Immunity and the Population Genetics of Malaria

Bruce A. Wedgwood-Oppenheim

Submitted for the degree of Doctor of Philosophy
The University of Edinburgh
1997

Table of Contents

Declaration

I hereby declare that this thesis has been composed by myself, and all work described within is my own, except where otherwise stated.

Acknowledgements

My unreserved thanks go to Drs. **Joel Peck** and **Ian Hastings** for their invaluable advice, support, friendship and patience. Thank you to Ian also for suggesting this work and to Joel for introducing me to the methods of population genetics theory. Without them, this work would not have been possible.

Thanks must also go to my supervisors Dr. **Richard Carter** and Prof. **David Walliker** for their support and faith in me.

I am very grateful to Dr. **Jane Carlton** for the daunting task of proof reading beyond the call of duty and to Dr. **Stuart Baird** for help in checking those final details.

I would also like to thank Ms. **Alice Cruickshank** and Mr. **Guillaume Barreau** for teaching me the intricacies of C and the joys of UNIX (not!); Prof. **John Haig** for statistical advice; the staff and students at both Edinburgh and Sussex Universities for their friendship and advice and making my studies in both places productive, thought provoking and enjoyable.

Finally I would like to thank my family, my flatmates (in Brighton and Edinburgh) and friends for their kindness and understanding through the difficult times and my ever increasing mood swings as this thesis progressed!

Abstract

There has been much recent interest in the effects of host immunity, and in particular 'strain-specific' immunity, on the structure of malaria populations. However, the number of studies incorporating both epidemiological and genetic consequences have been limited. In this study, a theoretical model of malaria populations is constructed in an attempt to rectify this situation. The model is constructed to reflect the important features of the human-mosquito-malaria system under the key assumption that strain-specific immunity occurs, and is controlled by alleles at a single locus (to be referred to as the **immuno-allelic** locus) in the parasite genome. The model is more realistic with regard to the nature of malaria than previous models of strain-structured malaria populations, and includes such features as a short period of immune memory and competition between parasites growing in hosts. The model is complex, and is thus designed to be examined through computer simulation.

Both the epidemiological and genetic effects of strain structure on malaria populations are studied (the former is also examined through an analytical model). For example, the effects of the number of alleles at the immuno-allelic locus on the proportion of hosts infected are examined. Also, the degree of genetic heterogeneity of malaria parasites inside hosts compared to the total level of genetic heterogeneity of the parasite population is examined. This is done by calculating the value of the statistic G_{ST} for a neutral locus in the parasite genome (referred to as $G_{ST(n)}$). The degree of genetic heterogeneity in a parasite population is of interest as it determines the degree to which sexual recombination will lead to the generation of novel genotypes. The generation of such novel genotypes is important as it may allow evasion of a new vaccine, or the development of resistance to anti-malarial drug.

In the simulations examined, $G_{ST(n)}$ is found to decrease (the degree of genetic heterogeneity increases) with increasing numbers of alleles at the immuno-allelic locus. It also decreases with increasing levels of transmission in the population. This latter result appears to agree qualitatively with the findings of recent population-genetic field studies.

In a recent paper, Hastings proposed that if the G_{ST} value calculated for the immuno-allelic locus (to be referred to as $G_{ST(d)}$) is compared to $G_{ST(n)}$, then the former will be found to be larger. This could be used as a

method for identifying immunogenic loci, which may in turn help in the search for vaccine candidates. However, this prediction came from an analytic model of malaria populations with many simplifying assumptions, some of which may be important to the predictions.

In the simulation study presented, the values of $G_{ST(n)}$ and $G_{ST(d)}$ are compared under a wide variety of conditions. It is shown that the qualitative predictions of Hastings's study holds for a wide variety of conditions, and for a more realistic set of assumptions as to the nature of malaria populations. Bootstrap analysis shows that this qualitative prediction can be examined in the field using practicable sample sizes.

While the research presented here largely confirms the conclusions of Hastings work, there are some profound differences between the two studies, and these differences are described and discussed. The importance of taking into account the epidemiology of a parasite population when examining its genetics is also highlighted. Finally, the importance of the effects of a strain structure on methods of control of malaria populations are discussed.

Parameters of the model

Symbol	Standard value	Description
D	10	Number of alleles at the Immuno-allelic locus
N	6,000	Number of hosts
γ	5×10^{-5} per day	Host Death Rate
$\mathbf{M_{min}}$	0.05	Immune Threshold
ρ	0.025 per day	Probability of immune response being mounted
β	2.74×10^{-3} per day	Probability of loss of immunity
$\mathbf{b_0}$	12.3 per 2 days	Parasite Growth Rate
$\mathbf{k_d}$	b_0	Density-dependant Death Rate
s	0.2 per day^2	Rate of Increase in Immune Killing
T	0.1 per day	Double Bite Rate
c	1.0	Coefficient of Vector Infectivity
Θ	2×10^{-5}	Density of parasites (if any) transmitted in a double bite.

Statistics collected

Symbol	Description
$G_{ST(d)}$	G_{ST} value for the immuno-allelic locus
$\bar{G}_{ST(d)}$	Average, with time, of $G_{ST(d)}$ during the stationary distribution
$\bar{\bar{G}}_{ST(d)}$	Average of $\bar{G}_{ST(d)}$ across more than one simulation run
$G_{ST(n)}$	G_{ST} value for the neutral locus
$\bar{G}_{ST(n)}$	Average, with time, of $G_{ST(n)}$ during the stationary distribution
$\bar{\bar{G}}_{ST(n)}$	Average of $\bar{G}_{ST(n)}$ across more than one simulation run
t	Ratio of $G_{ST(d)}/G_{ST(n)}$
$\bar{t}$	Average, with time, of t during the stationary distribution
$\bar{\bar{t}}$	Average of $\bar{t}$ across more than one simulation run.
P_{inf}	Proportion of hosts infected
$\bar{P}_{inf}$	Average, with time, of P_{inf} during the stationary distribution
$\bar{\bar{P}}_{inf}$	Average of $\bar{P}_{inf}$ across more than one simulation run
W	Mean level of host immunity
$\bar{W}$	Average, with time, of W during the stationary distribution
$\bar{\bar{W}}$	Average of $\bar{W}$ across more than one simulation run
P_V	Proportion of double bites, from infected hosts, that transmit parasites
$\bar{P}_V$	Average, with time, of P_V during the stationary distribution
$\bar{\bar{P}}_V$	Average of $\bar{P}_V$ across more than one simulation run
Z_{inf}	Mean number of zygotes per transmitting double bite
$\bar{Z}_{inf}$	Average, with time, of Z_{inf} during the stationary distribution
$\bar{\bar{Z}}_{inf}$	Average of $\bar{Z}_{inf}$ across more than one simulation run

1) Introduction

1.1) Malaria

Malaria describes a group of diseases found in humans and other vertebrates caused by parasitic protozoa of the genus *Plasmodium*. Human malaria is caused by four species of *Plasmodium*: *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale*.

