1 ml with distilled water and denatured by boiling for 3 minutes. After snap-cooling on ice for 10 minutes the probe was added to the pre-hybridising solution and the blot hybridised overnight at 60 °C. The blot was subsequently washed twice at 0.5 x SSC, 0.1% w/v SDS at 60 °C for one hour and autoradiographed with intensifying screens at -70 °C.

2.2.11 Screening for genomic PCR-RFLPs

Three approaches were used to isolate PCR-RFLPs, as follows:

2.2.12a Random sequences from a genomic library

A random genomic DNA library was constructed as for section 2.2.8a except that the size selection step was modified to select fragments between 1 000-2 000 bp. Single colonies were picked from the library and suspended in 100 μl LB Ampicillin. medium. Sizes of the inserts of these colonies were determined by PCR amplification of plasmid DNA using M13 -21 and M13 reverse primers. 9 colonies with inserts of between 500 - 1500 bp were prepared (section 2.2.8a) and partially sequenced for 300 bp from each end. PCR primers were designed to these sequences and used in non-radioactive PCR reactions (section 2.2.9a).

The loci were surveyed for polymorphisms by amplifying each locus from template DNA made from two lines, line 29, line 32 and two isofemale lines, isofemale line 5 and isofemale line 54 (Table 2.1). 10 μl aliquots of the PCR-amplified DNA were independently digested with restriction endonucleases. All digestions were performed overnight at 37 °C using 10 units of the restriction endonuclease in 1X PCR buffer (Cambio) in 0.5 μl microcentrifuge tubes overlaid with mineral oil. Digestions were separated on 2 % w/v agarose gels containing ethidium bromide. Digestions showing putative polymorphic restriction sites were repeated to confirm the polymorphism.

2.2.12b PCR-RFLPs of RAPD-PCRs

Primers designed to isolated microsatellite sequences (section 2.2.8d) were used as RAPD primers in PCRs. A single primer was used in a standard non-radioactive PCR and the reaction optimised as described in section 2.2.9d. Primers that generated single amplification products were subsequently used to amplify a range of template DNA as described above (section 2.2.12a). These were subsequently surveyed for polymorphisms with restriction endonucleases (section 2.2.12a).

2.2.12c Calculations on nucleotide polymorphism

Data on restriction-site polymorphisms was used to infer amounts of genetic variation at the nucleotide level. For each locus, the conditional probability, $\hat{p}$, of a single base being polymorphic was calculated according to Weir (Weir 1996) where:

$$\hat{p} = \frac{k}{2jm} \qquad \text{(eqn. 2.1)}$$

where k = number of polymorphic restriction sites

j = the number of bases in the endonuclease recognition site

m = number of observed restriction sites.

2.2.12d PCR amplification by homology to *C. elegans* sequences

Nucleic acid databases were searched for microsatellites in previously sequenced *C. elegans* DNA. $(CA)_{10}$, $(GA)_{10}$ and $(AT)_{10}$ query sequences were used to search GenBank and EMBL databases by FASTA (Pearson 1988) using the GCG8 package (Genetics computer group, 1991) and primers designed to sequences identified. Universal primers to the 18S rDNA intragenic spacer were also obtained (donated by M. Blaxter). Non-radioactive PCRs were optimised for each locus (section 2.2.8d) and used to amplify template DNA from each of the *S. ratti* lines. The PCR products were subsequently screened for RFLPs (section 2.2.12a).

2.2.13 Genetic analysis of the segregation of genomic PCR-RFLPs

A genetic analysis of the progeny of naturally mated free living females was performed to ensure that markers isolated by the methods described in sections 2.2.9-2.2.12 segregated in a Mendelian manner and were due to alleles of single loci. Single worm DNA preparations were made using L3s from lines 5 Heterogonic, 54 and 132 Heterogonic. These were typed by PCR-RFLP at each polymorphic locus. Lines that were polymorphic were allowed to mate naturally in faecal cultures. Gravid free living females were collected from 3-day-old faecal cultures and isolated in individual wells of a 96 well microtitre plate. After hatching of the larvae, the female and her larval progeny were collected individually and single worm DNA preparations made. Preparations of a female and her progeny were then analysed by PCR-RFLP for the locus of interest.

2.2.14 Screening for mtDNA variation

Two methods were used to search for mtDNA variation: 1) screening for PCR-RFLPs and 2) direct sequencing of mtDNA PCR products.

2.2.15a mtDNA RFLPs

PCR primers were designed to regions of the mtDNA genome conserved between *A. suum* and *C. elegans* (Okimoto *et al.* 1992) in order to amplify fragments of approximately 1Kb in length. The positions of these primers within the mtDNA genome are illustrated in Figure 2.1. Non-radioactive PCR reactions were optimised as described in section 2.2.9d. PCR products were cloned using the TA Cloning kit

(Invitrogen) following the manufacturers' instructions, and colonies containing inserts identified by blue/white screening. 300 bp of the ends of each insert were sequenced using M13 -21 and M13 reverse primers (section 2.2.8c) and the sequences compared to the published *A. suum* and *C. elegans* mtDNA genomes. Primers were designed to the *S. ratti* -specific mtDNA sequences and the PCRs re-optimised. Template DNA from *S. ratti* lines 29, 32, isofemale line 5 Heterogonic, isofemale line 54 and additional lines B352, B362, G2924, G2932 were amplified. PCR products were subsequently surveyed for RFLPs (section 2.2.12a).

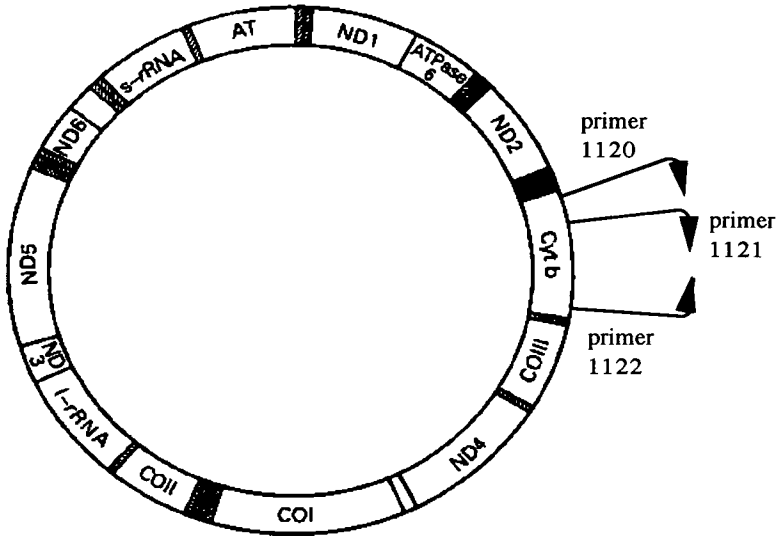


Figure 2.1 Map of the *C. elegans* mitochondrial genome (adapted from (Okimoto *et al.* 1992). The arrows represent the position and orientation of primers used to amplify fragments from *S. ratti.*

2.2.15b Direct sequencing of mtDNA PCR products

PCR products generated by the above protocol (section 2.2.15a) were purified using QIAquick spin columns (Quiagen) to remove unincorporated PCR primers and the DNA resuspended in 20 μl of distilled water. Each PCR fragment was cycle sequenced in both directions using 3.2 pmole of either of the appropriate primer together with 2.5 μl of template DNA and 8 μl ABI PRISM dye terminator mix (Perkin Elmer) according to the manufacturers' instructions. The samples were analysed on an ABI PRISM 377 automated sequencer (Applied Biosystems). Sequences from each of the template DNAs were aligned using assemblyline software

(Macintosh). These sequences were then analysed for variations in mtDNA sequence between lines or individual parasites.

2.3 Results

2.3.1 Isolation of Microsatellite loci

2.3.1a Number and types of microsatellites found in *S. ratti* DNA

12 positive colonies were identified and sequenced. Eight of these contained microsatellites (Table 2.2). Of the eight microsatellites, five were simple $(AC)_n$ dinucleotide repeats (locus CII, n=23; F, n=24; BIA, n=20; 5Ht, n=24; E, n=16). The remaining sequences showed complex combinations of $(AC)_n$, $(TA)_n$ and $(GA)_n$ repeats (locus B, $(TC)_{27}(TG)_{15}$; CI3, $(TA)_7(TG)_8(GT)_{17}$; G, $(GA)_7$ $(GA)_{17}$ $(CA)_{26}$). Two sequences (E and G) were not suitable for amplification by PCR due to the proximity of the microsatellite to the end of the insert preventing the design of primers. Primer pairs were designed to amplify the remaining six microsatellites. Sequences of the inserts are given in Table 2.2. Primer sequences are shown in Table 2.3.

2.3.1b Optimisation of non-radioactive PCRs

Attempts were made to optimise the PCR reactions for each locus by sequentially varying (1) the annealing temperature of the PCR, (2) the $MgCl_2$ concentration of the PCR and (3) the temperature cycle of the reaction.

It soon became apparent that, despite successful amplification of the cloned inserts (positive controls) in all experiments, these loci were not readily amplified from *S. ratti* DNA. Immediate success was met with primers to locus 5Ht (p19 + p20) in PCRs using genomic DNA as a template (Figure 2.2). However, visible products were unobtainable in reactions using single worm DNA as a template. For all the other loci either none or multiple amplification products were seen depending on the PCR conditions.

Results were similar for reactions using either genomic or single worm DNA preparations. Typically, during attempts to optimise the PCRs for these loci, low reaction stringencies (such as low temperature, high $MgCl_2$ concentrations) gave rise to multiple bands in the PCR products. Sequential increases in the stringency of these parameters reduced the number of bands in the PCR products but, with the exception of 5Ht, none of the of the reactions consistently resulted in a single band of the

Table 2.2. Sequences of eight microsatellite loci isolated from *Strongyloides ratti* isofemale line 43. Primer positions are underlined.

B: ccctaaccatgctaagtgctg atgtctgacatccagtcctcaaaggacccctctcaggagaggactaacttatcagggtcactttgcccatgtaactattaagacc**(TC)$_{27}$**
p1 p3 p5
(TG)$_{15}$tctgtgtgtcagaggacaactttggatggtattcttccagagacaaccacatttcttttgagacagggtctctcactgacctggga
p4 p6 p2

CII: gacaacacggtgtttgtatctatccttacaaacactacaagatgttccagtcataaataactcattgttattttaagtggttcaactcgttgaagttaccgtcgtcttaaacg
p7 p9 p11
atcaaaactaacaaat**(CA)$_{23}$**aaattcagttccagaatgatacacaccgacggaggcttgacggtacatccagtccgactggagcttgagtgtctctaggtgaactgaaac
p12
cgaggctaacgacgtcagtttatatacacgatgggactggacgactcttggtgtaagaac
p10 p8

F: agtcggacagactcctagactttggtctcccgtctccgacgaggtgaggatgagg**(AC)$_{24}$**agaaagtctggtgtacaccctgttggaacccttgatcaggaatggaagat
p15 p13 p16
ggaataaactttgctccaga
p14

BIA: ggagaagatattcaatccatagtaatcaaaatttttt**(CA)$_{20}$**tacacacagacgcacatacacggagacacatatatttcaataacagggtaactacttggactatgtgtc
p17
atcagtgagtataagtaactctttgattcacttgtgtaagggtatataaattattattttgtataagcaaacctaaaaatatataccctgaggaaggttccgtttgaaaacaa
ggatactacgataaacattgtaaatgtttttcaattacggaggacatttcaacatttcctctgggataacaatgagtacttctgagt
p18

5HT: cctctgccgtaagtgtaatgtaacatgtgttccaaaggaaacagatgaattg**(CA)$_{24}$**cctagaaagagagagagaaagagaaacagggagcgagagagacatagaagagaca
p19 p20
gaggcagagagat

CI3: catgacttttgagatatttgcttggattaataatgttcagatattagtaacatttacttgtgttttacatgtgtatgttttacatg**(TA)$_{7}$(TG)$_{8}$**ttt**(GT)$_{17}$**gt
p23 p21
agtctcatgtatcaagtcacatatggtgtgggtacacctgcacatatgtatgagagaggatagcaatatcttagacattcatactaaaaggttatatgaaacgtggtatttta
p24
gattaatcacgtagaatgagagatgcaacattaactaaaattaagggatgggcaaatggagttcatctg
p22

E: gaagcacgt**(AC)$_{16}$**agaacccccaccccccctcctaatacagaggagttaactggattagatcttttagggagtttcctgtccgtgtcgccaaactgagcatctactgct

G: aagtttcgtcactcgttcagacatacctcttttcctcctcctcatgacgggggagggagaggggaagggaaaggggaaggggaagagaaggagggaagaagaagagaa
(GA)$_{7}$gga**(GA)$_{17}$**ggggg**(CA)$_{26}$**gacgcacc

Locus	*Repeat type*	*Primer sequence (5'-3')*	*Expected size (bp)*
B	$(TC)_{27}(TG)_{15}$	p1 : cct aac cat gct aag tgc tg	173
		p2 : cag gtc agt gag aga ccc	
		p3 : atg tct gac atc cag tcc	227
		p4 : gaa atg tgg ttg tct ctg g	
		p5 : ctc aaa gga ccc ctc tca gga gag gac	219
		p6 : tgt ctc aaa aga aat gtg gtt gtc tct gg	
CII	$(CA)_{23}$	p7 : cac ggt gtt tgt atc tat cc	321
		p8 : ctt aca cca aga gtc gtc c	
		p9 : caa aca cta caa gat gtt cc	258
		p10 : tga cgt cgt tag ccc tcg	
		p11 : agt ggt ttc aac tcg ttg aag tta ccg tcg	159
		p12 : cgg act gga tgt acc gtc aag cct cc	
F	$(AC)_{24}$	p13 : gtc gga cag act cct aga ctt tgg	158
		p14 : ctt cca ttc ctg atc aag ggt tcc	
		p15 : tcc tag act ttg gtc tcc cg	141
		p16 : tct gga gca aag ttt att cc	
BIA	$(CA)_{20}$	p17 : gag aag ata ttc aat cca t	390
		p18 : cag aag tac tca ttg tta tcc	
5Ht	$(CA)_{24}$	p19 : gct tgg att aat aat gtt cag	146
		p20 : tct cta tgt ctc tct cgc tcc c	
CI3	$(TA)_7(TG)_8$ttt $(GT)_{17}$	p21 : gct tgg att aat aat gtt cag	315
		p22 : atg aac tcc att tgc cca tcc	
		p23 : tga ctt ttt gag ata ttt gct tgg	235
		p24 : tga atg tct aag ata ttg cta tcc	

Table 2.3 Sequences of the primers designed to amplify microsatellite loci.

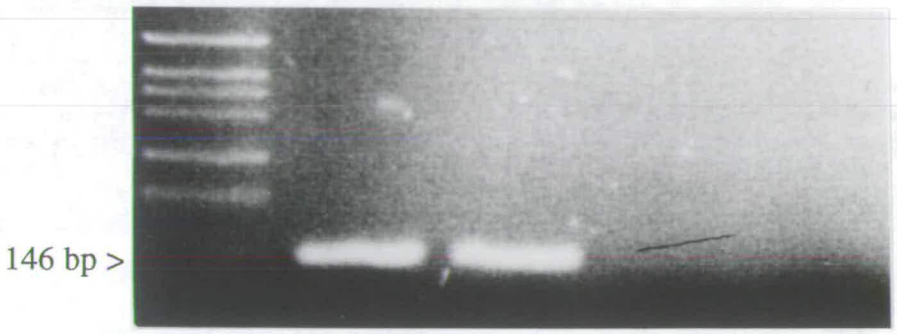
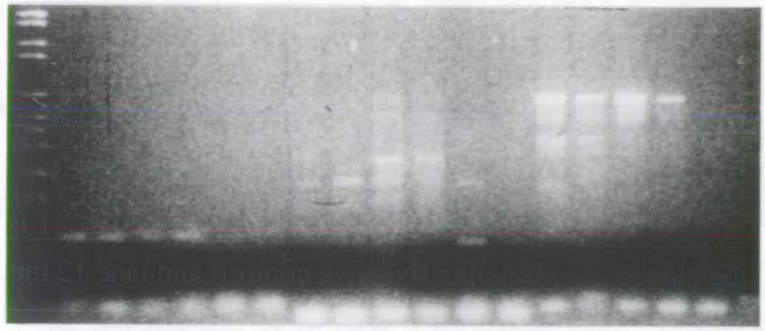

146 bp >
Lane : 1 2 3 4 5

Figure 2.2

PCR amplification of genomic DNA from isofemale line 5 Heterogonic (lane 2) and 54 (lane 3) using primers p19 + p20 to microsatellite locus 5 Ht. Lanes 4 and 5 are negative controls; lane 1 contains size standards.

Figure 2.3

The effect of increasing $MgCl_2$ concentration on the amplification of microsatellite locus CII from line 5 Heterogonic genomic DNA with primers p9 + p10. Increase in $MgCl_2$ concentration in μM is shown beneath the lane number. Lane 11 is the amplification of the cloned microsatellite (positive control) and lane 12 the negative control. Lane 1 contains size standards.

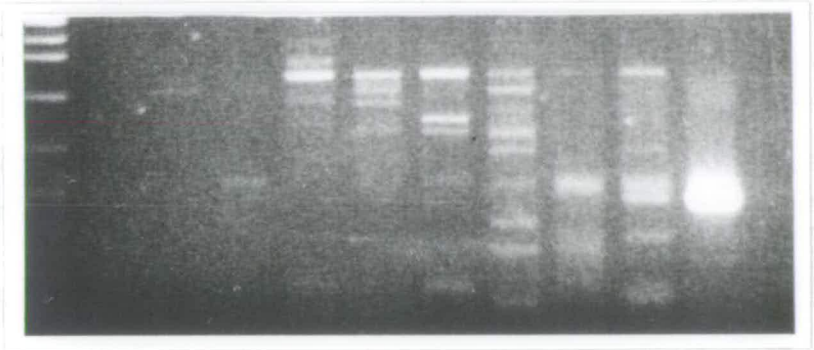


Lane: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Amplified locus: 5 Ht cI3 cII

Figure 2.4

Nonradioactive PCR amplification of microsatellite loci 5 Ht (primers p19 + p20; lanes 2 - 7), CI3 (primers p23 + p24; lanes 8 - 13) and CII (primers p9 + p10; lanes 14 - 19). For each locus, the first two lanes are amplifications of isofemale line 5 Heterogonic genomic DNA, the second two lanes are amplifications of line 54 genomic DNA and the final two lanes are negative controls. Lane 1 contains size standards.


672 bp >
300 bp >
Lane : 1 2 3 4 5 6 7 8 9 10 11 12
[MgCl2] : 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 2.5

predicted size from genomic or single worm DNA templates. For example, Figure 2.3 shows the effect of increasing $MgCl_2$ concentration using primers p9 + p10 to locus CII. A high molecular weight band is seen at low $MgCl_2$ concentrations (high reaction stringency), which disappears as the $MgCl_2$ concentration is reduced; smaller products appear in its place, some of which are of a similar size to the positive control. The reduction in reaction stringency has the effect of amplifying many non-specific products and a single product is not obtained. Amplification of locus CI3 with primers p23 + p24 consistently produces a product 650 bp in length using a step in annealing temperature from 48 - 52 °C after 10 temperature cycles (cycle C, section 2.2.9b). However this band was 415 bp longer than the expected product size (Figure 2.3). A similar reaction using primers p9 + p10 to locus CII with this temperature cycle shows a similar pattern to amplification of locus CI3; bands are produced but not within the expected size range. Locus 5Ht produces low molecular weight bands of the expected size, 146 bp, Figure 2.4.

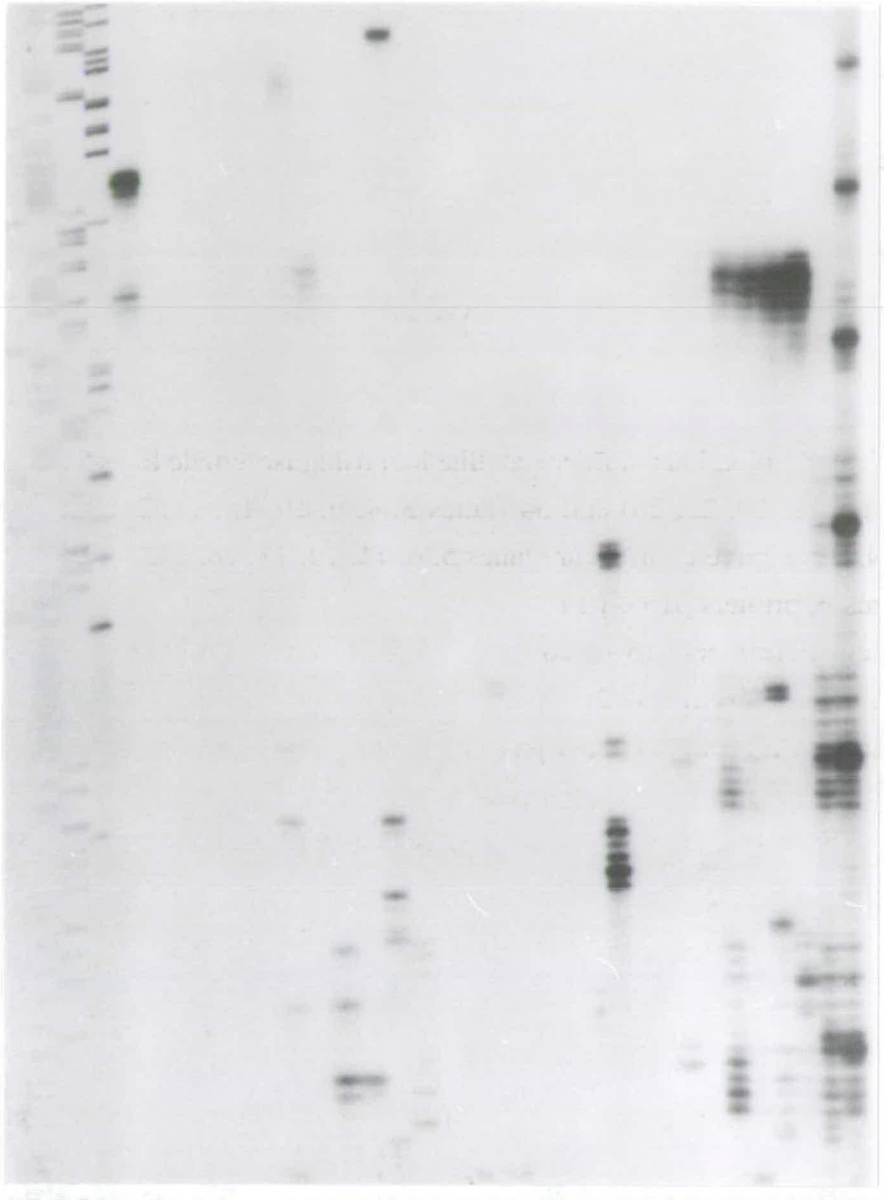
2.3.1c Optimisation of radioactive PCRs

The results for the optimisation of radioactive PCRs were essentially the same as above. Locus 5Ht was successfully amplified from genomic DNA with primers p19 + p20. The autoradiograph of the electrophoresed products shows a 'stutter' pattern that is characteristic of amplifications of microsatellite loci (Figure 2.5). Successful amplification of this locus from single worm DNA preparations was also achieved, in contrast to amplification by non-radioactive PCRs. This illustrates the increased sensitivity of radioactive PCRs in detecting low intensity amplification products. Extensive attempts were made to optimise these reactions to the stage where single amplification products could be reliably obtained and scored from individual parasites. These attempts were not successful in that the PCR products obtained were consistently faint and of insufficient focus to ensure reliable genotyping of the locus. The use of additives and hot start/touch-down temperature cycling was attempted, but none of these treatments increased the efficiency and specificity of the reaction to the state where locus 5Ht could be used to genotype single parasites.

For all other loci, multiple bands were amplified in low stringency reactions, but in no case was it possible to achieve a single reproducibly amplifiable fragment of the predicted size from genomic or single worm DNA preparations. Figure 2.5 illustrates a characteristic electrophoretic gel exposure of PCRs using primers to three loci, F, CI3 and 5Ht. Genomic DNA was used as a template in these reactions and two different primer pairs were used for the F and CI3 loci. While amplification products were seen

Figure 2.5

Radioactive PCR amplifications of microsatellite loci using isofemale lines 5 Ht (lanes 1, 2, 7, 8, 13, 14, 19, 20, 25, 26) and 54 (lanes 3, 4, 9, 10, 15, 16, 21, 22, 27, 28) as template DNA. Negative controls are lanes 5, 6, 11, 12, 17, 18, 23, 24, 29, 30)
Lanes 1-6: Locus F, primers p13 + p14
Lanes 7-12: Locus F, primers p15 + p16
Lanes 13-18: Locus CI3, primers p21 + p22
Lanes 19-24: Locus CI3, primers p23 + p24
Lanes 25 - 30: Locus 5Ht, primers p19 + p20
Lanes A, B, C, D are sequencing reactions used as size-standards.


146 bp
Lane: A B C D 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 2122 2324 25 26 27 28 29 30
Primers: p13 + p14 p15 + p16 p31 + p22 p23 + p24 p19 + p20
Locus: F F cI3 cI3 5 Ht

for these two loci for certain primer pairs (F, primers p15+p16; CI3, primers p23+p24), the products were not reproducible between duplicate reactions. The successful amplification of locus 5Ht from genomic DNA in this set of PCRs was a positive control for the reagents and quality of the DNA template.

In order to observe whether or not the amplification reactions would work better within another PCR system, primers to loci F, CI3, CII, B and 5Ht were given to a separate laboratory that regularly genotypes other organisms using radioactive microsatellite PCR amplifications. They formed the same conclusions, that locus 5Ht could be amplified from genomic DNA but not reliably from single worm DNA preparations, and that the other loci were not amplifiable using these primers (S. Paterson, unpub. observations).

2.3.1d Redesign of primer sequences

In order to test whether the primer sequences were responsible for the failure of these loci to amplify, second, and in some case third, generation primers were designed for all loci, except 5Ht. The primers were designed to recognise regions of sequence that had not been included in the initial primer design (Tables 2.2 and 2.3). The optimisation procedure was followed for each new pairs of primers; non-radioactive and subsequently radioactive PCR reactions were varied over a range of annealing temperatures, $MgCl_2$ concentrations and temperature cycle types. These conditions are shown in Table 2.4. Locus-specific amplification was not achieved for any loci using the new primer pairs showing that inability to amplify the loci was not due to inappropriate primer design.

2.3.1e Nested PCR reactions

The range of primer-pairs available for each locus meant that nested PCR reactions could be attempted, a method that can reduce the production of non-specific products and increase the sensitivity of the PCR (McPherson *et al.* 1995). This technique has been used successfully with other loci which have proved hard to amplify loci (section 2.3.3). Both radioactive and non-radioactive nested PCRs were attempted for loci B, CII, F and C13 using a range of annealing temperatures for the step 2 reactions (section 2.2.9b). The nested primer sequences are shown in Table 2.3 and the results of the PCR reactions in Table 2.4. The reactions generated a variety of small amplification products at low annealing temperatures, which disappeared as the annealing temperature was increased. Specific amplification of a single PCR fragment was not achieved.

Table 2.4

PCR reaction conditions for the attempted amplification of the microsatellite loci using radioactive and non-radioactive PCR, and the outcome of the reactions.

Notes to table: [a]**Primer** refers to the primers shown in Figure 2.3.1a. [b]**Expected product size** is the length of the expected PCR fragment, in bp [c]**Annealing temp.** is the temperature used in the temperature cycle (T_A see below) which was varied through this range at 2 or 4°C intervals. [d]**Cycle:** A = 93°C 3 min.; 40 x (93°C 1 min., T_A 1 min., 72°C 1 min.): B = 40 x (93°C 1 min., T_A reduced in 2°C increments every second cycle from 65°C to the primer Tm, 72 °C 1 min.): C = 93°C 3 min.; 10 x (93°C 1 min., T_A- 5°C 1 min., 72 °C 1 min.) 30 x (93 °C 1 min., T_A 1 min., 72 °C 1 min.): H= Cycle A + hot start (polymerase added to reaction at first 93°C denaturation). [e]**Additives:** 0 = none, P = Perfect match (Promega), Bs= BSA at 0.2 % w/v, N = 0.75 mM 7 deaza-GTP (Boehringer). [f]**Polymerase:** T = Taq polymerase (Boehringer), D = Dynazyme (Flowgen). [g]**$MgCl_2$** was added in the range shown with 0.5 mM increments. [h]**Result:** refers to the consensus of the PCR reactions, N = no bands, M = multiple bands, S = single bands.

[i]**Nested PCR reactions** were performed as described in section 2.2.9d using cycle A for step 1 and 2 with the difference that only 35 cycles were used for the step 2 reaction.

Locus	Primer[a]	Expected product size[b]	Annealing temp[c]	Cycle[d]	Additive[e]	Polymerase[f]	$MgCl_2$ mM[g]	Result[h]
B	P1 / P2	173	45 - 55	A,B,C,H	0/P/Bs/N	D,T	0.5-4.5	N/M
	P3 / P4	227	45 - 70	A,B,C,H	"	"	"	N/M
	P5 / P6	219	45 - 60	A	0	"	1.5	N
Nested B[i]	P1 / P2 P3 / P4	173	40 - 59	A	0	D	1.5	N
CII	P7 / P8	321	40 - 56	A,B,C,H	0/P/Bs/N	D,T	0.5-4.5	N/M
	P9 / P10	258	40 - 55	A,B,C,H	"	"	"	N/M
	P11 / P12	159	45 - 60	A	0	D	1.5	N
Nested CII[i]	P7 / P8 P9 / P10	258	40 - 56	A	0	D	"	N
F	P13 / P14	158	40 - 60	A,B,C,H	0/P/Bs/N	D,T	0.5-4.5	N
	P15 / P16	141	40 - 55	A,B,C,H	0/P/Bs	"	1.5	N
Nested F	P15 / P14 P13 / P16	141	45 - 55	A	0	D	1.5	N
BIA	P17 / P18	390		A,B,C,H	0/P/Bs/N	D,T	0.5-4.5	N
5HT	P19 / P20	146	40 - 58	A,B,C,H	0/P/Bs/N	D,T	1.5	S
C13	P21 / P22	315		A,B,C,H	0/P/Bs/N	D,T	0.5-4.5	N/M
	P23 / P24	235	40 - 55	A,B,C,H	0/P/Bs	"	"	N/M
Nested C13[i]	P24 / P23 P22 / P21	235	45 - 55	A	0	D	1.5	N

2.3.1f Reconstitution experiments

The inability to amplify the microsatellite loci may have been due to the presence of a contaminant of the genomic DNA that specifically inhibited amplification. To test this hypothesis, reconstitution experiments were performed for each locus (section 2.2.9e). Here, low concentrations of the cloned microsatellite (1 or 10 ng) was used as a template in PCR reactions with increasing concentrations of *S. ratti* genomic DNA. The expectation was that if the genomic DNA contained a PCR inhibitor, then amplification of the cloned locus would show reduced efficiency at high concentrations of genomic DNA. Reconstitution experiments were performed for all loci and in none of them did the addition of high concentrations of genomic DNA inhibit the amplification of the cloned microsatellite (Figure 2.6). This shows that the presence of a DNA inhibitor was probably not the cause of the failure of these loci to amplify.

2.3.2 Southern blot analysis

Amplified CI3 locus was used as a probe against a Southern blot of *S. ratti* genomic DNA (isofemale line 5 Heterogonic). A single band was observed (Figure 2.7) for *Eco* RI and two bands for *Rsa* I digested DNA. This confirms that the CI3 locus is present in the *S. ratti* genome.

2.3.3 Screening for PCR-RFLPs from a random genomic library

Nine clones were arbitrarily chosen from a random genomic library with insert sizes of between 500 and 1500 bp, sequenced and appropriate primers designed. The primer sequences are shown in Table 2.5. Five of the nine clones were successfully amplified by PCR from genomic template DNA, producing a single amplification product of the expected size (Table 2.5). Due to low efficiency of amplification of two loci (CM-2 and 29), nested primers were designed and these loci subsequently amplified by nested PCR (section 2.2.9d).

For the preliminary survey, amplified PCR products were digested with a wide range of restriction endonucleases. 19 - 33 enzymes were used to survey each locus for RFLPs (Table 2.6). An average of 8 enzymes recognised at least one cleavage site within each PCR-amplified locus. The total number of bases surveyed was calculated as the number of observed cut sites multiplied by the number of bases in the enzyme recognition sequence (Nei 1987). Summing all loci, 111 bases were surveyed out of a total of 6880 bp of DNA, *i.e.* 6.7% of the total sequence (Table 2.7). These estimates are conservative as some restriction fragments may have been too small to be

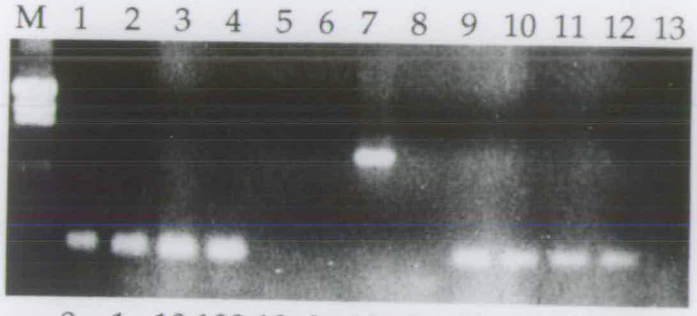
Figure 2.6

Reconstitution of PCR reactions with cloned microsatellite F. DNA added to the reactions was 10 ng (lanes 1-4) or 1 ng (lanes 9-12) of cloned microsatellite F (expected size 159 bp) with 0, 1, 10 or 100 ng of *S. ratti* genomic DNA (lanes 1-4 and 9-12, respectively). Lanes 5 and 13 are *S. ratti* genomic DNA only with no cloned microsatellite. Lane 6 is a negative control. Lanes 7 and 8 are PCR reactions for a non microsatellite locus (BSP 8, expected size 900 bp, M. C. Fisher, unpublished observations), lane 7 with 10 ng *S. ratti* genomic DNA, lane 8 is a negative control. M are size markers.

Figure 2.7

Southern blot of *S. ratti* genomic DNA. Microsatellite locus CI3 is used as a probe against *S. ratti* genomic DNA digested by *Rsa* I (lanes 1 and 2) and *Eco* RI (lanes 2 and 3). Lane M contains size standards (not visible) and lane 5 no DNA.

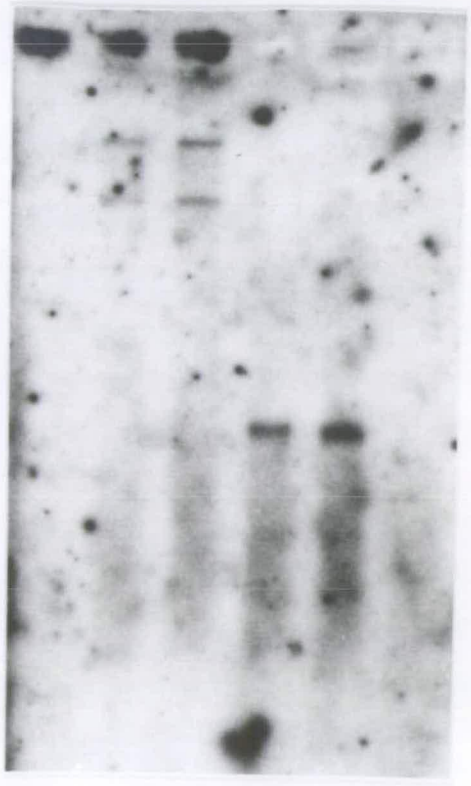
Figure 2.6



0 1 10 100 10 0 10 0 0 1 10 100 10

quantity of genomic DNA template (ng) in PCR

Figure 2.7



Lane: M 1 2 3 4 5

seen on a gel (fragments < 80 bp) and would therefore not have been scored. Four of these restriction sites were polymorphic. Assuming that each polymorphism is a result of a single base difference then 0.45 % of the total sequence surveyed was polymorphic (range 0 - 2.8 % across loci). Two loci (XP-1 and 27) had no RFLPS in 3550 bp screened and three loci contained polymorphic restriction sites; CM-2 (two polymorphic *Hae* III restriction sites), 29 (one polymorphic *Alu* I restriction site) and 24 (one polymorphic *Cfo* I site).

2.3.3a Locus CM-2

During the preliminary RFLP survey of locus CM-2 it became apparent that some single worm DNA preparations did not produce amplified products. This was especially noticeable within isofemale line 132 Heterogonic where all single worm and genomic DNA preparations did not amplify. Examples of non-amplifying DNA preparations are shown in Figure 2.8. This suggested that there was a non-amplifying (null) allele at this locus within line 132 Heterogonic. To investigate this possibility, single worm DNA preparations were made from five iL3s taken from isofemale lines 5 Heterogonic and 132 Heterogonic. These DNA samples were amplified by PCR at i) the Actin locus (section 2.3.6a) and ii) the CM-2 locus using all possible combinations of forward and reverse primers F+R, F+R2, F2+R, F2+R2 in the first round of PCR followed by nested amplification using primers F3+R3 (Figure 2.8 and Table 2.5).

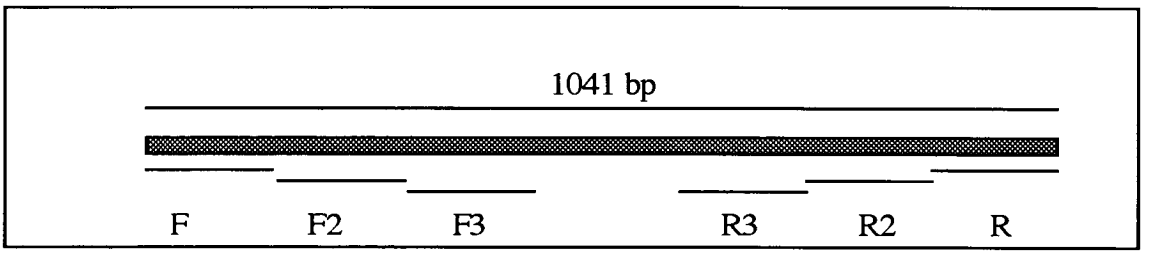


Figure 2.8 Positions of the primers used to amplify locus CM-2, F = forward, R = reverse. Primers F3 + R3 are the nested primers.

DNA from single worms amplified at the Actin locus (Table 2.8) showing that the preparations were of amplifiable quality. In addition, the worms from isofemale line 5 Heterogonic amplified at locus CM-2 for all primer combinations. However no worms from isofemale line 132 Heterogonic amplified at the CM-2 locus. This suggests that 132 Heterogonic worms were homozygous for a null allele (heterozygotes would be expected to amplify a band). Further, the null allele is probably a result of a deletion of the entire locus and not a nucleotide substitution(s) in

Locus, size (bp)	Primer sequence (5'-3')	Homology to sequences in GenEMBL database	PCR cycle	Number of PCR cycles
CM-2, 1000	F: tgg tga tgt taa cgc att cg R: tta tgc att agt ttc agc agg F2:cag caa tac gca ata cct gg R2:tga aat ctc ttg aag gca gc F3:tat aac cca ctt cat tat cg R3:tca gtt gaa ggt aaa gac g	83%: *C. elegans* cosmid D2085	**Round 1:** 95 °C 1 min. 52 °C 1 min. 72 °C 2 min. **Round 2:** 95 °C 1 min. 52 °C 1 min. 72 °C 2 min.	30 cycles 35 cycles
29, 1100	F: cta gtt att aaa aga tgg ctg R: caa gaa tca taa tgt gtt ctg F2:ctg tta gat gtg att atg aag g R2:aag gga tct aac aaa ttt atc cc	63%: *C. elegans* cosmid BO336	95 °C 1 min. 54 °C 1 min. 72 °C 3 min.	35 cycles
27, 1230	F: cgt atc ttg cga tga tca tc R: gat atc agt tgc aaa act acc	90% (30 bp overlap): *C. elegans* cosmid M02A10	95 °C 1 min. 54 °C 1 min. 72 °C 3 min.	40 cycles
18S rDNA tandem repeat, 1800	18P: tga tcc agc tgc agg ttc agg ttc ac 18A: aaa gat taa gcc atg cat g	100%: C. elegans rDNA tandem repeat	95 °C 1 min. 54 °C 1 min. 72 °C 3 min.	35 cycles
24, 1230	F: aac tgc aaa att aag aaa tgc g R: gtt cga tgc tgt gag tat gag	58%: *C. elegans* cosmid ZK622	95 °C 1 min. 54 °C 1 min. 72 °C 3 min.	35 cycles
XP1, 850	F: taa gga aac tcc cat ctg g R: act tgg aca ttt cga att gg	51.7% identity in 725 bp overlap: *S.cerevisiae* mitochondrion DNA	95 °C 1 min. 52 °C 1 min. 72 °C 3 min.	35 cycles

Table 2.5 Primer sequences and conditions used to amplify nuclear loci. Homology to sequences in GenEMBL are shown.

Locus, size (bp)	*Restriction endonucleases used to screen loci*[a]	*Endonucleases recognising polymorphic restriction sites*
CM-2, 1000	***Alu* I, *Rsa* I, *Hinf* I, *Mnl*I, *Hae* III, *Msp* I, *Sau* 3A, *Cfo* I, *Dde* I,** *Eco* RI, *Hind* III, *Xho* I, *Xba* I, *Pvu* I, *Pvu* II, *Dpn* I, *Dpn* II, *Sin* I, *Sac* I, *Sac* II, *Sma* I, *Bam* HI, *Asn* I, *Hind* II, *Nsi* I, *Dra* I, *Sfu* I, *Cla* I, *Apa* I, *Sph* I, *Ssp* I, *Ase* I, *Nco* I, *Nci* I, *Not* I	***Hae* III**
29, 1100	***Alu* I, *Rsa* I, *Hinf* I, *Mnl* I, *Msp* I, *Hae* III, *Sau* 3A, *Dde* I,** *Eco* RI, *Hind* III, *Xho* I, *Xba* I, *Pvu* I, *Pvu* II, *Sin* I, *Sma* I, *Bam* HI, *Asn* I, *Hind* II, *Dra* I, *Cla* I, *Sph* I, *Ssp* I, *Ase* I, *Nco* I, *Not* I	***Alu* I**
27, 1230	***Alu* I, *Rsa* I, *Hinf* I, *Mnl* I, *Msp* I, *Hae* III, *Sau* 3A, *Cfo* I, *Dde* I,** *Eco* RI, *Hind* III, *Xho* I, *Xba* I, *Pvu* I, *Pvu* II, *Dpn* I, *Dpn* II, *Sin* I, *Sac* I, *Sac* II, *Sma* I, *Bam* HI, *Asn* I, *Hind* II, *Nsi* I, *Dra* I, *Sfu* I, *Cla* I, *Apa* I, *Sph* I, *Ssp* I, *Bgl* II, *Ase* I, *Nco* I, *Bgl* IGQ, *Nci* I, *Not* I	-
18S rDNA tandem repeat, 1800	***Sau* 3A, *Alu* I, *Hae* III, *Mnl* I,*Msp* I, *Cfo* I, *Hinf* I, *Dde* I,** *Eco* RI, *Pvu* II, *Sin* I, *Xba* I, *Hind* III, *Sma* I, *Nco* I, *Dpn* II, *Dpn* I	-
24, 1230	***Alu* I, *Hinf* I, *Msp* I, *Hind* III, *Sau* 3A, *Mnl* I, *Cfo* I,** *Eco* RI, *Pvu* I, *Rsa* I, *Xho* I, *Xba* I, *Sac* I, *Dra* I, *Bam* HI, *Xho* I, *Xba* I,*Sac* I,*Dpn* I,*Nsi* I,*Sfu* I,*Sph* I, *Apa* I, *Nco* I	***Cfo* I**
XP1, 850	***Mnl* I, *Hinf* I,***Eco* RI, *Ase* I, *Sin* I, *Sac* I, *Xba* I, *Xho* I, *Nco* I, *Bam* HI, *Asn* I, *Sma* I, *Bgl* II, *Hind* III, *Kpn* I	-

Table 2.6 Restriction endonucleases used to survey PCR fragments for RFLPs.
[a] Enzymes restricting at 4-base recognition sequences are highlighted in bold. The remaining endonucleases restrict at 6-base recognition sequences.

Locus	Fragment size (bp)	No. of non-cutting enzymes	No. of cutting enzymes	Total No. of sites observed	Total No. of bp surveyed[a]	% sequence surveyed	No. of sites polymorphic	$\hat{p}$ [b]
CM-2	1000	23	10	16	70	7	2	0.0140
29	1100	14	10	29	92	8.4	1	0.0055
XP-1	850	21	4	8	36	4.2	0	0.0000
18S rDNA	1800	10	9	26	118	6.6	0	0.0000
27	900	21	9	16	76	8.4	0	0.0000
24	1230	15	8	16	66	5.4	1	0.0075
Overall	**6880**	**17.3[c]**	**8.3[c]**	**111**	**458**	**6.7[c]**	**4**	**0.0045[c]**

Table 2.7 Summary of the preliminary screen for RFLPs at six nuclear loci.