Malaria in humans is found in much of the tropical and semi-tropical world with approximately 40% of the world's population at risk (World Health Organisation 1995b). This leads to an estimated 1.5-2.7 million deaths a year (World Health Organisation 1995a). Most clinical cases and almost all deaths due to malaria occur in tropical Africa. It is one of the three leading causes of morbidity and mortality in the third world and has been described as "the single outstanding tropical disease control priority" by the World Health Organisation (World Health Organisation 1995a).

1.2) Malaria species and life-histories

The genus *Plasmodium* comprises a group of parasitic protozoa which infect both mosquitoes and vertebrates. The parasites are transmitted from one vertebrate host to another via mosquito bites. The degree of host specificity of *Plasmodium* species varies. However, those infecting humans are largely specific to humans in nature. Human malaria parasites are transmitted exclusively by members of the mosquito genus *Anopheles* (Garnham 1988).

Plasmodium has a haploid genome for most of its life history, although it is transiently diploid in the mosquito (Beale and Walliker 1988).

Much of a human malaria infection consists of parasites dividing inside red blood cells. A single parasite invades a red blood cell and multiplies asexually to produce between 8 and 32 daughter parasites depending on the species (Garnham 1988). These parasites burst out from the infected blood cell and then invade other red blood cells. Frequently this invasion, multiplication, and release is synchronous in an infection, and the parasite release is then associated with many of the clinical symptoms of the disease (Harinasuta and Bunnag 1988).

Some asexual blood stage parasites develop into gametocytes. These are haploid cells which can be clarified by their morphology as male and female forms (Garnham 1988). Gametocytes are generally found circulating in the infected blood of a host at extremely low densities (Taylor and Read 1996). If a mosquito takes a blood meal which includes mature gametocytes, these then develop into gametes, which may fuse to form zygotes. This diploid stage of the parasite migrates across the wall of the mosquito gut and undergoes what appears to be a single meiosis followed by many mitoses (Garnham 1988; Sinden and Hartley 1985) to form an oocyst. The oocyst then ruptures to release mitotic, haploid cells known as sporozoites. These migrate to the mosquito salivary glands, from where they are injected into the next host when the mosquito takes a new blood meal.

The sporozoites, once injected into a host, migrate to the liver and invade the liver cells. They then divide asexually many times. The daughter cells are then released into the blood stream where they invade red blood cells thus, completing the life-cycle. For a more detailed description of the life-cycle see Garnham (1988).

The life-cycle of malaria is complicated. The salient features of the cycle relevant to this study are: (i) it is a vector-born parasite; (ii) it is haploid for most of its life history; (iii) it multiplies asexually in its vertebrate host and (iv) it undergoes obligate sex in the vector, where it undergoes a transient diploid stage. It should also be remembered that, as genetically identical parasites can produce both male and female gametes, fertilisation between gametes does not necessarily correlate with outcrossing.

1.3) Immunity and Strain Theory in Malaria

1.3.1) Definitions of immunity

Immunity to malaria can be divided into two broad categories: anti-disease immunity and anti-parasite immunity.

Anti-disease immunity is seen as protection of the host from the symptoms of the disease (e.g. fever, anaemia, death) whilst not interfering with the growth of the parasites in that host. One possible explanation for this is that the human immune system removes parasite products which induce the host disease symptoms (Riley *et al.* 1994).

Anti-parasite immunity is seen as an immune response that interferes directly with the course of the malaria parasites' life-cycle. There is experimental evidence that different stages of the parasite life history are recognised differently (Riley, Hviid and Theander 1994), and thus we can break down anti-parasite immunity into two functional sub-groups, asexual immunity and transmission-blocking immunity.

In the case of asexual immunity, the immune system interferes with the development of asexual parasites, and thus suppresses the asexual parasitaemia. In some cases, asexual parasites are quickly and completely

cleared from the host, or fail to develop in the host at all after the initial inoculation of parasites. This will be referred to as 'sterile immunity'. Although anti-parasite immunity does not directly reduce disease, it is likely that this will also occur as the severity of disease is likely, in part, to correlate with the numbers of asexual parasites present (Kwiatkowski 1991).

Transmission-blocking immunity occurs when the immune system disables the activities of the sexual, transmission stages of the parasite (e.g. gametocytes, gametes, or zygote) (Riley, Hviid and Theander 1994). If only transmission-blocking immunity occurs, then there is no reduction in disease, or asexual parasitaemia within that host. However, no transmission can occur from that host to another.

1.3.2) Evidence for strain-specific immunity

There is clear epidemiological evidence from hyper-endemic regions that people do develop both anti-parasite and anti-disease immunity to malaria. However, it appears to develop slowly. Very young children do not generally suffer from malaria. This is the case even when they are born of a malarious mother (McGregor 1986). In the very young, clinical impact increases with age, and in studies in The Gambia a peak in disease severity is found in children 1-2 years old (McGregor 1986). However, after this period, the amount of severe and mild disease due to malaria decreases markedly with age (McGregor 1986). This decrease is coupled to a somewhat slower decline in parasite densities and parasite prevalence in these older age groups (Gupta and Day 1994b). However, even in adults, complete, sterile immunity is uncommon. Adults in endemic areas

are likely to show occasional periods of parasitaemia throughout their lives, though often in the absence of any clinical symptoms (Riley *et al.* 1994).

The reasons for the initial absence of parasites in the very young child appear to be at least three fold. Firstly, during very early childhood, foetal haemoglobin persists in the red blood cells and this provides some protection against malaria. Secondly, maternal antibodies can cross the placenta into the foetal blood stream. Thus, in endemic areas, a child may be born with circulating antibodies protective against malaria (McGregor 1986). Thirdly, malaria parasites need pABA (p-amino benzoic acid) for their growth. Though this is found in humans of most ages through dietary intake, it is absent from breast milk and is thus not found in very young children. This may also contribute to protection (Marsh 1992).

The decline in both disease and parasite burden with age from early childhood is often interpreted as the slow acquisition of anti-disease and anti-parasite immunity with increased exposure to the parasites. There has been recent controversy as to this commonly held view, and there is some evidence that increasing age alone may, at least in part, contribute to protection against malaria (Baird 1994; Baird 1995; Gupta and Day 1994a; Roberts *et al.* 1994).

In a study in a hyperendemic region of Irian Jaya, Baird (1995) examined the development of immunity in non-immune migrant workers from Java. He found that very quickly (within two years) the migrant workers developed levels of immunity comparable to that found in the local population who had a very similar age profile. Thus, although immunity clearly requires exposure to malaria to develop, this may well occur quite quickly, with the level of immunity being age dependent. However, the lack of age structure in susceptibility to malaria found in low endemic areas, and

amongst people with no previous exposure to the disease has been used to argue against this (Gupta and Day 1994a). Clearly more work needs to be done in this field to separate the relative importance of age-specific and exposure-specific immunity.

A common explanation for the age profile of malaria immunity in endemic populations is that it is exposure-specific and that this is caused by strain-specific immunity. The argument goes that there are many different 'strains' of malaria parasite in the population. Each strain is seen by the immune system differently, such that previous exposure to one particular strain causes a protective immune response against future exposure to that particular strain, but either no response, or a reduced response to a different strain. Thus, for a maximal level of immunity, an individual must experience all the strains in the population (Day and Marsh 1991). This takes time, and as many intermediate levels of immunity are passed through as there are strains. This creates the immunity profile seen in hyper-endemic regions. If there is a significant degree of cross-reactivity between strains, one would still expect immunity to increase with age, although one would expect the rate of acquisition of immunity to be greater, than if there is no cross reactivity between strains.

The earliest supporting evidence for strain-specific immunity comes from the use of malaria therapy to treat neurosyphilis. It was found that repeated inoculation with the same (homologous) strain of malaria parasite (in most studies *Plasmodium vivax* was used, but studies with *P. falciparum* produced broadly similar results, e.g. Powell *et al.* 1972) would quickly produce signs of immunity in the host - reduced periods of fever and fewer parasites in the blood. However, this was frequently not the case when a heterologous strain was used. In many cases when a heterologous strain

was used for a second inoculation, the reduction in fever and parasite densities was much less than for the homologous challenge. However, in some cases this was not so, and the heterologous challenge showed a very similar profile to the homologous challenge (Jeffery 1966).

It was also found that, even with a single strain, the level of immunity to that strain increased over several repeated challenges (Ciuca *et al.* 1934). Thus even in the case of a single strain, exposure-specific immunity appears to occur.

1.3.3) Possible bases for strain-specific immunity

1.3.3.1) Antigenic polymorphism

The basis of strain-specific immunity is not known. However there is a large degree of genetic polymorphism amongst malaria parasites in the natural world (Kemp *et al.* 1990). It is easy to assume that some aspect of this polymorphism accounts for the different responses of the immune system to different invading parasites.

Genes encoding many antigens that are putatively involved in immune protection have now been cloned and sequenced. These have been isolated from the different stages of the malaria parasite that are found in the vertebrate host. There are two striking features of many of these genes (e.g. the genes for the merozoite surface protein-1, the merozoite surface protein-2, the circumsporozoite protein and the S-antigen). First, many of these genes are highly polymorphic, with a large degree of sequence diversity between different alleles. Second, the immuno-

dominant epitopes of the gene products are, in many cases, found to have a repetitive sequence (Kemp, *et al* 1990; Mendis *et al.* 1991; Schofield 1991).

There is likely to be strong selection for a high degree of sequence divergence between alleles at an immunologically important locus. A new variant that is not recognised by an immune system that has seen a different variant before is likely to survive in that host. Thus, new or rare variants are likely to be at a selective advantage in a population and increase in frequency The greatest advantage to a new variant occurs if its sequence is sufficiently distinct for no cross-reaction in immune response with a previous variant to occur. Thus, one would expect a gene involved in strain-specific immunity not only to be polymorphic but to have a large degree of sequence diversity between alleles. This diversity should be specifically in regions coding for immunodominant epitopes. Indeed, there is evidence from the analysis of nucleotide sequences (the examination of non-synonymous versus synonymous base changes) that several malaria surface antigens (e.g. CSP, MSP2, MSP1) are under positive diversifying selection (Hughes 1992; Hughes and Hughes 1995).

Repetitive, coding elements are found in many pathogenic organisms (e.g. Meyer *et al.* 1990; Schofield 1991). It has been postulated that they are associated with immune evasion. This is because repetitive DNA sequences have higher mutation rates than non-repetitive regions. This increases the probability that the pathogenic organism will generate a novel allele which can evade the immune system (Meyer *et al*1990).

The fact that repetitive elements have been found in many antigenic molecules of *Plasmodium* has been used to argue that they are involved in immune suppression. The suggestion is that these regions induce inappropriate and ineffective immune responses and prevent the host from

mounting more effective responses against other parasite epitopes (Anders *et al.* 1988; Schofield 1991). However, this view cannot explain the large degree of sequence diversity between alleles in many of these genes.

1.3.3.2) Antigenic variation

Not only are there several polymorphic antigens that may be involved in determining strain-specific immunity, there is also a family of genes known as the *var* genes (Baruch *et al.* 1995; Smith *et al.* 1995; Su *et al.* 1995) that may be involved. These genes appear to be involved in 'antigenic variation' in the malaria parasites.

Antigenic variation is a phenomenon by which a pathogen evades the host's immune system. This has been described in Trypanosomes (Borst 1991) and is also found in *Neisseria* (Borst 1991) and malaria (Brown and Brown 1965). It is a term used to describe the regular replacing of antigens visible to the immune system with different forms, which are not recognised by the antibodies that recognised previous forms. Thus, a parasite clone may evade a host immune response by changing its antigenic profile once the host starts to mount an effective immune response against the previous profile.

Antigenic variation of this type was first shown in malaria with *P. knowlesi*, a malaria species that infects monkeys (Brown and Brown 1965). It was shown that infection occurred in a series of recrudescences. Parasites from each recrudescence are not recognised by sera from the host collected before or during that recrudescence, but are recognised by sera collected from later recrudescences. These results suggest that the

parasite is indeed changing its antigenic profile to avoid the host immune system.

Recent work has described some of the genes involved in antigenic variation in *P. falciparum* (Smith *et al.* 1995; Su *et al.* 1995). These genes have been grouped together as the *var* genes. It has been estimated that there are at least 50 members of this family found in a single parasite genome, and perhaps as many as 150 (Su *et al.* 1995). The *var* genes, and their products PfEMP1 (Baruch *et al.* 1995), have excited particular interest, not only because of their importance in antigenic switching, but because they may also be important in defining the virulence of *P. falciparum.*

Many of the deaths caused by malaria are caused by a complication in *P. falciparum* known as cerebral malaria (White and Warrell 1988). This is believed to occur through the sequestration of infected erythrocytes into brain capillaries. This sequestration may well occur through the binding properties of PfEMP1 (Borst *et al.* 1995). Thus, expression of different members of the *var* gene family may well be effective in evasion of host immune responses, and also involved in determining the clinical course of an infection (Baruch *et al.* 1995; Borst *et al.* 1995).

1.3.4) The period of immune memory

The period of time that an individual maintains either an anti-disease or an anti-parasite immune memory against malaria, after clearing a malaria infection, is important for a clear understanding of the epidemiology of malaria and for an intelligent implementation of control strategies. Many epidemiological models of diseases assume immunity is sterile and life-

long and, for many disease agents (e.g. measles), this is a reasonable assumption (Anderson and May 1992). Some mathematical models of malaria populations have also made this assumption (e.g. Gupta and Day 1994b; Hastings 1996). However, there is reasonably strong epidemiological and experimental evidence against this.