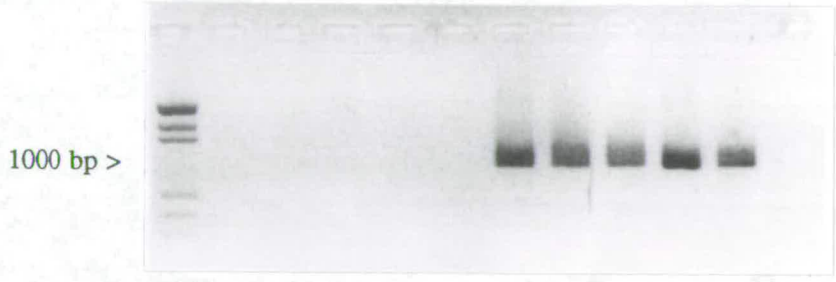
[a] Total number of bases surveyed calculated as the number of cut sites multiplyed by the number of bases in the enzyme recognition sequence (Nei 1987).

[b] $\hat{p}$ is the conditional probability that a nucleotide is polymorphic, calculated as described in section 2.2.12c.

[c] Means of overall values

Figure 2.9

Amplification of single iL3s from isofemale line 132 Heterogonic (lanes 2 - 6) and 5 Heterogonic (lanes 7 - 11) by first round primers F + R and second round (nested) primers F3 + R3 to the CM-2 locus. Lane 1 are size standards and lane 12 a negative control.



Lane: 1 2 3 4 5 6 7 8 9 10 11 12

132 heterogonic

5 heterogonic

one of the primer sites. The possibility that a base substitution was present in one of the primer sites for CM-2 primers F3 and R3 was tested by using nested PCR on 132 Heterogonic worms using primers F+R followed by F2+R2, no amplification being observed. Worms from lines 29 and 32 produced amplified bands as expected, suggesting that the null allele was either not present or at a low frequency within these populations of the parasite. On this basis the locus was included in the population survey described in chapter 3. A restriction map of locus CM-2 is shown in Figure 2.12.

	Locus				
Line	Actin	CM-2			
		F+R	F+R2	F2+R	F2+R2
5 Heterogonic	5	5	5	5	5
132 Heterogonic	5	0	0	0	0

Table 2.8 Numbers of worms amplifying at the Actin and CM-2 loci.

2.3.4 Screening for PCR-RFLPs of RAPD-PCRs

Five RAPD primers successfully amplified anonymous fragments (Table 2.9). The PCR products from lines 29, 32, 5 Heterogonic and 54 were screened for polymorphisms with a range of restriction endonucleases. The restriction endonucleases used to survey the sequences and the numbers that successfully restricted are shown in Table 2.9. A total of 19 restriction sites were observed and 90 bases surveyed out of a total of 4570, this corresponds to 2% of the total sequence. None of the loci were polymorphic among the parasite lines screened.

2.3.5 PCR amplification by homology to *C. elegans* sequences

Searches of the GenBank and EMBL databases found four *C. elegans* entries with dinucleotide repeats greater than 7 repeats long. Primers were designed to flanking sequences for three of these repeats. The locus identification, accession number and repeat-type are shown in Table 2.10. PCR using *S. ratti* genomic DNA was unsuccessful in amplifying these loci even after attempts to optimise the reactions. It was concluded that the genetic distance between *S. ratti* and *C. elegans* was too great to warrant using this technique further.

Table 2.9.

Primers and conditions used to amplify RAPD loci from *S. ratti*. Restriction endonucleases used to screen the sequences are shown.

[a]Endonucleases that restrict at 4-base recognition sequences are highlighted in bold. The remaining endonucleases restrict at 6-base recognition sequences. [b] Total number of bases surveyed calculated as the number of cut sites multiplied by the number of bases in the enzyme recognition sequence (Nei 1987).

Primer (sequence)	PCR conditions	Length of PCR product (bp)	Restriction endonucleases used to screen loci	Endonucleases that restrict (number of sites)[a]	Total No. of bp surveyed[b] (% sequence surveyed)	Endonucleases recognising polymorphic sites
4F 5'- agt tga tcg gtc ttg cag c	95 °C 1 min. 45 °C 1 min. 72 °C 2 min. } 40 cycles	1200	*Alu* I, *Rsa* I, *Hinf* I, *Mnl* I, *Msp* I, *Cfo* I, *Hae* III, *Pvu* I, *Pvu* II, *Sau* 3A, *Xba* I	***Alu* I (1), *Rsa* I (1), *Hinf* I (1), *Mnl* I (1)**	20 (1.7)	-
5Ht1 5'- gct tgg att aat aat gtt cag	95 °C 1 min. 45 °C 1 min. 72 °C 2 min. } 40 cycles	700	*Alu* I, *Rsa* I, *Hinf* I	-	- (0.0)	-
4R 5'- ccg caa aac ata aac ctg gac c	95 °C 1 min. 45 °C 1 min. 72 °C 2 min. } 40 cycles	670	*Alu* I, *Rsa* I, *Hinf* I, *Mnl* I, *Msp* I, *Hae* III, *Eco* RI, *Sau* 3A, *Hind* III, *Xho* I, *Xba* I, *Pvu* I, *Pvu* II, *Cfo* I, *Dpn* I, *Dpn* II, *Sin* I, *Sac* I, *Sac* II, *Sma* I, *Nci* I, *Dde* I, *Not* I	***Alu* I (1), *Rsa* I (1), *Mnl* I (1), *Hinf* I (1),** *Eco* RI (2), *Hind* III (1)	34 (4.9)	-
F2 5'- tcc tag act ttg gtc tcc cg	95 °C 1 min. 45 °C 1 min. 72 °C 2 min. } 40 cycles	600	*Alu* I, *Rsa* I, *Hinf* I, *Msp* I, *Cfo* I,*Hae* III, *Eco* RI, *Pvu*I,*Pvu*II,*Sau*3A,	-	- (0.0)	-
B4 5'- tgt ctc aaa aga aat gtg gtt gct tct gg	95 °C 1 min. 45 °C 1 min. 72 °C 2 min. } 40 cycles	1400	*Alu* I, *Rsa* I, *Hinf* I, *Mnl* I, *Msp* I, *Hae* III, *Eco* RI, *Sau* 3A, *Hind* III, *Xho* I, *Xba* I, *Pvu* I, *Pvu* II, *Sin* I, *Dpn* II,*Sma* I, *Not*I	***Alu* I (1), *Rsa* I (1), *Hinf*I (1),** *Eco* RI (1), *Pvu* II (1), *Sin* I (1), *Dpn* II (2)	36 (2.6)	-

Locus	*GenEMBL accession No.*	*Microsatellite*	*Primer sequence*
Pal 1	X62782	(AT)8	F: tgc cgt ggc cct ggc gac gc R: ttg cgg gag gct gaa ttg gag g
CAPK1	M37119	(CT)8	F: tgt ccg tgt acg aac agc tga c R: gaa aac cat gtt caa aat gga tcc g
Lin 1	X60232	(GT)8	F: cct taa act gat aaa cat g R: tgt atc atg taa agt ata cgg gc

Table 2.10 *Caenorhabditis elegans* microsatellite sequences found within GenEMBL database.

Amplification of *S. ratti* DNA using universal primers to the 18S rDNA tandem repeat gene was successful. A fragment of 1800 bp was amplified (expected product size 1600 bp) from lines 29, 32, 5 Heterogonic and 54. 12 restriction endonucleases were used to screen the sequence, nine of which cut (Table 2.5 and Table 2.6). In this way 118 bp were surveyed, corresponding to 6.6% of the total sequence. None of these restriction sites were polymorphic.

2.3.6 *S. ratti*-specific PCR-RFLPs

Primers to three loci were donated by M. Viney; Actin, TJCA-2 and BSP-8 (Table 2.11). These loci had been previously observed to contain RFLPs between isofemale lines 5 and 54 (M. Viney, unpublished observations). The loci were amplified from lines 29, 32, 5 Heterogonic and 54 using the primers and conditions shown in Table 2.11. A survey of the amplification products by restriction digestion confirmed the presence of the RFLPs. Actin was polymorphic at a single *Mnl* I restriction site, TJCA-2 at single *Alu* I and *Rsa* I sites, and BSP-8 at a single *Hinf* I site.

Locus, size (bp)	Primer sequence (5'-3')	Homology to sequences in GenEMBL database	PCR cycle	Number of PCR cycles
Actin, 390	F: gga gat ggt gtc acc aca ccg t R: tcc ata cca aga agg atg gct	100%: *C. elegans* Actin 4 gene	95 °C 1 min. 62 °C 1 min. 70 °C 2 min.	42 cycles
TJCA - 2, 1050	F: ggt tgc gct cgt tgc ggg R: caa atg tga aat ccc cga gc	97%: *Acinetobacter* 16s rRNA gene	95 °C 1 min. 61 °C 1 min. 72 °C 2 min.	40 cycles
BSP-8, 1240	F: cct agt ggc att att ctc g R: cac cta gtg gaa atc cag	56.5%: *C. elegans* cosmid CO1F6	95 °C 1 min. 52 °C 1 min. 72 °C 2 min.	35 cycles

Table 2.11 Primer sequences and conditions used to amplify *S. ratti*-specific nuclear loci. Homology to sequences in GenEMBL are shown.

2.3.7 Polymorphism of loci within British samples of *S. ratti*

Six polymorphic loci were isolated by a combination of methods, described in sections 2.3.3 - 2.3.6. These loci were Actin, TJCA-2, BSP-8, CM-2, 29 and 24. For these loci to be of use in this study, it was essential that the polymorphisms be present within British populations of *S. ratti*. To establish this, two British isolates (lines 29 and 32) were examined. DNA from 30 single worms were amplified at each of the polymorphic loci from lines 29 and 32 and DNA from 20 single worms from lines 5 Heterogonic and 54. The products from each PCR were then digested with the restriction endonucleases that had previously identified RFLPs.

Table 2.12 shows the distribution of RFLPs within the *S. ratti* lines examined. Actin, TJCA-2, BSP-8 and CM-2 were polymorphic within and amongst all the lines examined. Loci 29 and 24 were polymorphic within line 5 Heterogonic, but monomorphic within and between all other lines. Due to the criteria that loci must be polymorphic within British samples of *S. ratti*, loci 29 and 24 were not used any further as genetic markers in this study.

Locus	*Polymorphism present within lines?*			
	USA (5 Heterogonic)	Japan (54)	Scotland (32)	England (29)
Actin	y	y	y	y
TJCA-2	y	y	y	y
BSP-8	y	y	y	y
CM-2	y	y	y	y
29	y	n	n	n
24	y	n	n	n

Table 2.12 Presence of RFLPs within *S. ratti* lines analysed. y = polymorphism present and n = polymorphism absent. within the line.

2.3.8a Genetic analysis of the segregation of genomic PCR-RFLPs

It has been shown in a previous study that alleles of the Actin locus are inherited in a simple Mendelian manner (Viney *et al.* 1994). iL3s from line 132 Heterogonic were shown to be polymorphic for PCR-RFLPs at each of the loci TJCA-2, BSP-8 and CM-2. This line was subsequently used to examine the segregation of alleles at the loci.

To determine whether allelic segregation was occurring, the progeny of naturally mated heterozygous free living females were analysed for each locus (section 2.2.13). Nine heterozygous females were observed, four were heterozygous at locus BSP-8, four at TJCA-2 and one at CM-2. The progeny genotypes are shown in Table 2.13. Progeny that had different genotypes with respect to their mothers were observed for the BSP-8 and CM-2 loci, homozygous progeny being produced by a heterozygote parent. This result shows that allelic segregation had occurred at these loci.

Locus	*Female genotype*	*Observed progeny genotypes*		
		--	+-	++
	+-	3	3	1
BSP-8	+-	2	3	3
	+-	5	4	0
	+-	0	6	3
CM-2	+-	0	2	2
	+-	4	0	0
TJCA-2	+-	4	0	0
	+-	0	9	0
	+-	0	4	0

Table 2.13 Observed genotype frequencies from natural matings. "+" = presence of the polymorphic restriction site and "-"= its absence. Thus, "+-" denotes worms heterozygous and "--" or "++" worms homozygous for the RFLP.

2.3.8b Segregation of alleles of locus TJCA-2

The results from the progeny of naturally mated female worms at locus TJCA-2 were contradictory and unexpected: In two cases heterozygous females produced wholly heterozygous progeny; no segregation was observed at the locus in a total of 13 progeny. For two other females the progeny were homozygous despite being produced by heterozygous females. This apparently non-Mendelian inheritance prompted further investigation of the locus. Sequence analysis revealed close homology to bacterial 16S RNA genes. The highest match was a 96% identity between 500 bp of locus TJCA-2 and the 16S rRNA gene of a common soil bacteria, *Acinetobacter* sp (Table 2.11). No close homologies were found between this sequence and any nematode sequence.

To test whether this result was due to bacterial contamination, samples of water from faecal cultures of rats without *S. ratti* infections were prepared using the single worm DNA-preparation protocol (section 2.2.6). Preparations were made from faecal cultures one, two, three and four days old. Preparations from one and two day old cultures did not amplify the locus while that from three and four day old cultures amplified a fragment of the correct size. This demonstrates that (i) amplification of the TJCA-2 locus is not *S. ratti* specific and (ii) amplification of the locus is only possible after a period of culture. These results strongly suggest that PCR amplification of this locus is from contaminating bacteria that multiply within faecal cultures. The difference between the genotypes of the mother *S. ratti* and her progeny can be explained by the existence of two or more bacterial genotypes. The female worm would be expected to

be colonised by bacteria from the original faecal culture, and her progeny the bacteria that had grown in the 96 well plate. Thus, the progeny would have the same 'bacterial genotype', while the mother would have the 'bacterial genotype' of the original faecal culture.

Due to the non-*S. ratti* associated nature of this locus, it was excluded as a marker from the subsequent study.

2.3.8c Restriction maps of polymorphic loci

Restriction maps and gel examples for each of the loci used in the subsequent screen of British *S. ratti* are shown for the Actin locus (Figure 2.10), for the BSP-8 locus (Figure 2.11) and for the CM-2 locus (Figure 2.12).

2.3.9a mtDNA PCR-RFLPs

An 800 bp product was successfully amplified by primers 1120 and 1122 designed to sequences conserved between *C. elegans*/*A. suum* mtDNA (Figure 2.1). The expected fragment size, predicted from the *C. elegans* sequence was 819 bp. The sequences of the primers and conditions used to amplify the locus are shown in Table 2.14. The PCR product was cloned and sequenced as described in section 2.2.15a (sequence given in appendix). Sequence homologies of 68.3 and 69.7% were found between the *S. ratti* and the *C. elegans* and *A. suum* sequences, respectively.

The primers were used to amplify the locus from lines 29, 32, 5 Heterogonic and 54. 30 restriction endonucleases were used to survey the amplified sequence for RFLPs. Five of these cut at eight restriction sites, screening 34 bases in total. This compromises 2.7% of the total sequence. PCR conditions and restriction endonucleases used are shown in Table 2.14. No RFLPs were detected between any of the lines surveyed by these enzymes.

2.3.9b Direct sequencing of mtDNA PCR products

Cycle sequencing using primer 1121 was successful, sequence being obtained consistently for a minimum of 304 base pairs from the primer. This corresponds to bases 4815-5119 of the *C. elegans* mtDNA genome. The other two primers (1120 and 1122) did not produce readable sequence and were not used further.

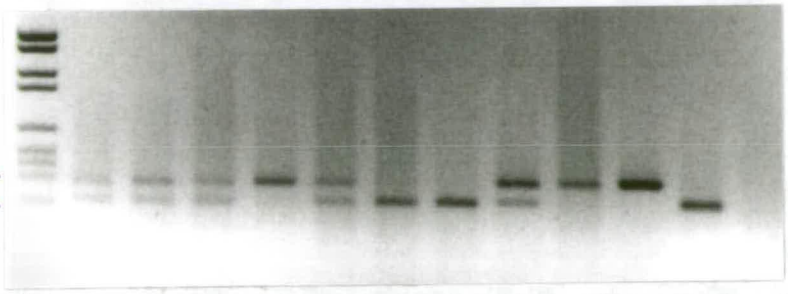
Primer 1121 was used to sequence single worm preparations from lines (5 Heterogonic, 68, 2924, 2932, Berks. 352 and Berks. 362) and a 5 Heterogonic

Figure 2.10

A. Genotyping *Mnl* I RFLPs of the PCR amplified Actin locus. Each lane is an amplification of a single iL3 from isolate 29. Lane 1 contains size standards, lane 13 the negative control PCR. Genotypes are shown beneath each lane; 11 = homozygote for absence of restriction site, 22 = homozygote for the presence of the restriction site and 12 = heterozygotes.

B. Map position of the variable *Mnl* I site within the Actin locus. Schematic restriction fragment profiles for each genotype are shown below.

A



size (bp)	2	3	4	5	6	7	8	9	10	11	12	13	: Lane
	12	12	12	11	12	22	22	12	11	11	22		: Genotype

B

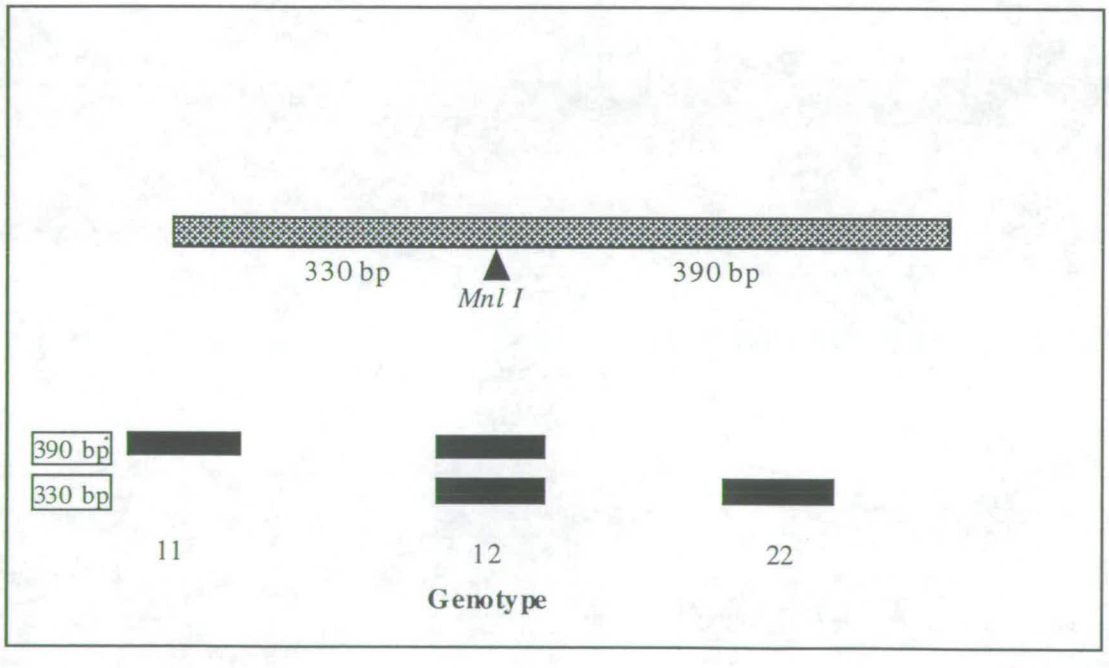
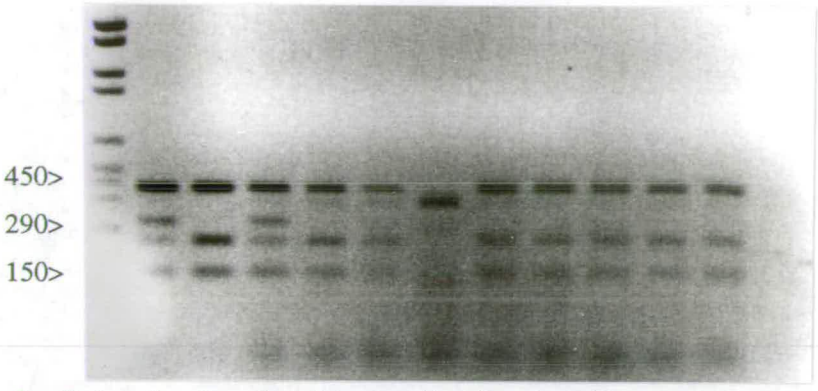


Figure 2.11

A. Genotyping *Hinf* I RFLPs of the PCR amplified BSP-8 locus. Each lane is an amplification of a single iL3 from isolate 29. Lane 1 contains size standards, lane 13 the negative control PCR. Genotypes are shown beneath each lane: 22 = homozygote for presence of restriction site, 12 = heterozygote. Lanes 2 and 4 are heterozygous worms, lanes 3, 5, 6, 8-12 homozygotes. The reaction in Lane 7 failed to amplify BSP-8, but did amplify a spurious smaller product which is also observed at low intensity in lanes 5 and 6.

B. Map position of conserved and variable *Hinf* I restriction sites with schematic restriction fragment profiles of each genotype shown below.

A



size (bp)	2	3	4	5	6	7	8	9	10	11	12	13	: Lane
	12	22	12	22	22	22	22	22	22	22	22		: Genotype

B

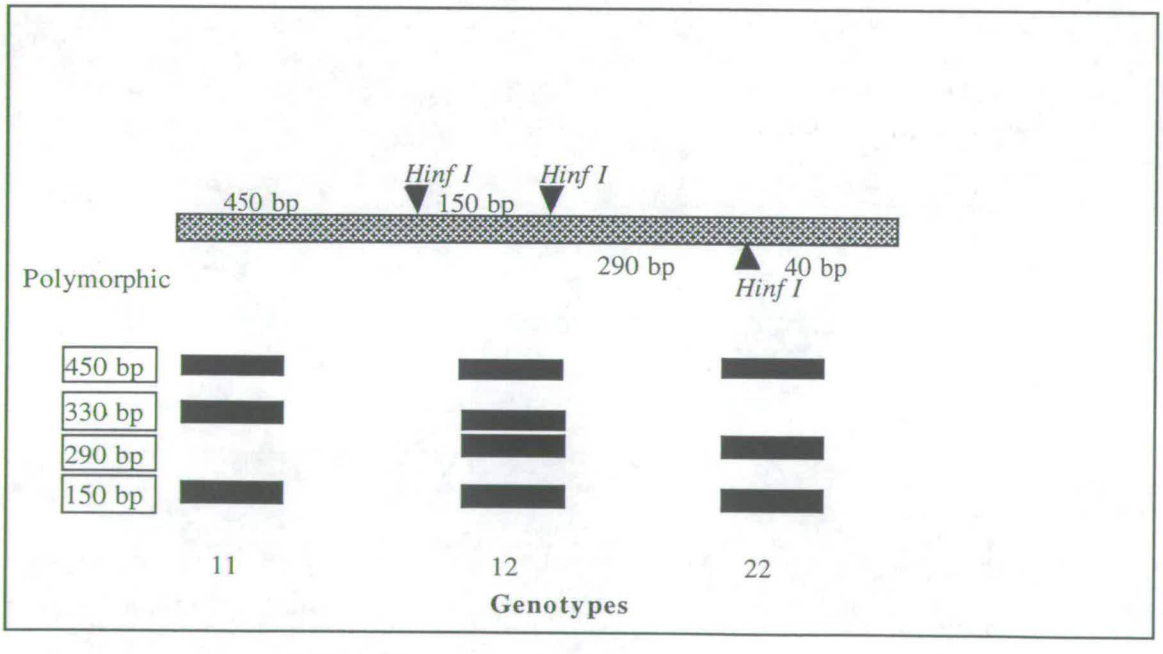
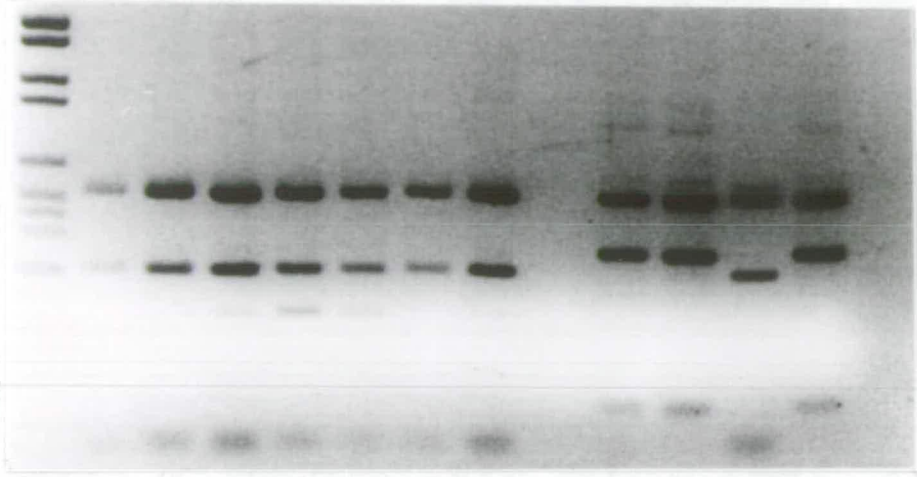


Figure 2.12

A. Genotyping *Hae* III RFLPs of the PCR amplified CM-2 locus. Each lane is an amplification of a single iL3 from isolate 29. Lane 1 contains size standards, lane 14 the negative control PCR. Genotypes are shown beneath each lane. Lane 12 (no amplification product) is a putative homozygote for a null allele. Lanes 3,4,5,6 and 8 contain faint non-specific amplification products.

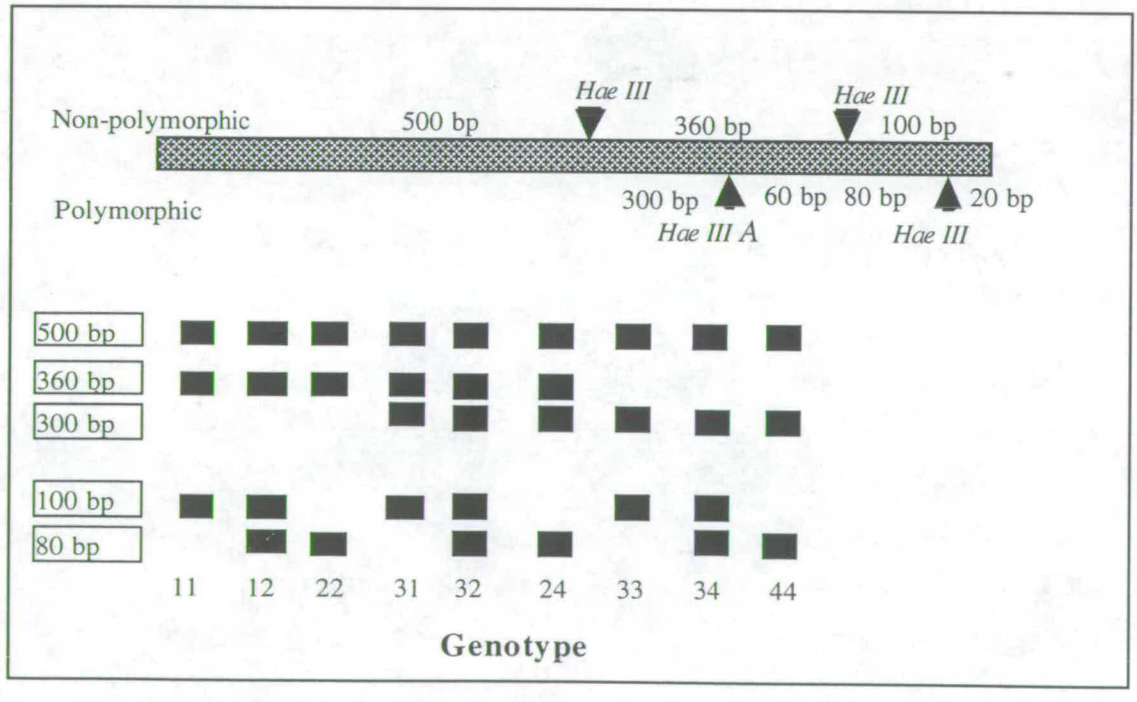
B. Map positions of the *Hae* III restriction sites within the CM-2 locus. Schematic restriction fragment profiles for each genotype are shown below.

A



size (bp)	2	3	4	5	6	7	8	9	10	11	12	13	14	: Lane
	44	44	44	44	44	44	44	-	11	11	44	11		: Genotype

B



genomic DNA preparation (Figure 2.13). The majority of bases were scored unambiguously although a small proportion were unreadable (denoted as "n" in Figure 2.13). These bases were unreadable due to coincident peaks of similar intensities. It is probable that these coincident peaks are due to the background 'noise', as only peaks of low intensity were affected. A total of 3040 bases of the *S. ratti* mtCytB gene were sequenced from nine single worms and one genomic template. One unambiguous base difference was observed (5 Heterogonic 1sw; position 190, Figure 2.1, this was a transversion from 't' to 'g'. This single variation corresponds to 0.00032 substitutions per site between the individuals tested or 0.032% of the overall sequence surveyed being variable within this group of worms.

2.4 Discussion

2.4.1 Microsatellites

Microsatellite sequences were successfully cloned from the genome of *S. ratti* and are the first such sequences to be described from a parasitic nematode. The presence of these sequences within *S. ratti* DNA supports the apparent ubiquity of microsatellites in eukaryotic genomes. It follows, therefore, that mechanisms causing the evolution of microsatellite sequences are active within the genomes of parasitic nematodes as well as other organisms.

In total, eight microsatellites were isolated using standard techniques to screen *S. ratti* genomic DNA libraries. Five were simple dinucleotide repeats with the repeat number ranging from 16-24 and the others compound microsatellites of two or more dinucleotide repeat types. The maximum and minimum size of the microsatellite repeats were 54 and 14 bp respectively. These sizes are within the range observed for microsatellites isolated from other species. In contrast, however, the distribution of microsatellite lengths from the genomes of *S. ratti* and *C. elegans* were almost non-overlapping. The maximum length of microsatellites from *C. elegans* was 16 bp compared to a minimum length of 14 bp from *S. ratti.* It is probable that the larger overall length of *S. ratti* microsatellites is a consequence of the sequences having been found by a specific screening protocol that is perhaps biased to selecting long sequences. *C. elegans* microsatellite sequences were isolated as a consequence of the *C. elegans* genome project and not specifically searched for. The small length of the sequences, therefore, is probably not a reflection of the size of *C. elegans* microsatellites as a whole. Whether or not the difference between the two species was an artifact of the screening technique could be confirmed by isolating microsatellites

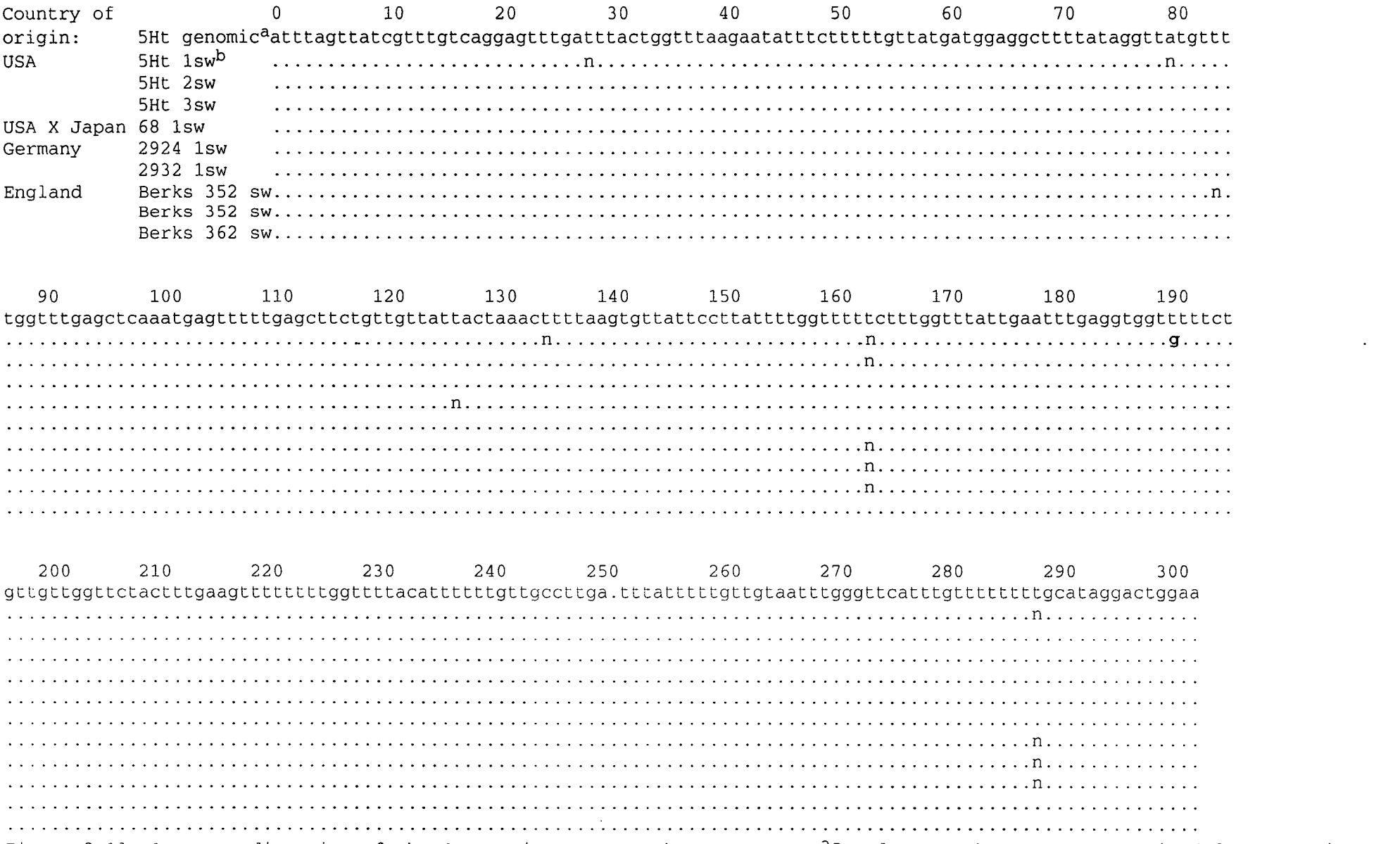


Figure 2.13. Sequence diversity of the *S. ratti* mtDNA Cytochrome B gene. [a]5Ht denotes the sequence obtained from genomic DNA for the isofemale line 5 heterogonic. [b] sw denotes the sequence obtained from single iL3s. n = denotes bases of ambiguous identity. Polymorphic bases are highlighted in bold.

from *C. elegans* by the method used in *S. ratti*. Indeed, this would provide useful data on comparative frequencies, length and types of microsatellites in the two species.

The attempts to use the *S. ratti* microsatellite sequences for population studies were inspired by the success of tandem repeats as polymorphic markers within other organisms. However it was found that, with a single exception, these loci could not be amplified from *S. ratti*. This result was surprising. DNA sequences isolated from an organism should, under appropriate conditions, be able to be amplified. Indeed, the cloning of other non-microsatellite sequences followed by PCR amplification was used routinely in other parts of this study. Only one locus, 5Ht, could be amplified from genomic DNA. However, even this locus could not be consistently amplified from single worms. Amplification from single worms was an absolute prerequisite for loci to be used in this study, so despite successful amplification this locus was not useful as a population genetic marker.

For the other microsatellite loci, either none or multiple PCR products were obtained. Extensive attempts were made to optimise these reactions but in no case was a single reproducible band obtained. A common reason why loci do not amplify is inappropriate primer design. In order to ensure that primer design was not at fault, second, and in some cases third generation primers were designed to flanking sequences of the loci. However this still did not result in specific amplification of the loci, suggesting that the failure to amplify was not due to inadequacies of the primers. Attempts were made to enhance the stringency of primer binding by using 'touchdown' (Don *et al.* 1991) and 'hot start' (Chou *et al.* 1992) PCR (Table 1, cycles B and H). Both techniques increase the specificity of the PCR reaction, touchdown PCR by ensuring only specific amplification of target DNA occurs during the first few thermocycles, and hot start PCR by reducing low-temperature mis-priming and primer dimerisation during the first round of amplification. However, although both techniques increased the effectiveness of PCR with primers to the locus 5Ht they did not have a noticeable effect on amplification of the other loci.

It may be that the failure to amplify is due to a higher order feature of the part of the genome containing these microsatellites which interferes with the efficiency of the amplification reaction. For instance, if these microsatellites were embedded in repetitive sequences there may be multiple priming sites which would probably result in the amplification of multiple bands. Indeed, multiple bands were commonly observed during attempts to amplify these loci (Table 2.4). Further work on these loci

should include the cloning and sequencing of these bands to determine whether or not they contain the microsatellite and/or homologous sequences containing the primer sites, or variants of the primer sites. In this way it would be possible to characterise the regions of the genome in which these loci occur more fully.

If microsatellites commonly occur in regions of a genome that are difficult to amplify, then failures in amplifying microsatellites should perhaps be a common observation for other organisms. While this does not seem to be generally the case, there is some indication that a significant proportion of isolated loci are not suitable for further use. A recent paper by Bell & Ecker (Bell & Ecker 1994) closely parallels my own experience. In *Arabidopsis thaliana*, only three of 22 $(CA)_n$ microsatellites isolated could be amplified despite extensive attempts at optimisation of the PCR conditions. Why this was so remained a mystery, but in the same study 30 of 37 $(GA)_n$ microsatellites could be amplified, suggesting that the type of dinucleotide repeat may be important in determining ease of amplification.

This first attempt to identify polymorphic loci in the genome of *S. ratti* found a number of non-amplifiable microsatellites. While it is not clear why this was so, the high redundancy that we have observed in cloning and successfully amplifying these loci meant that they were of no use as genetic markers for *S. ratti* in the subsequent population study.

2.4.2 Other nuclear sequences

Due to the lack of success in utilising microsatellite sequences as population genetic markers, three alternative methods were used to amplify and survey *S. ratti* genomic loci for polymorphisms. These were (i) PCR-RFLPs of anonymous *S. ratti* loci, (ii) RAPD PCR-RFLPs of anonymous *S. ratti* loci and (iii) amplification of homologous *C. elegans* loci from *S. ratti* DNA. Methods (i) and (ii) were used to amplify ten anonymous loci from *S. ratti* DNA, these sequences then being surveyed for polymorphisms by the use of restriction endonucleases. Sequences isolated by method (i) revealed four polymorphisms in three loci (24, 27 and CM-2), and method (ii) revealed none. These methods provided a rapid procedure for surveying loci in large numbers of individuals. The potential number of loci was unlimited and screening for polymorphisms was simple and unambiguous. No radionucleotides were required, and PCR-amplified DNA was able to be digested and polymorphisms scored within 24 hours.

2.4.2b Genetic variation in *S. ratti*

Of the loci amplified from *S. ratti* DNA, method (i) (including the 18S rDNA locus) revealed four nuclear DNA polymorphisms and method (ii) none. Why a difference exists between the two methods in detecting genetic variation is unclear. However, it is likely that sampling variance has an effect and that the differences observed are attributable to the relatively small number of sequences assayed by these methods.

From these observations, 0.25% of nucleotides, averaged across loci, were polymorphic in *S. ratti*. This is comparable to that found in free-living organisms, which ranges from 0.03 - 0.57% (Lynch & Crease 1990). A study of the introns of α-tubulin genes of *Haemonchus contortus* measured levels of nucleotide diversity of 9.4% and 9.1% at two loci from three Canadian parasite isolates (Beech *et al.* 1994). This is considerably higher than the range found in free-living organisms and higher than that found in *S. ratti*. Reasons that may account for this discrepancy are discussed in the final chapter of this thesis.

Method (iii) was unsuccessful with respect to amplifying *C. elegans* microsatellite containing loci from *S. ratti*. Microsatellite loci have been successfully amplified from species that have been separated for as long as 470 million years (Rico *et al.* 1996). However, a more general observation is that the longer the period of divergence between two organisms, the less chance there is of amplifying homologous loci. Estimates of the time of divergence between *C. elegans* and *S. ratti* are between 560 and 370 million years. (M. Blaxter, pers. comm.), estimated from sequence divergence between nematode globin genes using the mammalian clock rate. It may be that the period of divergence between the two species is too great to allow the amplification of homologous microsatellite loci. In comparison, amplification of the 18S rDNA sequence was successful using universal nematode primers. These primer sites are highly conserved between species and successfully amplify this locus in all nematode species so far studied (M. Blaxter, pers. comm.). Consequently the intervening sequence also shows a relatively high degree of conservation between species making this an unlikely sequence to find intraspecific genetic variation. This expectation was confirmed on screening the locus for RFLPs, as none were detected.

2.4.3a Distribution of variation between *S. ratti* lines

In addition to the polymorphisms isolated by the methods described above, primers to three loci previously observed to contain polymorphisms were donated by

M. Viney, namely Actin, BSP-8 and TJCA-2. Analysis of the distribution of the RFLPs between lines showed that, of the six loci containing polymorphisms (Actin, BSP-8, TJCA-2, CM-2, 29 and 24), two were monomorphic within lines isolated from Britain and Japan (29 and 24). These two loci were as a consequence removed from the suite of markers used to further analyse British samples of the parasite.

2.4.3b Segregation of alleles at each locus

The Mendelian nature of alleles at each locus was examined by analysing the genotypes of naturally-mated females and progeny. With one exception, alleles at each locus showed segregation. A single locus, TJCA-2, exhibited a non-Mendelian pattern of inheritance. The locus was subsequently shown to be amplified from the 16S RNA gene of a non-*S. ratti* associated bacteria and was not used further within this study.

2.4.4 Polymorphism of mitochondrial DNA sequences

Measurements were made of sequence diversity within the *S. ratti* mtDNA Cytochrome B gene. Genetic diversity was low within this locus. A 1250 bp fragment showed no variation when surveyed by restriction endonucleases. This result was corroborated by direct sequencing of a 304 bp segment of the CytB gene, one nucleotide being polymorphic out of a total of 3040 surveyed corresponding to a nucleotide diversity of 0.00032 substitutions per site between the individuals tested.

Direct sequencing of the ND4 and LNC (variable-length long-noncoding) regions within *O. ostertagi* mtDNA found within population diversities of 0.027 substitutions per site (Blouin *et al.* 1992). High levels of nucleotide variation were also found for the mtDNA of *Haemonchus contortus* (0.026%), *Teladosargia circumcincta* (0.024%), *Haemonchus placei* (0.019%) and *Mazamastrongylus odocoileus* (0.028%) within the USA (Blouin *et al.* 1995). Lower values were recorded for *Ascaris suum* (0.004%) (Anderson *et al.* 1993) in Guatemala. The observed *S. ratti* diversity is one to two orders of magnitude less than that of these nematodes. This may be attributable to functional constraints on the Cytochrome B gene. Indeed, there has been shown to be considerable variability in mutation rates between different mtDNA genes within a species (Hillis & Moritz 1990). However studies on *A. suum* have found extensive variation within the Cytochrome B gene (Anderson *et al.* 1993) illustrating that functional constraints on these sequences within nematodes, if present, are not absolute.

A general conclusion of these studies is that while genetic diversity within mtDNA is present within *S. ratti* samples, it is appears low relative to that of other species. Thus, lines initiated from isolates collected in Japan and the USA are remarkably similar despite the large geographical distances separating these samples. Conservatively, assuming that the diversity of *S. ratti* mtDNA is half that observed within *O. ostertagi*, we would expect to find, on average, 40 nucleotide substitutions among ten individuals within a 304 bp sequence. Only one was observed.

Reduced levels of diversity may be a consequence of historical processes such as population bottlenecks or random genetic drift within small populations. The effect of both processes is to reduce variability within populations and to increase differentiation among populations (Tajima 1989). Indeed, mtDNA will be expected to be affected more by these processes due to a smaller effective population size relative to nuclear loci (Karl *et al.* 1992). It may be that at some time in the past, *S. ratti* populations underwent a population bottleneck followed by a rapid range expansion and that this may explain the relatively low levels of genetic diversity within globally-distributed isolates of both mtDNA and nuclear loci. Indeed, it is possible that the spread of *Rattus norvegicus* in association with human movement may have caused a rapid recent expansion of the *S. ratti* populations. The present distribution of *Rattus norvegicus* is thought to have arisen from a focus near the Caspian sea, to have arrived in Europe in the 1600s and to have colonised Britain in the 1720s (Kroyer 1991). Thus, it is possible that the population of parasites found worldwide, and in Britain, may have originated from a relatively few individuals. Investigation of the population genetic structure of *S. ratti*, described in Chapter four, will test this hypothesis.

2.5 Summary

This chapter has described the methods leading to the isolation of polymorphic DNA sequences from the genome of *S. ratti.* The isolation of *S. ratti* microsatellites was successful but amplification of these loci from iL3s proved to be unsuccessful. Surveys for alternative polymorphic anonymous genetic loci successfully identified three loci containing RFLPs. A group of six polymorphic loci became available for analysis of the genetic structure of *S. ratti* populations. Two loci (29 and 24) were monomorphic within lines isolated from British samples of the parasite and were excluded from further study. A third locus (TJCA-2) proved to be a consequence of bacterial contamination and was also excluded. Studies on the inheritance of the remaining markers showed that they were inherited in a conventional manner, undergoing genetic segregation during the sexual phase of the *S. ratti* life cycle. RFLP surveys and direct sequencing of a segment of the mitochondrial genome were used as a method of characterising mtDNA polymorphisms. Levels of polymorphism in this molecule were found to be too low for it to be used successfully as a means of differentiating individual parasites and it was excluded from further study.

Thus, Actin, BSP-8 and CM-2 were used as genetic markers in the rest of this study.

Chapter three

Sampling wild populations of *Rattus norvegicus* for *Strongyloides ratti.*

3.1 Introduction and rationale

This chapter describes the sampling of natural populations of *Strongyloides ratti*. The samples provide a basis for an analysis of the structure of *S. ratti* populations by the use of polymorphic genetic loci.

To describe patterns of genetic variation within a species, allele and genotype frequencies are measured within a number of samples. Populations of *S. ratti* were sampled to allow an analysis of the distribution of genetic variation to be made at four levels: within individual parasites, within samples from a single host, between hosts from a single sample site and between sample sites. By a comparative analysis of variance within and between each of these four sampling levels, the extent of genetic differentiation was measured for the parasite samples. F-statistics and analyses of genotypic proportions were used to measure inbreeding. The genetic analysis is described in Chapter 4.

Such data need to be considered together with information on other key aspects of the parasites' life history in order to understand fully the observed patterns of genetic variation. As described in Chapter 1, factors of particular importance are (i) the amount of sexual reproduction occurring within natural populations of the parasite, and (ii) the frequency distribution of *S. ratti* within rats. Laboratory studies are described which measure these parameters.

Finally, this study has used a point sampling methodology. In order to show that this is a valid approach, data are presented on longitudinal studies of the variance of allele and genotype frequencies of iL3 samples from artificially and naturally infected rats.

3.1.1 Sampling methodology

Previous studies have established that *S. ratti* occurs widely in the UK (Viney *et al.* 1990). Surveys of brown rats recorded *S. ratti* infections in both England (3 of 20 rats positive, collected in West Dean, Sussex) and Scotland (2 of 2 rats positive,

collected in Leith and Craigmillar, Edinburgh). My work has aimed to enlarge upon these studies by sampling *S. ratti* from rats collected from a variety of geographical locations across Britain.

3.1.1a Sampling natural populations of *S. ratti*

The basic unit from which samples were taken was individual rats. There are three methods by which *S. ratti* may be sampled:

(i) Direct sampling of the parasite infra-population (the parasites within a single host) of hosts. Rats are sacrificed and the gut removed. Parasitic females of *S. ratti* are then dissected away from the intestinal tissues (Viney 1994).

(ii) Indirect sampling of individual rats. Faecal samples from rats are cultured (section 2.2.4). Free living adults and iL3s are subsequently recovered from culture.

(iii) Indirect sampling from a rat population: Faecal samples are collected from the ground of rat infested sites which are then cultured as in (ii).

Method (iii) is a simple method and provides samples of *S. ratti* iL3s from an undefined rat population, but obviously it does not allow comparison of the parasites between individual hosts. Method (i) potentially provides the most accurate sample of the parasites of a single rat. However the small size of the adult parasite (1.85 - 3.30 mm) makes successful and complete recovery from the gut a difficult task. As a consequence, surveys of rat parasites do not always record *S. ratti* infections even though their prevalence is high within Britain (Webster & MacDonald 1995; Owen 1976). Another limitation of method (i) is that sacrifice of the host means that only single samples can be made, thereby rendering longitudinal sampling impossible. The choice of this study was method (ii). This method is non invasive, provides samples of infections from individual hosts and is highly sensitive, being able to detect infections of single parasitic females (Viney *et al.* 1992). However, this method samples the progeny of parasitic females, and not the adult parasites themselves. A consequence is that as the intensity of infection decreases, the probability that the progeny from a single rat are derived from a single parasitic female increases. This necessitates particular statistical considerations when analysing population genetic data from single hosts, and this is described in Chapter 4.

3.1.1b Sample sizes

The question of how many parasites and hosts need to be sampled to estimate population allele frequencies was first considered. Estimates on the frequency of an allele *a* has a standard error that is dependent on sample size (Figure 3.1.1). Standard errors were calculated according to Weir (Weir 1996):

$$Var(\tilde{a}) = \frac{1}{2n}a(1-a) \qquad \text{(eqn 3.1)}$$

which is variance of a binomial distribution with parameters a and $2n$, where n is the sample size. The magnitude of the standard error also varies with the frequency of a within the population, estimates made when a is at a high frequency being less accurate than those made when the frequency of a is low. As a consequence, a lack of knowledge of the frequency of alleles within a population before the study commences limits the accuracy of estimating the sample sizes required. However, it is possible to calculate the standard errors associated with particular sample sizes and allele frequencies as Figure 3.1.1 illustrates.

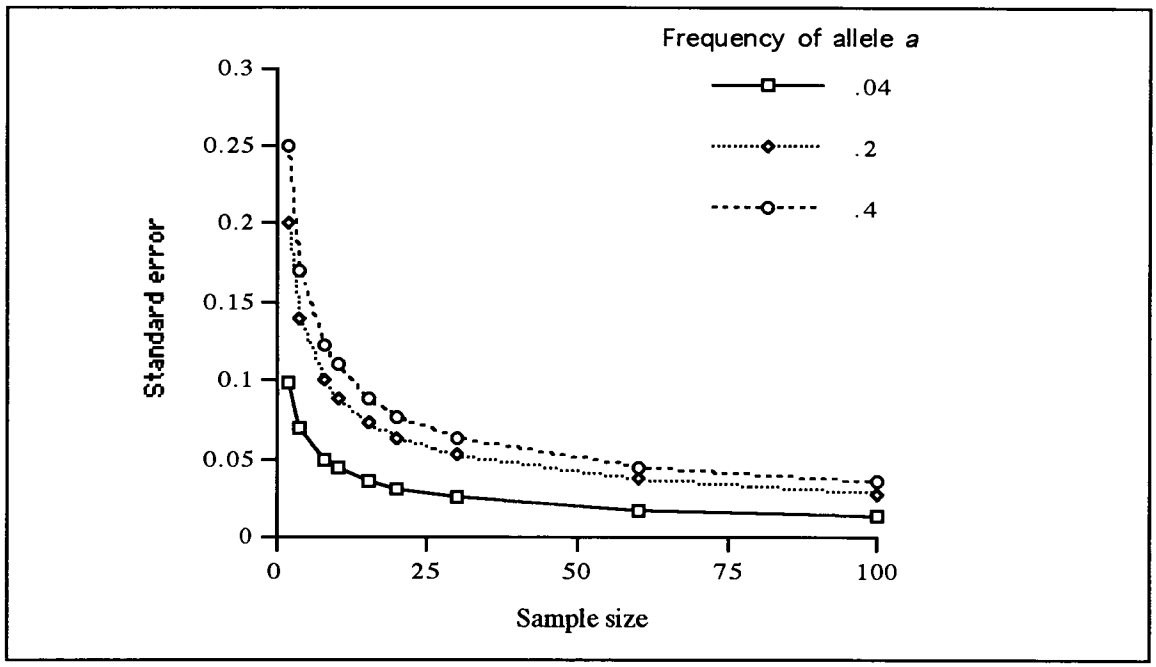


Figure 3.1 The relationship between standard error and sample size for populations with frequencies of allele *a* of 0.04, 0.20 and 0.40.

It can be seen that in order to estimate the allele frequencies of a population with an associated standard error of less than 10% (for alleles with frequencies of up

to 0.5), sample sizes of ten or more are necessary. Doubling the size of the sampled population to 20 individuals reduces the standard error by only 3%, 2% and 1% where the allele frequencies are 0.4, 0.2 and 0.04 respectively.

In the present study, the sample size of iL3s to be collected from rats was set at a minimum of ten for infections producing ten or more larvae. For infections that produced less than ten larvae, all were collected.

In order to compare the frequencies of alleles in the parasites infecting different hosts, the minimum number of sampled rats required is two (*i.e.* two sampled infections). However, in order to measure population structure, several infected rats have to be sampled within at least two sample sites. As it was not known in advance what the prevalence of infection would be in wild rat populations, and what amounts of genetic variation would be found within natural populations of the parasite, it was impractical to specify the number of individual rats that needed to be sampled. Therefore, for each rat colony as many individuals as possible were trapped.

Finally, there is a balance between how intensely a site is sampled and the number of sites sampled within the time available. Wilson (Wilson 1996) used computer simulations to estimate the power of available statistical tests to detect population subdivision for different numbers of sample sites and individuals within sites. He concluded that the precision of the estimates was best when large numbers of sample sites were sampled for a few individuals, rather than when intensive sampling was carried out at a few sites. Therefore, in this work I attempted to sample from as many sites as possible.

3.1.2 Frequency of sexual reproduction within populations of *S. ratti.*

The relative frequency of sexual versus asexual modes of reproduction will affect the population genetic structure of *S. ratti. S. ratti* has the ability to develop either heterogonically, which includes sexual reproduction, or homogonically, in which sexual reproduction is absent (Figure 1.1).

If the frequency of sexual reproduction within natural populations of the parasite is low or absent, *i.e.* homogonic development predominates, there will be little or no genetic segregation. This will result in a clonal population structure characterised by linkage disequilibria between genetic markers and deviations from Hardy-Weinberg expectations for genotype frequencies in a randomly mating

population (Tibayrenc *et al.* 1990). If sexual reproduction is frequent, then samples of the parasite will show linkage equilibria, and Hardy-Weinberg expectations.

Studies by Viney (Viney *et al.* 1992) have shown that the development in the laboratory of four *S. ratti* lines isolated from rats in Britain was almost exclusively homogonic. To determine whether this is a general observation for *S. ratti*, the frequency of homogonic development was examined for a subset of the isolates collected in this work.

3.1.3 Frequency distribution of *S. ratti* infections

The distribution of macroparasites within a host population is characteristically overdispersed, with a minority of hosts harbouring the majority of parasites (Anderson *et al.* 1992). This distribution is mathematically well described by the negative binomial distribution (Crofton 1971b). In order to investigate whether natural populations of *S. ratti* conform to this distribution, a method to measure the numbers of *S. ratti* within a host was necessary. Measurements on intensity of infection may be made by either; (i) direct counts of the numbers of parasitic females in *post mortem* examinations (Stear *et al.* 1996) or from faeces of anthelminthically treated hosts (Anderson *et al.* 1993) or (ii) indirect counts of parasite numbers by relating egg or larval faecal counts to numbers of parasites on the basis of expected adult parasite fecundity (Sinniah *et al.* 1983). As discussed in section 3.1.2, method (i) was not appropriate or practicable for use in this study. Therefore, counts of larvae in faecal cultures from naturally infected rats were made. To relate such faecal larval counts to the numbers of parasitic female *S. ratti* giving rise to them, fecundity was measured in experimental infections by counting the numbers of larvae produced each day from infections of single worms. Estimates of the numbers of parasitic female *S. ratti* within wild caught infected rats was then made from these data.

There are limitations to this technique. It is well documented that individual parasite fecundity decreases as intensity of infection increases *i.e.* that negative density dependent effects exist on parasite fecundity (Anderson & May 1992). This 'crowding effect' may be due to a combination of limited resources for high numbers of parasites and increased host immune responses. Such effects may result in an underestimation of the number of parasites within heavily infected hosts if measurements of *single* parasite fecundity are used as a measure of *average* parasite fecundity. However, it has been suggested that measurements demonstrating density dependent effects on fecundity result from sampling aggregated parasite populations (Keymer & Slater

1987). This is due to the majority of samples coming from hosts with low intensity infections, variations in individual parasite fecundity creating an illusion of density dependence. The accuracy of faecal egg counts in determining the intensity of infection has been empirically tested in several studies. A general conclusion is that 'crowding effects' are unimportant and that the use of faecal egg counts is a robust technique for predicting the numbers of adult worms harboured by a host (Stear *et al.* 1995; Bundy 1986; Haswell-Elkins *et al.* 1987b).

3.1.4 Longitudinal analysis of *S. ratti* infections

Conclusions on population genetic structure made from a single faecal sample have the caveat that the observed pattern of egg production of an adult worm may vary over time. Some studies have shown that faecal egg counts in nematode species tend to remain stable over time (*e.g.* Nawalinski *et al.* 1978). However, there are no studies that have directly measured longitudinal variation in the fecundity of a single parasite. If there are large temporal variations in individual parasite fecundity, then allele frequencies made from single samples of larvae may vary over time. This effect will be especially pronounced at low intensity infections where contributions by individual parasites to the number of sampled larvae is greatest.

To address this question, measurements were made on longitudinal variation in larval production from individual parasitic female *S. ratti* in the laboratory. Subsequently, small samples taken from an *S. ratti* line known to be polymorphic at a locus were used to infect laboratory rats. Variation in genotype frequencies over the course of the infections were followed for iL3 samples from each rat. Finally, variation in allele frequencies between samples of iL3s taken at two time points, six months apart, were measured from naturally infected rats.

3.2 Materials and Methods

3.2.1 Sampling of Brown rat populations

Wild brown rats were trapped from rural UK farmsteads using Bledbury style live-catch cage traps (Killgerm), 25 traps were positioned around farms identified as having rat infestations. These were pre-baited for one week with a grain/lard mixture and caught animals removed each day after their sex and weight were recorded. Animals were subsequently kept isolated within wire-mesh bottomed steel cages. Rats were sampled from nine English farms separated by distances ranging from 10 - 340

Km (Figure 3.2). Two farms were sampled within Scotland 40 km apart from each other and samples were acquired from sites within Germany.

Trapping was carried out between the dates January 1995 and July 1995 at farms at Edgefield, Nether Fala and Oxford, Tubney Manor. The Ministry of Agriculture and Fisheries (M.A.A.F. Central Science Laboratories) collected rats from farms in Norfolk, Surrey, Berkshire, Wiltshire and Dorset between June and August 1995. Rats were trapped in Germany by Dr. H. Pelz (Federal Biological Research Center, Munster) between February and September 1995 and maintained in isolation, faeces being provided for the study.

3.2.2 Sampling *S. ratti* from rats

Samples of faecal material were collected from individual captured rats over a 12 hour period, and 6g cultured (section 2.2.4). After three days of culture at 19 °C, iL3s were removed from each faecal sample. These were washed twice and placed singly in 0.5 ml microcentrifuge tubes in 5 μl of distilled water, given a unique identification number and stored at -20 °C. If less than ten larvae developed from an infected rat, then all the larvae were collected. If more than ten developed, a maximum of twenty larvae were collected.

3.2.3 Sexual reproduction within populations of *S. ratti.*

iL3s cultured from wild-caught rats were used to infect female Wistar rats by subcutaneous inoculation (section 2.2.3). Infections were made using parasites collected from individual rats from Surrey A (one rat), Wiltshire (one rat), Berkshire (five rats), Norfolk (three rats), Dorset (two rats) and Germany (two rats).

To determine the homogonic index, faecal cultures were made from the infected rats and maintained at four different temperatures; 4 °C, 13 °C, 19 °C and 30 °C. The numbers of larvae developing by the heterogonic and homogonic routes were counted in each culture after three days. For each line the homogonic index (Hi) was calculated. This value corresponds to the proportion of larvae developing by the homogonic route of development (Viney *et al.* 1992) and ranges from 0 (all larvae develop into free living males and females) to 1 (all larvae develop directly into iL3s). A minimum of 100 worms were counted for each culture. The infections and worm counts were undertaken in collaboration with A. Gemmill.

3.2.4 Frequency distribution of *S. ratti* infections

6g of faecal material from an overnight collection were cultured (section 3.2.2) from all rats from the Berkshire sample site. The number of larvae in three day old cultures were counted for each sample and the numbers of larvae/g faeces calculated. Maximum log-likelihood estimates were used to fit two models to the data; (i) a Poisson (random distribution) model with mean x, and (ii) a negative binomial model with mean x, variance s and overdispersion factor k. The parameter k was estimated by maximum likelihood. Analysis of variance (ANOVA) was used to calculate the deviance of both models relative to the observed data. All calculations were performed using the S-plus statistical package (AT&T).

The effects of rat weight and sex on counts of larvae/g faeces were investigated. Counts of larvae/g faeces were normalised by $\log_{10}$ transformation. Regression analysis was used to determine correlations of rat weight with larval counts for (i) the pooled sexes and (ii) each sex independently. Differences in the distribution of larval counts between sexes were then examined by a one-way unstacked ANOVA.

3.2.5 Estimation of fecundity of individual parasitic females

Single parasite infections were made by inoculating female Wistar rats with single iL3s from two different lines;

Group 1: Eight rats were infected with iL3s from line 5 Heterogonic. Seven rats developed patent infections. Individual parasite fecundity was determined by making cultures from the total overnight faecal production of rats at days 5, 6, 7, 8 and 9 pi. All worms were removed from three day old cultures and the numbers present were counted. Each count corresponded to the overnight fecundity of the parasitic female in each respective rat.

Group 2: 20 rats were infected with iL3s from line 132 Heterogonic, 13 of which developed patent infections. Individual parasite fecundity was determined from the overnight faecal production of rats at days 8, 9 and 12 pi. Collections were weighed and 6 g, or all the faecal material produced if less than 6 g was present, cultured (section 2.2.4). All worms were removed from three day old cultures and the numbers present counted. For rats that produced less than 6g of faeces, the count corresponded to total overnight parasite fecundity. For rats that produced more than

6g, the count was multiplied by the factor (total weight of faeces/6) to account for the uncultured faeces.

Mean overnight parasite fecundity was calculated for each infection and 95% confidence intervals estimated from the *t* distribution, calculated as $x \pm t_{.95}s/\sqrt{n}$, with six degrees of freedom for group 1 and 12 for group 2.

3.2.6 Estimation of the number of parasitic females from faecal larval counts

Larval counts/g faeces were calculated for each wild rat sampled from Berkshire (section 3.2.4). The number of parasites harboured by these rats was estimated as follows:

The mean overnight parasite fecundity/g of faeces (Fg) was calculated from:

$$Fg = \frac{Fo}{P} \qquad \text{(eqn. 3.2)}$$

where Fo = mean overnight parasite fecundity and P = average overnight faecal production per rat in g. The quantity P was determined for twenty 90g (weight on delivery) rats on two consecutive nights. Fg was then used to estimate the intensity of infection by simply dividing the observed numbers of larvae/g faeces, measured for each naturally infected rat, by Fg.

This method makes assumptions that (i) variation in the weight of faeces produced by rats of different weights has little effect on faecal larval counts, (ii) parasites are of equal fecundity within large and small rats, (iii) parasites are of equal fecundity within wild-caught and laboratory rats and (iv) density-dependent effects on fecundity do not occur. However, while there are expected to be large errors associated with such estimates of intensity infection, it will be used here to provide a first approximation of parasite load.

3.2.7 Longitudinal analysis of *S. ratti* infections

The temporal variation in genotype frequencies of *S. ratti* infections within single hosts was measured in laboratory infections (section 3.2.7a) and natural infections (3.2.7b) of rats.

3.2.7a Laboratory infections

From studies described in Chapter 2, line 5 Heterogonic was known to be polymorphic at the Actin locus with frequencies of 0.43 for allele 1 and 0.57 for allele 2. This line was used as a founding population for low intensity infections. Three female Wistar rats were infected with iL3s from line 5 Heterogonic in the following proportions:

Rat A: Infected with 12 iL3s
Rat B: " 12 iL3s
Rat C: " 6 iL3s

Faeces were collected on days 8, 11, 15, 19, 25 and 29 post infection, and cultures set up from 6g amounts. Worms were removed from three day old cultures and the numbers present counted. Ten iL3s were randomly sampled from each culture and single worm DNA preparations made (section 2.2.7). Each worm was then genotyped at the Actin locus (section 2.3.3).

Allele frequencies and their associated 95% confidence intervals were calculated according to Weir (Weir 1996) (section 4.2.2).

3.2.7b Natural infections

A number of wild caught rats were kept confined over a period of six months. All rats that had *S. ratti* infections at the beginning of this time period retained positive infections. 12 of these rats were sampled at two time points, three months apart, from the following sites:

Wiltshire (W), rats; W313, W332, W330, W331
Dorset (D), rats; D335, D338, D339
Surrey A (SA), rats; SA27, SA31, SA28, SA26
Surrey B (SB), rats; SB28

Ten iL3s were genotyped at the Actin and BSP-8 loci (section 2.3.3) for each rat. The genotype frequencies at each time point were compared by χ^2 tests.

3.3 Results

3.3.1 Sampling of rat populations

123 brown rats were trapped and sampled at 11 sites; 27 rats at two sites in Scotland and 96 rats at nine sites in England. Samples from 19 rats collected from eight sites in Germany were provided by Dr. H. Pelz. The overall prevalence of infection within Britain was 62% (standard deviation = 40%); none of the 27 rats sampled in Scotland were positive for *S. ratti* infections. 79% (standard deviation = 26%) of the English rats were positive for *S. ratti* with prevalences of infection varying by site from 20 - 100% (Table 3.1, Figure 3.2). Three of the 19 German rats were positive for *S. ratti* (prevalence of infection = 11%, standard deviation = 22%), with prevalences ranging from 0 - 66% (Table 3.2).

From the trapped rats, a total of 1472 iL3s were collected with a mean of 19 iL3s sampled per rat. These were stored at -20°C for subsequent genetic analysis.

3.3.2 Sexual reproduction within populations of *S. ratti.*

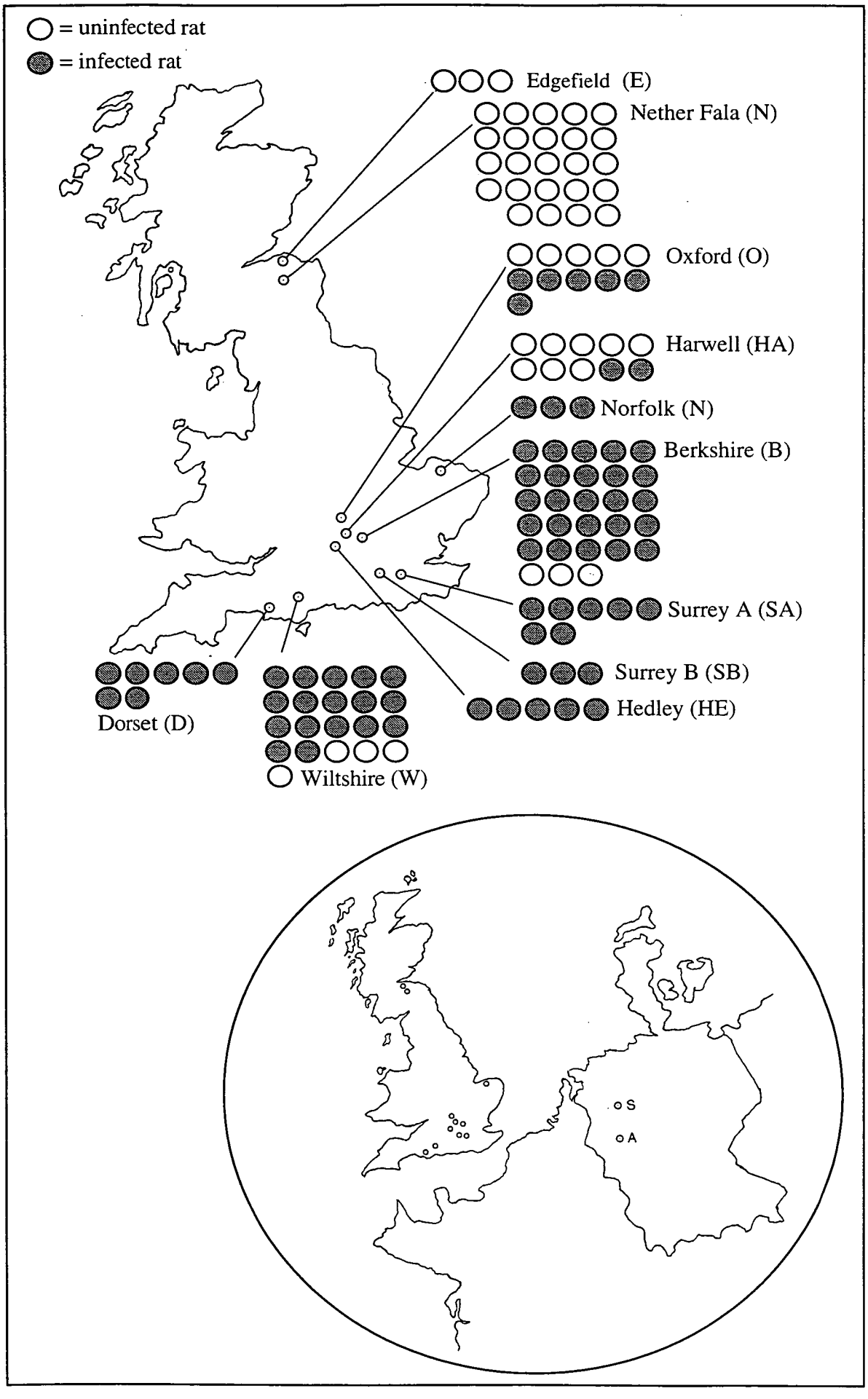
Frequencies of larvae developing by the homogonic route, measured by the homogonic index (Hi) are presented in Table 3.3. There was no significant effect of temperature on Hi in any isolate (tested by the χ^2 statistic). It was therefore statistically appropriate to pool the data from different temperature treatments for each rat, the mean value being shown in the table.

Sample site (Rat No.)	Hi	*Sample site (Rat No.)*	Hi
Surrey A SA 31	0.995 (0.031)	Norfolk N383	1.000 (0.000)
Wiltshire W313	0.992 (0.017)	Norfolk N384	1.000 (0.000)
Berkshire B352	0.995 (0.023)	Norfolk N385	1.000 (0.000)
Berkshire B362	0.965 (0.022)	Dorset D355	0.998 (0.006)
Berkshire B370	0.995 (0.024)	Dorset D337	0.985 (0.006)
Berkshire B369	0.995 (0.009)	Germany G25	1.000 (0.000)
Berkshire B377	0.995 (0.036)	Germany G32	1.000 (0.000)
			Mean[a] = 0.994 (0.018)

Table 3.3 The homogonic index (Hi) for each *S .ratti* population. For each site, the Hi is combined for all temperatures.[a] Mean value for all rats. Values in parentheses are standard deviations of the mean. Data supplied by A. Gemmill.

Figure 3.2

Sampling locations within Britain and Germany. Positions shown are the sample sites. Circles represent trapped rats, while filled circles signify rats infected with *S. ratti*.
Inset shows sample sites in Germany. S = Sendenhorst, A = Asbeck.


= uninfected rat
= infected rat
Edgefield (E)
Nether Fala (N)
Oxford (O)
Harwell (HA)
Norfolk (N)
Berkshire (B)
Surrey A (SA)
Surrey B (SB)
Hedley (HE)
Dorset (D)
Wiltshire (W)
S
A

Sample site	*Site code*	*Map reference*[b]	*No. of Rats trapped*	*No. of Rats infected*	*% of rats positive*	*Number of iL3s sampled*
Edgefield	E	NT 37	3	0	0	0
Nether Fala	NF	NT 35	24	0	0	0
Norfolk	N	TF 80	3	3	100	54
Oxford	O	SP 52	11	6	55	159
Harwell	HA	SU 48	10	2	20	78
Hedley	HE	SU 37	6	6	100	83
SurreyA	SA	SU 94	7	7	100	147
SurreyB	SB	SU 84	3	3	100	30
Wiltshire	W	SU 02	21	17	81	368
Dorset	D	SY 69	7	7	100	164
Berkshire	B	SU 59	28	27	89	359
Total	-	-	123	76	x [a]=62(40)	1442

Table 3.1 Prevalence of infection of *S. ratti* within British rat populations.
[a] mean prevalence of infection with standard deviation of the mean in parentheses.
[b] British national grid reference coordinates.

Sample site	*Site code*	*Map reference*[b]	*No. of Rats trapped*	*No. of Rats infected*	*% of rats positive*	*Number of iL3s sampled*
Altenberge	AL	52.02N 7.29E	3	0	0	0
Asbeck	A	50.40N 7.50E	4	1	25	10
Bad Bentheim	BB	51.56N 7.34E	1	0	0	0
Billerbeck	BI	52.52N 7.19E	1	0	0	0
Gievenbeck	G	52.50N 7.18E	1	0	0	0
Affenhaus	AF	51.58N 7.37E	4	0	0	0
Nordhorn	NO	52.27N 7.05E	2	0	0	0
Sendenhorst	S	51.50N 7.50E	3	2	66	20
Total	-	-	19	3	x [a]=11(22)	30

Table 3.2 Prevalence of infection of *S. ratti* within German rat populations.
[a] = mean prevalence of infection with standard deviation of the mean in parentheses.
[b] International coordinates.

The proportion of parasites developing by the Heterogonic route is low, or absent, in all sites with a mean value of 0.6% of the parasites developing into free living adults. Thus, it is clear that British and German populations of *S. ratti* are predominantly asexual in their mode of reproduction.

3.3.3 Frequency distribution of faecal larval counts within a group of naturally-infected rats

The number of larvae developing per gram of faeces (iL3s/g) was measured for the 27 rats sampled from the Berkshire site. Counts of iL3s/g ranged from 0 - 173.3 with a mean (x) of 36.95 iL3s/g (Figure 3.3A). The variance of the mean (s^2) was 1859 and the variance to mean ratio (s^2/x) was 50.3, illustrating that the distribution of iL3s/g of faeces is aggregated between hosts. A histogram of the distribution of infection (Figure 3.3B) shows that the distribution of the data is skewed with most rats showing low faecal larval counts and relatively few rats with high counts. The hypothesis that the data follow a negative binomial distribution was tested by fitting negative binomial and Poisson (random distribution) models to the data by generalised linear models (GLIM). Maximum log-likelihood estimates were used to fit each model to the data using the observed mean, for the Poisson distribution, and the observed mean and variance for the negative binomial distribution. An ANOVA with 26 degrees of freedom (d.f.) was used to compare each model to the observed distribution. Residual deviances were 343 for the Poisson and 32 for the negative binomial distributions. The test statistic in question is the difference between the two deviances and is distributed as a chi-square with 1 d.f. (Crawley 1993) $\chi^2 = 311$. This was highly significant and it could be concluded that the data would be described most accurately by a negative binomial distribution with a mean = 36.95 and $k = 0.77$ rather than a Poisson distribution with a mean = 36.95.

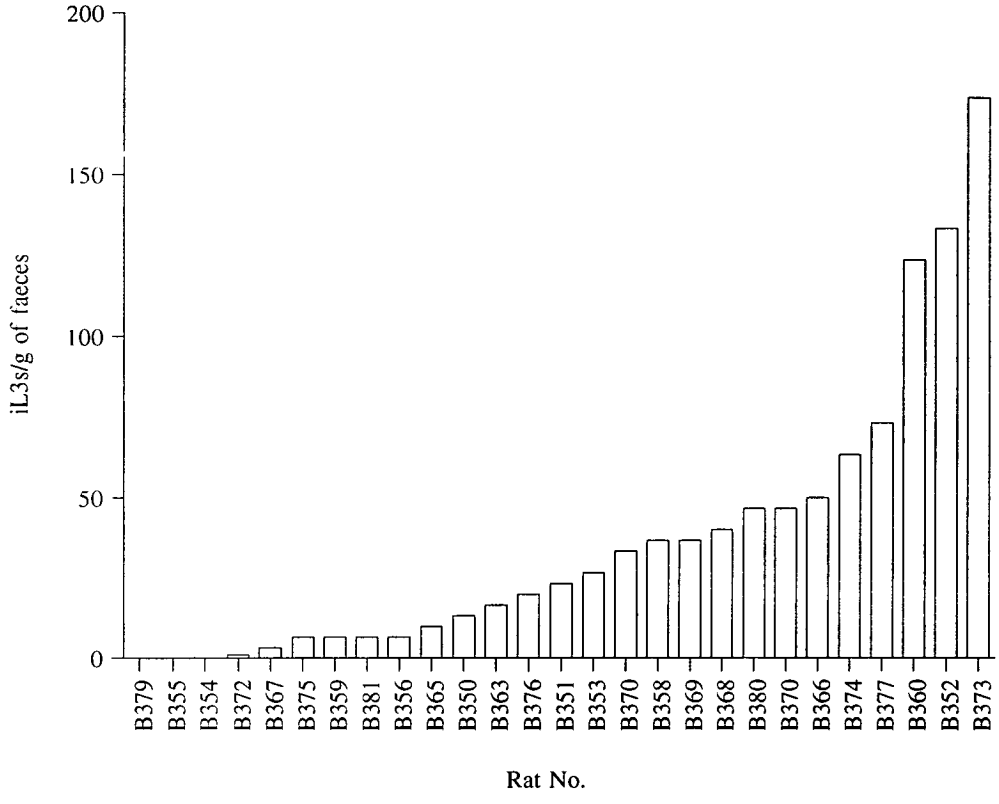
Within the Berkshire sample, 18 rats were weighed by M.A.A.F. Data from the other nine individuals was missing from M.A.A.F. records. The mean weight of male rats was 203g (n = 10, s = 10.8) and the mean weight of female rats 134g (n = 8, s = 4.3). Male rats had higher iL3/g counts relative to females (male mean iL3/g = 41.2, s = 45, female mean iL3/g = 33.3, s = 43). ANOVA of the distribution of $\log_{10}$ larvae/g faeces between the two sexes was not significant ($F_{1, 17} = 0.06$ p= 0.805) showing that sex was not an important factor in determining faecal larval counts. There was no correlation between rat weight and $\log_{10}$ larvae/g faeces for all rats (Pearson correlation coefficient r = -0.014, $F_{1, 17} = 0.20$, p = 0.665, Figure 3.4) or

Figure 3.3

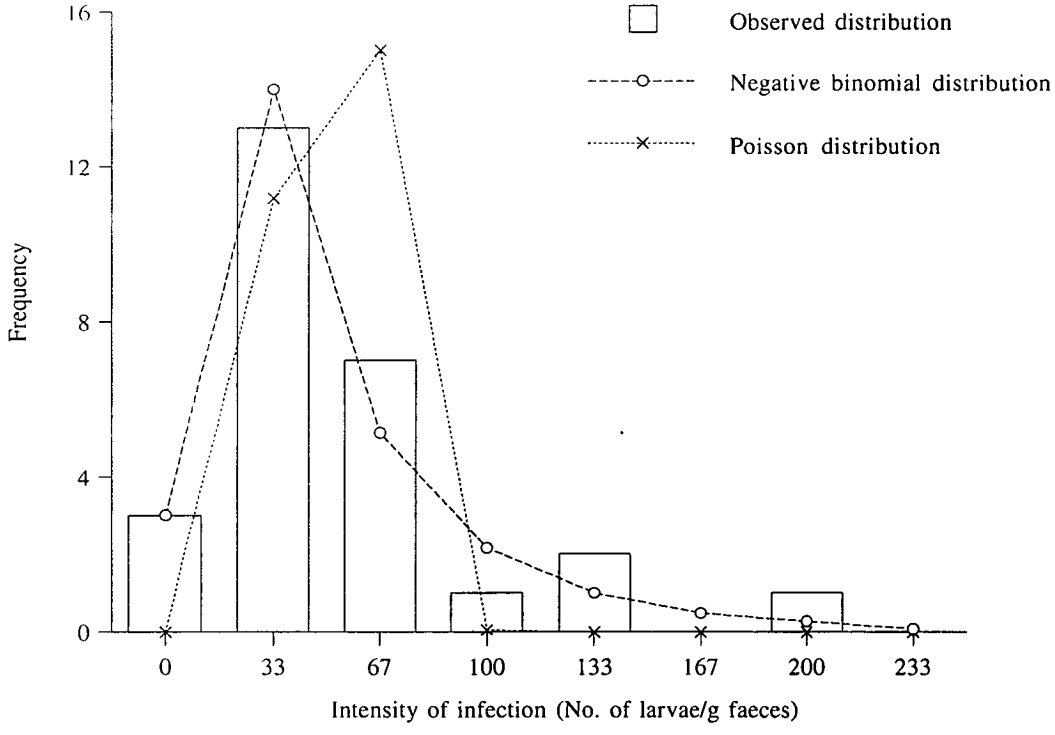
A. Numbers of larvae/g faeces for 27 rats collected from the Berkshire sample site. $x = 36.95$, $s^2 = 1859$

B. Frequency distribution of larvae/g faeces for 27 rats collected from the Berkshire sample site. The distribution following the negative binomial distribution ($k = 0.77$) is shown by circles and a Poisson distribution shown by crosses.

A.


200
150
100
50
0
iL3s/g of faeces
B379
B355
B354
B372
B367
B375
B359
B381
B356
B365
B350
B363
B376
B351
B353
B370
B358
B369
B368
B380
B370
B366
B374
B377
B360
B352
B373
Rat No.

B.


16
12
8
4
0
Frequency
Observed distribution
Negative binomial distribution
Poisson distribution
0
33
67
100
133
167
200
233
Intensity of infection (No. of larvae/g faeces)

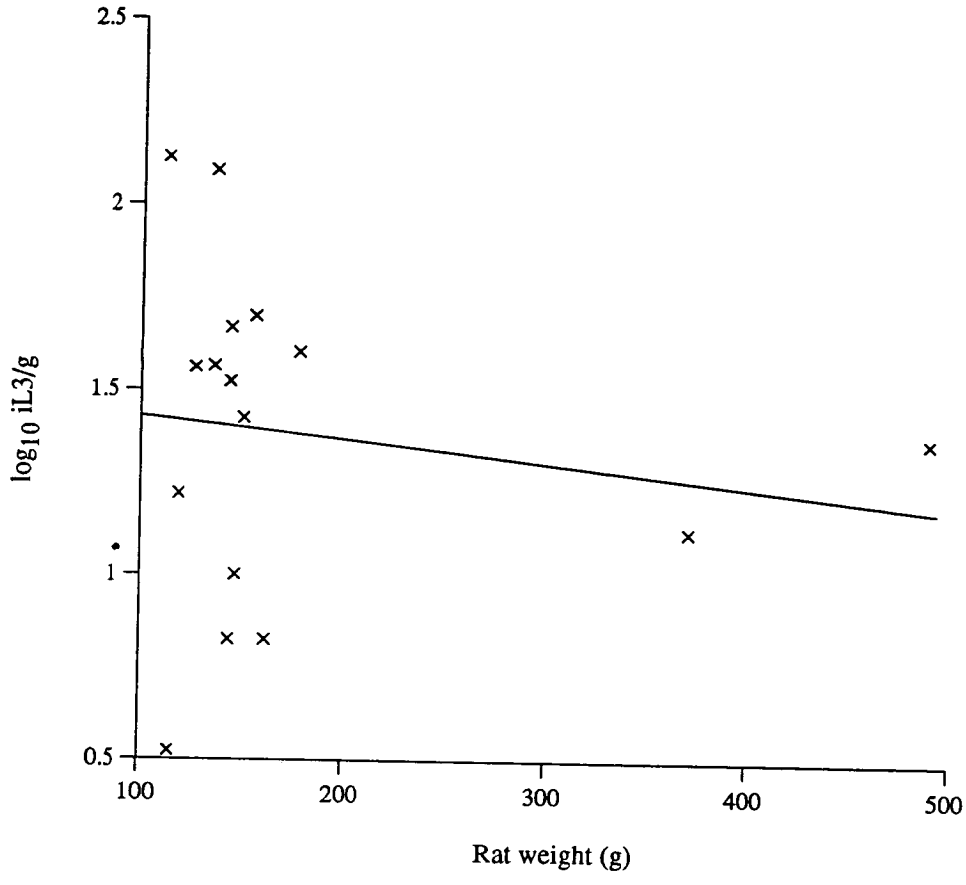


Figure 3.4. Numbers of iL3s/g faeces plotted against weight (g) of individual Berkshire rats (males and females). Pearson correlation coefficient r = -0.014, $p = 0.67$ (not significant).

for male and female rats separately (males; r = -0.019, $F_{1, 9}$ = 0.20 p = 0.707, females; r = 0.05, $F_{1, 7}$ = 0.02 p = 0.895). This suggested that rat weight does not influence larval production in any simple linear fashion.

3.3.4 Fecundity of single parasitic females

The total overnight larval production from rats infected with line 5 Heterogonic (group 1) was determined on days 5, 6, 7, 8 and 9 post infection and from rats infected with line 132 Heterogonic (group 2) on days 7, 8 and 12 (Table 3.4).

GROUP1; line 5 Heterogonic			GROUP 2; line 132 Heterogonic		
Rat	Mean overnight fecundity	*s*	Rat	Mean overnight fecundity	*s*
100	1.8	2.5	259	9	9.6
102	6.2	5.8	260	58	59.7
103	31	12.0	261	6	7.2
104	21.3	13.8	262	2.3	5.4
105	12	10.0	263	0.3	0.6
106	13	10.8	265	0.3	0.6
107	12	8.5	270	2	1.7
			271	1	1.2
			272	14	8.5
			273	10	9.7
			274	1	2.3
			276	1	1.0
			277	2	2.8
Overall	12.5	11.6	Overall	8.22	20.9
Max.[b]	52		Max.[b]	126	

Table 3.4 Overnight fecundity of single parasitic adult *S. ratti*. [b]Max. is the maximum number of larvae produced by any one parasite on a single night. *s* is the standard deviation of the mean.

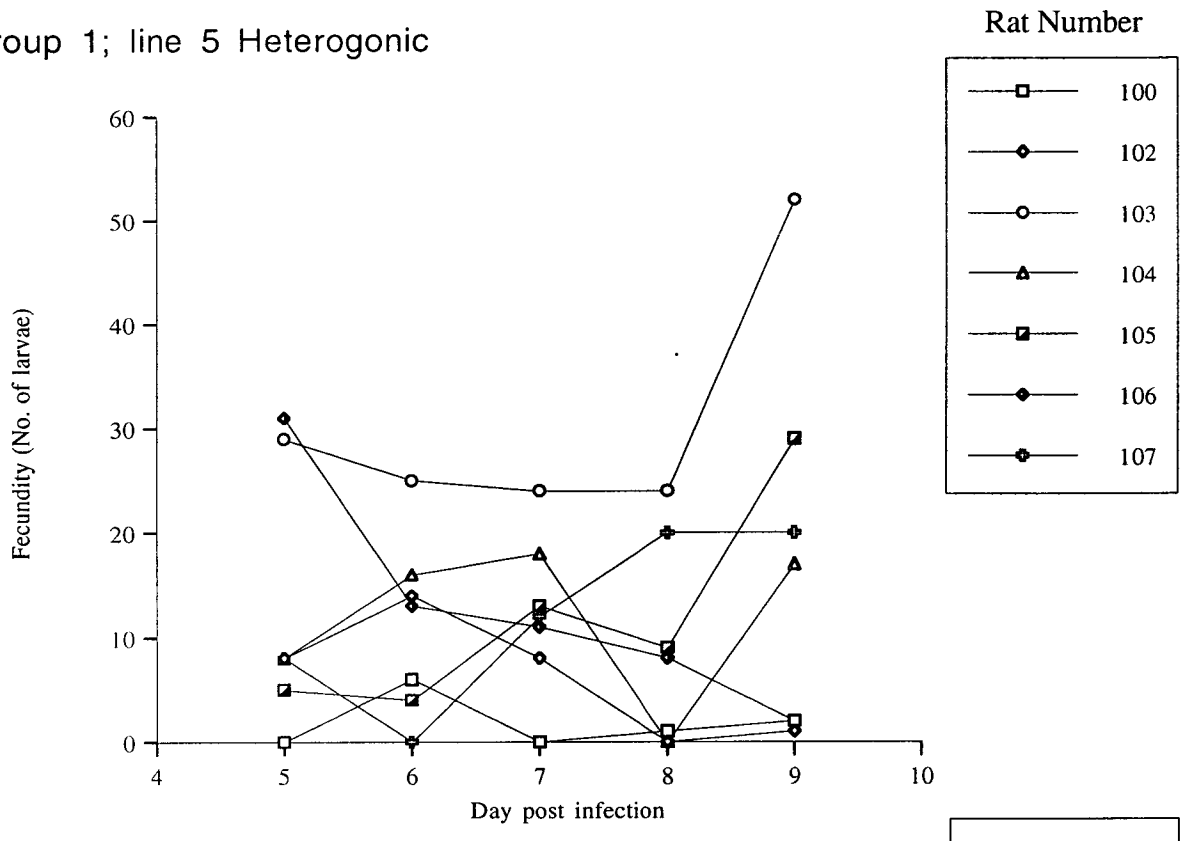
Within group 1, larval production was generally constant over the five days of sampling. One rat showed a significant negative correlation between day post infection and intensity of infection (rat 106, Pearson correlation coefficient r = -0.915, p = 0.03, Figure 3.4). Within group 2, there was no correlation between day p.i. and intensity of infection showing that in this group also, the fecundity of parasitic females was constant. Mean fecundity for all samples was 12.5 (95% c.i.= ± 3.32) for group

Figure 3.5

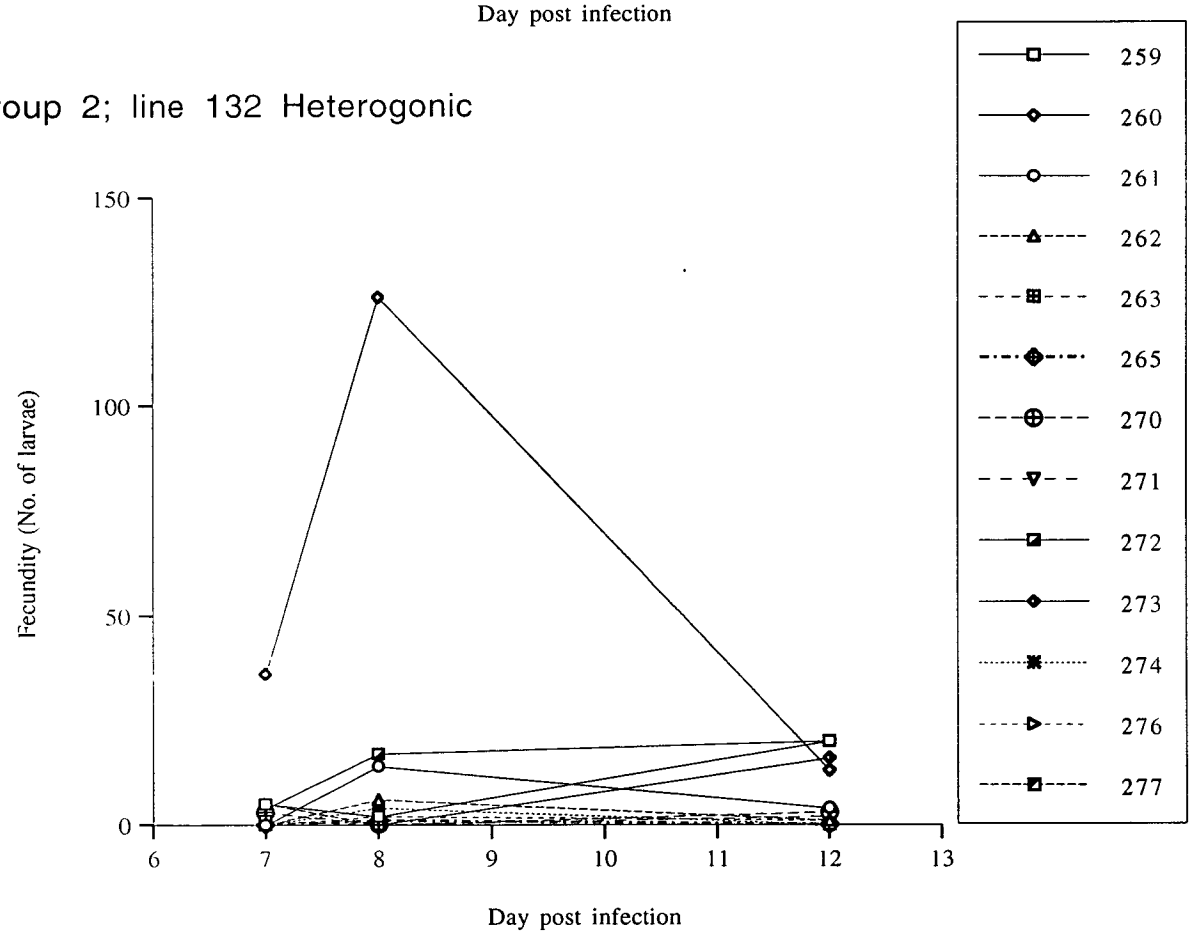
Group 1; fecundity of individual parasitic females from isofemale line 5 Heterogonic over five consecutive days.