Epidemiological evidence of an immune response against disease symptoms is conflicting, with reports both suggesting a short period of immune memory (Molineaux 1988), and at least one report suggesting that anti-disease immunity may last as long as 40 years (Deleron and Chougnet 1992). However, there is no evidence of anti-parasite immunity lasting for such a long time.

Malaria therapy studies have also shown some degree of disagreement in their results. In these studies, neurosyphilis patients were deliberately infected with malaria parasites. Boyd and Matthews showed some degree of both anti-parasite and anti-disease immunity lasting up to 7 years with *P. vivax* (Boyd and Matthews 1939). However, Jeffery found immunity to be much shorter lived (Jeffery 1966), and this difference has been suggested to be due to differences in the treatment regimes between the two studies (Jeffery 1966). In a study on *P. falciparum*, when human volunteers were infected with malaria parasites, sterile immunity has been shown to last less than one month, although anti-disease and a less strong anti-parasite response was shown to last for at least 3 years (Powell *et al.* 1972). Studies on transmission-blocking immunity in *P. vivax* have suggested that this form of immunity may last for only 3-4 months (Mendis and Carter 1992).

1.3.5) A comment on the term 'strain'

The term 'strain' has many connotations, and, though it was used in much of the past literature on malaria (e.g. Jeffery 1966), recently, it has often been replaced by more carefully defined terms such as 'line' or 'clone' (Kemp *et al* 1990).

The traditional view of the term "strain" is to describe a group of organisms within a species with a similar set of phenotypic characteristics, which are presumably caused by co-segregating alleles at genetic loci. This can occur in sexual species through geographic isolation (and genetic drift or local adaptation (Crow and Kimura 1970)) or through a lack of recombination between the loci (Hastings and Wedgwood-Oppenheim 1997). In parasite populations, the presence of co-segregating strains in a population is often referred to as a "clonal population structure" (Tibayrenc *et al.* 1990).

However, this definition of strains is not necessary to explain the observations of 'strain-specific immunity'. Here the only phenotypic characteristic of importance is the response of the immune system to the parasite. Thus the co-segregation of other characteristics, whether it occurs or not, is unimportant, and should not be implied by either the evidence for, nor the name 'strain-specific immunity'.

The antigenic profile of a parasite may be caused by differences in its genotype at either one or many loci and genetically inherited (this will be referred to as *haplotype-specific immunity*), or by differences in expression of genes and thus, epigenetically inherited (which will be referred to as *expression-specific immunity*).

Some recent theoretical studies on malaria have examined the effects of strain-specific immunity on the epidemiology of the parasite (e.g. Gupta and Day 1994b). An important feature of these studies has been the assumption that strains are transmitted independently. Strains can be seen as having independent values for R_0 (Gupta *et al.* 1994a), where R_0 is the *basic reproductive rate*, defined as the number of secondary infections one would expect to arise from a single primary infection, if surrounded by a very large number of fully susceptible potential hosts (Bailey 1982). Under all definitions of strain structure, the independent transmission of strains is unlikely to be the case. The only situation which would support such an assumption is as follows:-

a) when each strain is transmitted by a different mosquito population.

and

b) when growth of one strain in a host does not interfere with the growth or transmission of a second strain in the same host.

If either of these assumptions are not true for malaria populations, then the number of strains present in a population is likely to have a strong effect on the R_0 of each strain. Thus, there is clearly a need to examine the effects of strain structure on the epidemiology of malaria under the assumption that strains do interfere with each other's transmission.

1.4) Population Genetics of Malaria Parasites

1.4.1) Previous studies

The first practical population genetic studies of *P. falciparum* were done in the early 1970's (Carter and McGregor 1973; Carter and Voller 1975) and studies have continued up to the present day. These studies have, in general, had two foci: The degree of genetic diversity amongst parasite populations across the world, and the degree of genetic diversity in individual hosts.

Studies of the first type (e.g. Carter and Voller 1975; Creasey *et al.* 1990) are important for examining the degree to which *P. falciparum* populations are subdivided on a large geographic scale. This is important for understanding and predicting the spread of novel advantageous mutants across the world's malaria populations (e.g. the spread of drug resistance). It may also be useful for the study of the spread of an epidemic, specifically if it arose from a genetically distinct parasite population. A smaller geographic scale has also been examined for similar reasons, e.g. where the level of subdivision has been defined as coastal versus inland villages in the Madang region of Papua New Guinea (Paul *et al.* 1995).

The second group of studies has examined the degree to which genetically distinct parasite populations are found in either human hosts (e.g. Carter and McGregor 1973; Hill and Babiker 1995) or in the mosquito host (e.g. Babiker *et al.* 1994; Hill *et al.* 1995). There are several reasons for this type of study.

Firstly, the degree of genetic diversity inside individual human hosts determines the degree to which outcrossing can occur. *Outcrossing* is

defined here as the frequency at which zygotes are formed from genetically non-identical gametes. *Selfing* is defined here as the probability that a zygote is formed by the fusing of gametes that are genetically identical and the daughters of the same zygote. If only a single parasite genotype is present in an individual, then, since the parasite is haploid, only selfing can occur during a transmission event. Thus, no novel combinations of genes can be produced. However, if more than one genotype is present, then outcrossing can occur. Outcrossing, through meiotic recombination of non-identical haploid genotypes, can lead to the generation of novel genotypes. The degree to which novel genotypes can, or cannot be created is of great interest and importance, as novel genotypes have the potential to deal with changes in the parasite environment which the parental genotypes might not be able to survive (Charlesworth 1989; Crow 1992; Kondrashov 1993). Changes in the parasites' environment could be the presence of a novel drug treatment, or an anti-malaria vaccine.

Examining the degree of genetic diversity of oocysts found in mosquitoes (the oocyst is a more convenient stage to examine than the zygote but is likely to reflect the genetics of the zygote) allows one to measure directly the outcrossing rate. Direct measures of the outcrossing rate may be better than inferring it by examining asexual parasites in the human blood. It may well be that the frequencies of asexual parasites do not directly reflect the degree of outcrossing that occurs in the mosquito because the proportion and degree of infectivity of gametocytes of particular parasite genotypes varies. However, in the one case where blood stage and oocyst measurements have been compared they did appear to match up well (Hill and Babiker 1995).