Group 2; fecundity of individual parasitic females from isofemale line 132 Heterogonic over six days.

Group 1; line 5 Heterogonic



Group 2; line 132 Heterogonic



1 and 8.2 for group 2 (95% c.i.= ± 5.46) (Table 3.4). As the confidence intervals overlapped, these estimates may be pooled. Combining the data gives a mean overnight fecundity of 10.4 (95% c.i.=±3.89), with a range of 0 - 126. Variation in mean fecundity among parasitic females was high for both groups, ranging from 2 - 31 among parasites in group 1 and 1 - 58 among parasites in group 2 (Table 3.4). However, the pattern of variability showed a skew towards low fecundity with only three out of 20 rats containing a parasite with average fecundities of more than 14 larvae per night.

Variation in fecundity of individual parasitic females was also high during the course of an infection. For example rat 103 (group 1) contained a parasite showing a maximum fecundity of 52 progeny on day 9 and minimum fecundity of 24 progeny on day 8, a two-fold difference in productivity between consecutive days. However, Figure 3.5 shows that parasites in group 1 with initially high fecundity tend to remain high over the sampling period (*e.g.* rat 103) while parasites with low fecundity remained low (*e.g.* rat 100). The same trend is observed with infections in group 2; rat 260 contained a parasite which was highly fecund except for day 12 p.i, producing 126 larvae on day eight post infection. The majority of the other infections were constant, and low, relative to rat 260 over the sampling period.

In summary, these data show that (i) there is a degree of variability in *per capita* egg production between parasitic females and between day post infection, however, (ii) the majority of parasites have a low fecundity with a few parasites having much higher fecundity and (iii) while variation in parasite fecundity is high within a host, highly fecund parasites tend to remain highly fecund and *vice versa.* The highest measurement of *S. ratti* overnight fecundity was 126 larvae (measured for rat 260) and the average fecundity was measured as 10.4 larvae/parasitic female/night.

3.3.5 Estimations of intensity of infection from data on parasite fecundity

The mean overnight parasite fecundity, Fo, measured from the experiments described in section 3.3.4 was 10.4 larvae/parasitic female/night (95% c.i. = ±3.89). Overnight faecal production per rat, P, was 9.8g; therefore Fg, the mean overnight parasite fecundity/g faeces, was 10.4/9.6 ≈ 1. Intensities of infection within each naturally infected Berkshire rat were estimated by dividing larval counts/g faeces by mean overnight parasite fecundity/g faeces (Fg). As Fg = 1, this did not alter the characteristics of the distribution described in section 3.3.3. Therefore, estimated

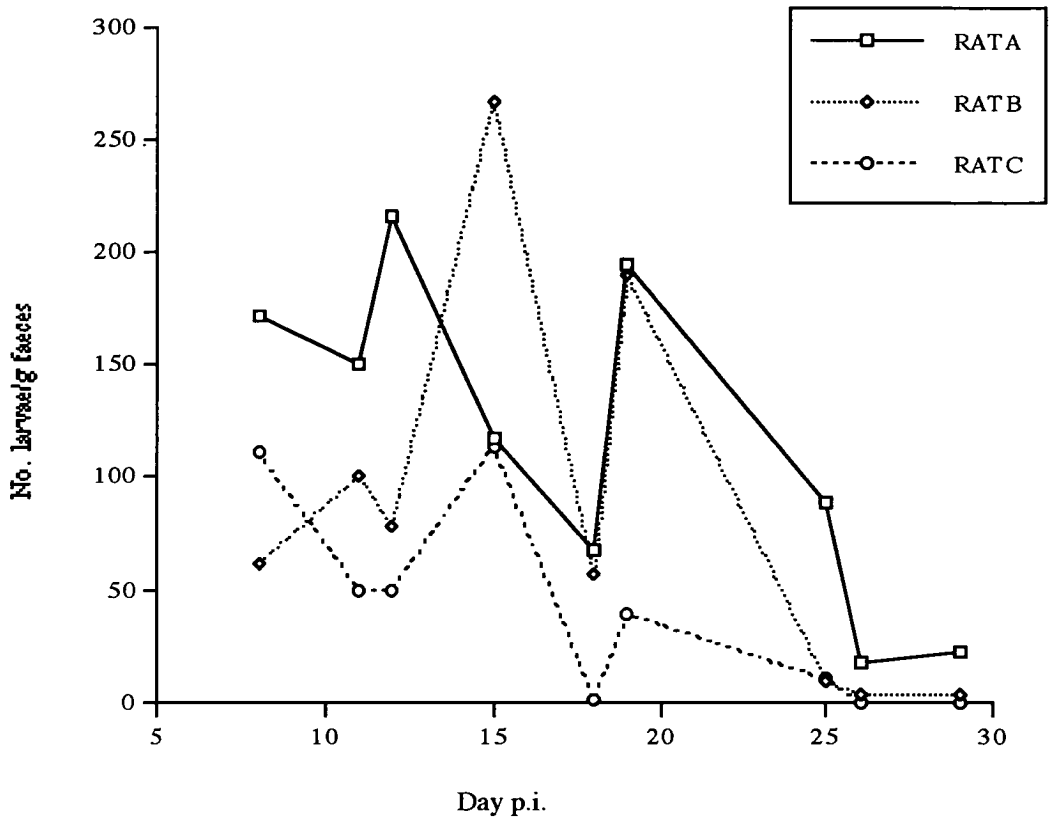
intensities of infection within the Berkshire rat population were aggregated (s^2/x = 50.3) with a range of 0 - 173 parasites per host. The distribution was described best by the negative binomial with a mean parasite population size of 36.95 and overdispersion factor k =0.77 (section 3.3.3).

3.3.6 Longitudinal analysis of *S. ratti* infections

3.3.6a Laboratory infections

Figure 3.6 shows the progression of infections of *S. ratti* in each of the three rats. As is typical for laboratory animals, the infections were almost resolved by day 29 post infection (p.i.) in the rats infected with 12 iL3s (A and B), and was resolved by day 26 for the rat infected with 6 iL3s (C). Rat C produced slightly under half the total number of worms of the rats infected with 12 iL3s suggesting that this rat had established an infection about half the size of the other two; faecal larval counts were also consistently lower from this rat relative to the other two (except for day 8 p.i.). There was a high degree of variation in larvae/g faeces/day within each rat. Day 18 showed a drop in faecal larval counts of all rats suggesting an environmental effect on parasite hatchability/survivability on this day.

183 iL3s were genotyped from all three rats at the Actin locus (68 from rats A and B, 47 from rat C, table 3.5) on days 8, 11, 15, 19, 25 and 29. Frequencies of Actin allele 1 in the total populations of worms collected from each rat were as follows; Rat A, 0.39 (95 % c.i. ± 0.080), rat B, 0.51 (95 % c.i. ± 0.084) and rat C, 0.38 (95 % c.i. ± 0.1) compared to a frequency of 0.43 in the parent parasite population. The frequencies of genotypes between the three rats were not significantly different (χ^2 = 7.94, d.f = 4, p = 0.094).



Total number of iL3s produced by each rat:	A = 4865
	B = 4106
	C = 1947

Figure 3.6 Faecal larval counts/g faeces produced over 29 days from three rats; rat A and B infected with 12 iL3s, rat C infected with 6 iL3s. The total numbers of larvae counted from each rat are shown in the lower inset box.

Genotype	Rat A	Rat B	Rat C
11	0.09	0.18	0.06
12	0.60	0.68	0.72
22	0.31	0.15	0.22
n	68	68	47

Table 3.5 Frequencies of homozygous ('11' and '22') and heterozygous ('12') genotypes for the Actin locus from the total collection of each of the three rats. *n* = the number of iL3s sampled.

Figure 3.7 shows the numbers of each genotype within each sample of the three rats over time. It is apparent that there is a degree of stability in genotype frequencies between each time point over the course of the infection; the proportion of heterozygotes remains high within each rat over the course of the experiment. This is independent of the drop in the number of larvae/g faeces/day during the infection suggesting that this reduction is due to a diminishing fecundity of the total worm population and not death or temporary reduction in fecundity of individual parasites.

In order to test whether the relative proportions of each genotype remained constant over the course of an infection, the null hypothesis "genotypes are not significantly different over time" was tested for each rat. As counts were less than five for some cells, the numbers of homozygous genotypes were pooled for each rat for each day. Also, the numbers of genotypes on days 8 and 11, 15 and 18, 19 and 25 were combined for both heterozygous and homozygous genotypes. This makes the assumption that the genotype frequencies were the same on these pairs of days and that there was no significant difference between the frequencies of homozygous genotypes over time. Thus, a 2 X 4 contingency table was constructed and observed and expected frequencies compared by χ^2 tests. Values for each rat were as follows; rat A $\chi^2 = 5.32$, p=0.15; rat B $\chi^2 = 6.5$, p=0.09; rat C $\chi^2 = 0.675$, p=0.713. Combining data for each class between rats resulted in a χ^2 value of $\chi^2 = 11.2$,p=0.14. As none of these values were significant the null hypothesis was accepted and it was concluded that the genotype proportions do not vary significantly during the course of an infection.

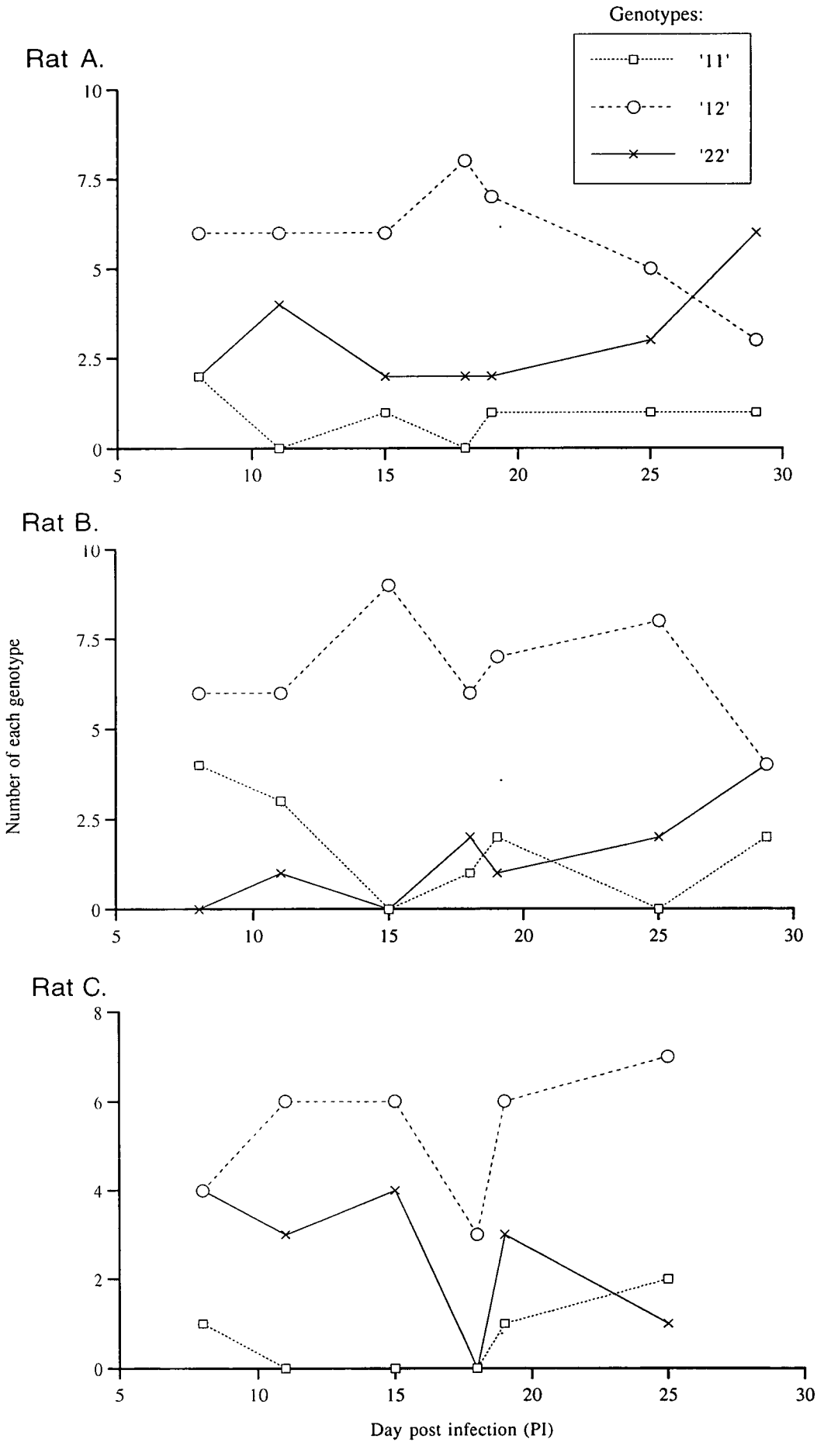
3.3.6b Infections in wild-caught rats

12 rats were sampled in both July and October. These rats came from Wiltshire (4 rats), Dorset (3 rats), Surrey A (4 rats) and Surrey B (1 rat). A total of 85 iL3s were genotyped at the BSP-8 locus (38 in July and 47 in October) and 97 iL3s at the Actin locus (55 in July and 42 in October).

All samples were fixed for the '22' genotype at both time-points at BSP-8. This shows that, at this locus, there were probably no genetically variable parasites within these infections and that a single sample was sufficient to characterise the parasite infra-populations. The frequencies of genotypes at the Actin locus are shown in Table 3.6 for the pooled data from all rats at each sample site. For three of the sites, there was no significant difference in genotype frequency between the two sampling

Figure 3.7

Frequencies of homozygous ('11','22') and heterozygous ('12') genotypes over 21 days of sampling from three rats infected with line 5 Heterogonic:
Rat A; infected with 12 iL3s
Rat B; infected with 12 iL3s
Rat C; infected with 6 iL3s.


Genotypes:
'11'
'12'
'22'
Rat A.
Rat B.
Rat C.
Number of each genotype
Day post infection (PI)

times. There was significant variation in genotype frequencies within one site (SA) between the two time points ($\chi^2 = 6.3$, d.f = 2, $p = 0.04$). However, this value was only marginally significant at $p = 0.05$.

	Site											
	W			D			SA*			SB		
	Genotype											
	(11	12	22)	(11	12	22)	(11	12	22)	(11	12	22)
Jul.	0.52	0.32	0.16	0.36	0.64	0.00	0.64	0.04	0.32	1.00	0.00	0.00
Oct.	0.88	0.06	0.06	0.58	0.42	0.00	0.85	0.14	0.00	1.00	0.00	0.00

Table 3.6 Genotype frequencies at the Actin locus for the pooled iL3 samples from each of four sites at two time points. Jul. = July and Oct. = October samples. * signifies sites that have significantly different genotype frequencies between the two time points at $p = 0.05$.

Taken together, these data show that out of eight comparisons (four sample sites and two loci), seven showed no significant variation in genotype frequencies between two temporally separated samples. This shows that, for the pooled data for each sampling site, measurements of genetic variation based on genotype frequencies of samples of iL3s from wild rats are generally constant over the time period observed.

3.4 Discussion

Infection distribution and prevalence

This study has shown that there is a spatial variation in the prevalence of *S. ratti* among wild rats from different sample sites. No parasites were found in either of the two sites sampled in Scotland (n = 27) whereas positive infections were found at all sites sampled in England, with prevalences ranging from 20 - 100%, with a mean of 79% (n = 96). This is higher than previously reported from other countries (14.7%, Nigeria (Udonsi *et al.* 1989); 62%, Malaysia (Sinniah *et al.* 1979); 36%, Israel, (Wertheim, 1970)). It is unlikely that the present geographical distribution does not cover Scotland as *S. ratti* has previously been recorded from other sites there (Viney 1990). However, the data may be indicative of a cline in reduction of prevalence of infection with increasing latitude. Further sampling of a greater number of sites in Scotland and Northern Britain would be necessary to resolve this question.

The incidence of infection was much lower in Germany with a mean prevalence of 11%. This corresponded to three infected rats from two sites. As these samples had traveled by air to be investigated in Edinburgh, it is possible that temperature and/or pressure changes during transport had reduced the viability of *S. ratti* eggs, this accounting for the apparent low prevalence of infection.

Frequency of heterogonic development.

The frequency of heterogonic development in 14 isolates has been shown to be low with a mean of 0.6 % of worms developing by this route. Thus, these populations are overwhelmingly asexual in their mode of reproduction. There is little evidence for any geographical differences in the frequency of sexual reproduction, suggesting that variation in the genetic component of this trait is relatively homogeneous across these populations. Interestingly, development in these lines was not affected by temperature unlike other lines of *S. ratti* which show increases in heterogonic development with a rise in temperature (Viney 1996). Temperature responsivity seems to be characteristic of species of *Strongyloides*, having been observed in *S. fuelleborni* (Premvati 1958) and *S. papillosus* (Premvati 1963) where heterogonic development is favoured at higher temperatures. If responsiveness is an ecological adaptation to allow sexual reproduction during clement conditions, then European *S. ratti* seems to have lost this ability. It may be that the cost of sex within a relatively cold climate (*ie.* Europe relative to the tropics) is high and for this reason homogonic development has been selected almost to fixation within these populations. Indeed, it has been suggested that species that are at the edge of their geographic range tend towards parthenogenesis (Hughes 1989; Lynch 1984a). It would be useful to select these European isolates in a similar manner to that done by Viney (Viney 1996) in order to determine (i) whether these isolates respond to selection and (ii) whether they respond in a similar manner to heterogonically developing isolates. In this way it would be possible to determine how much genetic variation exists in determining this trait within these natural populations.

The influence of the frequency of sexual reproduction on population genetic structure would be predicted to be great. However, no comparable studies have been carried out on the amount and subdivision of genetic variation in facultatively sexual parasitic nematodes, so the effect is unpredictable. It has been demonstrated that *Ascaris suum* has a female-biased sex ratio (males : females = 0.44 : 1) (Nadler *et al.* 1995). This has the effect of reducing the effective population size of this species and may well account for the relatively high population differentiation that has been

observed in this species (Nadler *et al.* 1995; Anderson *et al.* 1993). In extreme situations, the effective population size of *S. ratti* may be as low as one (for a population founded by a single parasitic female) and, if sexual reproduction is absent, the population will be fixed for the multilocus genotype of the founder. Thus, asexually reproducing populations of *S. ratti* may be expected to contain non Hardy-Weinberg equilibrium proportions of genotypes, and pronounced population subdivision as a consequence of stochastic effects on parasite transmission. Even low levels of sexual reproduction may mitigate these effects. This is discussed further in Chapter 4.

Intensities of infection.

The frequency distribution of *S. ratti* faecal larval counts within Berkshire rats was shown to be consistent with studies on other nematode parasites in that the distribution between hosts was aggregated and described by the negative binomial, with an overdispersion factor (k) of 0.77. A method was described where the mean fecundity of parasitic female worms was used to calculate worm burdens, and thus the size of the parasite infra-population, within these naturally infected rats. This showed that, based on laboratory estimates of parasite fecundity, the distribution of intensities of infection was the same as that of faecal larval counts and consequently described by the same parameters x, s, and k. However, it should be recognised that any attempt to relate faecal larval counts to the numbers of adult parasites will have associated errors. These stem from a multitude of possible sources, those with the greatest expected effect being (i) within and between parasite variation in fecundity, (ii) variation in fecundity between parasites in naive rats and rats that have a degree of acquired immunity (iii) possible density dependent effects on fecundity and (iv) differences in size between rats causing variation in faecal output. Due to the sampling regime used in this study, it was not possible to dissect hosts in order to test predictions of worm burden. However, the accuracy of the technique has been assessed in other parasite systems. It has been shown that faecal egg counts provide good estimates of the numbers of adult parasites within hosts in infections of *Trichuris trichuris*.(Bundy 1986) and *A. suum* and *T. trichuris* (Haswell-Elkins *et al.* 1987b). Also, variation in faecal volume has been shown to have little effect on final egg counts in infections of *Haemonchus contortus* (Roberts & Swan 1981). For the purposes of this study, estimates of intensity of infection were required such that measurements of within host genetic variation could be correlated against worm burden. If positive correlations are observed, this in itself is a test of the validity of the technique. Such data will be presented in Chapter 4.

Longitudinal analysis of S. ratti *infections*

The validity of the assumptions that (i) genotype frequencies of iL3 samples reflect the genetic composition of the parasite infra-population and (ii) that the genetic composition of parasite infra-populations are stable over time, were addressed. In order to do this, empirical observations were made on (a) genotype frequencies from longitudinal samples of small genetically heterogeneous parasite populations and (b) genotype frequencies from iL3 samples of wild rats sampled at a two time points.

It was shown in the laboratory studies of three rats with low intensity infections that genotype frequencies were stable over time. As sample genotype frequencies did not vary significantly within rats, it can be concluded that the relative fecundities of the parasites within each host were constant throughout the period of these infections. These data corroborate the observations on single-parasite infections. Here, it was shown that a general trend is for highly fecund parasites to remain highly fecund and *vice versa* for low fecundity parasites. Although day-to-day variation in fecundity was observed in single parasite infections, this did not seem to have a large effect on the genotype frequencies of low intensity infections. Thus, over short time scales and low intensity infections, it can be concluded that inferences made on the genetic variation of a parasite infra-population are insensitive to the day that the sample was taken.

In order to observe the variation of sample frequencies made over a longer time scale, genotypes at the Actin and BSP-8 loci were scored for iL3 samples at two time points, three months apart in wild infections. Genotype frequencies, between time points, of the iL3 sample pooled for the rats within each site were significantly different in only one out of seven comparisons. Thus, it seems that the parasite infrapopulations within these rats are relatively stable in the genotypic proportions produced over time. This suggests that there is probably a relatively stable long-term infection within these rats and that parasites are not being rapidly eliminated. The observation that some wild rats maintained infections for six-months in captivity, suggests that natural *S. ratti* infections are not rapidly cleared by anti-parasitic immune responses or old age of the parasites. That these infections are being maintained by re-infection is unlikely. The rats were kept in gridded wire-bottom cages that allow no accumulation of faeces. Furthermore, re-infection would be expected to cause genotype frequencies to drift rapidly due to severe bottle-necking of the parasite populations.

Samples of iL3s from a single rat provide a measure of the genetic diversity of the infective output of that rat. Measures from a series of rats within a single geographic area provide a measure of the diversity of the infective 'landscape' that the rat encounters. This is an important parameter to measure in its own right as it will, to a large extent, determine the genetic composition of the parasite infrapopulations. However, my concern was whether an iL3 sample reflects the genetic composition of the parasite infra-population. Several lines of evidence now suggest that it does; (i) faecal egg-counts from a wild rat population were overdispersed as expected for a natural parasite distribution (Anderson & May 1992), (ii) genotype frequencies of iL3 samples were the same from rats infected with the same genetically heterogeneous parasite line, (iii) these samples showed little variation over time and (iv) natural infections show little genetic variation over time. It is therefore reasonable to conclude that point samples provide good estimates of the parasite population genotype frequencies even within hosts with low intensity infections, and that the data are consistent over time. Therefore, it is valid to estimate the population genetic structure of adult parasites from samples of infective stages.

Taken together, these results suggest that the genetic structure of natural *S. ratti* populations will be stable over short time-scales. However, nothing is known about the long term population dynamics of *S. ratti* infections. Studies by Rollinson *et al.* (Rollinson *et al.* 1986) on the population genetics of *Schistosoma mansoni* infection in black rats (*Rattus rattus*) found that prevelances of infection were stable over one year, and that allele frequencies of parasite samples were also stable over this time period. However, over longer time periods, prevalences of infection varied showing fluctuation in population sizes of *S. mansoni.* Short term temporal variations in the prevalence of infection have been documented in infections of helminths in Wood mice (Montgomery & Montgomery 1989). That the prevalence of infection of *S. ratti* in Scotland appears to be lower than was observed in 1989 (Viney 1990), suggests that *S. ratti* may undergo temporal variations in population size. It is possible that changes in prevalence of infection will be linked to the temporal dynamics of the Brown rat population, however this topic requires further investigation.

3.5 SUMMARY

This chapter has described the methodology and sites used to sample *S. ratti* from wild brown rats. 123 rats were trapped and sampled from 11 sites; 27 rats at two sites within Scotland, 96 rats at nine sites within England. Prevalence of infection were high within all sites except those in Scotland, and a total of 1472 iL3s were sampled. Amounts of Heterogonic development were shown to be low within these populations, with a mean homogonic index of 0.994. It was concluded that sexual reproduction was exceptional within these isolates. Measurements of faecal larval counts were made at one site (Berkshire) and the distribution found to be overdispersed and described by the negative binomial distribution. Estimates of single parasite fecundity were used to calculate worm burdens within these rats. On this basis, the mean size of parasite infra-populations within the sampled rats was 37, and the greatest infection contained 173 parasitic females.

Finally, an analysis was made of the accuracy of point samples of iL3s in estimating the frequencies of genotypes of infra-populations. Genotype frequencies of iL3 samples from three low-intensity artificial infections were shown to be stable over 21 days. Genotype frequencies calculated from the pooled iL3 samples from natural infections of rats from separate sample sites were similar between two time points, three months apart. These data justify the technique of point sampling iL3s as a method of determining the genetic composition of the adult intestinal parasite infra-populations.

Chapter four

The genetic structure of *Strongyloides ratti* populations determined by analysis of RFLP allele frequencies.

4.1 Introduction

In the previous chapters I have described how molecular biological methods were used to identify polymorphic loci of *Strongyloides ratti*. I have also shown that *S. ratti* is widespread in Britain and Germany, and have described how a number of *S. ratti* infections of *Rattus norvegicus* were sampled from rat colonies in these countries. In this chapter I give the results of an analysis of the population structure of *S. ratti* by using these genetic markers and the Europe-wide parasite samples. I then describe what information this genetic analysis provides with regard to understanding the genetic structure of geographically separated samples of the parasite. Finally, I detail the breeding structure and epidemiology of the *S. ratti* infections within these sample sites.

4.2 Genetic data analysis: Methods

4.2.1 Sample genotypes

The sampling of rats is described in section 3.2.2. For each iL3, DNA preparations were made using the single worm digestion protocol described in section 2.2.7. Aliquots of the preparations were used immediately as templates in PCR reactions or stored at -20 °C for subsequent use. Each individual iL3 DNA preparation was screened for RFLPs at the Actin, BSP-8 and CM-2 loci as described in section 2.3.3. Genotypes of each iL3 were stored in an Excel database (Microsoft) for genetic analysis.

4.2.2 Allele frequencies

Allele frequencies were calculated from the genotypes of iL3s at three levels: (i) the iL3 samples of a single rat, (ii) the total iL3 sample from each sample site and iii) the total iL3 sample of all sample sites using:

$$p_u = (2P_{uu} + P_{uv})/2n \qquad \text{(Eqn 4.1)}$$

where for two alleles A_u and A_v, (where $u \neq v$, u and v are integers between 1 and j, where j is the number of alleles in the sample), p_u is the frequency in the sample of the A_u allele, P_{uu} is the frequency of the A_uA_u genotype, P_{uv} is the frequency of the A_uA_v genotype and n is the sample size (Weir 1996).

The variance in allele frequencies was calculated from the sample allele and genotype frequencies according to Weir, (Weir 1996). Confidence limits for estimates of allele frequencies were calculated from the variance (assuming an approximately normal distribution):

$$Var(p_u) = (p_u + P_{uu} - 2p_u^2)/2n \qquad \text{(Eqn 4.2)}$$

$$95\ \%\ \text{confidence limits} = 1.96\sqrt{Var(pu)} \qquad \text{(Eqn 4.3)}$$

4.2.2 Heterozygosity

Heterozygosity (h) of a locus l with alleles A_u and A_v was determined from the numbers of A_uA_v heterozygotes (n_{luv}) (Weir 1996):

$$h_l = \sum_u \sum_{v \neq u} \frac{n_{luv}}{n} \qquad \text{(Eqn 4.4)}$$

where n= the number of individuals genotyped at locus l. The variance of h_l within a sample site was calculated from the sample heterozygosity at locus l (Weir, 1996) by:

$$Var(h_l) = \frac{1}{n} h_l(1 - h_l) \qquad \text{(Eqn 4.5)}$$

The average heterozygosity (H) for m loci was calculated using;

$$H = \frac{1}{m} \sum_{l=1}^{m} h_l \qquad \text{(Eqn 4.6)}$$

Heterozygosities were first calculated using the pooled data from each sample site. Where relevant, and where sample size allowed, heterozygosities were subsequently calculated for the iL3 samples from single rats. Heterogeneities in heterozygosity between sample-sites and between loci were investigated by analysis of variance (ANOVA). Loci for each individual iL3 were coded 1 if heterozygous or 0 if homozygous (Weir 1996). A split-plot analysis was used where loci (factors) were

applied to each sample site (a plot). The total variance therefore had three components; between loci, between loci and populations and between populations. Significant differences within any one of the levels were determined from the F-ratios of the mean squares. The ANOVA was performed using general linear models (PROC GLM) in the SAS statistical package (SAS institute).

4.2.3 Hardy-Weinberg equilibria

Expected numbers of homozygotes and heterozygotes were calculated from the observed allele frequencies for Hardy-Weinberg expectations of genotype frequencies in a randomly mating population. The null hypothesis "random assortment of alleles" was tested by comparing observed and expected genotype frequencies for each locus within each sample site and the pooled data set by Fisher's exact test (Louis & Dempster 1987; Guo *et al* 1992) using Levenes' (Levene 1949) correction for small sample sizes (section 4.2.7). Unbiased estimates of exact p values were made using a Markov chain method, carried out by the Hardy-Weinberg analysis program in GENEPOP version 1.2 (Raymond & Rousset 1995). Sequential Bonferroni corrections were used as suggested by Rice (Rice 1989) to correct for the probability of type I errors resulting from multiple comparisons.

4.2.4 Linkage disequilibria between loci

Whereas deviations from Hardy-Weinberg expectations can be viewed as single-locus disequilibria (*ie.* non-random associations between alleles at a single locus), non-random patterns of pairwise associations between loci are known as linkage disequilibria. For the purposes of this study, disequilibria between loci was defined as follows. For two segregating loci with alleles P and Q at one locus, R and S at the other, with frequencies p, q, r and s, linkage disequilibria occurs if the observed frequency of the PR genotype is significantly different from its expected frequency pr. This is known as genotypic linkage disequilibria. From the data obtained in this study, the gametic phase of alleles in double heterozygotes could not be deduced. Consequently, an individual with genotype PS/QR is indistinguishable from an individual with PR/QS. Therefore, the calculated values for genotypic linkage disequilibria are a composite of two separate possible associations; cross-gamete and within-gamete disequilibria (Weir 1996).

Genotypic linkage disequilibria between loci were estimated for each locus-locus combination. Due to the small within-host sample sizes, data from all hosts were pooled within sample sites to calculate the observed genotype combinations. This

makes the assumption that there is no significant variation in genotype frequencies between parasites from different hosts within sites. Whether this is a valid assumption will be discussed later (section 4.4.6). Contingency tables of the numbers of genotypes at each pair of loci within each sample site were tested against the null hypothesis "genotypes at each loci are independent of one another" by exact tests using a Markov-chain method (section 4.2.7). The tests were performed using the program GENEPOP version 1.2. Overall significance of multiple tests was estimated by the use of Fisher's combined probability test.

4.2.5 Maximum-likelihood estimates of the frequency of null alleles

Data on the amplification of locus CM-2 demonstrated that there was a non-amplifying (null) allele at this locus within isofemale line 132 Heterogonic (section 2.3.3). During the genotyping of iL3s collected in section 3.2.2 for use in this genetic analysis, it was noted that (i) it was impossible to amplify DNA at this locus from certain iL3s despite several repeats of the PCRs and (ii) DNA from these iL3s was amplifiable by PCR for both other loci. It was thus suspected that there was a null allele present at this locus within the British samples of *S. ratti*.

Allele frequencies in each sample site were analysed for the presence of null alleles by re-coding the data set. Individuals that amplified at two loci but not the third were designated homozygous for a null allele at the non-amplifying locus and coded as such (instead of being designated 'missing data' by a '0000' code). For instance, if an iL3 amplified successfully at the Actin and BSP-8 loci, but not at CM-2, then the CM-2 genotype was coded homozygous for a dummy allele '0303' instead of '0000' in the dataset. Estimates were made on the frequency of the putative null alleles (coded with dummy alleles) using the EM algorithm of Dempster *et al.* (Dempster *et al.* 1977). The method calculates expected genotype frequencies from observed allele frequencies (the E step) and then calculates maximum likelihood estimates of allele frequencies assuming the expected genotype frequencies are the true genotype frequencies (the M step). The process is iterated until it converges to stable values, these being the maximum likelihood frequencies of alleles (including the null allele) within the population. Maximum likelihood calculations were performed on the dataset using the program GENEPOP.

4.2.6 Population genetic structure

To determine the extent of genetic subdivision between *S. ratti* samples, two methods were used. The first tests for significant variation in *allele* frequencies at each locus by the use of Fisher's exact test. The second estimates inbreeding coefficients for each locus (F-statistics) using *allele* and *genotype* frequencies, then tests for significant variation between samples by the use of permutation tests.

4.2.7 Between sample differentiation in allele frequencies

It is possible to estimate the significance of variation between allele frequencies of two samples by the use of probability tests. Contingency tables are drawn for individual loci where *Nij* represents the numbers of allele *j* within sample *i* (table 4.1)

	Alleles			
Sample	1	*j*	....	*k*
1	N_{11}	$Nj1$	....	$Nk1$
i	$Ni1$	Nij		Nkj
....	....	....	....	....
r	N_{r1}	....	....	N_{kr}

Table 4.1 Contingency table for testing the variation in allele frequencies between samples.

Chi-squared tests may then be used to test for independence between the cells in a contingency table and therefore acceptance of the null hypothesis, "independence between row and column variables". However, with multiallelic loci and/or low sample sizes, alleles are often at such low frequencies that the numbers recorded within some cells invalidate the use of the test. An alternative method of testing for independence between row and column variables is by the use of Fisher's exact test. Given the null hypothesis, the probability of the observed contingency table is given by:

$$\prod = \frac{\prod_{i=1}^{r}(Ni.!)\prod_{j=1}^{k}(N.j!)}{(N..!)\prod_{i=1}^{r}\prod_{j=1}^{k}(Nij!)} \qquad \text{(Eqn 4.8)}$$

The exact test statistic, p, is the sum for all tables having the same row and column sums, that have smaller probabilities of being found than the observed table. (Fisher 1930). For population data sets of normal size, this test is impossible to perform due to the numbers of tables concerned. Instead, an unbiased estimate of p may be formed by sampling a subset of the total number of contingency tables using a Markov-chain method (Guo & Thompson 1992; Raymond & Rousset 1995). For this sample, the test statistic is an unbiased approximation of p.

The method provides test statistics for each locus and has been shown to be accurate and unbiased for small samples and low-frequency alleles (Raymond & Rousset 1995) and to perform well in comparisons with other methods (Wilson 1996). However, it is assumed that the genotypic combinations follow from the observed allele frequencies, *ie.* the populations are in Hardy Weinberg equilibrium. If this is not the case, and alleles within samples are not independent, then exact tests may lead to incorrect results. To test for allelic differentiation between populations, the program GENEPOP was used. In all cases, the Markov chain was set to 100 000 steps and 1000 dememorisation steps.

The exact test may also be used to estimate (i) deviation of genotype frequencies from Hardy-Weinberg equilibrium and (ii) genotypic linkage disequilibria. These tests are performed as above except that for (i) the contingency tables contain single locus genotype frequencies and (ii) the contingency tables contain genotype counts for two-locus combinations (*ie.* r = genotype r at locus 1, k = genotype k at locus 2).

4.2.8 F-statistics

Wrights (Wright 1931) inbreeding coefficient is defined as the probability that two alleles at a locus are identical by descent (Hartl & Clarke 1989). It is normally calculated by comparing the observed numbers of heterozygotes with the expected numbers under conditions of random mating;

$$F = \frac{Ho - \overline{H}}{Ho} \qquad \text{(Eqn 4.9)}$$

Where $\overline{H}$ is the observed frequency of heterozygotes and Ho the frequency expected from Hardy-Weinberg equilibria. Inbreeding-like effects may occur as a consequence

of three processes. These are (i) selection, (ii) non-random mating and (iii) population subdivision.

In this project, the genetic markers used are assumed to be selectively neutral until otherwise proven, so it is only necessary to consider (ii) and(iii). As defined above, F is a measure of the amount of inbreeding within a population. However, if two populations have been separated, and are different due to drift or founder effects, they will exhibit a reduction in heterozygosity compared to that predicted from the total random mating population. This is known as a Wahlund effect. A measure of F can therefore be defined as:

$$F = \frac{H_T - H_S}{H_T} \qquad \text{(Eqn 4.10)}$$

Where H_T is the genetic diversity within the total population and H_S the diversity within the subpopulations. These two different aspects of inbreeding can be considered together as the hierarchical F-statistics F_{IS}, F_{IT} and F_{ST}, where the hierarchies are T, the total population, S, the subpopulation and I, the individual. Therefore, F_{IS} is the inbreeding coefficient due to assortative or disassortative mating within populations. F_{ST} is the inbreeding coefficient due to population subdivision and F_{IT} is the total inbreeding coefficient. These three coefficients are related by:

$$(1 - F_{IT}) = (1 - F_{IS})\,(1 - F_{ST}) \qquad \text{(Eqn 4.11)}$$

The value F_{ST} may also be considered as the standardised variance of allelic frequencies between populations (Weir & Cockerham 1984), with possible values ranging from 0 (no variation between populations) to 1 (complete differentiation). Methods of estimating F_{ST} from finite numbers of samples and for multiple loci with more than two alleles were developed by Nei (G_{ST}) (Nei 1975) and Weir and Cockerham (θ) (Weir & Cockerham 1984). Both estimators calculate values with the same expectation as F_{ST}, but the methods differ in that θ includes terms that correct for the number of sites sampled, and the sample size from each site. In this sense θ can be seen to be an unbiased estimator of F_{ST} whereas G_{ST} is not. However, qualitative comparisons between the two methods (Slatkin & Barton 1990) based on their relative accuracy in measuring gene flow between demes showed that they performed equally well. Due to the small sample sizes at the within-host level of sampling, θ was henceforth used here as the estimator of F_{ST} due to corrections for sample size that are included within the formulae. Simulations by Wilson (Wilson 1996) on the relative

power of the two tests, and that of Fisher's exact test, found that methods (i) and (ii) gave similar results, and that exact tests were the most powerful of the three. These simulations also showed that the strategy of sampling small numbers of individuals from many populations detected population subdivision more readily than large sample sizes from few populations.

4.2.9 An analysis of variance method for measuring population differentiation

Cockerham (Cockerham 1973) and Weir and Cockerham (Weir & Cockerham 1984) developed a method of estimating inbreeding coefficients from an analysis of variance (ANOVA) of allele and genotype frequencies. For a locus *j* in a sample from population *i*, a variable x_{ij} is defined, for an allele *A*, as x_{ij} =1 if *A* is present, or x_{ij} =0 if *A* is absent. For each locus, sums of squares are calculated from allele and genotype frequencies. ANOVA is then used to calculate the variance components between the observed and expected mean squares. These variance components are used to estimate the appropriate F-statistics as follows:

The samples collected in this study may be considered in a four-level hierarchy (i) I, the individual iL3s, (ii) R, the individual rats, (iii) S, the sample sites and (iv) T, the total sample. There are thus four sources of variation within the hierarchy of relatedness of alleles at a locus; within individual iL3s (σ_I^2) (ie. the variation between alleles within worms); among iL3s within rats (σ_R^2); among rats within sample sites (σ_S^2) and among sample sites (σ_T^2). From these components of variance, F-statistics are estimated as correlations between alleles within each level of the hierarchy, defined by Peakall *et al.* (Peakall *et al.* 1995) as:

(i) the correlation between alleles within an individual iL3 compared to those randomly drawn from the sampled rat,

$$F_{IR} = \frac{\sigma_R^2}{\sigma_I^2 + \sigma_R^2} \qquad \text{(Eqn 4.13)}$$

(ii) the correlation between alleles within an individual iL3 compared to those randomly drawn from the total data set,

$$F_{IT} = \frac{\sigma_R^2 + \sigma_S^2 + \sigma_T^2}{\sigma_I^2 + \sigma_R^2 + \sigma_S^2 + \sigma_T^2} \qquad \text{(Eqn 4.14)}$$

(iii) the correlation between alleles within a sampled rat compared to those randomly drawn from the sample site,

$$F_{RS} = \frac{\sigma_S^2}{\sigma_I^2 + \sigma_R^2 + \sigma_S^2} \qquad \text{(Eqn 4.15)}$$

(iii) the correlation between alleles within a sample site compared to those randomly drawn from the total data set,

$$F_{ST} = \frac{\sigma_T^2}{\sigma_I^2 + \sigma_R^2 + \sigma_S^2 + \sigma_T^2} \qquad \text{(Eqn 4.16)}$$

The relationship between these F-statistics is similar to that described in eqn. 4.11 and defined as;

$$(1-F_{IT}) = (1-F_{IR})(1-F_{RS})(1-F_{ST}) \qquad \text{(Eqn 4.17)}$$

For cases where there are several alleles at a locus, the variance components are combined as weighted averages. Similarly, data for several loci may be combined in this manner, the results being essentially unbiased (Weir & Cockerham 1984). The variance in sample size between populations is accounted for within the analysis by weighting variance components on the basis of sample size. Similarly, specific corrections are made to account for small numbers of sampled 'populations'.

Variance components and F-statistics were calculated using the Genetic Data Analysis package (GDA, Weir 1996).

Significance testing of F-statistics

In order to assess the significance of F-statistics, a permutational approach was used. This approach is described in detail by Excoffier *et al.* (Excoffier *et al.* 1992) and Goudet (Goudet 1994), and determines whether F-values are greater (or less than) zero by generating null distributions from the observed data. For F_{IS} and F_{IT}, this is done by randomly resampling *alleles* (with replacement) from the sample site (F_{IS}) or the total sample (F_{IT}) a thousand times, new values of $F_{IS/T}$ being re-calculated each time. The *p*-value is the proportion of permutations that result in $F_{IS/T}$ > the observed $F_{IS/T}$. If $p<0.05$ or 0.01, then the null hypothesis that $F_{IS/T}$ are equal to zero is

rejected. Similarly, values of F_{RS} and F_{ST} are tested in this way, except that the units of permutation are (i) the *genotypes* of individual iL3s from rats randomly re-sampled within a sample site (F_{RS}) and (ii) the *genotypes* of individual iL3s from within a sample site randomly re-sampled between sites. Permutation tests were performed using the program FSTAT (Goudet 1994).