Secondly, there has been debate in the literature as to whether malaria populations show a clonal population structure (i.e. whether the population can be divided into a population of genetically distinct co-segregating strains) (e.g. Dye 1991; Tibayrenc *et al.* 1990; Walliker 1991). The degree to which outcrossing in a population of parasites occurs will determine, at least in part, the degree to which a clonal population structure can exist (Hastings and Wedgwood-Oppenheim 1997). The degree to which clonality is present in a population is of great importance in the tracking of epidemics (Tibayrenc 1995), the diagnosis of different clinical forms of the disease, and the design of control strategies. However, it has been pointed out (Hastings and Wedgwood-Oppenheim 1997) that estimates of high levels of outcrossing from field data does not preclude the presence of co-segregating strains, for one has the potential to measure outcrossing *within* strains rather than *between* strains, depending on the choice of markers used.

Thirdly, it has been argued by Frank (Frank 1994; Frank 1996), that kin selection may be an important determinant for the level of virulence of a parasite population. If high virulence levels of a parasite correlate with a short-term transmission advantage, while lower virulence levels correlate with a longer-term transmission strategy, then it can be shown that if there are, on average, many different parasite clones in each infected individual, there would tend to be a greater level of virulence in that population than if there is, on average, only a few clones per individual.

Finally, studies of the genetic diversity of parasites per individual has lent extra weight to the concept of strain-specific immunity. A study by Ntoumi and co-workers (Ntoumi *et al.* 1995), showed a reduction in genetic diversity per individual with increasing age of the population. This fits with

the hypothesis that one is susceptible to fewer parasite types as age increases due to an increase in prior exposure.

1.4.2) A comment on terminology

The population genetics of malaria, though a small and recent field, has already generated a large number of terms to describe a very similar group of concepts or parameters. The symbols λ, μ, N_e, s, F, *f*, F_{IS} and G_{ST} have all been used, in one way or another, to describe or calculate the degree to which malaria parasites are restricted in their potential to outcross due to the subdivision of the parasite population into hosts.

The symbols λ and μ have both been used (Carter and McGregor 1973; Hill and Babiker 1995 respectively) to describe the mean number of genetically distinct clones of malaria parasite found in blood samples from infected human hosts. Both studies examined the number of alleles at polymorphic loci found in each host. Then, by measuring the allele frequencies across hosts and, assuming that genetically distinct clones have either a Poisson (Carter and McGregor 1973; Hill and Babiker 1995) or Negative Binomial Distribution (Hill and Babiker 1995) of the number of clones within hosts, it is possible to estimate the values of λ (or μ) for a particular population.

The number of genetically distinct clones found in infected hosts affects the potential for outcrossing in the population. The more clones present in a host, the greater the probability that any zygotes formed by a mosquito taking a blood meal from that host is formed by the fusion of two non-identical parasites. The more clones present in the donor host, the

greater the probability that a zygote will be heterozygous at a given polymorphic locus.

Consequently one can work back from the proportion of heterozygotes (in the zygotes) at a locus (and the allele frequencies at that locus) to produce a measure of the *effective* mean number of clones per infected host (N_e) (Hill *et al.* 1995). The values of N_e and λ (or μ) (i.e. the values obtained from oocysts and blood samples respectively) would be equivalent in value if the following assumptions are true for the population:

1) all parasite genotypes in a host are equally infectious to mosquitoes.

2) no disassortative or assortative mating occurs.

3) all parasite clones in a host are completely unrelated (as this is assumed in the method by Hill *et al* .(1995) for calculating N_e).

4) all infected hosts are equally infectious to mosquitoes*.

5) all infected hosts are infected by the same number of clones*.

The selfing rate, s, has been looked at by Read (Read *et al.* 1995; Read *et al.* 1992). Read used a theoretical model of malaria and ESS analysis (Maynard Smith 1982) to predict the selfing rate for a malaria population based on the sex ratio of gametocytes in that population. The model used by Read implicitly assumed that parasites in a host that are neither identical, nor daughters of the same zygote, are not related to each

*Assumptions 4 and 5 are interactive, i.e. if all hosts are not equally infectious, but the greater the number of clones present, the greater the infectivity, then it is not necessary for all hosts to have equal numbers of clones, and the exact distribution will depend on the degree of infectivity of each combination.

other in anyway. Under this assumption, s can be seen as the reciprocal of N_e.

Dye and Godfray (1993) extended the model of Read to take into account the possibility that parasites that are not identical sisters could be related. In this model, it was shown that Read's predicted values for s were in fact predicted values for F, the Inbreeding Coefficient.

The Inbreeding Coefficient is referred to in the malaria literature by the symbols F and *f* (Dye and Godfray 1993; Hill *et al.* 1995). It is defined as the probability that the two alleles at a locus in a diploid organism are *identical by descent.* Traditionally F is the calculated value and is seen as an estimate of *f*, the true value (Hartl and Clark 1989). It is normally calculated by counting the excess of homozygotes at a polymorphic locus and comparing this to the predicted value under conditions of random mating (Hartl and Clark 1989). If non-identical parasites are assumed never to be related (assumption 4 above) then the Inbreeding Coefficient is identical to the Selfing Rate. This is likely to be a reasonable assumption if the population of human hosts is large, and only a single haplotype is transmitted by an infected mosquito biting a human host.

The reasons for inbreeding can be two fold. It can be caused by (i) assortative mating (like preferentially mates with like) and (ii) population subdivision. In malaria, the population of parasites is subdivided into hosts, and mating, in almost all cases*, can only occur in mosquitoes between parasites from the same host. When only a few genetically distinct clones are found in a host, the chances of selfing (or mating with a closely related parasite) are increased which thus increases the degree of inbreeding.

* The only biologically plausible exception to this is interrupted feeding of a mosquito leading to a blood meal from more than one host.

Wright developed F statistics to deal with these two different aspects of inbreeding and produced a hierarchy of Inbreeding Coefficients: F_{IS}, F_{ST} and F_{IT}. F_{IS} is normally used to describe the coefficient of inbreeding caused by assortative or disassortative (collectively to be referred to as 'non-random') mating. F_{ST} is used to describe the coefficient of inbreeding for population subdivision and F_{IT} is used to describe the total Inbreeding Coefficient (and is thus equivalent to F and *f* above). For a full discussion of F statistics see Hartl and Clark (1989) chapter six.

Unfortunately, F_{IS} as used in malaria has had a different usage from that above, caused by using two different levels of subdivision. In a paper by Paul (Paul *et al.* 1995), a population of malaria parasites was considered at two different levels, that of coastal vs. inland villages and also between hosts in a group of villages. Thus, F_{IS} was used in this case to denote the degree of inbreeding found in zygotes caused by both population substructuring in hosts and any inbreeding effects due to non-random mating.

In a recent paper by Hastings (1996), the statistic G_{ST} was used to describe the degree of inbreeding due to the population subdivision of parasites into discrete hosts (the statistic G_{ST}, by convention, replaces F_{ST} when more than 2 alleles are found at the locus examined). Thus the use of G_{ST} by Hastings is equivalent to Paul's use of F_{IS} under the assumption that no non-random mating occurs. Under this assumption, and treating a population of individuals as the total population of hosts, G_{ST} (and F_{IS}) are equivalent to F and *f*.