One drawback of this approach to estimating the significance's of F-statistics is that differences in the magnitude of statistics between samples or sample sites cannot be compared. The bootstrapping approach may be used to generate variances for the observed statistics, and thus enable comparisons to be made. This method randomly resamples each sample site a large number of times, with replacement, to generate new F-statistics. From the variance of these values, confidence intervals are calculated that reflect the sample sizes for each sample site. To estimate significance's for all loci, 95% bootstrap confidence intervals across loci were calculated using the GDA computer package (Weir 1996).

4.2.10 Isolation by distance

The amount of differentiation between two populations may be used as a measure of genetic distance between them. Finite populations separated for a period of time will differentiate as a consequence of drift. Specifically;

$$d = -\ln(1- \theta) \qquad \text{(Eqn 4.18)}$$

$$\approx \frac{t}{2N_e} \qquad \text{(Eqn 4.19)}$$

where d is divergence due to drift over time t, N_e is the effective population size and θ the genetic differentiation between two populations (Weir 1996). If a roughly uniform migration rate is assumed, then in a species with many population subdivisions, genetic differentiation between subpopulations will show a correlation with increase in geographical distance (*ie.* isolation by distance, (Crow & Kimura 1970)).

To test for isolation by distance, pairwise values of F_{ST} were calculated for all British sample-site combinations. The German populations were excluded due to small sample-sizes. Geographical measurements used were the shortest distances between sites measured on a map. The relationship between F_{ST} and geographical distance was examined by calculating Spearman's rank correlation coefficient (Sokal & Rolf 1992).

To test for independence between geographical and genetic distances, Mantel tests were performed using the NTSYS statistical package with 10 000 permutations.

4.2.11 Within host genetic diversity

For each sampled rat, three measures of genetic diversity were calculated.

(i) Diversity = the number of genotypically distinct iL3s present in each rat sample. This measure used data from all loci together;

$$D = ni+nj+nk \qquad \text{(Eqn 4.20)}$$

where *ni, nj* and *nk* are the number of genotypes present within a rat at each of three loci *i, j* and *k*. If no genotypes were successfully scored at one or two loci, then the *D* was measured using data from one, or two loci. As each iL3 is genetically identical to its parent, *D* is a measure of the number of genetically unique parasites within the rat.

(ii) Diversity within a rat was calculated as the number of complete, unique three-locus genotypes.

4.2.12 Correlations of genetic diversity (corrected for intensity of infection) and worm burden

This analysis was performed in order to determine whether the number of genotypes observed per rat increased proportionally with the intensity of infection. The number of genotypes from each rat, corrected for intensity of infection, were calculated as the ratios;

R_a; (number of genotypes/rat)/(number of larvae/g faeces/rat)

R_b; (number of three-locus genotypes/rat)/(number of larvae/g faeces/rat).

These ratios were correlated against estimated worm burdens, significances being assessed by an ANOVA.

4.3 Results

4.3.1 Data set

A total of 741 iL3s collected from 72 rats and 11 sample sites were genotyped for RFLPs by restriction digestion of three PCR amplified loci, Actin, BSP-8 and CM-2. A total of 2223 PCRs were carried out, of which 1123 were successful (50.5%). Thus, a total of 2246 alleles were sampled from British and German *S. ratti* infections yielding 1123 single locus iL3 genotypes.

PCRs were performed once, and if they did not work, were performed a second time. There was a high degree of redundancy in the technique as 49.5% of PCRs did not yield any data. The PCR success varied for each locus; 59.5% and 52.6% of reactions worked for Actin and BSP-8 loci respectively. There was a high correlation between successful amplification of BSP-8 and Actin loci, a successful PCR for BSP-8 almost always being successful for Actin. Amplification of locus CM-2 was successful in 39.4% of reactions. There was little correlation between successful amplification of this locus and that of the other two. Successive attempts made to amplify this locus from DNA samples that amplified at the other loci were invariably unsuccessful.

4.3.2 Genetic diversity

4.3.2a Allele frequency distributions

The numbers of alleles at each locus observed within each population and the sample sizes are shown in Table 4.2. The allele frequencies at each locus within each sample site are shown in Table 4.3 and Figures 4.1 - 4.2.

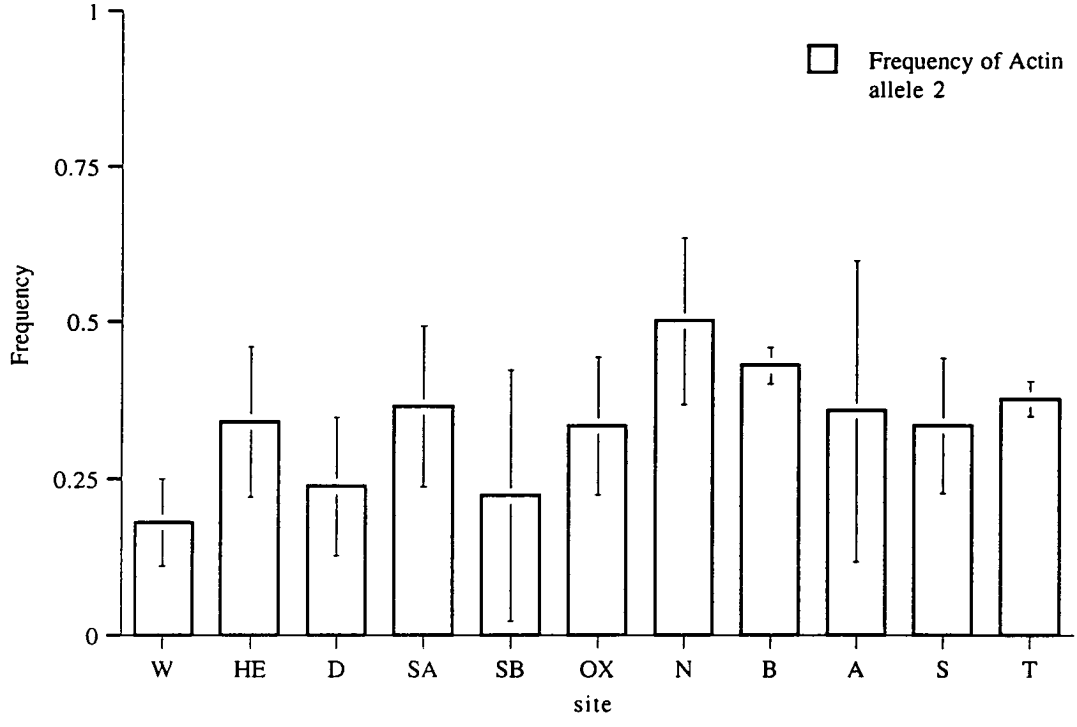
Actin locus: All populations were polymorphic, and diallelic at this locus (Table 4.2). Average frequencies of Actin allele 1 within the pooled sample was 0.626 (Table 4.3). The frequency of this allele ranged from a minimum of 0.500 at site N to a maximum of 0.820 at site W (Figure 4.1). Allele frequencies within the sampled German rats were similar to the mean of the British samples, with values of 0.642 and 0.677 recorded from Asbeck and Sendenhorst respectively.

BSP-8 locus: This locus has one common (allele 2) and one rare (allele 1) allele in five of the eight British sites (Table 4.2). The remaining three sites were fixed for allele 2. The average frequency of BSP-8 allele 2 across Britain was 0.943 with a

	Locus					
	Actin		BSP-8		CM-2	
Sample site	a	*n*	a	*n*	a	*n*
W	2	50	2	47	3	34
HE	2	25	2	18	3	16
D	2	19	0	24	2	7
SA	2	33	0	28	3	19
SB	2	9	0	7	1	2
OX	2	21	2	8	0	0
N	2	32	2	35	2	35
B	2	231	2	203	3	172
A	2	6	1	8	0	0
S	2	15	2	12	1	7
Total	2	441	2	390	4	292

Table 4.2 Number of alleles found "a", and numbers of iL3s genotyped "*n*", within
each sample site for each of the three loci. "Total" refers to the pooled data from all sample sites. W = Wiltshire, HE = Hedley, D = Dorset, SA = Surrey A, SB = Surrey B, N = Norfolk, B = Berkshire, A = Asbeck and S = Sendenhorst.

A.



B.

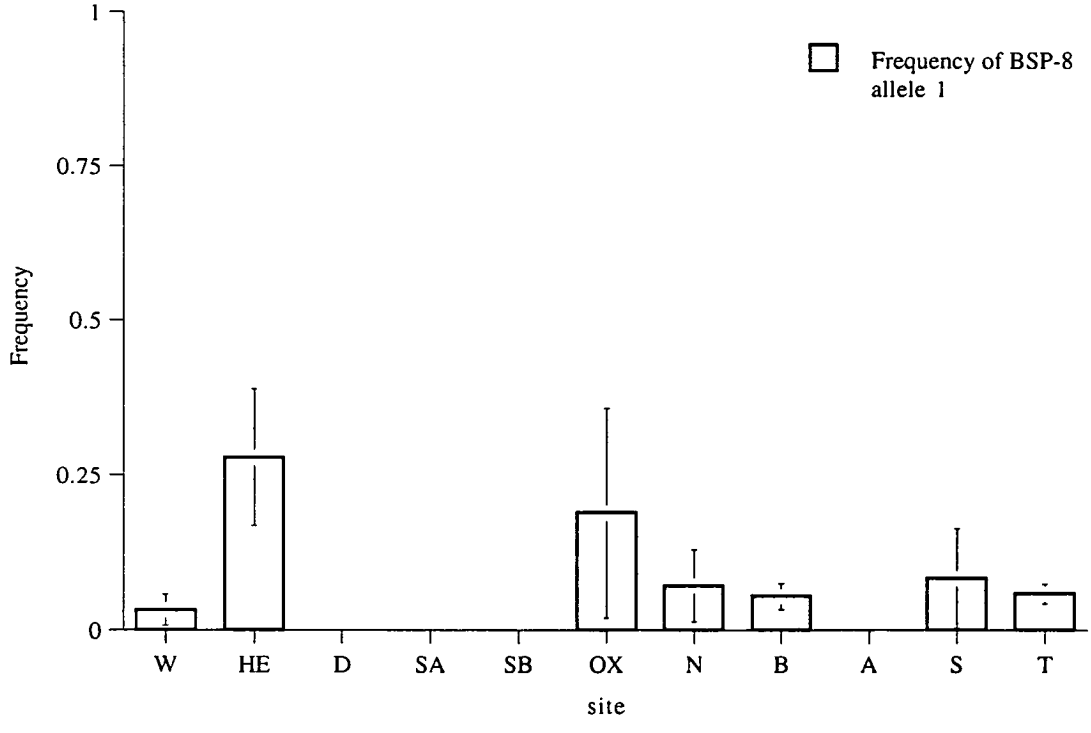
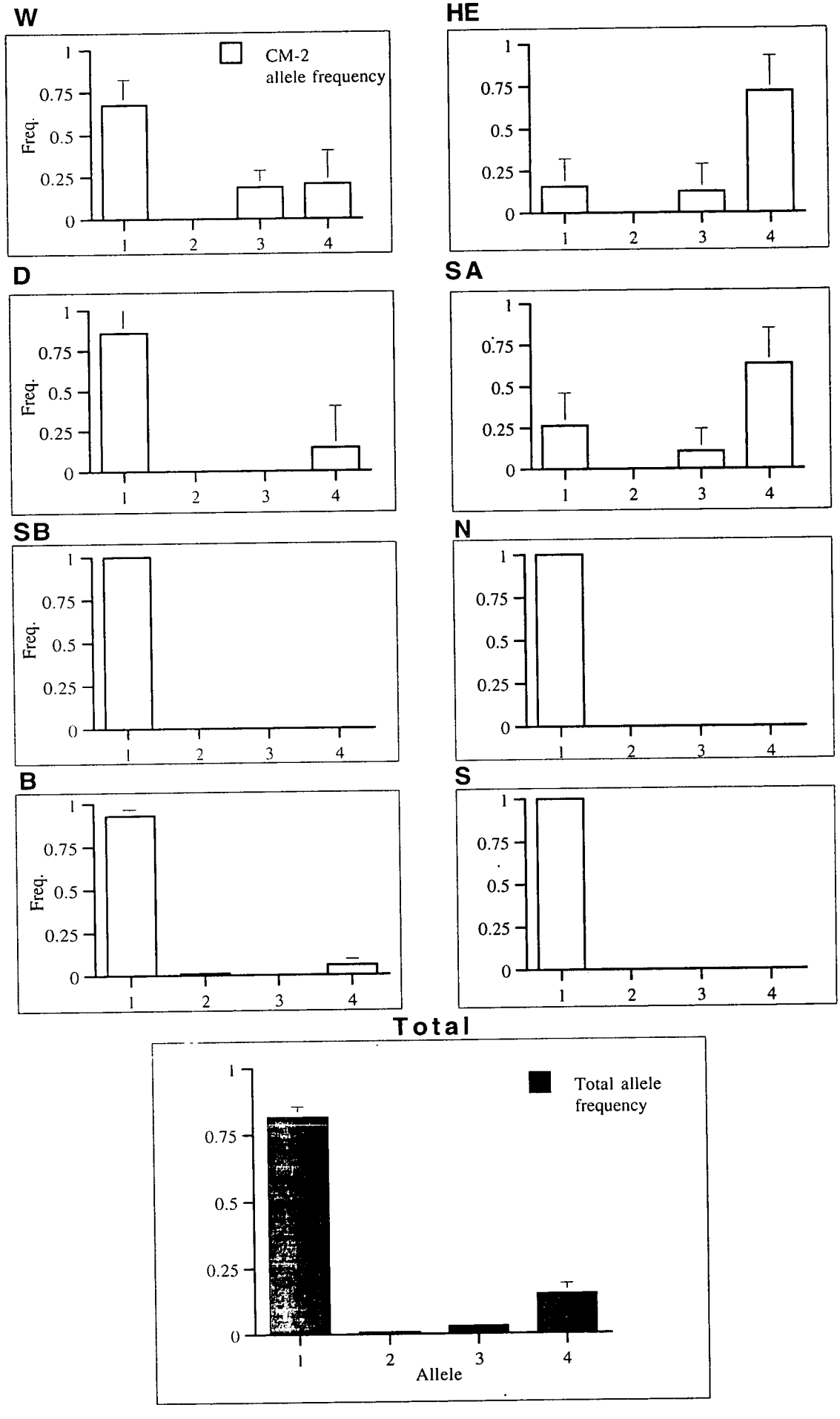


Figure 4.1 Allele frequencies at the Actin (A) and BSP-8 (B) loci within each sample site. W = Wiltshire, HE = Hedley, D = Dorset, SA = Surrey A, SB = Surrey B, N =Norfolk, B = Berkshire, A = Asbeck and S = Sendenhorst. 95% confidence intervals are shown as bars.

Figure 4.2

Frequencies of alleles 1, 2, 3 and 4 at the CM-2 locus within each sample site. W = Wiltshire, HE = Hedley, D = Dorset, SA = Surrey A, SB = Surrey B, N =Norfolk, B = Berkshire and S = Sendenhorst. "Total" refers to the pooled data for all sample sites. 95% confidence intervals are shown as bars.


W
CM-2
allele frequency
Freq.
HE
D
SA
SB
N
B
S
Total
Total allele
frequency
Allele

		Sample site										
Locus	Allele	W	HE	D	SA	SB	OX	N	B	A	S	Total
Actin	1	0.820 (± 0.070)	0.660 (± 0.120)	0.763 (± 0.110)	0.636 (± 0.128)	0.778 (± 0.223)	0.667 (± 0.110)	0.500 (± 0.133)	0.571 (± 0.029)	0.642 (± 0.240)	0.677 (± 0.108)	0.626 (±0.000)
	2	0.180 (± 0.070)	0.340 (± 0.120)	0.237 (± 0.110)	0.364 (± 0.128)	0.222 (± 0.223)	0.333 (± 0.110)	0.500 (± 0.133)	0.429 (± 0.029)	0.357 (± 0.240)	0.333 (± 0.108)	0.374 (±0.000)
BSP-8	1	0.032 (± 0.034)	0.278 (± 0.110)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.188 (± 0.168)	0.071 (± 0.058)	0.054 (± 0.021)	0.000 (0.000)	0.083 (± 0.104)	0.058 (±0.000)
	2	0.968 (± 0.034)	0.722 (± 0.110)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.813 (± 0.168)	0.929 (± 0.058)	0.946 (± 0.021)	1.000 (0.000)	0.917 (± 0.104)	0.943 (±0.000)
CM-2	1	0.676 (± 0.150)	0.156 (± 0.167)	0.857 (± 0.259)	0.263 (± 0.198)	1.000 (0.000)	-	0.971 (± 0.056)	0.930 (± 0.038)	-	1.000 (0.000)	0.820 (±0.000)
	2	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-	0.000 (0.000)	0.012 (± 0.016)	-	0.000 (0.000)	0.006 (±0.000)
	3	0.188 (± 0.106)	0.125 (± 0.165)	0.000 (0.000)	0.105 (± 0.138)	0.000 (0.000)	-	0.000 (0.000)	0.000 (0.000)	-	0.000 (0.000)	0.027 (±0.000)
	4	0.206 (± 0.198)	0.719 (± 0.211)	0.143 (± 0.259)	0.632 (± 0.217)	0.000 (0.000)	-	0.029 (± 0.056)	0.058 (± 0.035)	-	0.000 (0.000)	0.146 (±0.000)

Table 4.3 Allele frequencies for the Actin, BSP-8 and CM-2 loci within eight British and two German sample sites. The allele designations are those used in Chapter 2 (section 2.3.8c). Values in parentheses are 95% confidence intervals of the allele frequency estimates. "Total" refers to the allele frequency of each locus, for all sample sites pooled. "-" signifies that no data were collected.

range in frequency from 1.000 at sites D, SA and SB, to a minimum of 0.722 at site HE (Table 4.3, Figure 4.1). The mean frequency of allele 2 was similar within the German samples with mean values of 1.000 and 0.917 recorded from Asbeck and Sendenhorst respectively.

CM-2 locus: Of the four alleles at locus CM-2, alleles 2 and 3 were rare or absent in the sample sites (Table 4.3). Allele 1 was the common allele at five out of eight British sites (W, D, SB, N and B) with an average frequency of 0.816. At two sites (HE and SA) allele 4 was the most common with frequencies of 0.719 and 0.632 respectively (Figure 4.2). iL3s sampled from Sendenhorst in Germany were monomorphic at this locus, and like most British sites, had a high frequency of allele 1. No iL3s were successfully genotyped at this locus from the other German site.

4.3.2b Variation in allele frequencies between sample sites

The significance of heterogeneity in allele frequencies between sample sites was assessed by Fishers' exact test. This analysis was performed for the total British data set and between all pairwise combinations of sample sites. The results are presented in Table 4.4.

Tests between each pairwise combination of sample sites showed that 6/25, 10/25 and 12/25 combinations were significantly different in allele frequencies at the Actin, BSP-8 and CM-2 loci respectively. After adjusting significance levels for the effects of multiple comparisons using a sequential Bonferroni technique, 3/25, 5/25 and 11/25 sites remained significantly differentiated at the Actin, BSP-8 and CM-2 loci respectively. Heterogeneity in allele frequencies across the total dataset was significant for each locus at $p < 0.001$. Removing the most differentiated sample site for each locus (W for Actin, HE for BSP-8 and W for CM-2) resulted in a nonsignificant p value for Actin. However, the remaining sample sites remained significantly differentiated at the other two loci.

4.3.2c Variation in allele frequencies between rats within each sample site

Within each sample site, the statistical significance of variation in allele frequencies at each locus between individual rats was calculated by exact tests. The German sample from the Asbeck site was excluded from this analysis as only one rat was sampled.

Sample site combination	Locus	*p*
W-HE	Actin	0.040
W-SA	"	0.010
W-N	"	0.000*
W-B	"	0.000*
D-N	"	0.012
D-B	"	0.025
All sites	"	0.000*
W-HE	BSP-8	0.000*
W-OX	"	0.039
HE-D	"	0.000*
HE-SA	"	0.000*
HE-SB	"	0.040
HE-N	"	0.006
HE-B	"	0.000*
HE-S	"	0.010
D-OX	"	0.000
SA-OX	"	0.000
All sites	"	0.000*
W-HE	CM-2	0.000*
W-SA	"	0.000*
W-N	"	0.000*
W-B	"	0.000*
HE-D	"	0.000*
HE-SB	"	0.004
HE-N	"	0.000*
HE-B	"	0.000*
D-SA	"	0.000*
SA-SB	"	0.015
SA-N	"	0.000*
SA-B	"	0.000*
All sites	"	0.000*

Table 4.4 Sample site combinations that had significantly different allele frequencies as measured by exact tests. W = Wiltshire, HE = Hedley, OX = Oxford, SA = Surrey A, SB = Surrey B, N = Norfolk, B = Berkshire and D = Dorset. * indicates sample sites that remain significantly differentiated ($p < 0.05$) after application of sequential Bonferroni procedures to adjust significance levels.

Table 4.5 shows p values for each sample site, and indicates the sample sites for which variation in allele frequency was significantly different. Variation between rats was low for Actin and BSP-8; 3/9 and 2/6 sites showed significant differences in allele frequencies at these loci. Greater heterogeneity was observed between rats at the CM-2 locus where there were significantly different allele frequencies between rats at all sample sites.

	Locus		
Sample site	Actin	BSP-8	CM-2
B	0.334	0.043*	0.000**
N	0.031*	0.010**	0.026*
W	0.000**	0.359	0.000**
D	0.361	-	-
SA	0.002**	-	0.000**
SB	0.080	-	-
OX	0.060	0.730	X
HE	0.137	1.000	0.000**
S	0.693	0.448	-

Table 4.5 Within-site differentiation in allele frequencies between iL3s from individual rats in each sample site as measured by exact tests. The numbers in the table are the probabilities that samples are significantly different. "*" signifies the level of significance, *p<0.05, **p<0.01. "-" signifies sample sites that are fixed for one allele, "X" signifies no data.

Within the sites that show significant difference in allele frequency between rats, it is often the sample from a single rat that biases the estimation. For instance, within rat W325 from site W, all nine iL3s were of the genotype '12' for the Actin locus, whereas the most common genotype within this site is '11' for Actin within all but one of the other rats (Table 4.6). Similarly, a relatively few rats bias the estimates for the CM-2 locus. In particular, samples from rats SA 30 A, HE 267 and B 369 have high frequencies of the rare allele 4.

In order to quantify the influence of these 'aberrant' rats on the measures of differentiation, for each locus the single-most differentiated rat was removed from each sample site and exact tests recalculated. This had the following effect;

Table 4.6.

Genotypes of *S. ratti* iL3s typed from the faecal cultures of 72 rats.

[a] The loci are listed as the numbers of homozygotes and heterozygotes found for each locus. The genotypes "11" for BSP-8 and "13", "23", "24" and "34" for CM-2 are not shown because none of these genotypes were found. "-" indicates that no data was collected.

		LOCUS									
		Actin[a]			BSP-8		CM-2				
Site	Rat	11	12	22	12	22	11	12	33	14	44
B	B367	2	8	0	0	6	5	0	0	0	2
	B359	5	7	0	0	11	-	-	-	-	-
	B381	0	12	0	0	12	-	-	-	-	-
	B356	0	11	0	2	9	4	0	0	0	0
	B365	1	10	0	1	7	14	0	0	0	0
	B350	1	3	0	-	-	-	-	-	-	-
	B363	3	2	0	0	1	1	0	0	0	0
	B376	2	9	0	1	10	13	0	0	0	0
	B351	1	11	0	1	10	9	0	0	0	0
	B353	0	11	0	0	7	12	0	0	0	0
	B378	0	8	0	5	4	5	0	0	0	0
	B358	2	4	2	0	4	13	0	0	0	0
	B369	3	6	0	3	3	0	3	0	0	6
	B368	4	8	0	1	10	15	0	0	0	0
	B380	8	3	0	0	12	8	0	0	0	0
	B370	0	9	2	1	10	14	0	0	0	0
	B366	1	11	0	0	12	12	0	0	0	0
	B374	1	7	1	1	7	4	0	0	0	0
	B377	3	5	1	1	8	12	1	0	0	0
	B360	2	5	0	0	9	1	0	0	0	2
	B373	0	11	1	2	9	6	0	0	0	0
	B352	0	8	1	2	7	10	0	0	0	0
	B375	0	9	1	1	9	-	-	-	-	-
	B372	3	2	0	0	4	-	-	-	-	-
W	W325	0	9	0	0	10	-	-	-	-	-
	W313	1	0	0	-	-	-	-	-	-	-
	W328	6	0	0	0	1	-	-	-	-	-
	W316	4	0	0	1	5	-	-	-	-	-
	W332	2	0	0	-	-	2	0	1	0	2
	W330	3	3	1	0	2	0	0	3	0	0
	W317	1	1	0	0	5	-	-	-	-	-
	W323	1	2	0	1	9	-	-	-	-	-
	W320	4	0	0	0	5	-	-	-	-	-
	W318	1	1	0	-	-	-	-	-	-	-
	W331	5	0	0	1	1	-	-	-	-	-
	W315	5	0	0	0	6	0	0	0	0	4
	W321	-	-	-	-	-	9	0	0	0	0
	W322	-	-	-	-	-	7	0	0	0	0
	W324	-	-	-	-	-	5	0	0	0	1
HE	HE265	3	6	0	6	3	-	-	-	-	-
	HE294	0	1	0	-	-	1	0	0	0	1
	HE271	2	1	0	1	1	0	0	2	0	1
	HE285	0	1	0	-	-	-	-	-	-	-
	HE267	1	3	1	3	2	0	0	0	1	9
	HE287	0	0	1	0	1	-	-	-	-	-
	HE292	4	1	0	0	1	1	0	0	0	0
D	D336	4	4	0	0	8	-	-	-	-	-
	D334	4	1	0	0	4	-	-	-	-	-
	D335	0	3	0	0	6	-	-	-	-	-
	D338	2	1	0	0	4	-	-	-	-	-
	D339	-	-	-	0	2	6	0	0	0	1
SA	SA30	2	4	0	0	6	0	0	0	0	9
	SA27	3	0	0	-	-	3	0	1	0	0
	SA31	1	2	0	0	3	-	-	-	-	-
	SA28	1	0	0	0	6	1	0	1	0	0
	SA26	6	0	0	0	7	0	0	0	0	3
	SA29	0	6	0	0	5	-	-	-	-	-
	SA32	2	0	6	0	1	1	0	0	0	0
SB	SB25	0	0	1	0	1	-	-	-	-	-
	SB26	1	0	0	0	1	-	-	-	-	-
	SB28	5	2	0	0	5	2	0	0	0	0
OX	T1/2	3	5	1	0	2	-	-	-	-	-
	T1/3	1	3	0	1	2	-	-	-	-	-
	T1/5	0	0	1	-	-	-	-	-	-	-
	T1/7	1	2	0	2	1	-	-	-	-	-
	T1/8	4	0	0	-	-	-	-	-	-	-
N	N384	1	7	5	5	8	14	0	0	0	0
	N383	4	4	4	0	12	15	0	0	0	0
	N385	5	1	1	0	10	5	0	0	0	1
A	A2932	2	4	0	0	8	-	-	-	-	-
S	S2924	2	7	0	1	6	0	0	0	0	3
	S2932	3	3	0	1	4	0	0	0	0	4

Actin: Variation in allele frequencies between rats remained significantly different in two sites, W and SA. One site (N) was no longer significantly different.

BSP-8: Variation within all sites was no longer significantly different.

CM-2: Variation between rats was significantly different at all sites except N, which was no longer significantly different.

Therefore, removing the single most differentiated rat from those sampled at a given sample site caused between-rat variation in allele frequencies to become no longer significantly different in 4/10 comparisons.

4.3.3 Heterozygosity

Observed heterozygosities for each locus and sample site are shown in Table 4.7. Large differences in mean heterozygosity were observed between loci; this ranged from a mean frequency of 0.609 at Actin to a mean frequency of 0.115 at locus BSP-8 for values calculated from the pooled data set. The least heterozygous locus was CM-2 with a mean heterozygosity of 0.017 across the whole data set. This corresponds to only five out of the 285 iL3s genotyped at this locus being heterozygous.

Mean heterozygosity across all loci and sample sites was 0.247. The range in value of mean heterozygosity for sites was from a maximum of 0.379 at HE to a minimum of 0.074 at SB.

Heterozygosity varied between sites for each locus, with the highest range seen for Actin (± 0.279) and less for BSP-8 (± 0.189) and CM-2 (± 0.003). The significance of heterogeneity in heterozygosity (excluding Germany) was examined using ANOVA comparing variances at three levels; between loci, between sample sites and between loci and sample sites. The results of the analysis showed that the variance in heterozygosity between sample sites was significant for Actin and BSP-8 (p=0.0001), but not CM-2 ($F_{1, 7}$ = 0.695, p=0.691). Variance components were significant at the p=0.001 level for all other levels of the analysis, between individual loci and between loci and sample sites.

This analysis shows that there is heterogeneity in heterozygosities between all loci, and between sample sites for two out of three loci. However, the rank order of heterozygosities between loci is mostly consistent for all sample sites, *i.e.*

LOCUS		SAMPLE SITE										
		W	HE	D	SA	SB	OX	N	B	A	S	MEAN
Actin	H_o	0.320	0.520	0.474	0.364	0.222	0.476	0.375	0.779*	0.714	0.667	0.609*
		(0.063)	(0.090)	(0.103)	(0.070)	(0.150)	(0.098)	(0.080)	(0.029)	(0.064)	(0.045)	(0.023)
	H_e	0.295	0.440	0.361	0.463	0.346	0.444	0.500	0.490	0.459	0.444	0.468
	F_{IS}	0.084	-0.159	-0.310	0.214	0.357	-0.071	0.250	-0.591*	-0.556	-0.500	-0.299*
BSP-8	H_o	0.064	0.556	0.000	0.000	0.000	0.375	0.143	0.108	0.000	0.167	0.115
		(0.044)	(0.071)	(0.000)	(0.000)	(0.000)	(0.107)	(0.051)	(0.021)	(0.000)	(0.071)	(0.012)
	H_e	0.062	0.401	0.000	0.000	0.000	0.305	0.133	0.103	0.000	0.153	0.109
	F_{IS}	-0.033	-0.395	0.000	0.000	0.000	-0.231	-0.077	-0.057	0.000	-0.091	-0.061
CM-2	H_o	0.000*	0.063*	0.000	0.000*	0.000	-	0.000	0.023*	-	0.000	0.017*
		(0.000)	(0.033)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.010)	-	(0.000)	(0.008)
	H_e	0.486	0.443	0.245	0.521	0.000	-	0.056	0.131	-	0.000	0.305
	F_{IS}	1.000*	0.859*	1.000*	1.000*	0.000	-	1.000*	0.823*	-	0.000	0.944*
MEAN	H_o	0.128	0.379	0.158	0.121	0.074	0.284	0.173	0.304	0.238	0.278	0.247*
		(0.098)	(0.159)	(0.158)	(0.121)	(0.074)	(0.145)	(0.109)	(0.239)	(0.238)	(0.200)	(0.104)
	H_e	0.281	0.431	0.202	0.328	0.115	0.250	0.229	0.241	0.153	0.199	0.294
	F_{IS}	0.350*	0.098*	0.230*	0.317*	0.119	-0.151	0.391*	0.058*	-0.278	-0.300	0.198*

Table 4.7. Observed (H_o), expected (H_e) heterozygosities and F_{IS} for each locus at each sample site. Values in parentheses are standard errors. "*" signifie that Ho is significantly different from H_e at $p = 0.01$. "Mean" refers to the mean values for all sites, and all loci, pooled. "-" signifies that no data were collectec

heterozygosity for Actin is generally high, BSP-8 is intermediate between the other two loci and heterozygosity at CM-2 is always low. Exceptions are seen for sites W and He, where the heterozygosity of BSP-8 exceeds that of Actin. Whether the variation in heterozygosity is due to non-random mating is examined in section 4.3.4.

4.3.4 Hardy-Weinberg equilibria

Deviations from random mating expectations were tested for each locus within each sample site, and in the pooled British dataset, by exact tests. Six significant deviations from Hardy-Weinberg equilibria were observed ($p< 0.01$), representing 25% of the 24 tests performed. These sample sites and loci are shown in Table 4.8. After adjusting the significance levels by sequential Bonferroni procedures, five sample sites retained significant deviations. Of these five, one was due to a significant excess of heterozygotes at the Actin locus within the Berkshire sample site (site B, Table 4.8). All of the remaining departures from Hardy-Weinberg expectations were due to deficits of heterozygotes at the CM2 locus. Five out of seven sample sites showed significant heterozygote deficits at this locus.

Locus	Sample site	Deviation	*P*-value
Actin	B	Heterozygote excess	0.0001*
CM-2	W	Heterozygote deficit	0.0001*
CM-2	HE	Heterozygote deficit	0.0001*
CM-2	SA	Heterozygote deficit	0.0001*
CM-2	B	Heterozygote deficit	0.0001*
CM-2	N	Heterozygote deficit	0.01
Actin	Total	Heterozygote excess	0.0001*
CM-2	Total	Heterozygote deficit	0.0001*

Table 4.8. Sample sites and loci not in Hardy-Weinberg equilibria. Total refers to the exact test performed on the pooled British data. * indicates samples that remain significantly differentiated after application of sequential Bonferroni procedures.

Within sample site B, 180 heterozygous Actin genotypes were observed, compared to an expected value of 113. To see whether this excess of genotypes was

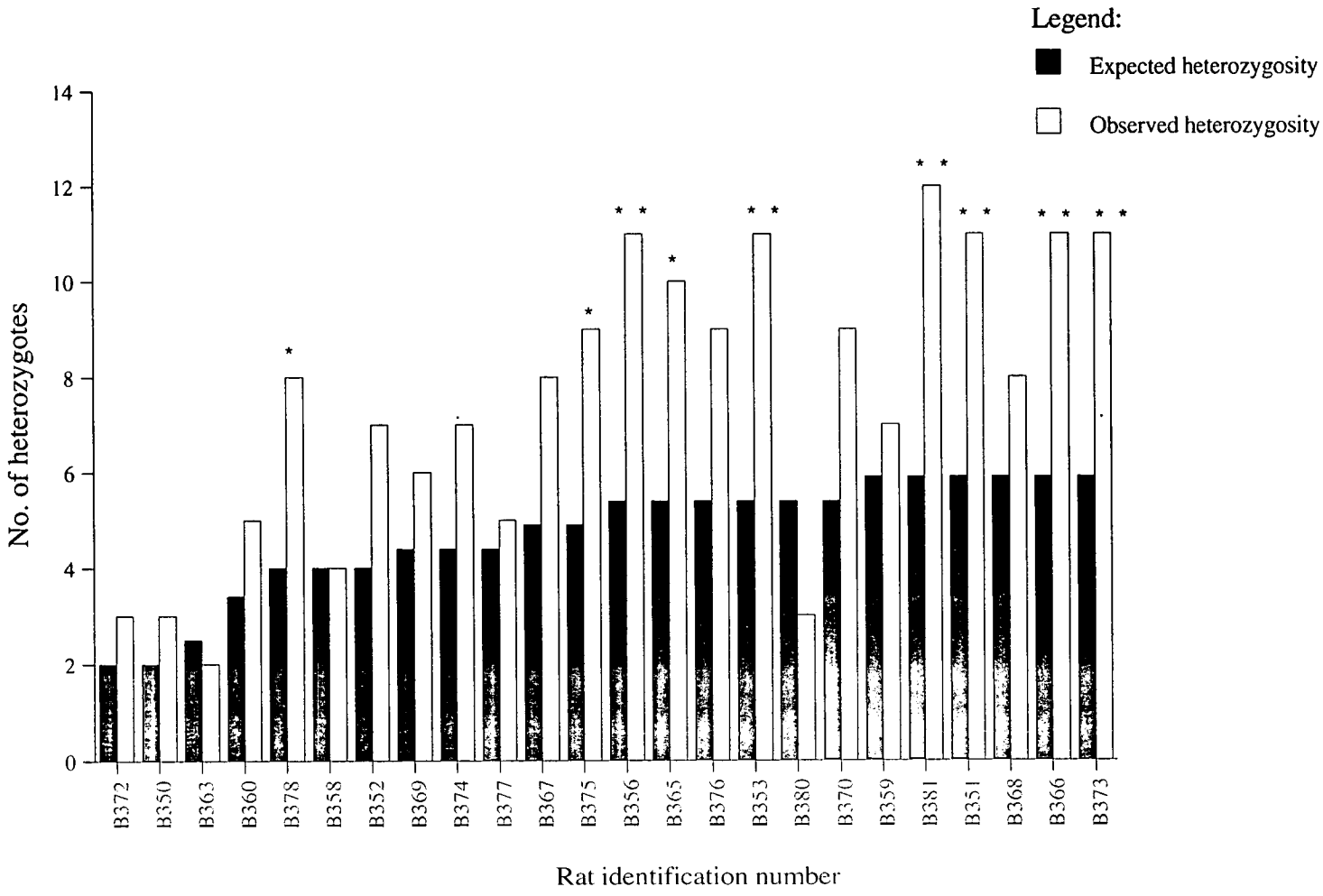


Figure 4.3. Observed and expected heterozygosity at the Actin locus for the sampled iL3 population from each Berkshire rat. "*" and "**" denote samples showing significant heterozygote excesses at $p = 0.05$, and $p = 0.01$ respectively.

due to the influence of a few 'aberrant' rats, the expected numbers of heterozygote iL3s were calculated for each rat. This was done by multiplying the total number of worms genotyped within each rat by the mean frequency of Actin heterozygotes within the Berkshire site. The number of heterozygote iL3s observed from each sampled rat relative to the expected numbers are shown in Figure 4.3. A large proportion of rats, 24 / 27, had more heterozygote iL3s than were expected. Nine of these were significantly different from Hardy-Weinberg expectations ($p < 0.05$). This shows that the heterozygote excess within the Berkshire sample site is a feature common to almost all rats, and does not result from a few unique infections that unduly bias the data from the whole sample site.

4.3.5 Maximum-likelihood estimates on the frequency of null alleles

Table 4.9 shows the estimated frequencies of putative null alleles in all sample sites for each locus. While there was no suggestion that null alleles existed for the Actin and BSP-8 loci, they were included in this analysis for the sake of completeness. All sample sites contained putative null homozygotes for locus CM-2 while only 4/10 and 5/10 sample sites contained iL3s that had putative null alleles at the Actin and BSP-8 loci. Frequencies of null homozygotes were demonstrably higher for locus CM-2 relative to the other loci within all sample sites. Maximum frequencies of the estimated null alleles within sample sites were 0.295, 0.316 and 1.000 for Actin, BSP-8 and CM-2, respectively. The corresponding frequencies for the pooled data set were 0.134, 0.257 and 0.616.

	Estimated frequencies of null alleles at each locus		
Sample site	Actin	BSP-8	CM-2
A	0.000	0.000	1.000
S	0.000	0.231	0.645
W	0.166	0.184	0.693
HE	0.000	0.231	0.649
D	0.000	0.000	0.837
SA	0.272	0.258	0.672
SB	0.295	0.000	0.866
OX	0.000	0.000	1.000
N	0.000	0.000	0.337
B	0.125	0.316	0.549
Pooled sample sites	0.134	0.257	0.616

Table 4.9. Estimated frequency of putative null alleles at each locus within each sample site.

While not conclusive, this analysis shows that PCR amplification of locus CM-2 is anomalous relative to that of the other two loci.

4.3.6 Tests of genotypic linkage disequilibria between loci

The null hypothesis that "genotypes at one locus are independent of genotypes at another locus" was tested for all pairwise-locus comparisons within each of the eight British sample sites using exact tests. Out of the 24 possible tests, 12 were not possible due to fixed alleles or missing (or insufficient) data for one of the loci. For the remaining 12 tests, two sample sites (HE and N) showed significant genotypic linkage disequilibria for the locus-combination 'Actin+BSP-8' at the $p < 0.05$, but not $p < 0.01$ level. Adjustment of the significance levels using Bonferroni procedures removed the significance of these values.

Pooling the data from all British sample sites found significant genotypic linkage disequilibria between two out of three 2-locus combinations. Disequilibrium was significant between Actin + BSP-8 at $p = 0.001$ ($\chi^2 = \infty$, d.f = 2) and Actin + CM-2 at $p = 0.002$ ($\chi^2 = 12.7$, d.f = 2). No disequilibria was found between BSP-8 and CM-2 ($p = 0.517$, $\chi^2 = 1.318$, d.f = 2).

Therefore, there is little evidence for non-random associations between genotypes at the three loci within sample sites, but pooling the samples from all sites (and thus increasing the sample size) showed that genotypes for two of the three loci were found in association more often than was expected by chance alone.

4.3.7a Hierarchical analysis of genetic diversity

The extent of genetic differentiation between samples from rats and sample sites was estimated by ANOVA of allele and genotype frequencies. F-statistics were estimated from each level of the sampling hierarchy, for each locus, by use of the variance components σ_I^2, σ_R^2, σ_S^2 and σ_T^2. These statistics were; F_{IT} = the total inbreeding coefficient, F_{IS} = the inbreeding coefficient of individual iL3s within sample sites, F_{ST} = the variation that was attributable to between-sample site differentiation and F_{RS} = the variation that was attributable to between-rat differentiation within sample sites. Due to low sample sizes, data from the two German sites was excluded from the analysis. The ANOVA data are presented in Table 4.10 and the F-statistics for each locus in Table 4.11.

The ANOVA for all loci showed that variation between sample sites was low, accounting for 1.4% of the total genetic variation, but significant at $p < 0.01$ (Table 4.10). For each locus separately, significant variation between sites was observed at BSP-8 and CM-2 (4.77% and 12.37% respectively), but not Actin (-6.04%). It is worth commenting here that values of F_{ST} and F_{RS} that are close to zero, using the Weir and Cockerham estimation procedure, have an equal chance of being below or above zero (Weir 1996). Therefore, unless proven otherwise, negative variance components were assumed to be equal to zero.

The null distributions of the variance component σ_T^2, created by a thousand permutations of individual iL3s between sample sites, are shown in figure 4.4. This is simply a visual representation of the distributions created by random permutations of data, with the positions of the observed data values indicated. Biases of σ_T^2 to the right of the null distributions reflects the degrees of significance. Therefore, the position of σ_T^2 for Actin in the middle of the null distribution shows that it is not significantly greater than zero.

The relatively high value of σ_T^2 for the CM-2 locus was mainly due to the influence of two rats, HE 267 and SA 30A, that showed unusually high frequencies of a rare genotype (genotype '44'). Removal of these two rats from the dataset reduced

Locus	Hierarchy	d.f	Variance	% of total variation
Actin	Between sample sites	7	-0.026	-6.04
	Between rats within sites	57	0.047	4.69**
	Within rats	781	0.611	95.31**
BSP-8	Between sample sites	7	0.005	4.77*
	Between rats within sites	49	0.004	8.65**
	Within rats	691	0.114	86.58**
CM-2	Between sample sites	6	0.033	12.37**
	Between rats within sites	32	0.143	66.10**
	Within rats	531	0.017	21.53*
Actin+BSP-8	Between sample sites	7	-0.021	-3.90
	Between rats within sites	58	0.051	3.81**
	Within rats	934	0.725	96.19**
All loci	Between sample sites	7	0.012	1.44**
	Between rats within sites	61	0.194	25.30**
	Within rats	1091	0.742	73.26**
All loci, with rats HE 267 and SA 30A removed.	Between sample sites	7	-0.09	-1.3
	Between rats within sites	59	0.179	12.8**
	Within rats	1055	0.743	87.2**

Table 4.10 Hierarchical ANOVA between iL3 samples from 69 rats and eight sample sites. d.f = degrees of freedom; Variance = mean squares variance at each level of the hierarchy. "*" signifies $p < 0.05$, "**" signifies $p < 0.01$.