The use of G_{ST} by Hastings is in a theoretical model of malaria. Field studies are unlikely to be able to separate the effects of non-random mating from the effects of population subdivision into hosts, thus field studies are

likely to measure F_{IT} (or G_{IT}) rather than G_{ST}. However, experimental studies by Ranford-Cartwright (Ranford-Cartwright *et al.* 1993) suggest that random mating occurs between parasites in a mosquito blood meal, and thus non-random mating seems unlikely. This is also suggested by the fact that when N_e has been calculated from both oocyst and blood data from the same population, the results are very similar (Hill and Babiker 1995).

In the study presented here, G_{ST} values are examined. A detailed description of the calculation of G_{ST} is described in Section 2.2. Here, it is enough to say that G_{ST} values can be between zero and one. When a value of zero is found, there is no inbreeding (reduction in heterozygosity) due to population subdivision. When a G_{ST} value of one is found there is complete inbreeding due to population subdivision. In this case, at the diploid stage, no heterozygosity at a polymorphic locus is expected to be found.

1.4.3) A study by Hastings

A demonstration of the potential usefulness of studying population genetics theory in the context of malaria is embodied in a recent study by Hastings (1996). In this study it was assumed that haplotype-specific immunity occurred and was defined by a single locus in the parasite genome. It was assumed that alleles at this locus elicited independent immune responses with no cross-reactivity. Hastings coined the term **'immuno-allelic'** to describe a locus behaving in this manner. Using his analytic model Hastings examined (under equilibrium conditions) the G_{ST} value for this immuno-allelic locus, and a neutral locus, with the subpopulations being defined as parasites within hosts. He showed that, under the assumptions of the model, the G_{ST} values at the immuno-allelic locus always had a larger value (the population appeared to be more substructured) than the value at the neutral locus for all parameter values examined. This qualitative result was found to be consistent irrespective of the degree of physical linkage between the two loci, provided the loci are in genetic linkage equilibrium.

There is an intuitive explanation for Hastings's results. If a mixture of parasite genotypes infects a host which is immune to some, but not all, alleles at the immuno-allelic locus, then the diversity of alleles at that locus, in that host, will be restricted. Across many hosts this immune restriction will appear as a high level of genetic substructuring at that locus. No such restriction occurs at a neutral locus, and thus a lower degree of substructuring is expected.

This result is important as it provides a potential method for identifying immunologically significant loci through population genetic

methods. Thus the importance of a locus's polymorphism for evasion of the host's immune responses can be examined. This may help in the finding of candidate vaccine genes and designing the subsequent vaccines.

Hastings's study also looked at drug resistance genes in experimental (e.g. murine) populations and it was shown that the G_{ST} values at a drug resistant locus (in a situation where both drug resistant and non-drug resistant parasites are present) would be higher than for a neutral locus assuming a drug treatment regime is involved. This could then be used to test whether a putative drug resistance locus is indeed involved in resistance.

Hastings's model is analytic and, as with all such analyses, some degree of biological realism must be lost for the sake of mathematical tractability. Thus, it may be important to examine analytic results through computer simulation studies, where more realistic assumptions can be employed.

An important assumption of Hastings is that sterile strain-specific immunity is life-long. This has been an assumption of other models (e.g. Gupta and Day 1994b). However, the evidence for any form of anti-parasite immunity generally supports the idea that it is short lived (see earlier in this chapter). The length of immune memory is likely to have large effects on the degree of immune restriction of the immuno-allelic alleles, and thus may have important effects on the ratio of G_{ST} at the two loci.

There are other simplifications in the model, which may have effects on its outcome of the model, e.g. the fact that parasites are assumed to grow in non overlapping generations and that the population of hosts is assumed to be infinite, but the effects of removing these simplifications are less clear.

The model used by Hastings has a defined number of parasite zygotes contributing to the infections in a host. However, several of the parameters used in the model are likely to have interactive effects on this number (e.g. the number of immuno-allelic alleles present in the population), but this is not examined. Maintaining such a fixed epidemiology is a useful way of examining the effects of various assumed parameter values on the values of G_{ST} for a particular malarious region, where the epidemiology is reasonably well known. Thus this is a sensible technique for seeing whether the examination of G_{ST} is likely to yield useful results in any particular field study. It is not, however, a useful technique for examining the effects of altering a particular malaria parasite population. Thus it cannot be used as a predictive study on the effect of, for instance, reducing the number of alleles present at an immuno-allelic locus (as might happen with the introduction of a vaccine), on the values of G_{ST} at the two loci. This is because there are likely to be both direct, and indirect (through changes in the populations epidemiology) effects of the number of alleles on the parasite population genetics.

1.6) ***The aim of this study.***

The aim of the study presented here is to create a model of malaria parasite populations to examine the population genetics of the parasites in relation to strain-specific immunity. It is intended to be more realistic in terms of the biology of malaria parasites than the model of Hastings (e.g. including a short period of immune memory), and to allow for the genetics to be examined in an epidemiological framework.

More specifically, the model is intended to examine the degree of population substructuring that occurs under various epidemiological conditions, and with different assumptions as regards the parasite's biology. It is also intended to examine the robustness of Hastings's predictions with more realistic assumptions as to the nature of malaria parasites. It is designed to examine the parasite genetics in a dynamic epidemiological framework, which might allow predictions as to the effects of different intervention strategies on both the epidemiology and genetics of the parasite populations.

This model is also intended to allow some understanding of the effects of strain-specific immunity, under a series of biologically plausible assumptions, on the interactions between strains that co-habit the same population of hosts. A simplistic, analytic model is also designed. This is intended to allow for the direct comparison between inter-strain and no inter-strain competition on the epidemiology of a malaria population.

2) Model and Methods

2.1) The Model

In this section, a mathematical model is presented that is intended to simulate the key dynamic features of the malaria-mosquito-human system. Like malaria, the hypothetical parasite is assumed to infect two animals. One of these animals, the *vector*, reflects the role of mosquitoes in the system. The other, the *host*, reflects the role of humans.

Host-parasite dynamics in the malaria system occur in continuous time. That is to say, there is no natural way to divide the life cycles into discrete parts. Nevertheless, for purposes of simulation, discrete time-steps are used. Clearly, this can introduce biologically unrealistic dynamic outcomes. However, it should be possible to eliminate such a possibility by choosing time steps to be sufficiently small. For most of the simulations presented here, 0.2 day is used as the time step, and this should be taken to be the time step unless otherwise stated. Thus, the parasite growth in hosts etc. are updated every 4.8 hours. It seems intuitively plausible that this constitutes a reasonable model of continuous time. However, smaller time steps are also used for some of the simulations to ensure that 0.2 day is a reasonable value, and data from these are presented in Appendix 1.