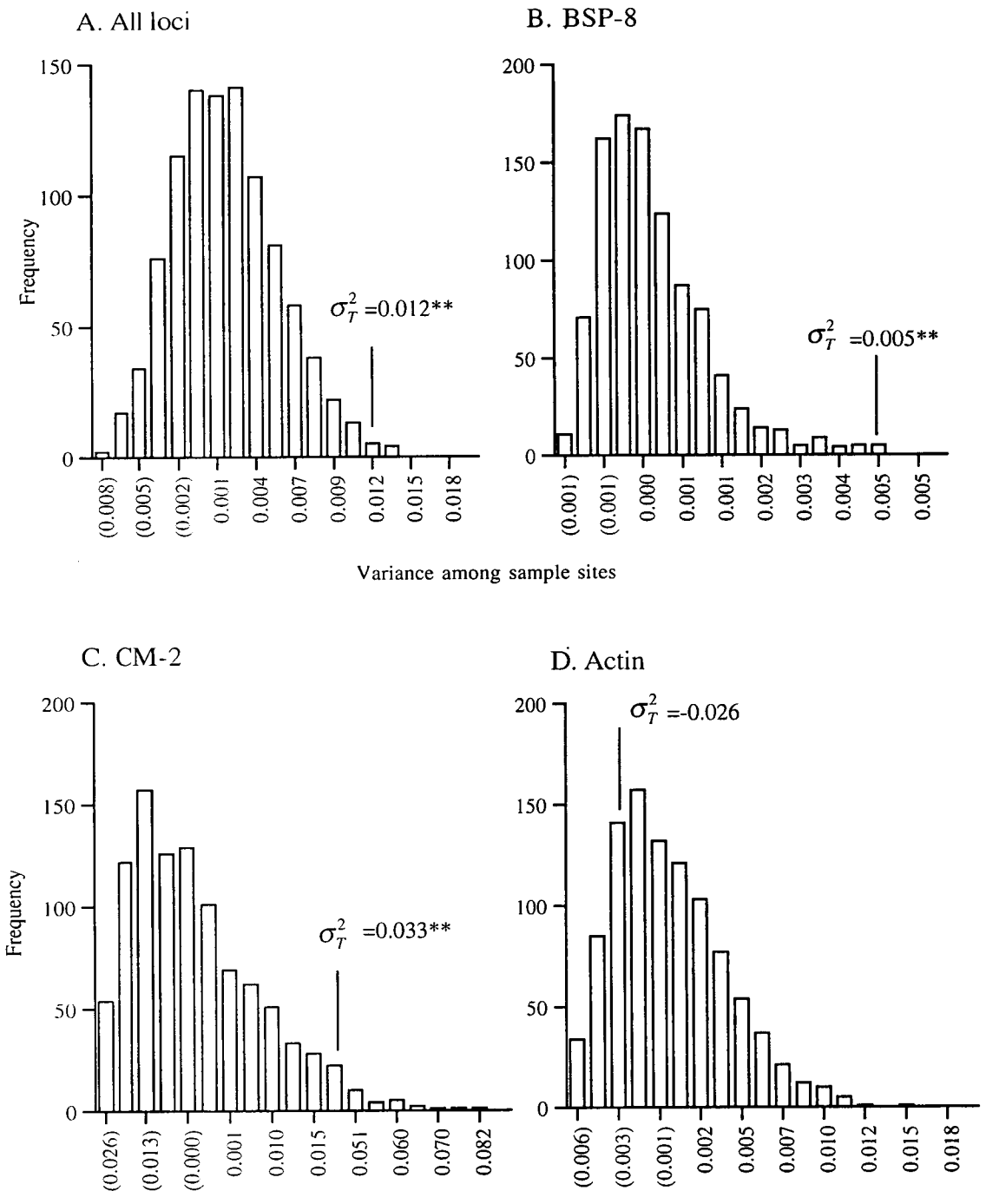


Figure 4.4. Null distributions of the variance components obtained by random permutations of rat samples between sample sites. Negative values are shown in parentheses. Observed among-site variance components are indicated on each graph.
A. Distribution for all loci combined, B. for BSP-8, C. for CM-2 and D. for Actin.
** significant at $p<0.01$

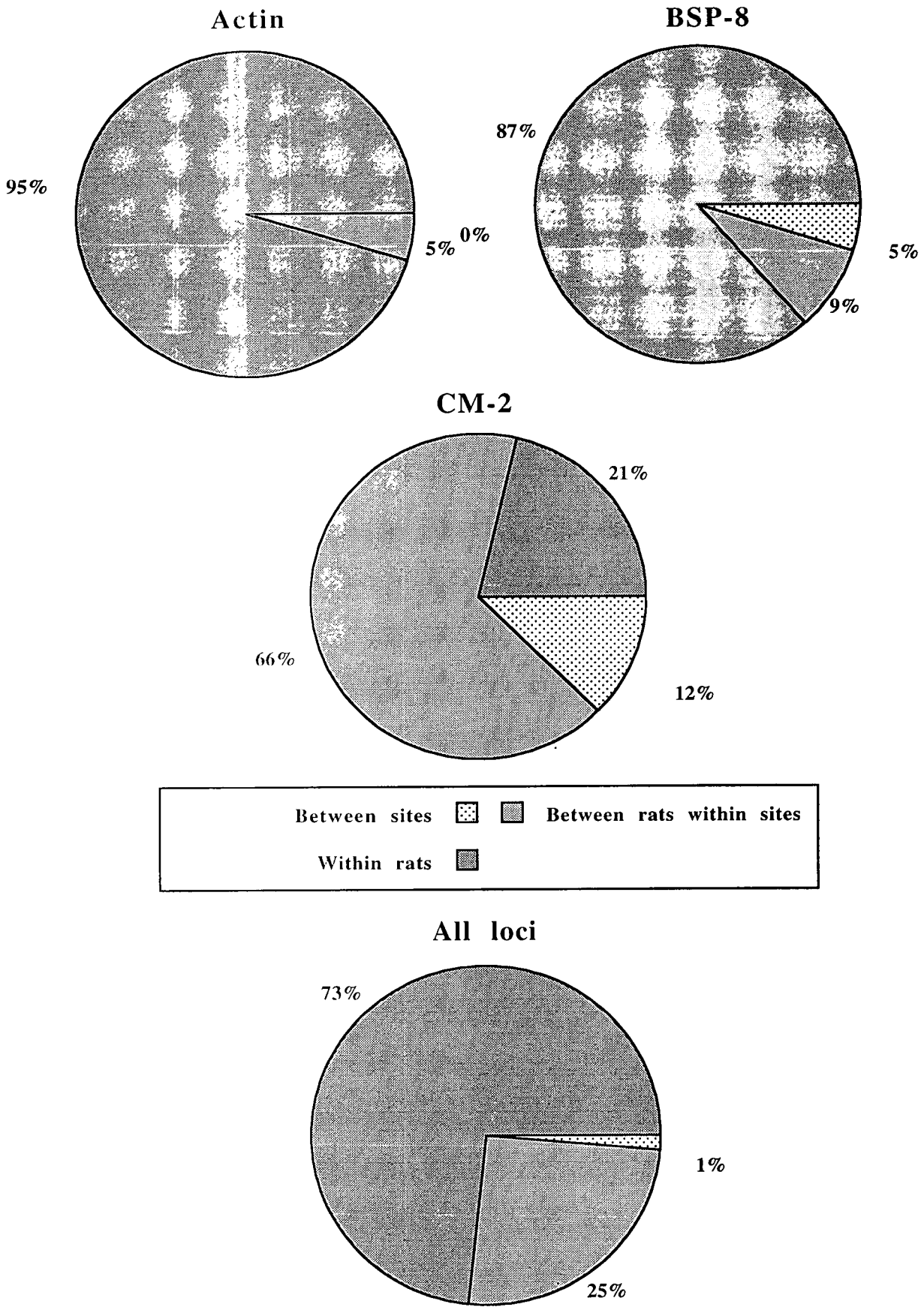


Figure 4.5. Pie charts of the variance components measured for each level of the hierarchical ANOVA for each locus separately and all loci together.

the between-site component of variation to a nonsignificant value (F_{ST} = -0.13). Thus, there was little evidence from these data that significant variance in allele or genotype frequencies occurs between sample sites.

Mean F_{RS} for all three loci showed that a significant component (25.3%, F_{RS} = 0.253) was attributable to between rat variation. A substantial proportion of this variation was due to the effect of the CM-2 genotypes from rats HE 267 and SA 30A, removal of which reduced the mean F_{RS} to 0.128 (12.8%). F_{RS} was considerably lower than that of CM-2, but still significant, for the Actin and BSP-8 loci considered independently, with 4.7% and 8.7% of the total variance due to between rat effects.

Therefore, from these data, it is apparent that the majority of genetic variation is found within individual rats (73%) (Table 4.10, Figure 4.5). The CM-2 locus shows more between-rat variation than the other two loci. This is mainly due to the effect of two rats (rats HE 267 and SA 30A). Removal of these rats from the dataset increases the measure of within rat genetic diversity to 87%. Removal of the CM-2 locus from the data set further increases this measure of within rat genetic variation to 96%. Thus, for the Actin and BSP-8 loci, only 4% of the total genetic variation is attributable to between rat and between site effects.

Locus	F-statistics			
	F_{IT}	F_{IS}	F_{RS}	F_{ST}
Actin[a]	-0.395**	-0.464**	0.047**	-0.060
BSP-8[a]	-0.063	-0.164**	0.086**	0.048*
CM-2[a]	0.934**	0.806**	0.658**	0.124**
Mean for all loci[c]	0.085 (0.12-⁻0.06)	0.198** (0.934-⁻0.394)	0.253** (0.659-0.047)	0.014** (0.123-⁻0.06)
Mean for all loci (no HE 267 and SA30A)[d]	-0.060 (0.08-⁻0.47)	-0.216** (0.90-⁻0.39)	0.128** (0.39-0.05)	-0.130 (0.02-⁻0.43)
Mean for Actin+BSP-8[b]	-0.329** (⁻0.06-⁻0.40)	-0.329** (⁻0.06-⁻0.04)	0.055** (0.047-0.087)	-0.039 (0.048-⁻0.06)

Table 4.11. [a] F-statistics for each locus independently, [b] for Actin and BSP-8 combined and [c] for all three loci combined.[d] F-statistics with rats HE 267 and SA 30A removed from the data set. "*" signifies significance at p=0.05, "**" significance at p=0.01. Values in parentheses are 95% confidence intervals estimated by bootstrapping across loci.

Values of F_{ST} and F_{RS} (Table 4.11) reflect the variance components shown in Table 4.10. Bootstrap confidence intervals across each locus provide 95% confidence intervals for the estimates. These intervals are in general agreement with significances estimated by permutation tests (shown in Table 4.11) except for the F_{ST} of all loci, where the confidence interval includes zero. This is attributable to the effect of the negative value of F_{ST} at the Actin locus.

F_{IS} and F_{IT} are the inbreeding coefficients of iL3s within sites and within the total sample respectively, F_{IT} being the sum of F_{IS}, F_{RS} and F_{ST} (eqn 4.17 section 4.2.9). F_{IS} values for Actin and BSP-8 are negative within most sample sites (Table 4.7) with mean values across the British data set of -0.464 and -0.164, respectively. These negative values indicate that more heterozygotes were being found than were expected under random mating predictions. Values of F_{IS} were significantly negative at the Actin locus, reflecting the heterozygote excess within the Berkshire sample site. Removing this sample site from the data set caused F_{IS} to become positive for both

Actin and BSP-8 (+0.0252 and +0.0223 respectively). These values are not significantly different from zero. Therefore, it is the influence of the Berkshire site that causes negative inbreeding coefficients at these two loci.

In comparison, F_{IS} and F_{IT} at the CM-2 locus are uniformly positive, ranging from 0.859 - 1.000 between sample sites. Heterozygote deficits are characteristic of this locus (section 4.3.4, Table 4.7) and have an inbreeding-like effect due to the high coefficients of relatedness of alleles within individual iL3s, resulting in positive values of F_{IS}.

Considering all loci, the mean value of F_{IS} was 0.198, this value being significantly greater than zero. F_{IT} for all loci was 0.085, and not significant. These values were positive, but highly biased by data from the CM-2 locus. Removing rats HE 267 and SA 30A caused the mean values of F_{IS} and F_{IT} to become negative, and significantly less than zero for F_{IS}. This shows that it is the CM-2 genotypes of iL3s from these two rats which bias the data set most.

4.3.7b Variation in allele and genotype frequencies between British and German samples

The significance of variation in allele and genotype frequencies between sample sites in Britain and Germany was assessed by Fisher's exact test and F_{ST}. This analysis was performed by comparing the pooled British data set with the pooled German data set.

Table 4.12 shows that no significant difference was found in allele or genotype frequencies between the British and German samples at any loci, or for all the loci considered together.

Locus	p	F_{ST}
Actin	0.743	-0.005
BSP	1.000	-0.012
CM-2	0.341	0.020
All loci	0.809	0.001

Table 4.12. Genetic variation between pooled samples from Britain and Germany. p is the probability that the observed allele frequencies are the same, as assessed by exact tests. F_{ST} is the variation that is attributable to between-sample site differentiation in genotype frequencies, no values being significantly greater than zero as assessed by permutation tests.

4.3.8 Within sample site variation

Genetic variation between sampled rats was considered separately for each British sample site, and values of F_{RS} calculated (Table 4.13). For the CM-2 locus, all but two of the values of F_{RS} were significantly greater than zero in permutation tests. This demonstrates that high levels of genetic differentiation between individual rats are seen with this locus.

	Locus		
Site	Actin	BSP-8	CM-2
B	0.038**	0.050**	0.528**
D	0.098	fixed	fixed
HE	0.119	-0.041	0.571**
W	0.256**	0.021	0.692
SB	0.602	fixed	fixed
SA	0.323**	fixed	0.740**
N	0.108	0.155	0.131**
OX	0.170	0.003	-

Table 4.13. Within site/between rat genetic variation measured by F_{RS}. "*" signifies values of F_{RS} that are significantly greater than zero at $p < 0.05$ and "**" $p < 0.01$. "-" signifies missing data and "fixed", loci that had only one allele at a particular sample site.

The data from the Actin and BSP-8 loci showed that, while high values for between-rat differentiation were observed, only 4/13 permutation tests were significantly greater then zero. The high, but non-significant, values observed in some sites (*ie.* F_{RS} for Actin is 0.602 within site SB, n = 9 iL3s, 3 rats sampled) were due to the small sample sizes that are apparent when considering the numbers of iL3s from single rats. Unlike the CM-2 locus, there was little evidence for significant genetic subdivision between the iL3s sampled from different rats at these two loci. However, the Berkshire sample site shows significant genetic differentiation at all three loci. This is the most intensively sampled site and most rats were genotyped for at least ten iL3s, showing that there is some genetic structure between infections of different rats. The variance components that were due to between-rat variation within this site were 3.8%, 5% and 53% for Actin, BSP-8 and CM-2 respectively. This illustrates the heterogeneous pattern of genetic diversity seen for alleles of the CM-2 locus.

4.3.9 Isolation by distance

Isolation by distance was tested for by calculating pairwise estimates of F_{ST} between British sample sites for (a) the combined data for all three loci and (b) the combined data for Actin and BSP-8 only. In total, F_{ST} values were calculated for 28 pairwise comparisons between sample sites. Positive correlations were found for both comparisons (a) and (b), with Pearsons correlation coefficients of +0.023 and +0.216 observed respectively, suggesting that isolation by distance may occur. However, these correlations were non-significant, with Mantel's test values for rejecting independence between genetic and geographic distance of (a) p= 0.492 and (b) p=0.227. Therefore, the hypothesis that there is genetic isolation by distance in *S. ratti* is rejected.

4.3.10 Within host genetic diversity

The number of distinct genotypes (measure (i), section 4.2.11) were calculated for each rat within the dataset, including those in Germany, and the results shown in Figure 4.6. For all rats, 76% were infected with two or more genotypes, with a maximum of five genotypes detected in a single rat. The mean number of unique genotypes per rat was 2.34 (Table 4.14). This shows that mixed infections are common for *S. ratti*, and found at a greater frequency than single-genotype infections. The numbers of genotypes per rat was greatest at the Actin locus, with a mean of 1.77 and least at the CM-2 locus with a mean of 1.36. This was expected from estimates of F_{RS}, which showed that more genetic diversity is found between than within rats at locus CM-2.

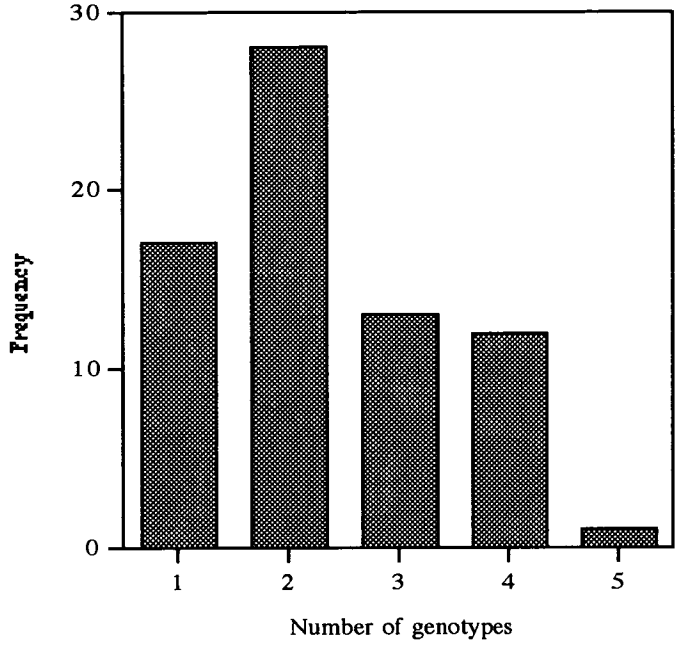


Figure 4.6. Histogram of the number of three-locus genotypes (measure (ii), section 4.2.11) found in samples from British and German rats combined. Mean = 2.34, s.d = 1.07, n = 72.

	Locus			
Sample site	Actin[a]	BSP-8[a]	CM-2[a]	all loci[b]
W	1.42 (0.64)	1.33 (0.47)	1.50 (0.76)	1.73 (0.92)
HE	1.71 (0.69)	1.60 (0.49)	1.75 (0.43)	2.17 (0.69)
D	1.75 (0.43)	1.00 (0.00)	2.00 (0.00)	1.80 (0.40)
SA	1.43 (0.49)	1.00 (0.00)	1.25 (0.43)	1.71 (0.45)
SB	1.33 (0.47)	1.00 (0.00)	1.00 (0.00)	1.33 (0.47)
OX	1.80 (0.75)	1.67 (0.47)	-	1.80 (0.75)
N	3.00 (0.00)	1.33 (0.47)	1.33 (0.47)	3.33 (0.47)
B	1.95 (0.53)	1.56 (0.49)	1.21 (0.41)	3.20 (0.98)
A	2.00 (0.00)	1.00 (0.00)	-	2.00 (0.00)
S	2.00 (0.00)	2.00 (0.00)	1.00 (0.00)	2.00 (0.00)
Total	1.77 (0.67)	1.39 (0.49)	1.36 (0.53)	2.34 (1.07)

Table 4.14. Numbers of different genotypes per rat within each of the sample sites for [a]each locus (measure (i), section 4.2.11) and [b]three-locus genotypes (measure (ii), section 4.2.11). Figures in parentheses are the standard deviation (*s*) of the estimate. "-" signifies no data collected.

4.3.11a Genetic diversity of iL3 samples from Berkshire rats

Within the sample of 24 Berkshire rats, 338 iL3s were analysed. This resulted in 231, 203 and 172 genotypes for the Actin, BSP-8 and CM-2 loci, respectively. Data were complete for all loci for 167 of the iL3s and consequently a three-locus genotype described for each worm (measure (ii), section 4.2.11). The numbers of observed genotypes for each locus and complete three-locus genotypes are shown in Table 4.15, compared to the theoretical maximum numbers of genotypes.

	Locus			
	Actin[a]	BSP-8[a]	CM-2[a]	3-locus genotype[b]
Obs. number of genotypes	3	2	5	14
[a]Max. number of genotypes	3	3	10	90

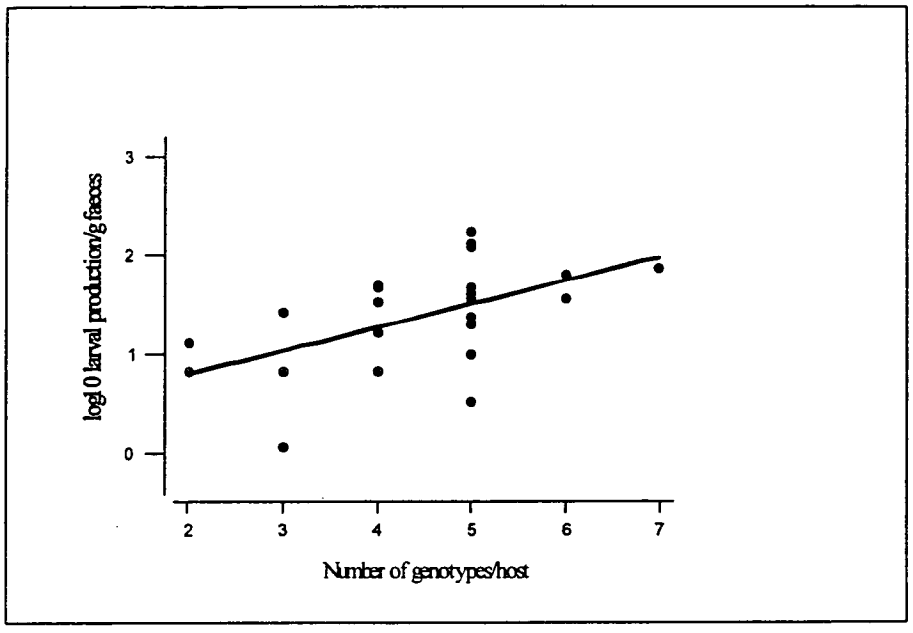
Table 4.15. The number of genotypes scored from iL3s sampled from rats collected at the Berkshire site.[a]= measure (i), section 4.2.11,[b]= measure (ii) section 4.2.11. [c]The maximum number of genotypes is calculated as the maximum possible number of combinations of the alleles observed within Britain (not accounting for the gametic phase of heterozygotes).

Of the 90 possible genotypes possible from combinations of alleles at these loci, 14 were observed from the iL3s sampled in Berkshire. All but one of the hosts (rat B381) harboured multiple parasite genotypes with a maximum diversity of 5 genotypes observed within rat B369. Mean diversity was 3.2 genotypes per host with a standard deviation (s) of 0.98. The maximum repetition of identical genotypes within a single host was ten (rat B381), with only a single genotype recorded for each locus.

4.3.11b Correlations between genetic diversity and intensity of infection

The intensity of infection in Berkshire rats, estimated by faecal larval counts, was overdispersed and distributed according to the negative binomial (section 3.3.5). Regression analysis was performed between the variables 'estimated worm burden' and 'within-host iL3 genetic diversity'. Data on worm burden was normalised by $\log_{10}$ transformation and used as the response variable within an ANOVA. Two measures of within-rat iL3 genetic diversity, D, were used in separate analyses. These were (a) the total number of different genotypes at each locus per rat (measure (i), section 4.2.11) and (b) the total number of different three-locus genotypes per rat

A.



B.

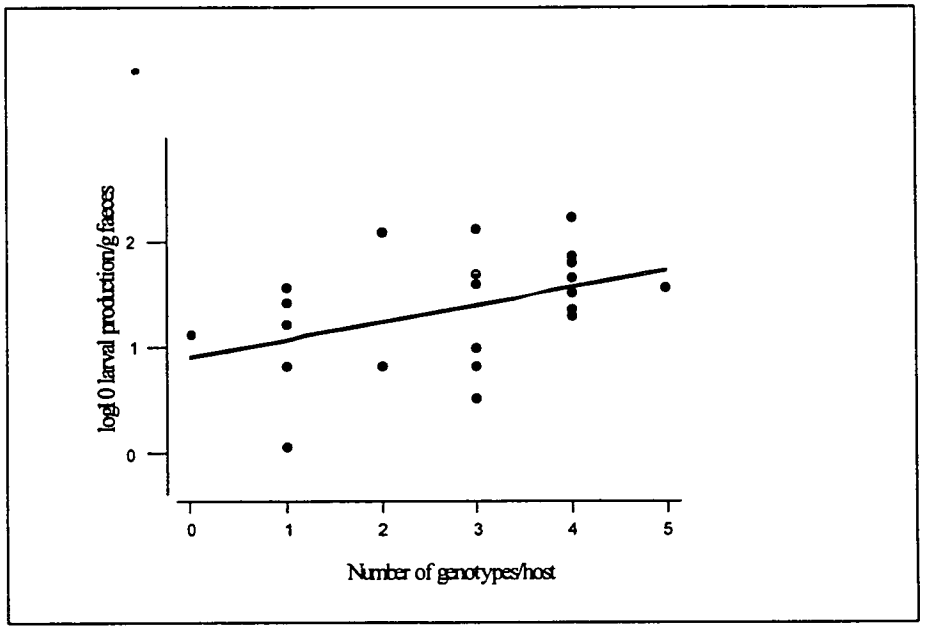


Figure 4.7. The relationship between $\log_{10}$ faecal larval count/g faeces and A. total number of observed different genotypes per host and B. total number of observed three-locus genotypes per host. Each datapoint represents one rat and the fitted line is the correlation between variables.

(measure (ii) section 4.2.11). For analyses (a) and (b), all rats that had less than three genotypes scored from the sampled iL3s were omitted from the analysis. This was done to remove the bias that low sample sizes would have on measurements of the numbers of genotypes per rat. This removed three rats (B350, B363 and B358) from the data set for analysis (b) and no rats for analysis (a).

Significant positive correlations between genotype diversity per rat and $\log_{10}$ worm burden were found for both analyses (a), the number of different genotypes per rat (R = 0.533, ANOVA; $F_{1, 23} = 8.79$, $P = 0.007$) and (b), the number of different three-locus genotypes per rat (R = 0.486, ANOVA; $F_{1, 20} = 5.9$, $P = 0.025$). The regression plots are shown in Figure 4.7. This shows that rats with greater intensities of infection harbor more genetically diverse infections.

4.3.11c The effect of host sex and weight on within-host genetic diversity

The effect of host sex and weight on within-host genetic variation were examined. Six rats were omitted from this analysis due to incomplete sex/weight data. Sex of host was treated as a class effect and host weight as a continuous variable within the ANOVA. Host sex or weight were not significantly correlated with within host genetic diversity for analysis (a) genotype/weight (ANOVA; $F_{1, 17} = 0.94$, $P = 0.349$), genotype/sex (ANOVA; $F_{1, 17} = 0.13$, $P = 0.719$) and analysis (b) genotype/weight (ANOVA; $F_{1, 14} = 0.03$, $P = 0.870$), genotype/sex (ANOVA; $F_{1, 14} = 0.01$, $P = 0.928$). While sample sizes are small, this result suggests that host sex and weight has no effect on within host genetic diversity.

4.3.11d Correlations of genetic diversity (corrected for intensity of infection) and worm burden

Significant negative correlations were observed for analyses of R_a, (R = -0.800, ANOVA; F = 39.32, d.f. = 23, $P = 0.000$) and R_b, (R = -0.851, ANOVA; F = 32.08, d.f. = 22, $P = 0.000$), these correlations remaining significant when the two most extreme values were omitted. This showed that the ratio of genotypes decreases as the intensity of infection increases. For instance, in a rat producing seven iL3s $R_a = 0.42$, this reduces to $R_a = 0.005$ for a rat producing 1040 iL3s. This shows that increases in genetic diversity with intensity of infection is a linear effect; hosts with high intensities of infection do not show proportionally higher levels of genetic diversity. This conclusion is not surprising. Intensities of infection may be enormous, while within-host genetic diversity has a maximum resolution of 90 genotypes within

this system (Table 4.15). Many of these genotypes will be very rare, or even non-existent within these samples and therefore not represented within even the most heavily infected hosts. Consequentially, the distribution of genotypes will not be negatively binomially distributed, hence the strong negative correlation observed.

4.4 Discussion

4.4.1 Bias in the dataset due to sampling multiple progeny from a single parasite

This study has measured the genetic diversity of the infective output of rats, and not the genetic diversity of *S. ratti* infections *per se*. Therefore, a potential bias exists when making inferences on population genetic structure as multiple (genetically identical) progeny may be sampled from individual parasites. This is because progeny may be sampled from rats infected with a single parasitic female, or from an exceptionally highly fecund parasite in a mixed genotype infection. Samples from such infections have the effect that the genotypes of these parasites are over-represented in the data set. Such infections may be the cause of situations where iL3s from rats are (i) fixed for one genotype (such as iL3s sampled from rat W 325) or (ii) biased for high frequencies of normally rare genotypes (such as those from rats SA 30A and HE 267 at the CM-2 locus). The magnitude of this source of error on measurements of genetic differentiation will be large if these types of infections predominate within the dataset. However, several lines of evidence suggest that they do not.

Most helminth-infected animals have low intensity infections (Anderson & May 1992), but in this study it is not known what proportion of the rats sampled contain single parasites. The evidence argues that single, or low intensity, infections do not predominate in the rats. First, estimates of the intensity of infection were made for one sample site (Berkshire, section 3.3.5). The mean estimated worm burden was 37. As this distribution was overdispersed, 63% of the estimated worm burdens were in the class 0-37, but only 4% of these were estimated to be single worm infections. Second, mixed infections were usual with a mean of 2.34 (3.2 in Berkshire) genetically different parasites per rat. 24% of all rats contained genetically uniform infections and thus, potentially, were infected with a single parasite. However, within the Berkshire site, only one rat had a genetically pure infection and was possibly infected with a single parasite. Third, 96.2% of the observed genetic variation of the Actin and BSP-8 loci was found within rats. If single parasite infections were

common, then the between-rat component of variation would be expected to be much greater due to the biases of sampling multiple, genetically identical progeny.

Thus, based on these lines of evidence, it seems unlikely that there was significant bias in the dataset due to multiple sampling of progeny of single parasitic females.

4.4.2 PCR reactions and non-amplifying alleles

DNA templates from iL3s sampled from the faecal cultures of some rats did not amplify at some, or any, of the loci. Why DNA from certain iL3s did not amplify is not known. It was observed that water from faecal cultures could act as a PCR inhibitor. Presumably, iL3s that retained adherent faecal matter through the washing process may have inhibited PCRs by the transfer of an inhibitor. Samples from some rats in which no worms at all were typed suggests that these iL3s came from faecal cultures that may have been especially contaminated.

The efficiency of PCR amplification of the different loci varied. Amplification products of the BSP-8 locus tended to be more faint than those at the other two loci. The CM-2 locus consistently produced the brightest PCR products. This was expected due to the nested nature of this reaction. Often, DNA from iL3s amplified the CM-2 locus but not the other two. As this locus is amplified by nested PCR, the second round of amplification probably overcomes the effects of PCR inhibitors. Therefore, amplification by primers for this locus frequently shows more sensitivity than amplification by primers to the other two.

A repeated observation throughout this study was that DNA from many worms could be amplified at the Actin and BSP-8 loci, but not the CM-2 locus. This observation was striking given the relative sensitivity of the CM-2 PCR. Within any batch of PCR reactions, a certain number will inevitably fail due to undefined causes. Repeating the PCR was a strategy used to increase the success rate. However, this did not achieve a significant increase in success in amplifying CM-2. The possibility of the existence of a null allele at CM-2 was considered as an explanation for the non-amplification of this locus. Such null alleles in PCR reactions may occur due to mutations in primer sites preventing primers from binding or even to deletions of whole DNA sequences (Callen 1993) .

Several observations were made that strengthen this possibility. First, laboratory observations for isofemale line 132 Heterogonic were consistent for the presence of a null allele at CM-2 (section 2.3.3). Here, an experiment was performed using five different combinations of six primers. None of these amplified DNA from isofemale line 132 Heterogonic, although all amplified DNA from isofemale line 5 Heterogonic. This shows that there was a deletion of the CM-2 locus in line 132 Heterogonic rather than a single-nucleotide mutation in a primer site. The second observation was that recoding the data set for all loci on the assumption that PCRs which worked at two, but not three, loci were due to null alleles showed that the CM-2 locus had a much higher frequency of putative null alleles. Although there were some PCR reactions that worked at all loci except Actin or BSP-8, these were infrequent and were most likely due to the normal failure rate of PCR reactions. Moreover, the frequency of the putative null allele at locus CM-2 was consistently high within all sites except one, Norfolk (N). This observation was in agreement with conclusions, drawn from F_{ST}, that there was little variation in allele frequencies between sample sites. Therefore, the putative null allele appears to be conforming to the same geographic distribution observed for the other loci as well as for the other alleles at the CM-2 locus. The third observation was that the CM-2 locus showed strong heterozygote deficits in four out of eight sample sites. This was contrary to the pattern of the other two loci for which genotypes were generally found in, or in excess of, Hardy-Weinberg proportions. The occurrence of heterozygote deficits at all loci is characteristic of inbreeding (Hartl & Clarke 1989), but the occurrence of heterozygote deficits at a single locus is characteristic of null alleles (Pemberton *et al.* 1995).

Heterozygote deficits at a single locus may result from processes other than null alleles. One of the aims of this study was to determine the effect of parthenogenetic reproduction on the genetic structure of *S. ratti*. Sample sites that have only low, or no, sexual reproduction may show strong non-random distributions of genotypes, and these many be manifested as heterozygote deficits at some loci (Hughes 1989). It is therefore possible that the observed heterozygote deficit at this locus was due to the high frequency of parthenogenetic reproduction in *S. ratti* at all sample sites, rather than a null allele. This issue will be examined further (section 4.4.3).

4.4.3 Evidence for genetic recombination in *S. ratti*

This section examines the data from the molecular markers to test for the clonal population structure expected if *S. ratti* is wholly parthenogenetic (non-recombining)

in nature. Data will be used on (i) heterozygosity, (ii) Hardy-Weinberg equilibria and (iii) correlations between genotypes at different loci (genotypic linkage disequilibria).

(i) Heterozygosity

High and low levels of heterozygosity may occur as a consequence of apomictic parthenogenesis as has been shown in many studies (Hughes 1989). The mean heterozygosity of *S. ratti* across all loci and populations was 0.247. Previous studies of genetic diversity in obligately sexual species of parasitic nematodes found mean heterozygosities of 0.1 with a range of 0 - 0.21 in 23 species (Nadler 1990). Heterozygosity of *S. ratti* was therefore higher than that recorded from any other parasitic nematode.

Estimates of heterozygosity for *S. ratti* were heavily dependent on the data from the Actin locus, where heterozygosity was uniformly high within sample sites with a mean of 0.609. Although there was significant variation in heterozygosity at this locus between sample sites, all sites showed values higher than the range found for other nematodes. Mean heterozygosity at the BSP-8 locus was within the range found in other helminths (0.115), although two sites (HE and OX) had heterozygosities higher than the range normally seen. That Actin has a higher heterozygosity than the other loci is expected given the balanced nature of allele frequencies at this locus (frequencies of Actin allele 1 and 2 were 0.626 and 0.374, respectively in the total data set).

In contrast, values of heterozygosity for the CM-2 locus are uniformly low with a mean of 0.017. Six out of eight sample sites had no heterozygous iL3s whatsoever. It is possible that the aberrant characteristics of this locus relative to the other two are a consequence of the presence of a null allele at high frequencies within the samples. However, it is equally possible that low heterozygosities at this locus are a consequence of a parthenogenetic mode of reproduction. Agatsuma and Habe showed that fixed heterozygosity at some loci, and fixed homozygosity at others, occurs in certain forms of the trematode *Paragonimus westermani* (Agatsuma & Habe 1985b). In this species, hybridisation appears to have created a triploid form that reproduces exclusively by apomictic (non-meiotic) parthenogenesis. Similarly, in the bivalve mollusc *Lasaearubra*, fixed heterozygosities occur at 12 isoenzyme loci (Crisp & Standen 1988). Both of these species are obligate parthenogens and hence show non-recombining population structures. The low frequency of sexual reproduction observed in these samples of *S. ratti* is more comparable to that of cyclical

parthenogens where clonal reproduction is punctuated by bouts of mictic reproduction. Populations of cyclical parthenogens, like obligate parthenogens, are often characterised by non-equilibrium levels of heterozygosity (Hughes 1989), but show increased clonal diversity as a consequence of occasional segregation due to sexual reproduction. The effect of segregation on genotype frequencies is ably demonstrated by studies of cyclical and obligately parthenogenetic populations of *Daphnia magna*. Populations showing frequent sexual reproduction have many clones with genotypes found close to Hardy-Weinberg expectations (Hebert 1974a). However, in populations were sexual reproduction is rare (such as those in permanent ponds (Hebert 1974b-a) or Arctic populations (Weider & Hobaek 1994)), genotype frequencies fluctuate widely from random-mating proportions and stable clones are found that are frequently associated with heterozygous excesses.

It is probable that high, and low, levels of heterozygosity in *S. ratti* are a consequence of parthenogenetic reproduction amplifying specific genotypes, without the mitigating effect of sexual reproduction. The proportion of larvae developing into sexual forms (*i.e.* undergoing heterogonic development) is low within lines derived from isolates of the sampled *S. ratti* (section 3.3.2). Therefore, the majority of development is by the homogonic route, which is exclusively parthenogenetic. If parthenogenesis is maintaining high and low levels of heterozygosity, then deviations from Hardy-Weinberg equilibrium expectations and linkage disequilibria between loci are expected within sample sites (Hughes 1989).

(ii) Hardy-Weinberg equilibria

Analysis of Hardy-Weinberg proportions within sample sites found significant deviations for (i) the Actin locus where a heterozygous excess was observed within a single site (sample site B) and (ii) the CM-2 locus where heterozygous deficits were observed within four of eight sites. Departures from random mating were reflected by the inbreeding coefficients F_{IS} and F_{IT}, positive values of F_{IS} indicating inbreeding populations. Total F_{IS} values at the Actin and BSP-8 loci for all sample sites were predominately negative with mean values of F_{IS} of -0.299 and -0.061 respectively. These were significantly different from zero at the Actin, but not BSP-8 locus, and therefore reflect higher levels of heterozygosity at this locus than are expected for a randomly mating population. Samples from one site, Berkshire, unduly bias the data set. Removing this sample site from the data set causes (a) genotypic proportions for the pooled data set at the Actin locus to conform to Hardy-Weinberg expectations and (b) values of F_{IS} at Actin and BSP-8 to become not significantly different from zero.

These data show that, at these two loci, the majority of sample sites contain randomly-breeding parasite populations. Therefore, there is little evidence at these two loci for the existence of a clonal population structure.

In contrast, F_{IS} values were uniformly positive at locus CM-2 with a range of 0.823 - 1.000 reflecting the highly 'inbred' nature of this locus resulting from an overall heterozygote deficit. This pattern may reflect a lack of segregation due to parthenogenesis, but it could also indicate of high rates of selfing. If so inbreeding may be a mechanism causing low levels of heterozygosity at this locus. A study by Viard *et. al.* (Viard *et al.* 1996) on populations of the freshwater snail ***Bulinus truncatus*** found high levels of genetic diversity at four microsatellite loci, but large heterozygote deficits at all loci. This snail, while hermaphrodite, also has a self fertilising aphallic morph. In the populations sampled, the ratio of aphallic individuals to hermaphrodites was high. Therefore, the observed heterozygote deficits (and positive values of F_{IS}) were consistent with mainly selfing populations of snails. Here, two results argued that low levels of heterozygosity at the CM-2 locus were not a consequence of inbreeding. Firstly, if inbreeding was occurring, then data from the three loci would be expected to be relatively consistent. This was not so, the genotypes of Actin and BSP-8 showed that high rates of selfing were not occurring due to normal-to-high frequencies of heterozygotes at these loci. Secondly, mixed infections predominated within infections of *S. ratti* arguing that, unless preferential mating with siblings was occurring, inbreeding was unlikely. Therefore, observations of low heterozygosities at the CM-2 locus are consistent with two hypotheses. The first is that genotype frequencies reflect low levels of recombination due to parthenogenesis. The second is that there is a null allele at this locus causing strong heterozygote deficits.

(iii) Genotypic linkage disequilibria

When considering each sample site separately, significant associations between genotypes were not observed. However, this cannot be used as absolute proof that *S. ratti* is undergoing recombination in nature. This is because the power of tests to identify linkage disequilibria are low with these numbers of loci, often requiring larger sample sizes than are found in these samples (Thompson *et al.* 1988). Considering the entire data set together (excepting sites in Germany) increases the sample size. Under these conditions, significant disequilibria are found for two out of the three locus combinations, Actin+BSP-8 and Actin+CM-2. That these samples were drawn from geographically separated samples, however, may cause an illusion of linkage disequilibria due to population subdivision (Hartl & Clarke 1989). Therefore, it is not

possible to conclude that *S. ratti* shows a recombining population structure based on these tests.

When (i), (ii) and (iii) are considered together, it is clear that these samples are from recombining populations of *S. ratti*. However, there is strong evidence that there is some deviation from panmixia and that not all samples are from sites that are at a genetic equilibria. This is mainly illustrated by the Actin locus, where convincing heterozygote excesses occur within the Berkshire population. Additionally, heterozygote deficits at CM-2 may be due to low levels of sexual reproduction. But, there is good evidence that there is a null allele at this locus and that the observed deficits result from an artifact of PCR rather than reflecting a biological phenomena.

4.4.4 Heterozygous excess at the Actin locus in Berkshire

There are two explanations why there is an excess of heterozygotes seen for the Actin locus in the Berkshire site.

(1) This study assumes that the molecular markers used are selectively neutral and that their distribution is only influenced by stochastic processes, such genetic drift, founder effects and gene flow. However, if loci are under selection then they will not reflect such processes. Heterozygote excesses may occur as a consequence of several processes, such as preferential mating between unlike genotypes (negative assortative mating) or heterozygote advantage. Heterozygote advantage, exemplified by overdominance of the *S* allele of ß-globin in West African humans (Willcox *et al.* 1983), has been argued to maintain polymorphism at the Major Histocompatibility Complex (MHC) locus in both human (Markow *et al.* 1993) and Ovine species (Paterson *et al.* 1997). Furthermore, hybrid vigor (heterosis) has been argued as a mechanism maintaining heterozygote excesses in species of parthenogens such as *Daphnia pulex* (Hebert *et al.* 1982). It is possible that the heterozygote excess seen in the Berkshire site is due to a fitness advantage for this genotype within this site. While I am unable to rule out this possibility, it appears unlikely. If heterozygosity at this locus was associated with increased fitness, then the effect may be expected to be more wide-spread. However, only a single site shows this effect.

(2) It is probable that the heterozygous excess at this locus is due to amplification of this particular genotype by parthenogenetic reproduction. That most sample sites contain genotypes, at Actin and BSP-8, in Hardy-Weinberg proportions shows that genetic recombination is occuring at a rate that is able to randomise

genotypes within most sites. However, that there is a heterozygous excess within this site may be interpreted as evidence that amplification of genotypes by parthenogenesis is able to bias genotypic frequencies. This effect is probably transient as bias in genotype frequencies due to parthenogenesis will be expected to be randomised eventually by sexual reproduction. Such a balance is often seen in populations of *D. pulex*. Here, the genotypes of those populations undergoing regular bouts of sexual reproduction remain close to Hardy-Weinberg equilibrium proportions while those exhibiting infrequent sex deviate widely from expected frequencies (Hebert 1974b; Hebert 1974a).

If the Berkshire *S. ratti* population originated from a small sample of parasites that were principally, by chance, heterozygous for the Actin locus then amplification of this genotype would occur by the homogonic cycle. Then, if rat to rat transmission occured, the progeny from the original infection may be expected to spread the 'founder genotype' to all members of the rat group. Thus, deviations from Hardy-Weinberg equilibrium would persist for several generations until sexual reproduction eventually restored genotypes to equilibrium frequencies. It may be argued that if non-random associations between alleles at this locus are due to limited sexual recombination, then multilocus linkage disequilibrium would also be found within the sample of iL3s. This would be seen as a predominating 'clone' within the sample. However, such a clone was not observed. But, given the sample sizes and that some intragenic recombination will have occured, the original founding multilocus genotype would have been expected to have formed new combinations. This will obscure the statistical detection of multilocus linkage disequilibria. Thus, the population structure seen is intermediate between that of a fully recombining and a clonal population.

It is impossible to show whether selection or recombination is causing the heterozygote excess by using these data. However, given that it is known that sexual reproduction is infrequent in these isolates of *S. ratti* it appears most likely that the effect is due to parthenogenesis and not selection. But, it would be expedient to test the fitness of iL3s that are heterozygous for Actin in this site against other genotypes. Given that these isolates are stored in liquid nitrogen, it would be possible to revive the isolates and test isofemale lines of each genotype for variation, and thus 'fitness', for characters such as fecundity, longevity and infectivity.

4.4.5 Significance testing of genetic variation

The *S. ratti* sampled in this study were not mating randomly in all sites, and so the use of exact tests may not be valid (Raymond & Rousset 1995). Although it was not strictly necessary to use exact tests, they were used principally to act as a control for the estimates of differentiation derived from F-statistics. This was because Raymond *et al.* (Raymond & Rousset 1995) showed that significance testing of variation components by the bootstrap method suggested by Weir (Weir 1996) was inaccurate for small numbers of loci. However, in the majority of cases, the significance of F-statistics and exact tests were the same. In the sites that show non-concordance in results for F-statistics and exact tests, more biological relevance is attributed to the result from the F-statistic due to the non-assumption of Hardy-Weinberg proportions used in these calculations.

In addition to bootstrapping across loci, this study used a permutational method to estimate significances of F-statistics for individual loci. A comparison of bootstrap confidence intervals and significances of permutation tests show a general agreement. However, there are three exceptions. Bootstraps include zero for two F_{IS} values and one F_{ST} value, whereas permutation tests show that the observed distributions are significantly greater than zero (F_{IS} for Actin + BSP-8; F_{IS} for all loci; F_{ST} for all loci). These contradictory significance values are due to bias by the locus with the most extreme value creating wide bootstrap confidence intervals, which are not reflected by the null (permutational) distribution of the data. Therefore, this study agrees with the general conclusion by Raymond (Raymond & Rousset 1995) and Weir (Weir 1996) that minimum samples from five or more loci are necessary to produce informative bootstrap confidence intervals.

4.4.6 Spatial genetic differentiation of *S. ratti*

Mean F-statistics for all loci showed that a minor, but significant, proportion of the total genetic variation (1.4%) was attributable to differentiation between sample sites. F-statistics for two loci, Actin and BSP-8, described a similar pattern. The component of variation due to between-site differentiation for the two loci considered together was not significantly greater than zero, although F_{ST} for BSP-8 alone was significant. At the CM-2 locus there was greater, and significant, genetic differentiation between sites, accounting for 12.4% of the observed variation.

The reason that these results are not completely consistent with one-another may be due to the presence of null alleles at locus CM-2. The presence of a null allele

at a high frequency at this locus will cause variation in genotypic frequencies between sample sites and rats within sample sites. This is because the null allele would affect the calculations of F-statistics for the following reason: If an iL3 was heterozygous for the null allele, then from the genotyping system that I have used, the heterozygote would be scored as homozygous for the visible allele, and two copies of the allele recorded in the data set instead of one (the real value). Thus, there is a possibility that rare alleles will have an artificially high frequency. Therefore, small variations in the frequency of the null allele would be expected to have a disproportionate effect on the frequency of rare alleles. This is a potential source of error when F values are subsequently estimated from these data.

However, the general pattern described by CM-2 is similar to that at the other two loci. Allele 1 is consistently the most frequently observed in 7/9 sample sites, and alleles 2 and 3 are consistently rare or absent within all sites. Inspection of the data shows that allele 4 causes the majority of sample site differentiation because of its high frequency within the HE and SA sites. Furthermore, it is the high frequencies of this allele from the samples of just two rats, SA 30A and HE 267, that bias the genotype frequencies in these two sites. Removal of these two rats from the data set causes the mean value of F_{ST} at this locus to become not significantly greater than zero, and the estimate for between-rat variation to reduce to 12.8%. This implies that there is a homogeneous geographic distribution of alleles at this locus, and that CM-2 is conforming to the same pattern observed for the other two loci.

Recalculating F_{ST} for all loci with the omission of rats SA 30A and HE 267 shows that no genetic variation is attributable to between-site genetic differentiation in genotype frequencies. This is evidence that there is no genetic subdivision between sample sites. A corollary of this is that significant gene flow between sites is occurring. If gene flow between these sample sites is not limited, then an isolation-by-distance model predicts that no correlation between geographic and genetic distance will occur (Crow & Kimura 1970). This was tested by comparing pairwise F_{ST} values and geographic distance. While positive associations were observed, these were non-significant. It was thus concluded that isolation by distance was not occurring. This result lends weight to the emerging conclusion that British *S. ratti* is geographically genetically unstructured. Furthermore, there was no evidence for genetic subdivision between British and German samples of the parasite. While these samples are small and more data is required, it appears that *S. ratti* may show no genetic subdivision within a Europe-wide scale.

However, it cannot be concluded from these data that *S. ratti* has a panmictic population structure (*i.e. S. ratti* within Britain exists as a single interbreeding population). This is for two reasons. The first is that, despite low levels of genetic differentiation between sites, the between-site component of variation was significant when all loci were considered together (although the significance of this value is dependent on samples from two rats, SA 30A and HE 267). The second is that significant deviations of genotype frequencies from random mating proportions are observed in several sites. These are principally due to heterozygote deficits at the CM-2 locus and a heterozygote excess at the Actin locus within the Berkshire sample site. These deviations are reflected in the ANOVA of heterozygosities for individual loci across the data set, which show significant variation between sites. This argues that the parasites within the Berkshire sample site are to an extent isolated from those in other sites. This may be taken to show that transmission is not occurring at sufficiently high rate to homogenise genotypes between sites, and that there is a degree of genetic subdivision between the sites. The rates of transmission of *S. ratti* between sites is considered in the final chapter of this thesis.

4.4.7 Within-site genetic variation

Mixed genotype infections were found to be characteristic of natural *S. ratti* infections, with 77% of sampled rats harbouring infections of two or more genotypes. 14 different multilocus genotypes were found in the whole data set, of which a maximum of five were found within a single rat. Mixed infections have been recorded in other species of helminths. Surveys of *Ascaris suum* in Guatemala found that most hosts contained multiple genotype infections, with a maximum of five detectable in a single host (Anderson *et al.* 1995). Rollinson *et al.* (Rollinson *et al.* 1986) observed a maximum of five different genotypes/rat in studies of *Schistosoma mansoni* using enzyme electrophoresis as a method of discriminating parasites. A subsequent study on the same foci of infection, but using RAPD's, increased this estimate to a maximum of 28 genotypes/rat (Barral *et al.* 1996) showing that use of multiple loci can significantly increase estimates of within host genetic diversity.

Variation between rats within sites was low at both the Actin and BSP-8 loci, with most sites showing non-significant values for permutation tests of F_{RS} and exact tests. 3/8 and 1/8 sites found significant values of F_{RS} for Actin and BSP-8 respectively, and 3/8 and 2/8 found significant variation by exact tests. In all sites except one (Berkshire, Actin), significant exact tests corresponded to significant values

of F_{RS}. Variation between rats was higher at the CM-2 locus with 5/5 and 4/5 sites showing significant differentiation by both exact tests and F_{RS} respectively.

The effect of unusual single rat infections on between-site differentiation has previously been demonstrated; removing two rats from the data set significantly reduced between-site variation for locus CM-2. To determine whether single rats were causing significant within-site differentiation, rats with the most extreme genotype frequencies were removed and exact tests recalculated. This had the effect of causing 3/10 significant values to become non-significant. Therefore, the majority of sites showing significant variation between rats are doing so because of the effect of more than one rat. Most of the within site variation is attributable to CM-2 as heterogeneity in allele and genotype frequencies between rats is greatest at this locus. Why this locus shows more heterogeneity than the other two is not known, although the argument that null alleles are causing errors in the number of rare alleles scored applies equally to the within-site situation as it does the between-site.

If the effect of this locus is discounted, it may be concluded that between rat variation is small. This is quantified by the mean value of F_{RS} for Actin and BSP-8 which shows that 5.5% of the total genetic variation is attributable to between rat variation. While this value is significantly greater than zero, it shows that the genotype distributions of infections of different rats within the same sample site are not very different from each other.

Analysis of the distribution of genotypes within the Berkshire population of rats found significant correlations between faecal egg counts and genetic diversity. Rat sex or weight did not correlate with the genetic diversity of an infection. Thus, genetic diversity within hosts is correlated with the intensity of infection. This shows that rats with high faecal larval counts have greater worm burdens than those with lower faecal larval counts and validates the use of faecal larval counts as a measure for estimating worm burden. It also shows that high faecal larval counts do not result from an infection of a few uncharacteristically fecund parasites. Barral *et al.* (Barral *et al.* 1996) demonstrated a significant aggregation of diversity in *S. mansoni* genotypes within rats (variance to mean ratio = 73.6) and showed that this correlated with intensity of infection. Here, the distribution of genotypes between rats in this site, while correlated with intensity of infection, was not aggregated (variance to mean ratio = 0.48). However, this may be an artifact of the number of genotypes that are resolved by the use of a limited number of loci. It may be that if a larger number, or more polymorphic

loci were used, then an aggregated pattern of genetic diversity would be found mirroring the distribution of intensity of infection.

There are several implications of these data. The first is that, for the majority of rats, there is no genetic subdivision between the *S. ratti* of one rat and another. This shows that the infection rate within sites is sufficient to homogenise genetic diversity between rats. Secondly, as genetically diverse infections are normal, genetic subdivision is not a mechanism that will be expected to cause inbreeding *S. ratti* within the faeces of individual rats. Thirdly, as rats having large infections contain genetically more diverse worm populations than hosts with light infections, these rats will be expected to provide the main opportunity for outbreeding. The greater number of diverse genotypes within the faeces of heavily infected hosts will lead to an increased probability of crossing between unlike genotypes. Thus, the infections of overdispersed rats may be expected to give rise to most of the genetic variation within the local parasite population .

4.5 Summary

The genetic diversity of three nuclear loci was examined in samples of *S. ratti* iL3s from 72 rats from ten sample sites. Most genetic variation was found within individual rats with the majority of rats harbouring genetically mixed infections. Little genetic differentiation was found between rats within sites showing that rates of transmission within sites is sufficient to homogenise allele frequencies.

Positive correlations were found between intensity of infection and the numbers of iL3s produced by a rat overnight. This suggests that high faecal larval counts most probably result from a large population of parasites, rather than a few highly fecund worms. Moreover, infections in heavily infected rats may provide an opportunity for greater outbreeding in rat faeces.

Observations of Hardy-Weinberg equilibria and low levels of genotypic linkage disequilibria are consistent with *S. ratti* undergoing genetic recombination in nature. Deviations of genotype frequencies from random mating expectations in some sites are evidence that sexual recombination is limited and that the homogonic phase of the life cycle may bias the frequencies of some genotypes within sample sites. The local genetic structure of *S. ratti* therefore appears to be intermediate between fully recombining and clonal.

Small amounts of genetic differentiation were observed between sample sites, and no isolation by distance was observed. This shows that there is little population genetic structure. Differentiation between samples from Britain and Germany was not significant suggesting that there may be no population structure at this scale.

Chapter five

Conclusions and final comment

5.1 The population genetic structure of *S. ratti*

Significant differences in the frequencies of neutral alleles in samples from geographically distinct locations show the existence of a subdivided gene pool. Such non-random distributions of alleles demonstrate whether or not distinct, inbreeding populations of parasites exist. This study systematically and intensively sampled *S. ratti* within Britain. A genetic analysis of these samples using molecular markers provide several lines of evidence showing that *S. ratti* does not have a subdivided population structure, but rather shows unrestricted gene flow on the scales examined here.

F-statistics showed that there was little genetic variation between samples of British and German *S. ratti.* Measurements found that, for the combined estimates of F_{ST} from all three loci, only 1.4% of the total genetic variation was attributable to between-site variation. The majority of this originated from samples from two rats which, when removed, resulted in a measure of F_{ST} that was not significantly greater than zero.

At two out of three loci there was no evidence of heterozygote deficits within sites compared to heterozygosities expected from the allele frequencies of the total dataset. Therefore, reduction in heterozygosity due to population subdivision (the Wahlund effect) was not in evidence. One locus (CM-2) did show significant heterozygote deficits in many sites, however these were due to null alleles and not population subdivision. The observed genotypes were in Hardy-Weinberg proportions in most sample sites thereby suggesting that the individual parasites from separate sample sites are close to a state of panmixis (*i.e.* each parasite has an equal chance of breeding with another independent of where they are in Britain). In the Berkshire site there was a strong excess of heterozygote genotypes at the Actin locus demonstrating that there was not a completely uniform distribution of genotypes between sites. It is probable that this heterozygote excess resulted from a founder effect that has been

amplified by parthenogenetic reproduction within this sample of parasites, rather than natural selection of the heterozygote genotype.

It is concluded that the basic reproductive unit of *S. ratti* is larger than the scale sampled here and that the parasite is a single interbreeding population within Britain. Panmixia may extend to include the parasites in Europe. These data suggest a large effective population size for *S. ratti* resulting from unrestricted dispersal of the parasite. These results agree well with the findings from studies on other species of nematode parasites. Here, essentially no difference was found for the distributions of mtDNA haplotypes between widely-separated sites for *Haemonchus contortus* (1.0% of variation due to population subdivision), *Trichostrongylus circumcincta* (2.0%) and *Haemonchus placei* (4.0%) (Blouin *et al.* 1995). In comparison, samples of *Mazamastrongylus odocoileus* from white tailed deer showed that a significant proportion of the observed genetic variation (31%) was found between sites (Blouin *et al.* 1995). It was argued that the restricted range of deer was reducing gene flow between parasite populations to the extent that random genetic drift, and hence differentiation, occured. That an isolation by distance effect was observed for populations of *M. odocoileus* was evidence for this hypothesis. A similar effect was observed for the trematode *Fascioloides magna,* also infecting white tailed deer, where samples from separate states showed substantial genetic differentiation at isoenzyme loci and an isolation by distance effect (Mulvey *et al.* 1991a). Thus, it appears that the observed population genetic structure of *S. ratti* conforms to a model where unrestricted host movement allows widespread dispersion of the parasite and a state of panmixia for the parasite population.

Of the above studies, all except that by Mulvey *et al.* (Mulvey *et al.* 1991a), used measurements of mtDNA diversity. As described in Chapter 2, mitochondrial markers are more sensitive than nuclear loci for detecting population structure as their rate of gene flow is effectively a quarter that of nuclear loci. Thus, the data from studies using the mtDNA markers are not directly comparable to those utilising nuclear loci. It is therefore not known whether repeating this study of *S. ratti* by using mtDNA markers (if any could be found), would find more population structure than was revealed here. A recent study of the geographic population structure of the mosquito *Anopheles gambiae* using both microsatellite and mtDNA markers found no genetic

subdivision for either class of marker (Lehmann *et al.* 1997). Thus, in this species, the pattern of genetic variation was represented by the whole genome. It is possible that further characterising the mtDNA of *S. ratti* would not alter the conclusions obtained *via* the use of nuclear markers. Moreover, as studies on the population genetic structure of parasites could be used to understand the spread of important traits such as drug resistance, which are likely to be nuclear genes, characterising a population by variation at nuclear loci describes a more biologically relevant picture.

If rates of gene flow between sample sites in Britain and Germany are high enough to equilibrate allele frequencies, then the effective population size of *S. ratti* will be large. Moreover, from the high observed prevalences of infection of *S. ratti* within Britain, it would appear that the actual population size of *S. ratti* is large. Assuming that the average size of a rat colony is 20 individuals (although this value may vary by an order of magnitude) (Barnett 1952) and approximately 1.5 million properties contain rat infestations (5% of premises in Britain) (Institution of Environmental Health Officers 1993), then the census of rats in Britain is $\approx 3.0 \times 10^7$. The observed prevalence of infection in this study was 68% and the mean number of genotypically different *S. ratti* per rat, 2.3. Therefore, multiplying these values provides a rough estimate of the actual population size of British *S. ratti* of $\approx 4.7 \times 10^7$ parasites. The effective population size of *O. ostertagi* in the USA was estimated at between $4 - 8 \times 10^6$ parasites, using data from the observed divergence of mtDNA sequences (Blouin *et al.* 1992). The arguments of Beech *et al.* (Beech *et al.* 1994) and Blouin *et al.* (Blouin *et al.* 1992) predict that nematode parasites with large effective population sizes will have high genetic diversities. However, the nucleotide diversity of *S. ratti* was similar to that observed in free living organisms, and less than that found in other species of parasitic nematode at both nuclear and mitochondrial loci. Therefore, on this basis it appears that, despite the lack of population genetic structure and high estimated population size of *S. ratti*, the effective population size is lower than that of other parasitic nematodes.

5.2 The effect of historical processes on the population genetic structure of *S. ratti*

The amount and distribution of genetic variation in a species is determined by many factors, both contemporary and historical. Such factors are the rates of mutation

and recombination, recent and ancient effective population size, random genetic drift, population subdivision and natural selection (Tajima 1992). Why the effective population size of *S. ratti* appears to be low relative to other species of parasitic nematodes will depend on these factors.

Historical demography of *S. ratti*

Low nucleotide diversities of *S. ratti* and the homogeneous distribution of allele frequencies may be due to the historical demography of the parasite and its host. The long term effective population size of a species is the harmonic mean of its generational effective population size (Hartl & Clarke 1989). This means that population bottlenecks have a disproportionate effect on levels of polymorphism. For instance, the extremely low levels of genetic variation found in northern elephant seals are presumably due to the recent near-extinction that reduced the species to a few individuals (Bonnell & Selander 1974). The present worldwide distribution of *Rattus norvegicus* originated from a focus near the Caspian sea, from where it underwent a large range expansion through Europe in the 1600s (Kroyer 1991). A measure of the range expansion of *R. norvegicus* is perhaps provided by the reduction in plague outbreaks as *R. norvegicus* aggressively outcompeted the principle plague carrier, *R. rattus*, across Europe (Kroyer 1991). *R. norvegicus* was first recorded in Britain in the 1720s (Smith 1995). This is fairly recent and may correspond to the first introduction of *S. ratti* into Britain. However, *R. rattus* also carries the parasite (Udonsi 1989; Hasegawa *et al.* 1994) and the discovery of *R. rattus* bones in second-century A.D. remains in York show that the Black rat predates the Brown rat in Britain by several centuries (Smith 1995). Whether this population of *R. rattus* were infected with *S. ratti*, or if it was, whether the infection was transmitted into the *R. norvegicus* population, is not known.

That the present day distribution of *R. norvegicus* appears to mirror the recent expansion of modern man (Barnett 1952; Kroyer 1991) may (i) explain the lack of genetic differentiation between *S. ratti* samples and (ii) the low levels of genetic diversity. If the colonisation of Britain by *R. norvegicus* occured within the last 300-900 years, then that the parasite gene pool within Britain appears to be similar to that in Europe is not surprising given that substantial differentiation due to genetic drift is unlikely to have occured within this time. Even if the sampled parasite populations

were not of recent origin, then the natural movement of rats between colonies may have caused levels of gene flow that were sufficient to homogenise allele frequencies and to prevent differentiation by drift as shown by other studies (Blouin *et al.* 1995; Mulvey *et al.* 1991b). However, that genetic structure exists at some level in *S. ratti* is shown by examining the distribution of polymorphisms of nuclear DNA in globally distributed isolates of the parasite. Here, polymorphisms in two loci (24 and 29) were found only in the USA. This means that, if the population genetic structure of *S. ratti* was examined further, it is likely that population differences would be observed for the global scale.

Although little is known about the short-term dispersal of *R. norvegicus*, the animal does not appear to be characterised by a naturally high dispersal rate. Brown rats are territorial and live in groups with a strong degree of social structure (Barnett 1952). Brown rats also show marked neophobia, and will therefore re-use well known pathways and food sources (Webster & MacDonald 1995). This group structure may limit the dispersal of individual rats and result in a degree of intra-group isolation. Mark/recapture experiments have shown that colony home ranges tend to be small, with 80% of individuals recaptured within 20 - 25 meters of the original capture site (Barnett 1952; Glass *et al.* 1989). However, there is some indication that juvenile males may be ostracised by the colony alpha male, and disperse (Barnett 1952). These individuals may therefore act as the primary agents of gene flow for *S. ratti.* That longer term dispersal of *R. norvegicus* occurs is inarguable. Brown rats are endemic to every corner of the globe due to colonisation in association with humans. Therefore, it appears that while normal migration between rat groups is low, infrequent longer range movements of individual rats may occur. This is in agreement with population genetic theory that shows only low levels of migration are necessary to prevent genetic differentiation within a species (Slatkin 1987). Therefore, it may be that low levels of nucleotide diversity and a panmictic population structure in *S. ratti* reflect a relatively recent origin of the sampled populations.

5.3 Microepidemiology of *S. ratti*

This study has shown that there is little heterogeneity between the composition of *S. ratti* infrapopulations within sample sites. A high proportion of rats (76%) had mixed-genotype infections. Mean values of F_{RS} between rats within sites show that

6% (Actin) and 3% (BSP-8) of the total genetic variation was a consequence of between rat variation in genotype frequencies. Thus, at these loci, infrapopulations of *S. ratti* are genetically highly diverse with the majority of the total diversity within rats. The low level of genetic structure observed is probably due to infections of rats with large worm burdens being more genetically diverse (section 4.3.11b). The lack of genetic substructuring at the level of the individual host has several implications when considering the epidemiology and genetical interactions between rat and parasite.

A consequence of overdispersion of parasites amongst hosts is that clumping of parasite infective stages in the environment occurs. This leads to non-random transmission of parasites as hosts acquire infections from such 'patches'. If transmission is low, such that successive parasite generations are essentially non-overlapping, then inbreeding would be expected to occur. This study has shown that there is little evidence that *S. ratti* is inbreeding. Thus, it appears that rats are transmitting their infections to other rats at rates that are high enough to homogenise parasite infrapopulations and to prevent genetic differentiation. It has been suggested that in human populations, the household is the focus of transmission. This is based on the observation that individuals with heavy worm burdens tend to be from the same family (Forrester *et al.* 1988). Studies by Anderson *et al.* (Anderson *et al.* 1995) provide some support for this hypothesis. Here, it was shown that there was familial clustering of *Ascaris suum* haplotypes in some, but not all families. Moreover, it was demonstrated that the infrapopulations of *A. suum* were strongly structured at the level of individual hosts in both humans and pigs. This showed that co-transmission of related *A. suum* was occuring. Co-transmission of genotypically similar *S. ratti* was seen within the colony of rats sampled from within Berkshire, but there was little evidence for between host genetic structure. This may reflect differences in rates of transmission between *A. suum* and *S. ratti*. For instance, studies on the epidemiology of wildlife diseases show that the contact structure of infections in natural populations is usually spatially localised (Hudson & Dobson 1995). It is then expected that social species living in a colony (such as rats) will share common infections due to high rates of individual contact and use of communal resources. That Brown rats occupy well defined territories and defecate in communal latrines is probably a factor that contributes to high rates of rat to rat transmission of parasites (Hurst, pers. comm.). This will reduce the probability that sibling parasites will co-occur within rats.

The patterns of prevalence of infection are well known for natural infections of parasitic nematodes (Bundy 1988; Anderson & May 1992). Moreover, the importance of heterogeneity between hosts in their genetics, previous exposure to parasites, and behaviour has been shown in determining intensities of infection *e.g.* (Beh & Maddox 1996; Tanguay & Scott 1992). However, little attention has been paid to the genetic variation of parasites in their ability to infect particular hosts. Co-evolutionary interactions between hosts and their parasites are predicted by theoretical work (May & Anderson 1983; Lively & Apanius 1995) and are often observed in plant-pathogen systems where gene-for-gene interactions occur (Thompson & Burdon 1992). That this study found no obvious between-host differences in the genetic composition of parasite infrapopulations argues that there is not a strong interaction between host genotype and the genetic markers used in this study. However, these markers were selected as anonymous markers-they are presumably not immunodominant loci and as the sampled *S. ratti* are in linkage equilibria, then associations between the markers used and rat genotype will not be observed. In order to look for associations between rat and parasite genotype, then loci would have to be selected for on the basis of their antigenicity. This in itself would be a worthwhile study, and indeed, an appropriate starting point would be to use the Actin locus as there is some evidence that this locus may be under selection within the Berkshire sample site.

This study provides evidence that both macro and microgeographic population structure is not a rule in populations of nematode parasites. Predictions by Price (Price 1980) that the life-history of parasites predisposes them to small populations, high levels of population differentiation, high levels of inbreeding and low intrapopulation genetic diversity, are not borne out by this study. Thus, a general conclusion is that populations of parasitic nematodes are genetically unstructured. This appears to be especially true of parasite species infecting host species showing unconstrained dispersal.

This conclusion has several implications for considering the rate of spread of economically and medically important traits through populations of nematode parasite. It is debatable as to whether or not that the evolution of drug resistance will take longer in structured populations of parasites. In a simple deterministic model of allele

frequency change, the allele of greatest fitness will eventually replace all others. However, the probability of fixation of a mutation is a function of its frequency. In large populations there is a high probability that novel mutations will be lost by random genetic drift, relative to small (structured) populations. In structured populations, mutations will reach fixation more frequently, and also at a faster rate, due to the smaller effective population size (Nadler 1995). Furthermore, the alleles involved in anthelmithic resistance are probably recessive, as is the case with pesticide resistance genes (Plapp 1986). This has been shown for benzimidizole resistance in *Trichostrongylus colubriformis* (Grant 1994). As a consequence, expression of resistance will only occur in the homozygous state. Thus, in outbreeding populations of parasites the rate of increase in frequency of the allele will be slower relative to inbreeding populations, and resistance will take longer to emerge. Therefore, on this basis it may be argued that the evolution of resistance against chemotherapeutic control methods will occur more rapidly in structured populations. However, this would be a local effect. If a resistance allele was to arise in a structured population, a lack of gene flow would predict slow dispersal of the allele into susceptible populations. Therefore, the gene pool as a whole is expected to evolve more slowly in a species with a subdivided population structure.

Little is known about the importance of the effects of population structure on the evolution of drug resistance. However, the observation that most agriculturally important species of nematode show drug-resistant phenotypes when searched for, shows empirically that the parasite gene-pools are evolving fast (Prichard *et al.* 1980 ; Waller *et al.* 1996). That these species have unstructured populations argues that there are no barriers to the evolution, and spread, of genetic resistance.

5.4 Future work

It has been shown that using genetic markers which detect increased levels of polymorphism, such as RAPDs, significantly increases measures of within-host genetic diversity (Barral *et al.* 1996). Here, three loci were analysed of which two were diallelic. If *S. ratti* follows the trend observed in the above study, then the use of more polymorphic markers may uncover higher levels of within-host genetic diversity. The use of a multilocus fingerprinting technique on the samples in this study would be useful in providing an absolute estimate of the within-rat genetic diversity.

Extension of the range from which *S. ratti* is sampled is required to find out at what scale, if any, *S. ratti* begins to show genetic subdivision. In the first instance, further comparison of British and European isolates of *S. ratti* are necessary to confirm that panmixia extends to include European populations. That genetic subdivision occurs on a continental scale is suggested by the occurrence of two polymorphisms found only in a North American isolate of the parasite. Further analysis of American samples, and samples from other continents, would enable the amount of genetic subdivision on the worldwide scale to be analysed.

This study, as do many population genetic studies, provides only a single snapshot of the population genetic structure of *S. ratti* by the use of point samples. A method has been developed by which the genetic composition of infections can be non-invasively followed through time. This provides an opportunity for the longitudinal observation of infrapopulations. It is apparent within the Berkshire site that deviations from random mating occur. Longitudinal studies of this sample site provide an opportunity to observe whether this is a persistent, or short-lived effect. Further, it would be useful to imitate wild populations of parasites by creating semi-wild, enclosed populations of rats infected by *S. ratti* that are homozygote for a genetic marker. Introduction of a single rat infected by *S. ratti* of the opposite homozygote genotype would enable (i) the dynamics of infection within the 'wild' colony to be followed and (ii) the rate of appearance of heterozygote genotypes within the colony to be assessed. This would characterise the effect of sexual reproduction on the within-colony population genetic structure. Such an approach would be usefully complemented by the use of mathematical models.

The development of automated sequencing techniques has improved the efficiency of rapidly sequencing long stretches of DNA. Using these sequences, coalescent approaches for analysing gene genealogies enable past evolutionary processes to be inferred, this is 'phylogenetic epidemiology' (Harvey & Nee 1997). The use of such techniques will assist in determining whether the parasites sampled in this study are descended from a recently bottlenecked population of parasites, or are descended from lineages of large effective population size.

5.5 References

Abadie, S. (1963) The life cycle of *Strongyloides ratti*. *Journal of Parasitology* 49: 241-248

Agatsuma, T. and Habe, S. (1985b) Electrophoretic studies on diploid and triploid *Paragonimus westermani*. *Parasitology* 91: 489-497

Agrios, G. N. (1988) *Plant pathology.* Academic press, inc. 119-123

Anderson, J., Fuglsang, H. and Marshall, T. F. (1974) Studies of Onchocerciasis in the United Cameroon republic. Comparison of onchocerciasis in rain forest and Sudan-savannah. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 68:209-222

Anderson, R. M. and May, R. M. (1985) Helminth infections of humans: mathematical models, population dynamics and control. *Advances in Parasitology* 24:1-99

Anderson, R. M. and May, R. M. (1992) *Infectious disease of humans.* Oxford University Press, Oxford.

Anderson, T. J. C. (1994) The population genetics, epidemiology and evolution of the parasitic nematode *Ascaris suum.* Ph.D. thesis, University of Rochester.

Anderson, T. J. C., Romero-Abal, M. E. and M. E., Jaenike, J. (1993) Genetic structure and epidemiology of *Ascaris* populations: Patterns of host affiliation in Guatemala. *Parasitology* 107:319-334

Anderson, T. J. C., Romero-Abal, M. E. and Jaenike, J. (1995) Mitochondrial DNA and *Ascaris* microepidemiology: the composition of parasite populations from individual hosts, families and villages. *Parasitology* 110:221-229

Anderson, T. J. C. (1995) Ascaris infections in Humans from North America: molecular evidence for cross-infection. *Parasitology* 110:215-219

Avise, J. C. (1994) *Molecular markers, natural history and evolution.* Chapman and Hall, 2-6 Boundary Row, London.

Barnett, S. A. 1952. The biology of rat populations. *Surgo*:111-121

Barral, V., Morand, S., Pointier, J. P. and Theron, A. (1996) Distribution of schistosome genetic diversity within naturally infected *Rattus rattus* detected by RAPD markers. *Parasitology* 113:511-517

Bell, C. J. and Ecker, J. R. (1994) Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. *Genomics* 1;19(1):137-144

Beech, R., Prichard, R. and Scott, M. (1994) Genetic variability in the ß-tubulin genes in Benzimidizole-resistant and -susceptible strains of *Haemonchus contortus. Genetics* 138:103-110

Beh, K. J. and Maddox, J. F. (1996) Prospects for development of genetic markers for resistance to gastrointestinal parasite infection in sheep. *International journal of Parasitology* 26:879-897

Blattman, A. N., Hulme, D. J. and Beh, K. J. (1993) A search for associations between MHC RFLPs and resistance to *Haemonchus contortus infection* in sheep. *Animal Genetics* 24:277-282

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

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

Boggs, J. F., McMurray, S. T. and Leslie, D. M. (1991) Influence of habitat modification on the community of gastrointestinal helminths of cotton rats. *Journal of Wildlife Disease* 27:584-593

Bonnell, M. L. and Selander, R. K. (1974) Elephant seals: Genetic variation and near extinction. *Science*:908-909

Bullini, L. and Nascetti, L. (1986) Genetic variation of Ascaridoid worms with different life cycles. *Evolution* 40:437-440

Bundy, D. A. P. (1986) Epidemiological aspects of *Trichuris* and trichuriasis in Caribbean communities. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 80:706-718

Bundy, D. A. P. (1988) Population ecology of intestinal helminth infections in human communities. *Philosophical Transactions of the Royal Society, London, series B* 321:405-420

Burt, A., Carter, D. A., Koenig, G. L., White, T. J. and Taylor, J. W. (1996) Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *Proceedings of the National Academy of Sciences, USA*. 93:770-773

Callen, D. F. (1993) Incidence and origin of 'null' alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics* 52:922-927

Chou, Q., Russell, M., Birch, D. E., Raymond, J. and Bloch, W. (1992) Prevention of pre-PCR mis-priming and primer dimerization improves low-copy- number amplifications. *Nucleic Acids Research* 20:1717-1723

Ciro, R.and Hewitt, G. 1996. 470 million years of conservation of microsatellite loci among fish species. *Philosophical Transactions of the Royal Society, London, series B* 263:549-557

Cockerham, C. C. (1973) Analysis of gene frequencies. *Genetics* 74: 679 - 700

Collura, R. V. and Stewart, C. B. (1995) Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* 378:485-492

Comins, H. N. (1977a) The managment of pesticide resistance. *Journal of Theoretical Biology* 65:112-120

Conway, D. J. (1997) Natural selection on polymorphic malaria antigens and the search for a vaccine. *Parasitology today* 13:26-29

Conway, D. J., Hall, A., Anwar, K. S. and Bundy, D. A. P. (1995) Household aggregation of *Strongyloides stercoralis* infection in Bangladesh. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 89:258-261

Cook, G. C. (1987) *Strongyloides stercoralis* hyperinfection syndrome: how often is it missed? *Quarterly Journal of medicine* 64:625-629

Crawley, M. J. (1993) *GLIM for ecologists (methods in Ecology).* London: Blackwell Scientific publications

Crisp, D. J.and Standen, A. (1988) *Lasaea rubra:* an apomictic crevice-living bivalve with clones separated by tidal level preference. *Journal of Experimental Marine Biology and Ecology* 117:27-45

Crofton, H. D. (1971a) A model of host-parasite relationships. *Parasitology.* 63:343-364

Crofton, H. D. (1971b) A quantitative approach to parasitism. *Parasitology.* 62:179-193

Crow, J. F. (1986) *Basic concepts in population, quantitative and evolutionary genetics.* W. H Freeman, New York

Crow, J. F. and Kimura, M. (1970) *An introduction to population genetics theory.* 1 st. ed. Harper and Row, New York, USA.

Curran, J., Baillie, D. L. and Webster, J. M. (1985) Use of genomic DNA restriction fragment length differences to identify nematode species. *Parasitology* 90:137-144

Curran, J., McClure, M. A. and Webster, J. M. (1986) Genotypic differentiation of *Melidogyne* populations by detection of restriction fragment length difference of total DNA. *Journal of Nematology* 10:323-332

Dalmasso, A. and Berge, J. B. (1983) Enzyme polymorphism and the concept of parthenogenetic species, exemplified by *Melidogyne*. In: *Concepts of nematode systematics*. Ed. Stone, A. N., Platt, H. M. and Khalial, L. F. 187-196. Systematics Association Special Volume 22, Academic Press, London.

Dawkins, H. J., Grove, D. I., Dunsmore, J. D. and Mitchell, G. F. (1980) *Strongyloides ratti*: susceptibility to infection and resistance to reinfection in inbred strains of mice as assessed by excretion of larvae. *International journal of Parasitology* 10:125-129

Dempster, A. P., Laird, N. M.and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, series B*:391-38

Dobson, C.and Tang, J. M. (1991) Genetic variation and host-parasite relationship: *Nematospiroides dubius* in mice. *Journal of Parasitology* 77:884-889

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. and Mattick, J. S. (1991) Touchdown PCR to cicumvent spurious binding during gene amplification. *Nucleic Acids Research* 19:4008-4010

Endler, J. A. (1977) *Geographic variation, speciation and clines.* Princeton University press, New Jersey.

Excoffier, L., Smouse, P. and Quattro, J. M. (1992) AMOVA inferred from metric distances among DNA haplotypes: Application to human mtDNA restriction data. Genetics 131:479-491

Fenwick, P. K., Aggett, P. J., Macdonald, D. C., Huber, C. and Wakelin, D. (1990) Zinc deprivation and zinc repletion: effect on the response of rats to infection with Strongyloides ratti. American Journal of Clinical Nutrition 52:173-177

Field, D. and Wills, C. (1996) Long, polymorphic microsatellites in simple organisms. *Philosophical Transactions of the Royal Society, London, series B* 263:209-215

Fincham, J. R. S. (1994) *Genetic analysis. Principles, scope and objectives.* Blackwell Science, Inc. Osney Mead, Oxford.

Fisher, R. A. (1930) *The genetical theory of natural selection.* Oxford university press, Oxford.

Flockhart, H. A., Cibulskis, R. E., Karam, M. and Albiez, E. J. (1986) *Onchocerca volvulus,* enzyme polymorphism in relation to the differentiation of forest and savannah strains of this parasite. *Transactions of the Royal society of Tropical Medicine and Hygiene* 80:285-292

Forrester, J. E., Scott, M. E., Bundy, D. A. P.and Golden, M. N. H. (1988) Clustering of *Ascaris lumbricoides* and *Trichuris trichuria* infections within households. *Transactions of the Royal society of Tropical Medicine and Hygiene* 82:282-288

Frenkel, M. J., Savin, K. W., Bakker, R. E. and Ward, C. W. (1989) Characterisation of cDNA clones coding for muscle tropomyosin of the nematode *Trichostrongylus colubriformis. Molecular and Biochemical parasitology* 37:191-200

Frost, D. R. and Wright, J. W. (1989) The taxonomy of uniparental species with special reference to parthenogenetic *Cnemidophorus* (Squamata: Teiidae). *Systematic Zoology* 37:200-209

Gemmil, A., Viney, M. E. and Read, A. F. (1997) Host immune status determines sexuality in a parasitic nematode. *Evolution* (In press)

Glass, G. E., Childs, J. E., Korch, G. W. and Leduc, G. W. (1989) Comparitive ecology and social interactions of Norway rat (*Rattus norvegicus)* populations in Baltimore, Maryland (USA). *Occasional papers of the Museum of Natural History, University of Kansas.* 130:1-33

Goudet, J. (1994) FSTAT, a programme for IBM PC compatibles to calculate Weir and Cockerams (1984) estimators of F-statistics (version 1.2). PhD thesis, Institiute de Zoologie et d'Ecologie Animale, Universitie de Lausanne, Switzerland

Goulart, E. G., Jourdan, M. C., Brazil, R. P., Brazil, B. G. and Cosendey, A. E. (1977) Ecological control of hookworm and strongyloidiasis. *Journal of Helminthology* 51: 131-132

Graham, G. L. (1936) Studies of *Strongyloides. S. rattil* in parasitic series, each generation in the rat established with a single homogonic larvae. *American Journal of Hygiene* 24:71-87

Grant, W. N. (1994) Genetic variation in parasitic nematodes and its implications. *International journal of Parasitology* 24:821-830

Gray, G. D., Woolaston, R. R. and Beaton, B. T. (1995) Breeding for resistance to infectious disease in small ruminants. A.C.I.A.R. monograph series No. 34, Canberra, Australia

Grenfell, B. T., Wilson, K., Isham, V. S., Boyd, H. R. G. and Dietz, K. (1995) Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode-ruminant interactions as a case study. *Parasitology* 111:s135-s151

Grove, D. I. (1989) *Strongyloidiasis, a major roundworm infection of man* (ed. D.I. Grove). London:Academic press.

Guo, S. W. and Thompson, E. A. (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 43:805-811

Hartl, D. L. and Clarke, G. L. (1989) *Principles of population genetics* (second ed.). Sinauer associates, Sunderland, Massachusetts, USA

Harvey, P. H. and Nee, S. (1997) Phylogenetic epidemiology lives. *Trends in Ecology and Evolution.* 9:361-363

Hasegawa, H., Kobayashi, J. and Otsuru, M. (1994) Helminth-parasites collected from *Rattus rattus* on Lanyu-Taiwan. *Journal of the Helminthological Society of Washington* 61:95-102

Hasegawa, H. and Shiraishi, S. (1992) *Tikusnema javaense*, new genus new species and other nematodes from *Rattus argentiventer* in west Java, Indonesia. *Journal of Parasitology* 78:800-805

Haswell-Elkins, M. R., Elkins, D. B.and Anderson, R. M. (1987b) Evidence for predisposition in humans to infection with *Ascaris,* hookworm, *Enterobius* and *Trichuriasis* in a South Indian fishing community. *Parasitology* 95:323-327

Hebert, P. D. N. (1974a) Enzyme variability in natural populations of *Daphnia magna.* III Genotypic frequencies in intermittent populations. *Genetics* 77:335-341

Hebert, P. D. N. (1974b-a) Enzyme variability in natural populations of *Daphnia magna.* II Genotypic frequencies in permanent populations. *Genetics* 77:323-324

Hebert, P. D. N. (1974b-b) Genotypic characteristics of cyclical parthenogens and their obligately sexual derivatives. In; *The Evolution of sex and its consequences* (ed. S.C. Stearns), Birkhauser, Verlag, Basel and Boston:175-195

Hebert, P. D. N. (1987b) Genotypic characteristics of cyclical parthenogens and their obligately asexula derivatives. In; *The evolution of sex and its consequences* (ed.S. C. Stearns), Birkhauser Verlag, Basel and Boston:175-195

Hebert, P. D. N., Ferrari, D. C. and Crease, T. J. (1982) Heterosis in Daphnia: a re-assessment. *American Naturalist* 119:427-434

Hillis, D. A. and Moritz, C. (1990) *Molecular systematics.* Sunderland: Sinauer Associates Inc.

Hirst, M., Grewal, P. and Davis, K. 1994. Precursor arrays for triplet repeat expansion at the fragile X locus. *Human Molecular Genetics* 3:1553-1560

Hominick, W. M. (1987) Population biology of hookworms in west Bengal: analysis of numbers of infective larvae recovered from damp pads applied to the soil surface at defaecation sites. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 81:978-986

Hudson, P. J. and Dobson, A. P. (1995) Macroparasites: observed patterns in naturally fluctuating animal populations. In; (Grenfell B. T. and Dobson, A. P., ed.) *Ecology of infectious disease in natural populations.* Cambrige University press, Cambridge.

Hudson, P. J., Newborn, D. and Dobson, A. P. (1992) Regulation and stability of a free-living host parasite system- *Trichostrongylus infection* in the red grouse 1. Monitoring and parasite reduction experiments. *Journal of Animal Ecology* 61:477-486

Hugall, A., Moritz, C., Stanton, J. and Wolstenholme, D. R. (1994) Low, but strongly structured mitochondrial DNA diversity in root knot nematodes (*Meloidogyne*). *Genetics* 136:903-912

Hughes, C. R. and Queller, D. C. (1993) Detection of highly polymorphic microsatellite loc in a species with little allozyme polymorphism. *Molecular Ecology* 2:131-137

Hughes, R. N. (1989) *A functional biology of clonal animals.* Chapman and Hall, London-New York

Jaenike, J. (1978) An hypothesis to account for the maintenance of sex within populations. *Evolutionary theory* 3:121-137

Jarman, A. P. (1989) Hypervariable minisatellites: recombinators or innocent bystanders? *Trends in Genetics* 5:367-371

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67-73

Karban, R. (1989) Fine-scale adaptation of herbivorous thrips to individual host plants. *Nature* 340:60-61

Karl, S. A., Bowen, B. W. and Avise, J. C. (1992) Global population genetic structure and male mediated gene flow in the Green turtle, *Chelonia mydas*: RFLP analysis of anonymous nuclear loci. *Genetics* 131: 121-132

Kawanabe, M., Nojima, H. and Uchikawa, R. (1988) Transmamarry transmission of *Strongyloides ratti. Parasitology Research* 75:50-56

Kennedy, C. R. (1987) Long term stability in the population levels of the eyefluke *Tylodelphis podicipina* in perch. *Journal of Fish Biology* 31:571-581

Kernodle, S. P., Cannon, R. E. and Scandalios, J. G. (1993) Concentration of primer and template qualitatively affects products in RAPD DNA PCR. *Biotechniques* 14:362-363

Keymer, A. E. and Slater, A. F. G. (1987) Helminth fecundity: Density dependance or statistical illusion? *Parasitology today* 3:56-60

Kimura, M. (1964) *Diffusion models in population genetics.* 1st ed. Methuen, London

Kimura, M. and Crow, J. F. (1963) The measurement of effective population number. *Evolution* 17:279-288

Kroyer, A. B. (1991) The Brown rat: shadow of humankind. *Fauna* (Oslo) 44:243-250

Lander, E. S. and Botstein, D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185-199

Lehmann, T., Besansky, N. J., Hawley, T. G., Kamau, L. and Collins, F. H. (1997) Microgeographic structure of*Anophololes gambiae* in western Kenya based on mtDNA and microsatellite loci. *Molecular Ecology* 6:243-253

LeJambre, L. F., Ractliffe, L. H., Uhazy, L. S. and Whitlock, J. H. (1972) Evidence that the seasonal alternation of polymorphs apparent in linguiformH. contortus cayugensis also occurs in the other phenotypes. *International journal of Parasitology* 2:171-173

Leslie, J. F., Cain, G. D., Meffe, G. K. and Vrijenhoek, R. C. (1982) Enzyme polymorphism in *Ascaris suum* (nematoda). *Journal of Parasitology* 68:576-587

Levene, H. (1949) On a matching problem in genetics. *Annals of Mathematics and Statistics.* 20:91-4

Lindo, J. F., Robinson, R. D., Terry, S. I. and Bundy, D. A. P. (1995) Age-prevalence and household clustering of *S. stercoralis* infection in Jamaica. *Parasitology* 110:97-102

Little, M. D. (1966a) Comparitive morphology of six species of *Strongyloides* (Nematoda) and redefinition of genus. *Journal of Parasitology* 52:69-84

Lively, C. M. and Apanius, V. (1995) Genetic diversity in host parasite interactions. *in* Grenfell B. T. and Dobson, A. P., ed. *Ecology of infectious disease in natural populations.* Cambrige University press, Cambridge.