2.1.1) Description of parasite genetics

Like malaria, the hypothetical parasites are assumed to be haploid when in the host. They are assumed to be genetically identical except at two loci where there may be more than one allele present in the population. One of these loci is **immuno-allelic**, with a maximum of D alleles ($D \geq 1$). The other locus is selectively neutral, and has two alleles.

Alleles at the immuno-allelic locus will be designated i, where i = 1,2,3,4...D. Alleles at the neutral locus will be designated j, where j = 1,2. If there are N hosts in the simulation, then hosts are designated k, where k = 1,2,3,4...N.

It is assumed that parasite numbers are limited by both resource competition and by aspects of the host immune response. Under the assumption of my parasite-growth model (to be described presently), it can be shown that if a single parasite genotype is present in a host and no immune response to that genotype is present, parasite numbers will eventually approximate K, where K is a very large integer. K is equivalent to the concept of the "carrying capacity" in ecological literature (Begon *et al.* 1990). The value of K is identical for all parasite genotypes. It is convenient to represent the number of individual parasites within a host as a proportion of K. This proportion is referred to as the parasite *density*. Let $V_{i,j,k}$ represent the density of parasites with genotype i,j in host k. Thus, the density of all parasites with allele i at the immuno-allelic locus in host k is referred to as V_{ik} where $V_{ik} = V_{i1k} + V_{i2k}$.

2.1.2) Description of hosts

The host population is assumed to be finite with a fixed size *N*. The host death rate per time-step (γ) is constant and independent of any parasites present (where $0<\gamma<1$). All deaths are immediately replaced by births (in the same time step) and thus the population size remains unchanging. New-born hosts are uninfected and susceptible to all parasite types.

If the density of parasites with a particular immuno-allelic allele is above a certain critical density (designated M_{min}) in a host, then there is a probability (ρ) per time-step that an immune response will be mounted. Otherwise the host will remain susceptible to parasites with that allele. The immune response is specific to allele i, which means that if an immune response is mounted against allele i, it has no effect on parasites containing immuno-allelic allele g (where g = 1,2.3..D and $g \neq i$). Loss of immunity to a particular immuno-allelic allele i does not occur so long as parasites with the immuno-allelic allele i are present in the host. Once all parasites of type i have been removed from a host that is immune to type i parasites ($V_{i1k} + V_{i2k} = 0$), immunity is lost with probability β per time-step (where $0<\beta<1$).

Parasites of type i,j,k are assumed to be lost from host k through stochastic processes when $V_{ijk} < 10^{-9}$. When this occurs V_{ijk} is caused to equal zero.

2.1.3) Parasite population dynamics

In this section the dynamics that determine the growth of the population of parasites with the i^{th} allele at the immuno-allelic locus, and the j^{th} allele at the neutral locus in host number k will be described. These parasites will be called, simply, parasites of type i,j,k.

The growth of parasites of this type is described by:

$$\Delta V_{ijk} = V_{ijk}[b_0 - st^{\bullet}_{ik} - k_d V_k] + U_{ijk} \qquad \textbf{(2.1.3.1)}$$

Where ΔV_{ijk} is the change in density of parasites with genotype i,j in host k in a single time step. All terms inside the square brackets determine the per-parasite growth rate per time-step, and this is multiplied by the parasite density (V_{ijk}). The excess of parasite births above parasite deaths in the absence of all other factors can be seen as the maximum parasite growth rate, and is described by b_0. All other terms inside the square brackets are aspects of the death rate and cause a reduction in the per parasite growth rate below b_0.

The term ($st^{\bullet}_{ik}$) is the immunity-induced death rate, where s is the rate of increase in immune killing of parasites and represents the intensity of the immune response (where 0<s<1). The value of $t^{\bullet}_{ik}$ represents the period of time since host k became immune to allele i (in the case where no immune response is mounted $t^{\bullet}_{ik} = 0$).

The next term ($k_d V_k$) in the equation represents the density-dependent parasite death rate. V_k iş the total density of parasites in host k.

$$V_k = \sum^{i=D}\sum^{j=2} V_{ijk} \qquad \textbf{(2.1.3.2)}$$

The term k_d denotes the parasite density-dependant death rate and its value determines the degree to which the total density of parasites (V_k) affects the death rate of parasites of type i,j. It is assumed that $k_d>0$, and thus, as the parasite density increases, the per parasite death rate also increases. Furthermore, k_d is defined as having a value equal to b_0. This has the effect that when V_k is equal to one, if there is no immune response mounted, the parasite growth rate is equal to zero. Thus, the density of parasites have a value of one when at the carrying capacity (K).

The last term in the equation is U_{ijk}. This represents the introduction of parasites of genotype i,j into host k by transmission events (i.e., from bites from the vector). The manner in which U_{ijk} is determined is described in the next section.

2.1.4) Transmission dynamics

Let us now turn to the transmission dynamics of the system, and examine the factors contributing to the term U_{ijk} in the growth equation above. U_{ijk} represents the increase during a given time step in the density of parasites of genotype i,j in host k as a result of the inoculation of parasites by the vector. The value of U_{ijk} is expressed as a fraction of K, the carrying capacity for parasites within a host.

The parasites that are transmitted to host k via the vector can come from any host in the population. Let the contribution from host h be denoted by Ψ_{ijh} (where h=1,2,3..N). Thus, we have:

$$U_{ijk} = \sum^{h=N} \Psi_{ijh} \qquad \textbf{(2.1.4.1)}$$

For a vector to transmit parasites from host h to host k, it is assumed that a vector must first bite host h and then bite host k within a single time step. It is assumed that the number of such "double bites", from host h to host k, during a given time step follows a Poisson distribution with parameter T/N per day, where T is known as the Double Bite Rate. Let B designate the number of double bites during a particular time step, *from* a particular host h, who is the potential source of parasites, and *to* a particular host k, who is the potential recipient. Furthermore, let b designate a particular double bite where b=1,2...B.

After a vector bites host h, and before it bites host k, sex occurs within the vector. That is, certain parasites from host h will fuse in pairs to produce diploid zygotes in the vector, which then undergo meiosis, followed by mitoses to produce many haploid progeny. The number of zygotes formed

within the vector in double bite b is represented by Z_b, and Z_b is assumed to follow a Poisson distribution with parameter cV_h, where c, the "coefficient of vector infectivity", is a positive constant that reflects the typical dosage, and infectivity of parasites transmitted to the vector in a bite. Thus, the expected number of zygotes formed is an increasing function of V_h, the density of parasites in host h. Note that $Z_b=0$ is possible, and this is the case where no zygotes are transmitted. It is assumed that, before sex, the distribution of parasite genotypes within the vector is identical to the distribution in host h. It is further assumed that, when a zygote is formed, haploid parasites are chosen at random from the population within the vector. Thus, if we pick at random one of the haplotypes in one of the zygotes within a vector, then the probability that this haplotype has genotype i,j is given by $\frac{V_{ijh}}{V_h}$. This is equal to the proportion of parasites in host h that have genotype i, j, and shall be referred to as P_{ijh}.