Louis, E. J. and Dempster, E. R. (1987) An exact test for Hardy-Weinberg and multiple alleles. *Biometrics* 43:805-811

Lydeard, C. M., Mulvey, J. M., Aho, J. M. and Kennedy, P. K. (1989) Genetic variability among natural populations of the liver fluke *Fascioloides magna* in white-tailed deer. *Canadian journal of Zoology* 67:2021-2025

Lymbery, A. J. and Hobbs, R. P. (1990a) Genetic diversity and differentiation in *Echinococcus granulosus* from domestic and sylvatic hosts on the mainland of Australia. *Parasitology* 101:283-9

Lymbery, A. J., Thompson, R. C.and Hobbs, R. P. (1990b) Genetic diversity and genetic differentiation in Echinococcus granulosus (Batsch, 1786) from domestic and sylvatic hosts on the mainland of Australia. *Parasitology* 101:283-290

Lymbery, A. J.and Thompson, R. C. A. (1996) Species of *Echinococcus:* Pattern and process. *Parasitology today* 12:486-491

Lynch, M. (1984a) Destabilizing hybridisation, general-purpose genotypes and geographical parthenogenesis. *Quarterly review of biology.* 59:102-106

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

Maizels, R. M., Bundy, D. A., Selkirk, M. E., Smith, D. F. and Anderson, R. M. (1993) Immunological modulation and evasion by helminth parasites in human populations. *Nature* 365:797-805

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular cloning, a laboratory manual.* Cold Spring Harbour laboratory, Cold spring Harbour, NY

Markow, T., Hedrick, P. W., Zuerlein, K., Danilovs, J. and Martin, J. (1993) HLA polymorphism in the Havasupai: evidence for balancing selection. *American journal of Human Genetics* 53:943-952

May, R. M. and Anderson, R. M. (1983) Epidemiology and genetics in the coevolution of parasite and host. *Proceedings of the Royal Society of London, series B* 219:281-313

Maynard-Smith, J., Smith, N., O'Rourke, M. and Spratt, B. G. (1993) How clonal are bacteria? *Proceedings of the National Academy of Science, USA* 90:4384-4388

McPherson, M. J., Hames, B. D. and Taylor, G. R. (1995) *PCR 2.* In: The practical approach series. Oxford University Press, Oxford

Meunier, J. R. and Grimont, P. A. D. (1993) Factors affecting the reproducibility of RAPD DNA fingerprinting. *Microbiological Research* 144:373-379

Montgomery, S. S. J. and Montgomery, W. I. (1989) Cyclic and non-cyclic dynamics in populations of the helmith parasites of wood-mice *Apodemus sylvaticus. Journal of Helminthology* 62:78-90

Mopper, S. (1996) Adaptive genetic structure in phytophagous insect populations. *Trends in Ecology and Evolution* 11:235-238

Moritz, C. (1989) The origin and evolution of parthenogenesis in *Heteronotia binoei*; extensive genetic diversity among parthenogens. *Evolution* 43:994-1003

Moxon, E., Rainey, P., Nowak, M. and Lenski, R. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Current Biology* 4:24-33

Muller, H. J. (1964) The relation of recombination to mutational advance. *Mutation Research* 1:2-9

Mulvey, M., Aho, J. M. and Lydeard, C. (1991a) Comparitive population genetic structure of a parasite (*Fasciolopsis buski)* and its definitive host. *Evolution* 45:1628-1640

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

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

Nadler, S. A., Lindquist, R. L. and Near, T. J. (1995) Genetic structure of midwestern *Ascaris suum* populations: a comparison of isoenzyme and RAPD markers. *Journal of Parasitology* 81 (3):385-394

Nawalinski, T., Schad, G. A., Chowdbury, A. D. (1978) Population biology of hookworms in children in rural West Bengal. II. Aquisition and loss of hookworms. *American journal of Tropical Medicine and Hygiene* 27:1162-1173

Nei, M. (1975) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Science, USA* 70:3321-3323

Nei, M. (1987) *Molecular evolutionary genetics.* Columbia University press, New York

Nevo, E. (1978) Genetic variation in natural populations: patterns and theory. *Theoretical Population Biology* 13: 121-177

Niamatali, S., Nolan, T. J. and Schad, G. A. (1992) Can autoinfection be provoked in the *S. ratti*-infected gerbil, *Meriones unguiculatus*? *Journal of the Helminthological Society of Washington* 59:149-152

Nolan, T. J., Aitkens, L. M. and Schad, G. A. (1988) Cryopreservation of first-stage and infective third-stage larvae of *Strongyloides stercoralis. Journal of Parasitology* 74:387-391

Boyd, R. (1993) The National rodent report. Institution of environmental health officers, *Pest control news* (35):1-12

Okimoto, R., MacFarlane, J., Clary, D. O. and Worlstenholme, D. R. (1992) The mitochondrial genomes of two nematodes, *C. elegans* and *Ascaris suum. Genetics* 130:471-498

Orr, H., Chung, M. U. and Banfi, S. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics* 4:221-226

Owen, D. (1976) Some parasites and other organisms of wild rodents in the vicinity of and SPF unit. *Laboratory animals* 10:271-278

Parker, E. D. J. (1979) Ecological implications of clonal diversity in parthenogenetic morphospecies. *American Zoology* 19:753-762

Paterson, S., Wilson, K. and Pemberton, J. M. (1997) MHC variation associated with parasite resistance in an unmanaged vertebrate population. *Philosophical Transactions of the Royal Society, London, series B* (In press)

Peakall, R., Smouse, P. E. and Huffs, D. R. (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides. Molecular ecology* 4:135-147

Pemberton, J. M., Slate, J., Bancroft, D. R. and Barrett, J. A. (1995) Non-amplifying alleles at microsatellite loci-a caution for parentage and population studies. *Molecular ecology* 4:249-252

Plapp, F. W. (1986) Genetics and biochemistry of insecticide resistance in arthropods: prospects for the future. In; *Pesticide resistance: Strategies and tactics for managment*, pp 74 - 86.Washington, D. C. National Academy Press.

Premvati, P. (1958) Studies on *Strongyloides* of primates II. Factors determining the 'direct' and 'indirect' mode of life. *Canadian journal of Zoology* 36:185-195

Premvati, P. (1963) Effect of temperature on the development of free living adults of *S. papillosus* (Weld, 1856). *Parasitology* 53:483-489

Price, P. W. (1980) *Evolutionary biology of parasites.* Monographs in population biology, Princeton, N. J. Princeton University press.

Prichard, R. K., Hall, C. A. and Kelley, I. C. A. (1980) The problem of anthelminthic resistance in nematodes. *Australian Vetenarian* 56:239-251

Raymond, M. and Rousset, F. (1995) An exact test for population differentiation. *Evolution* 49:1280-1283

Raymond, M. A., Callaghan, P., Fort, P. and Pasteur, N. (1991) World-wide migration of amplified resistance genes in mosquitoes. *Nature* 350:151-153

Read, A. F. and Skorping, A. (1995) The evolution of tissue migration by a parasitic nematode larvae. *Parasitology* 111:359-371

Rice, W. R. (1989) Analyzing tables of statistical tests. *Evolution* 43:223-225

Rico, C., Rico, I. and Hewitt, G. (1996) 470 million years of conservation of microsatellite loci ampng fish species. *Philosophical Transactions of the Royal Society, London, series B* 263:549-557

Rollinson, D., Imbert-Establet, D. and Ross, G. C. (1986) *Schistosoma mansoni* from naturally infected *Rattus rattus* in Guadeloupe: identification, prevalence and enzyme polymorphism. *Parasitology* 93:39-53

Rychlik, W. and Rhodes, R. E. (1989) A computer program for choosing optimal oligonucleotides for filter hybridisation, sequencing and in-vitro amplification of DNA. *Nucleic acids research* 17:8543-8551

Sandground, J. H. (1925) Speciation and specificity in the nematode genus *Strongyloides. Journal of Parasitology* 12:59-82

Sanger, F., Nicklen, S. and Coulsen, A. R. (1977) DNA sequencing with chain-termination inhibitors. *Proceedings of the National Academy of Sciences, USA* 74:5463-5467

Schad, G. A. (1989) Morphology and life history of *Strongyloides stercoralis*. In: Strongyloidiasis, *a major roundworm infection of man* (ed. D.I. Grove) London; Academic press:85-104

Schlotterer, C., Amos, B. and Tautz, D. (1989) Conservation of polymorphic simple sequence loci in cetacean species. *Nature* 354:63

Schlotterer, C. and Tautz, D. (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* 20:211-215

Scott, M. E. (1987) Regulation of mouse colony abundance by *Heligmosomoides polygyrus*. *Parasitology* 95:111-124

Selander, R. K. (1976) Genetic variation in natural populations. In F. Ayala (ed.), *Molecular Evolution.* Sinauer, Sunderland, Massachusetts 21-45

Shaw, D. J. and Dobson, A. P. (1995) Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* 111:s111-s133

Sinniah, B. (1979) Parasites of some rodents in Malaysia. *Southeast Asian Journal of Tropical Medicine and Public Health* 10:115-121

Slater, A. F. G.and Keymer, A. E. (1986) Epidemiology of *Heligmosomoides polygyrus* infection in mice: experiments on natural transmission. *Parasitology* 71:51-60

Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. *Science* 236:787-792

Slatkin, M. and Barton, N. H. (1990) A comparison of three methods for estimating average levels of gene-flow. *Evolution* 43:1349-1368

Smith, M. (1995) Pied piper at the docks. *Independant on Sunday* (1235) 32-34

Sokal, R., and Rohlf, F. (1995) *Biometry.* 3rd ed. New York : Freeman

Speare, S. (1989) In: Strongyloidiasis, *a major roundworm infection of man* (ed. D.I. Grove). London: Academic press.

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C.and Hildebrand, C. E. (1991) Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. *Genomics* 10:807-815

Stear, M. J., Bishop, S. C., Duncan, J. L. and Murray, M. (1995) The repeatability of faecal egg counts, peripheral eosinophil counts, and plasma pepsinogen concentrations during deliberate infection with *Ostertagia cicumcincta.* International Journal of *Parasitology* 25:375-389

Sved, F. J. and Ayala, F. J. (1970) A population cage test for heterosis in *Drosophila pseudoobscura. Genetics* 66:97-113

Tajima, F. (1989) The effect of change in population size on DNA polymorphism. *Genetics* 123:597-601

Tajima, F. (1992) Measurement of DNA polymorphisms; in *Mechanisms of molecular evolution*, (eds. Takahata, N. and Clarke, A. G.) Academic press, Osney mead, Oxford.

Takahata, N. (1985) Population genetics fof extrauclear genomes: a model and review. in: *Population genetics and molecular evolution*, (eds. Ohta, T., Aoki, K), Springer Verlag, Berlin

Tanguay, G. C. and Scott, M. E. (1992) Factors generating aggregation of *Heligmosomoides polygyrus* (Nematoda) in laboratory mice. *Parasitology* 104(3), 519-529

Targett, G. A. (1995) Malaria-advances in vaccines. *Current Opinions in Infectious Disease* 8:322-327

Thomas, W. K. and Wilson, A. C. (1991) Mode and tempo of molecular evolution in the nematode *Caenorhabditis*; cytochrome oxidase II and calmodulin sequences. *Genetics* 128:269-279

Thomaz, D., Guiller, A. and Clarke, B. (1996) Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proceedings of the Royal Society of London, series B.* 263:363-368

Thompson, E. A., Deeb, S. and Walker, D. (1988) The detection of linkage disequilibria between closely linked markers. *American journal of Human genetics*.113-124

Thompson, J. N. and Burdon, J. J. (1992) Gene-for-gene coevolution between plants and parasites. *Nature* 360:121-5

Thompson, R. C. A. and Lymbery, A. J. (1990) Intraspecific variation in parasites-what is a strain? *Parasitology Today* 6:345-348

Thompson, R. C. A. and Lymbery, A. J. (1996) Genetic variability in parasites and host-parasite interactions. *Parasitology* 112:s7-s22

Tibayrenc, M., Kjellberg, F. and Ayala, F. J. (1990) A clonal theory of parasitic protozoa. *Proceedings of the National Academy of Science, USA.* 87:2414-2418

Tindall, N. R. and Wilson, P. A. G. (1988) Criteria for proof of migration routes of immature parasites inside hosts exemplified by studies of *S. ratti* in the rat. *Parasitology* 96:551-563

Triantaphyllou, A. C. and Monocol, D. J. (1977) Cytology, reproduction and sex determination of *Strongyloides ransomi* and *S. papillosus. Journal of Parasitology* 63: 132-139

Uchikawa, R., Nojima, H. and Sato, A. (1989) The effects of single and repeated inoculations of various larval doses on *Strongyloides ratti* burden and distribution in rats. *Journal of Parasitology* 75:577-584

Udonsi, J. K. (1989) Helminth parasites of wild populations of the black rat, *Rattus rattus* , from urban, rural residential and other ecological areas of Nigeria. *Acta Parasitologica Polonica.* 34:107-116

Viard, F., Bremond, P., Labbo, R. and Jarne, P. (1996) Micosatellites and the genetics of highly selfing populations in the freshwater snail *Bulinus truncatus. Genetics* 142:1237-1247

Viney, M. E. (1990) The first report of *Strongyloides ratti* Sandground, 1925 in wild rats in Britain. *Transaction of the Royal Sociaety of Tropical Medicine and Hygiene* 84:173

Viney, M. E. (1994) A genetic analysis of reproduction in *Strongyloides ratti. Parasitology* 109:511-515

Viney, M. E. (1996) Developmental switching in the parasitic nematode *Strongyloides ratti*. *Proceedings of the Royal Society of London, series B* 263:201-208

Viney, M. E. and Ashford, R. W. (1990) The use of enzyme electrophoresis in the taxonomy of *Strongyloides*. *Annals of Tropical Medicine and Parasitology* 84:35-47

Viney, M. E., Matthews, B. E. and Walliker, D. (1992) On the biological and biochemical nature of cloned populations of *Strongyloides ratti*. Journal of Helminthology 66: 45-52

Viney, M. E., Matthews, B. E. and Walliker, D. (1993) Mating in the nematode parasite *Strongyloides ratti*: Proof of genetic exchange. *Proceedings of the Royal Society of London, series B* 254:213-219

Waller, P., J., Echevarria, F., Eddi, C., Maciel, S., Nari, A. and Hansen, J. W. (1996) The prevalence of anthelmintic resistance in nematode parasites of sheep in southern Latin America: general overview. *Veterinary Parasitology* 62:181-187

Weber, J. L. and Wong, C. (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* 2:1123-1128

Webster, J. P. and MacDonald, D. W. (1995) Parasites of wild brown rats (*Rattus norvegicus*). *Parasitology* 111:247-255

Weider, L. J. and Hobaek, A. (1994) Molecular biogeography of clonal lineages in a high Arctic apomictic *Daphnia* complex. *Molecular Ecology* 3:497-499

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

Weir, B. S. (1996) *Genetic data analysis II*. Sinauer Associates, Inc. publishers, Sunderland, Massachusettes.

Welsh, J. and McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18:7213-7218

Wertheim, G. and Greenberg, Z. (1970) Notes of helminth parasites of myomorph rodents from Southern Sinai. *Journal of Helminthology* 44:243-252

Willcox, M., Bjorkman, A. and Brohult, J. (1983) Falciparum malaria and beta-thalassaemia trait in northern Liberia. *Annals of Tropical Medicine and Parasitology* 77:335-347

Williams, J. C., Knox, J. W. and Hoerner, T. J. (1983) Seasonal changes of nematode populations in yearling beef cattle in Louisiana with emphasis of prevalence of infection in *Ostertagia ostertagi. International Journal of Parasitology* 13:133-143

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18:6532-6535

Wilson, I. J. (1996) Genetic differentiation in spatially structured populations with particular regard to the Atlantic salmon (*Salmo salar* L.). PhD. thesis, Edinburgh University.

Woolaston, R. R. and Baker, R. L. (1996) Prospects of breeding small ruminants for resistance to internal parasites. *International journal of Parasitology* 26:845-855

Wright, S. (1931) Evolution in Mendelian populations. Genetics 16:97-159

Yamada, M., Matsuda, S., Nakazawa, M. and Arizono, N. (1991) Species-specific differences in heterogonic development of serially transferred free-living generations of *Strongyloides planiceps* and *Strongyloides stercoralis. Journal of Parasitology* 77:592-594

Zaffagnini, F. (1973) Parthenogenesis in the parasitic and free-living forms of *Strongyloides papillosus* (Nematoda, Rhabdiasoidea). Chromosoma 40:34-37

Appendix 1. Composition of solutions and buffers

PCR dNTP solution (100X)

75 μM dATP
75 μM dTTP
75 μM dGTP
75 μM dCTP

acetate/EDTA)

TBE buffer (Tris/borate/EDTA)
0.09 M Tris-base
0.09 M boric acid
0.002 M EDTA

Sequencing gel solution
7 M urea
6 % (w/v) acrylamide (Anachem)
0.16 % (w/v) bis-acrylamide (Anachem)
1X TBE
TEMED and 25 % Ammonium persulphate (w/v) added to a final concentration of
0.6 μl ml^{-1}

Pre-hybridisation solution
for 1 litre:
100 g dextran sulphate solution (sodium salt, MW
~500 000, Sigma)
58 g NaCl
846 ml dH_2O)
Mixed at 65°C until dissolved, aliquoted and frozen at -20°C

20 X SSPE
3.6 M NaCl
0.2 M Sodium phosphate
0.02 M EDTA

Denaturing solution
1.5 M NaCl
0.5 M NaOH

Neutralising solution
1.5 M NaCl
0.5 M Tris-HCl pH 7.2
1 mM EDTA

20 X SSC
3M NaCl
0.3 M Na_3 citrate

Gel loading buffer (10X)

20% ficoll (w/v) (GIBCO)
0.25% (w/v) xylene cyanol FF

TAE buffer (Tris-

0.04 M Tris-acetate
0.002 M EDTA

TE buffer (Tris/EDTA)
10 mM tris-HCL
1mM EDTA

WILTSHIRE

W325.Actin BSP-8 CM-2
, 0000 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000

W316.Actin BSP-8 CM-2
, 0000 0102 0404
, 0101 0202 0000
, 0101 0202 0404
, 0000 0202 0101
, 0101 0202 0000
, 0000 0000 0000
, 0000 0000 0101
, 0000 0000 0303
, 0101 0202 0000

W323.Actin BSP-8 CM-2
, 0000 0102 0000
, 0000 0202 0000
, 0101 0202 0000
, 0000 0202 0000
, 0000 0202 0000
, 0000 0202 0000
, 0102 0202 0000
, 0000 0202 0000
, 0102 0202 0000
, 0000 0202 0000

W324. Actin BSP-8 CM-2
, 0000 0000 0101
, 0000 0000 0000
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0000
, 0000 0000 0404
, 0000 0000 0101
, 0000 0000 0000
, 0000 0000 0101

W313
, 0000 0000 0000
, 0101 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000

W322
, 0000 0000 0000
, 0000 0000 0000
, 0101 0000 0303
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0303
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0101 0000 0303

W320
, 0101 0000 0000
, 0101 0202 0000
, 0000 0202 0000
, 0000 0000 0000
, 0101 0202 0000
, 0101 0202 0000
, 0000 0202 0000
, 0000 0000 0000

W321
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0101

W322
, 0101 0202 0000
, 0000 0000 0000
, 0101 0000 0000
, 0101 0000 0000
, 0101 0000 0000
, 0000 0000 0000
, 0101 0000 0000
, 0101 0000 0000
, 0000 0000 0000
, 0000 0000 0000

W330
, 0102 0202 0000
, 0101 0000 0000
, 0101 0000 0000
, 0101 0000 0000
, 0000 0000 0000
, 0202 0202 0000
, 0000 0000 0000
, 0102 0000 0000
, 0000 0000 0000
, 0102 0000 0000

W331
, 0101 0000 0000
, 0000 0000 0000
, 0101 0000 0000
, 0000 0000 0000
, 0000 0202 0000
, 0000 0000 0000
, 0101 0102 0000
, 0101 0000 0000
, 0000 0000 0000
, 0101 0000 0000

W315
, 0000 0202 0404
, 0101 0202 0000
, 0101 0202 0404
, 0101 0202 0000
, 0000 0000 0404
, 0000 0000 0000
, 0000 0000 0000
, 0101 0202 0404
, 0101 0202 0000
, 0000 0000 0000

W318
, 0102 0000 0000
, 0000 0000 0000
, 0101 0000 0000

W317
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0102 0202 0000
, 0000 0202 0000
, 0101 0202 0000
, 0000 0000 0000
, 0000 0202 0000
, 0000 0202 0000

W322
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0000
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0101
, 0000 0000 0000

HEDLEY

HE 265
, 0102 0102 0000
, 0102 0102 0000
, 0102 0102 0000
, 0102 0102 0000
, 0102 0102 0000
, 0101 0000 0000
, 0102 0102 0000
, 0101 0202 0000
, 0101 0202 0000
, 0000 0202 0000

HE 294
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0101
, 0000 0000 0000
, 0102 0000 0404
, 0000 0000 0000

HE 271
, 0101 0000 0303
, 0000 0000 0303
, 0000 0000 0404
, 0101 0202 0000
, 0000 0000 0000
, 0102 0102 0000

HE 267
, 0102 0202 0404
, 0102 0102 0404
, 0000 0000 0404
, 0000 0000 0404
, 0102 0102 0404
, 0202 0202 0404
, 0000 0000 0104
, 0000 0000 0404
, 0000 0000 0404
, 0101 0102 0404

HE 287
, 0202 0202 0000

HE 285
, 0102 0000 0000

HE 292
, 0000 0000 0000
, 0102 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0101 0000 0000
, 0101 0000 0000
, 0101 0202 0000
, 0101 0000 0101

DORSET

D 336
, 0102 0202 0000
, 0101 0202 0000
, 0101 0202 0000
, 0102 0202 0000
, 0101 0202 0000
, 0101 0202 0000
, 0102 0202 0000
, 0102 0202 0000

D 334
, 0102 0202 0000
, 0000 0000 0000
, 0000 0000 0000
, 0101 0000 0000
, 0101 0202 0000
, 0000 0000 0000
, 0000 0000 0000
, 0101 0202 0000
, 0101 0202 0000

D 335
, 0000 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0000 0000 0000
, 0102 0202 0000
, 0000 0202 0000
, 0000 0000 0000
, 0000 0202 0000

D 339
, 0000 0000 0101
, 0000 0000 0404
, 0000 0000 0101
, 0000 0000 0101
, 0000 0202 0101
, 0000 0202 0101
, 0000 0000 0101

D 338
, 0102 0000 0000
, 0101 0000 0000
, 0000 0202 0000
, 0000 0202 0000
, 0101 0202 0000
, 0000 0202 0000

SURREY SITE A

SA 30
, 0101 0202 0404
, 0000 0000 0404
, 0102 0202 0404
, 0102 0202 0404
, 0101 0202 0404
, 0102 0202 0404
, 0000 0000 0404
, 0000 0202 0404
, 0102 0000 0404

SA 27
, 0000 0000 0000
, 0000 0000 0000
, 0101 0000 0000
, 0000 0000 0101
, 0000 0000 0000
, 0101 0000 0303
, 0000 0000 0101
, 0101 0000 0101
, 0000 0000 0000
, 0000 0000 0000

SA 31
, 0000 0000 0000
, 0000 0000 0000
, 0101 0202 0000
, 0000 0202 0000
, 0000 0000 0000
, 0102 0202 0000
, 0102 0000 0000
, 0000 0000 0000
, 0000 0000 0000

SA 32
, 0101 0000 0101
, 0000 0000 0000
, 0101 0202 0000
, 0202 0000 0000
, 0202 0000 0000
, 0202 0000 0000
, 0202 0000 0000
, 0202 0000 0000
, 0202 0000 0000
, 0000 0000 0000

SA 28
, 0000 0202 0303
, 0000 0202 0101
, 0000 0202 0000
, 0000 0202 0000
, 0101 0202 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0202 0000

SA 26
, 0101 0202 0000
, 0000 0202 0404
, 0101 0202 0000
, 0101 0202 0000
, 0101 0202 0404
, 0101 0202 0404
, 0101 0202 0000

SA 29
, 0102 0202 0000
, 0102 0202 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0000 0000

SURREY SITE B

SB 25
, 0000 0000 0000
, 0000 0000 0000
, 0202 0202 0000

SB 26
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0101 0202 0000
, 0000 0000 0000
, 0000 0000 0000

SB 28
, 0101 0202 0000
, 0102 0000 0000
, 0000 0000 0101
, 0101 0000 0000
, 0000 0202 0101
, 0101 0202 0000
, 0102 0000 0000
, 0101 0202 0000
, 0101 0202 0000

OXFORDSHIRE

T1/2	T1/3	T1/5	T1/6	T1/7
0102 0000 0000	0102 0202 0000	0000 0000 0000	0000 0000 0000	0000 0000 0000
0102 0000 0000	0102 0102 0000	0000 0000 0000	0000 0000 0000	0101 0202 0000
0102 0000 0000	0000 0000 0000	0000 0000 0000	0101 0000 0000	0102 0102 0000
0101 0000 0000	0000 0000 0000	0000 0000 0000	0101 0000 0000	0000 0000 0000
0101 0202 0000	0000 0000 0000	0000 0000 0000	0101 0000 0000	0000 0000 0000
0000 0000 0000	0000 0000 0000	0000 0000 0000	0101 0000 0000	0000 0000 0000
0202 0000 0000	0101 0202 0000	0000 0000 0000	0000 0000 0000	0102 0102 0000
0102 0000 0000	0000 0000 0000	0202 0000 0000	0000 0000 0000	
0102 0202 0000	0102 0000 0000	0000 0000 0000		
0101 0000 0000	0000 0000 0000			

NORFOLK

N384	N383	N385
0102 0102 0101	0000 0000 0101	0202 0202 0101
0102 0202 0101	0102 0202 0101	0000 0000 0000
0202 0202 0101	0102 0202 0101	0000 0202 0000
0102 0102 0101	0000 0000 0101	0000 0000 0000
0102 0102 0101	0202 0202 0101	0000 0000 0101
0202 0202 0101	0101 0202 0101	0101 0202 0000
0202 0202 0101	0202 0202 0101	0000 0202 0000
0202 0202 0101	0000 0000 0101	0101 0202 0000
0102 0102 0101	0101 0202 0101	0101 0000 0000
0000 0000 0101	0202 0202 0101	0101 0202 0000
0101 0202 0101	0102 0202 0101	0000 0202 0101
0102 0202 0101	0102 0202 0101	0000 0202 0101
0102 0102 0101	0101 0202 0101	0000 0000 0101
0202 0202 0101	0101 0202 0101	0101 0202 0000
	0202 0202 0101	0102 0202 0404

BERKSHIRE

367	359	381	352
0000 0202 0101	0000 0202 0000	0102 0202 0000	0102 0202 0101
0000 0202 0101	0102 0202 0000	0102 0202 0000	0102 0102 0101
0000 0202 0000	0102 0202 0000	0102 0000 0000	0202 0202 0101
0000 0000 0000	0101 0202 0000	0102 0202 0000	0102 0202 0101
0101 0202 0404	0102 0202 0000	0000 0000 0000	0000 0000 0101
0102 0000 0000	0101 0202 0000	0102 0202 0000	0102 0202 0101
0102 0202 0000	0102 0202 0000	0102 0202 0000	0102 0202 0101
0102 0000 0101	0101 0000 0000	0102 0202 0000	0102 0202 0101
0101 0202 0101	0101 0202 0000	0102 0000 0000	0102 0202 0101
0102 0000 0000	0102 0000 0000	0000 0202 0000	0102 0102 0101
0102 0000 0000	0000 0000 0000	0000 0202 0000	
0102 0000 0000	0102 0202 0000	0102 0202 0000	
0102 0000 0404	0102 0202 0000	0102 0202 0000	
0000 0000 0000	0101 0202 0000	0102 0202 0000	
0102 0000 0101		0102 0202 0000	

356
, 0102 0000 0000
, 0102 0202 0000
, 0000 0000 0000
, 0000 0000 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0101
, 0102 0102 0000
, 0102 0202 0000
, 0102 0102 0000
, 0000 0202 0000
, 0102 0202 0101
, 0102 0202 0101
, 0000 0000 0101

363
, 0102 0000 0000
, 0102 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0101
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0101 0202 0000
, 0000 0000 0000
, 0101 0000 0000
, 0101 0000 0000

353
, 0102 0000 0101
, 0102 0000 0101
, 0102 0000 0000
, 0102 0000 0101
, 0102 0000 0101
, 0000 0202 0101
, 0102 0202 0101
, 0102 0202 0101
, 0102 0202 0101
, 0102 0202 0101
, 0102 0202 0101
, 0102 0000 0000
, 0000 0000 0101
, 0000 0202 0101
, 0000 0000 0000

365
, 0102 0202 0101
, 0102 0202 0101
, 0102 0202 0101
, 0102 0202 0101
, 0102 0102 0101
, 0101 0202 0101
, 0000 0000 0101
, 0000 0202 0101
, 0102 0202 0101
, 0102 0000 0101
, 0102 0000 0101
, 0102 0000 0101
, 0000 0000 0101
, 0102 0000 0000
, 0000 0000 0101

376
, 0102 0202 0101
, 0000 0000 0101
, 0000 0202 0101
, 0102 0202 0101
, 0102 0202 0101
, 0102 0000 0101
, 0102 0102 0101
, 0102 0000 0101
, 0102 0202 0101
, 0102 0202 0101
, 0101 0202 0101
, 0000 0202 0101
, 0101 0202 0101
, 0000 0000 0000
, 0102 0202 0000

378
, 0000 0202 0000
, 0000 0202 0000
, 0102 0202 0101
, 0000 0000 0000
, 0102 0000 0101
, 0000 0102 0101
, 0102 0102 0101
, 0000 0000 0101
, 0102 0102 0000
, 0000 0102 0000
, 0000 0000 0000
, 0102 0000 0000
, 0102 0202 0000
, 0102 0102 0000
, 0102 0000 0000

350
, 0102 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0102 0000 0000
, 0102 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0000 0000 0000
, 0101 0000 0000

351
, 0000 0202 0101
, 0102 0202 0101
, 0102 0102 0000
, 0102 0202 0101
, 0102 0000 0101
, 0101 0202 0101
, 0102 0202 0101
, 0102 0202 0101
, 0000 0000 0101
, 0000 0000 0000
, 0102 0000 0000
, 0102 0202 0000
, 0102 0202 0101
, 0102 0202 0000
, 0102 0202 0000

358
, 0102 0000 0101
, 0101 0000 0101
, 0000 0202 0101
, 0102 0202 0101
, 0000 0000 0101
, 0000 0202 0101
, 0102 0202 0101
, 0101 0000 0101
, 0102 0000 0101
, 0000 0000 0101
, 0202 0000 0101
, 0000 0000 0000
, 0000 0000 0101
, 0202 0000 0101

375
, 0102 0202 0000
, 0102 0102 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0102 0202 0000
, 0202 0202 0000

372
, 0102 0202 0000
, 0000 0000 0000
, 0101 0000 0000
, 0101 0202 0000
, 0000 0202 0000
, 0101 0000 0000
, 0102 0202 0000

377
, 0101 0202 0101
, 0102 0000 0000
, 0102 0202 0101
, 0102 0102 0101
, 0000 0000 0101
, 0101 0000 0101
, 0101 0202 0101
, 0000 0000 0101
, 0000 0000 0000
, 0102 0202 0102
, 0000 0202 0101
, 0000 0202 0101
, 0000 0202 0101
, 0102 0202 0101
, 0202 0000 0101

369	380	370	360
, 0000 0000 0404	, 0101 0202 0101	, 0102 0202 0101	, 0102 0202 0404
, 0102 0000 0102	, 0000 0202 0101	, 0102 0202 0101	, 0102 0000 0101
, 0000 0202 0102	, 0101 0202 0101	, 0000 0000 0101	, 0101 0202 0000
, 0101 0202 0102	, 0000 0000 0101	, 0102 0202 0101	, 0000 0000 0000
, 0000 0000 0404	, 0101 0202 0000	, 0102 0202 0101	, 0101 0202 0000
, 0101 0102 0404	, 0000 0000 0000	, 0202 0202 0101	, 0000 0000 0000
, 0000 0000 0404	, 0101 0202 0101	, 0102 0202 0101	, 0102 0202 0000
, 0102 0102 0404	, 0000 0000 0101	, 0102 0202 0101	, 0102 0202 0404
, 0000 0000 0000	, 0101 0202 0101	, 0000 0202 0101	, 0102 0202 0000
, 0101 0000 0000	, 0102 0202 0101	, 0102 0000 0101	, 0000 0000 0000
, 0102 0102 0000	, 0102 0202 0000	, 0102 0202 0101	, 0000 0000 0000
, 0000 0000 0000	, 0101 0202 0000	, 0102 0102 0101	, 0000 0202 0000
, 0102 0000 0404	, 0101 0202 0000	, 0000 0000 0101	, 0000 0000 0000
, 0102 0000 0000	, 0102 0202 0000	, 0202 0202 0101	, 0000 0202 0000
, 0102 0202 0000	, 0101 0202 0000		, 0000 0202 0000
368	**366**	**374**	**373**
, 0000 0000 0101	, 0102 0202 0101	, 0102 0000 0000	, 0102 0000 0000
, 0101 0202 0101	, 0101 0202 0101	, 0102 0202 0101	, 0000 0000 0000
, 0000 0202 0101	, 0000 0202 0101	, 0000 0000 0000	, 0102 0102 0000
, 0102 0000 0101	, 0102 0202 0101	, 0000 0000 0000	, 0102 0000 0000
, 0102 0000 0101	, 0102 0000 0101	, 0000 0000 0000	, 0102 0202 0101
, 0000 0202 0101	, 0102 0000 0101	, 0102 0102 0101	, 0000 0202 0101
, 0101 0202 0101	, 0000 0202 0101	, 0202 0202 0000	, 0102 0102 0000
, 0102 0000 0101	, 0102 0202 0000	, 0102 0202 0101	, 0102 0202 0101
, 0102 0202 0101	, 0000 0202 0101	, 0000 0000 0101	, 0102 0202 0101
, 0102 0202 0101	, 0102 0202 0101	, 0102 0202 0000	, 0102 0202 0101
, 0102 0202 0101	, 0102 0202 0101	, 0000 0000 0000	, 0000 0000 0000
, 0101 0202 0101	, 0102 0000 0000	, 0102 0202 0000	, 0202 0202 0000
, 0101 0202 0101	, 0102 0202 0000	, 0000 0000 0000	, 0102 0202 0101
, 0102 0102 0101	, 0102 0202 0101	, 0101 0202 0000	, 0102 0202 0000
, 0102 0202 0101	, 0102 0202 0101	, 0102 0202 0000	, 0102 0202 0000

Appendix 3. Sequences isolated in the course of the study

Bsp-8f:

cacctagtgg aaatccagaa agaaatcgaa aaggatcagg tgaagatgat ccaaataatc ctaataagaa gctacttcaa tgattgatga tgattattat aagaacgttt taaaaaattt aacaaagaat
tatgttgaaa acactgctgt attaaaaata tagtttaatg gttttcccct tgaattcttt agcccaagag tcattttaaa cgaagtacgt gttaaattaa gagatttcta aattaggaaa aagagtggat gaagt

24f:

aaacaatgcaatatctctcaatattgcatcgaattctttgtttttattccattttctccatgtaaattcatacatgtacccatttaacatactacggtgggttccgttgtgacgtaagtt

24r:

aaataggataataacaattaattatttggtataaaatgtcgaatgaaattaatgatacaacagaaagataaaaacataattacgagattctcaaaactgcaaaattaagaaatgcgtaaccaatattttgaaaactagaaaaatttactt
aacaaaaaaaagggtactat

4f:

ctttgaatgaggcacaagcgcagttgatcggtcttgcagcaaccagtaaaaaactggcttatgtttatccagaaggtttaccgccaggtagtaatgtcgatttaaaaattgatcggaataaggcagaagctttaggtgtgaattttgctga
tgtg

4r:

aagcaaggttggtttagtttaatgagcaggctttaccctgatactcatagcgatttacctcataacttttaccatcccaacgatacaccgcaaaacataaacctggacctccattttcaatatcaggccaaccatctttgcctgtgctttt
tagaaaggtcggtata

27f:

ttgatcggtcttgcagcaattctttttgttgaatcctttagtttcatatataagatgcgtatcttgcgatgatcatctccatcatccactcggtcaagttcaaatgctgtgtcaatcttccagataaaatcatcaaatgtctttgtgtctg
agttgtaatggtcaaactcagtaacctttttcaatatatatgcttgtttcagccattttnactancgtgttttcagcaggtatatcagatgcttcactattatagtcttccaggtgttggtgaaccaatccttggataatgtaggtttaagt
agtcttctatctncntttgtgtattgctcccnctgtat gaagcaatttcttca

27r:

agttgatcggtcttgcagctattgagatcatgatcagcttaaactatataanaagacctttttaatggtgatatcagttgcaaaactacccatacctgtgtgctgcatattgctagtctgtctaagttaatatcattttatttaatttatt
taagttatnattatcgttattgtntctcanctcttattantgcagtatccttttcatattttatatcgtgcctatgccgtaacataaggtattctttggtatttccgantactttantattattattatcattatttnagggttaaacgaag
taacactactaaaagtattaatagattgggtgtnctgagagacttntgtttaagtaagatactttatataattgtaagtctttgatttntattgta

29f:

agttgatcggtcttgcagcgactttattattccacattatagaatttatttttttttatttagaaattatttatttttatttataaaactactattctagttattaaaagatggctgttagatgtgattatgaaggatgtaatgatatnggtg
tttttacacttctaacnaatacttattgtcttgtaggaataggtaattcacaaaatttcttcaggtaagtactttaaaatatttttataaaatataaaatttttttaacngtatttttgaggatgaacttggtgattctattcccgttatnc
atgcttcaatagcaggtncaaganttgttggaagaatggcagttggtnntagacatggtcttctagtaccanatgccacaac

29r:

anttgatcggtcttgcagcanataattgtttaccaagaatcataatgtgttctgatgtttncccaccaactttgtctgccaataatattatagtaaaataaagggatctaacaaatttatcccttcctgtatacgtagaaagtgtacgaacn
attctaggtaatgtaacttttttccatatttaaagtaaaaaanaataataaaaacacnaatgcacctttattaataaatgatatttngtnntggcgacatttatattattanatggtctgaataatcagaaccttgataacttgtagtgtact
atgtacattgtag

52-6f:

ctttctctatgtctctctcgctcccctcaataacatcttatgagtttaggtttcggtgcataaggcaatacccagggcataaatcatttccactgtcaccatagcccgcangtgaggtgagcaatccattagcttgcttctggaaacaagcc
tgtctgtatttgttccccccagcatctgcactcagcacactgactttgctcagtatatgtattggatggataagtgagtaaatgaatgaatgacataaagatcatagattgcctcagccttctattggaaaataacctaaagggaggggggaaa
attactacttaaaaccacaagtgccactcaagggatgcctcctatgaaaggaggantgctaactgccangtccccactggttcttcgtgctgcttatggcctggaagcctctcctgcagcca

52-6r:

tctctacntctctctcgctccctcttgttgccctctctcttactccctctttcaagttcccctctaagtccnctttcttangcaaagggantcagcgctctcnattcctgcgcctaaagttcccanactcaccccagantcancgttgcaca
ncccaancacatcnnggctcttggctgcnngataggccttccacnccatnancancacnaanaaccaatggggacctgncantttncactcctcctttcntaggangnatcccttgaatggcactngtggttntaantactaatttccccccc
tccctttaagttattngccnnnnaaaagctgaagcaatctatgatctttatgtc

CM-2f:

catcagcaaggtgatttttcaaattaatattctatttcaatccaaaaattaaacctaaaaaacgtgcaataaaaaaagagcctcaaatgaggccctttttttcagctcaatactaattatgcattagtttcagcaggaacttttggatagctt
acaccgcccatttgttcagcaatacgcaatacctggcagctataacccacttcattatcgtaccaaacataagcagttaagcggttgcctgaagtgatagtcgcttgcgcgtcaaatacacctgcagtacgtgaaccaataaagtcagaaga
tacaacttccagtcgagtttagtataaccaatttgaccttgaaggttagaattgatagagatttggcgaatgtttccatccctcctcccaatcaactccttgaccc

CM2-2r:

gcaaaaggtg gtgatgttaa cgcattcgtt gctgatgctg tgaaatctct tgaaggcagc gcaacttcag ttgaaggtaa agacgttgta ctttatggct tcggtcgtat cggtcgtatt cttgcacgtt tgatgat
cag ccagtcaggt ctaggccgtg gcttgaacct gaaacccccc nccttcnttt nttntctcac ngcccaccnc naatnntttn cntttntatn gcgtcgtgac tcaatccatg gtccattcgc gggtac
gatc tctgttnatg aaaaaaanaa acaatcattg cnaacggtcn gttcatccaa gtnatctang cntctaaccc ntccnanctg gantnacnct gcc

TJCA-2 (full seq):

caaatgtgaaatccccgagcttaacttgggaattgcattcgatactgggaagctagagtatgggagaggatggtagaattccaggtgtagcggtgaaatgcgtagagatctggaggaataccgatggcgaaggcagncatctggcctaatac
tgacgctgaggtacgaaagcatggggagcaaacaggattagataccctggtagtccatgccgtaaacgatgtctactagccgttggggcctttgaggctttagtggcgcantaacgcgatnagtagaccgcctggggagtacggncgcaatn
ctaaaactcaaatgaattgacgggggcccgcacaagcggtggagcatgtggtttaattcgatgcaacgcgaagaacctnacctggtcttgacatagtaagaatttcccagagatggattggtgccttcgggaacttgtcatacaggtgactg
catggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgccaacc

52-11f:

gatggatatctgcagaattcggctttctctatgtctctctcgctcccctcaataacatcttatgagtttaggtttcggtgcataaggcaatacccanggcataaatcatttccactgtcaccatagcccgcaagtgaggtgagcaatccatt
agcttgcttctggaaacaagcctgtctgtatttgttcccccagcatctgcactcagcanactgactttgactcagtatatgtattggatggataagtgagtanatnaatgaatgacataaagatcatagattgcctcagccttctattggaa
aataacctaaagggaggggaaaattactacttaaaaccacaagtgccactcaagggatgcctcctatgaaaggaggagtgctaactgccangtccccactggttcttcgtgctgcttatggcctgnaaagcctctcctgcngccaanaacc
ctgatgtgcttgggctgtgcaaccgtgaatctggggtgagtctgggaactttaagcgcangaatggagancgctgactccctttgcct

52-11r:

ctngttatcgntctcngcaccaactctgttgcccactcnnctannccctcttccnagtgnccntcnnattnctctcacttacgcaaagggagtcatgcncactatatnnctgccccnnaagtncccagattcaccccanatnctancgctgc
acaganccaaatcacctcntggctctgggtcgnaggtagangcnntccaacccattaacctcnccaataacnannggggncccggcnatttgcactccnctttcncaggaggtntctcntaagtggcactantggtntcnactantgttng
tcacccnccccttngatnatgtangtgtaaatgctcacgcnctccatgatgtgtatngtccccnctgtatnactcactttctccacccactacatgtaccgatcacttccngtgtntgtactgatcactcaggnggaaagaantacangata
ggattctacncacaatactcnctatnntactgntcttgtcccact

18s-P:

tgatccagctgcaggttcac

18s-A:

aaagattaagccatgcatg

43-:

gagcgttgagatcgcctccccgaccatgcgctc?gc?ccgac?gaaccggtgcgttgcgccgcctgtcaaccgtcgcgattgaggt?gatga

43+:

aagcaggatgccgaggacggcggggaccccgacgcatgaacagcccttcgatcggtttgaggcaaac?gcgttggtgacgg?agcggatggcgaacgcggcgaattgctggtgatttgg

XP1 (fullseq):

aaagtgaaataagaaataacatttataaatataaaaagcaacaatataatatcaaaagataaaaagcatattttaataaaataagtacagtttttcaaaaaagaaagacatagagtaaaagaaaattgataaagtagagaaaaaaaaaattg
ggcagtttatgtttaatctttatagaaaagaagtaaagaaaattataaaaatattaacttgaacatttcgaattggttttagatttcaagaacacagacaaatttatttttattaaaaatatgttactatctatacnattgaaagacacttat
gatgtttactattaaaaatagtttaaaaaaaggaggaacagaaataatttttaatgttcccacaatttaatacacttcaatatgttattataaacgaatttttttttataaatataaaaagttaatatttaatattaccttctttttaatgtt
taaattaaatttatttgagattaatatttttatttgggagataataccctttaaataaattgggaccttatttaaaagggaagtataaaaatccaatataaaataaaaatttaataaatgttaaaaagatttttctacggatatattttaatc
cactaatagatatatattatataatctactttcccagataattttacacaataatatgataaccatttaaaaaaattgctaaacttgagatgtggcaatgaataataattaatttaaataaataaaaaaaaattaccgtaaaaattaatat
ttaatagattaaaaaaaaataatattatttgaggtaaatttaaatgttaaattttgagaaaaatatatagtaaaataactggcaatatttaatagtgaaatagaaacccaaatgtatgattaacttctattcaagatatattggaatatcatt
cattttatgactaaatttactggtttttatcattcttcaaatttatgattcaataaacagttgattgtatttaccagatgggagtttccttattattatcgtagatgtgttgt

A: ggtcacttaacaaagtttagacgtttagtaatatgaaatttgaaaaattaagaatatccagatttcttagaaaaacttccaacattttatcataacattaatcgttacttttttcgtttttcccagtaaacaacaatcctgttgggaaaagt

D: gaaagtttttgatttatataacttgctgtttatttacatgatttttatatttaaaaaagtagaaaactatatattcctacattatttggtggtagacaaggtaattaattactacggactctaaatgaggtggtacaaccaccttctac
taa
tagtggtaaggtccgttattaaa

C: aataggaaattgaggtattattagttaaagtgaaagtttttgatttatataacttgctgtttatttacatgatttttatatttaaaaaagtagaaaactatatattcctacattatttggtggtagacaaggtaattaattactacgga

IA: cccttgaatcaaccaaacactgattgcatccagcctcttgcttcatcttcacataagcccataggcatcactggagcatcaagtgccaatcctaaccatttgggaacattctgtttcaggtgctatcagtcataggttccaatgctcaacat
gattacgaaattttgatcttgaatttaatgttcaggactgttttgatttttcataaagctgagtctaaaaagattgaacttattctattaaatcgagggtaacaccatttttcaagctgatagttaattcattaaatatcaattctgtttga
ctcaatctaatcaagtaagtatatacgtatag