It is assumed that, after zygote formation, all haploid parasites remaining within the vector die. The next step is formation of new haploid parasites within the vector. This occurs by means of recombination during meiosis between the two haplotypes that make up the zygote. Each zygote produces only four meiotic products. In a system with 2 loci, such as this one, there are four possible haplotypes of these meiotic products. There are the two *parental* haplotypes with genotypes identical to the two parasites that fused to form the zygote, and there are the two *recombinant* haplotypes where the alleles at each locus are inherited from a different haploid parent. If the parental haplotypes share common alleles, then the parental and recombinant haplotypes may appear the same.

It is assumed that the 2 loci in our model are unlinked and that therefore, on average, these four haplotypes occur at equal frequencies.

However the four possible progeny haplotypes need not be found in the product of a single meiosis. The products of single meioses have been extensively studied in fungal genetics, and this has shown that such meiotic products, when examined at two loci may appear in one of three groups. In the first group, all four progeny types are present. In the second group, only the parental types are present with two progeny of each type and in the third group, only the recombinant types are present with two progeny of each type. The relative frequency of these three groups can be calculated. If it is assumed that our loci are unlinked with respect to each other, and are unlinked with respect to their chromosomal centromeres, and that recombination occurs without interference, then they are found to occur with frequencies of 2/3, 1/6, and 1/6 respectively. (For a detailed explanation of the products of meiosis see Appendix 2).

It is also assumed that the haploid progeny of a zygote multiply mitotically to produce a very large numbers of haploid parasites. It is further assumed that each meiotic product of any zygote produces the same number of haploid parasites within a vector. The relative frequencies of the parasites transmitted to host k are assumed to be equal to the frequencies of the haploid progeny genotypes after meiosis. If $Z_b=0$ (which means no zygotes formed in the vector), then it is assumed that no parasites are transmitted from host h to host k. If $Z_b>0$ (so that zygotes are present in the vector), then it is assumed that the number of parasites transmitted to host k takes a fixed value irrespective of the number of zygotes that have been formed, and this value is the same for all possible values of h and k. The value is given by θ (the inoculum density), which represents the number of parasites transmitted to host k as a fraction of K, the carrying capacity, for parasites living within hosts.

Using the foregoing assumptions, one can now write an expression for Ψ_{ijh}. It is:

$$\Psi_{ijh} = \sum_{b=1}^{B} \sum_{z=1}^{Z_B} \frac{\theta}{Z_B} \left(\delta_4 + \frac{\delta_2}{2} + \frac{\delta_1}{4} \right) \qquad \textbf{(2.1.4.2)}$$

where $\delta[4]$, $\delta[2]$, and $\delta[1]$ are all functions that can have a value of one or zero, and signify whether a zygote produces all, two or one meiotic product with genotype i, j respectively.

In cases where all four meiotic products have the same genotype (in this case i,j,.), the zygote would have to be homozygous at both loci (caused by the fusion between gametes with identical genotypes). Thus, the probability of $\delta[4]$ having a value of 1 ($P_{[\delta[4]]}$) is:

$$P_{[\delta[4]]} = P_{ijh}^{\,2} \qquad \textbf{(2.1.4.3)}$$

There are two ways in which only one of the four meiotic products has the genotype i,j. This can occur when the zygote is heterozygous at both loci, with one copy of allele i at the immuno-allelic locus, and one copy of allele j at the neutral locus. It can also arise when genotype i,j fuses with genotype a,b,(where a is an allele at the immuno-allelic locus and $a \neq i$, and b is an allele at the immuno-allelic locus where $b \neq j$), or when genotype a,j, fuses with genotype i,b. As has been stated above and in Appendix 2, the four meiotic products of such a zygote may contain zero, one or two products with the genotype i,j, depending on whether the parental ditype, non parental ditype or the tetratype is produced. As is explained in Appendix 2, the probabilities of these three possibilities occurring in a particular meiosis are respectively 1/6, 1/6 and 2/3

Thus the probability that $\delta[1]$ equals one ($P_{\delta[1]}$) is:

$$P_{\delta[1]} = \frac{2}{3}\left(\left(\sum_{b\neq j}^{b=2} P_{ibh} \sum_{a\neq i}^{a=D} P_{ajh}\right) + P_{ijh}\left(\sum_{b\neq j}^{b=2}\sum_{a\neq i}^{a=D} P_{abh}\right)\right) \qquad \textbf{(2.1.4.4)}$$

The probability of $\delta[2]$ being equal to one is more complex, as it can occur due to a non-parental ditype being formed from the double heterozygote or due to the production of a zygote homozygous for either i or j and heterozygous for the other. Thus:

$$P_{\delta[2]} = \frac{1}{6}\left(\left(\sum_{b\neq j}^{b=2} P_{ibh} \sum_{a\neq i}^{a=D} P_{ajh}\right) + P_{ijh}\left(\sum_{b\neq j}^{b=2}\sum_{a\neq i}^{a=D} P_{abh}\right)\right) + P_{ijh}\sum_{a\neq i}^{a=D} P_{ajh} + P_{ijh}\sum_{b\neq j}^{b=2} P_{ibh} \qquad \textbf{(2.1.4.5)}$$

Note that the sum of $P_{\delta[2]}$, $P_{\delta[1]}$ and $P_{\delta[4]}$ is the probability that the progeny of a zygote from host h contains parasites with genotype i,j. This is likely to be less than one and in many cases can be zero, depending on the frequencies of the different parasite genotypes in host h.

2.2) Statistics Collected

2.2.1) $G_{ST(n)}$, $G_{ST(d)}$ and t

The main statistics collected are $G_{ST(n)}$, $G_{ST(d)}$ and *t*. This section describes how each of these statistics is calculated. $G_{ST(n)}$, and $G_{ST(d)}$ are measures of the degree of genetic heterogeneity at the neutral and immuno-allelic locus, respectively. The value of *t* gives the ratio $G_{ST(d)}/G_{ST(n)}$. In the next section, the other statistics collected from the model will be discussed.

As stated in the introduction the aim of this work is to examine the degree of genetic heterogeneity of the simulated parasite populations in their hosts using Wright's F statistics. F_{ST} is the statistic normally used to examine the genetic structure of a subdivided population. However F_{ST} is normally used to describe a two allele system, and G_{ST} to describe loci where more than two alleles occur (Hartl and Clark 1989).

At the immuno-allelic locus, in many situations more than 2 alleles will be used, and thus G_{ST} is the more appropriate term. At the neutral locus, only two alleles are used and F_{ST} should thus be used. The two terms are, however, equivalent where 2 alleles are involved, and thus to prevent an expansion of terminology, I follow Hastings (1996) in using G_{ST} to describe both loci.

In diploid systems, the value of G_{ST} for a locus is normally calculated by measuring the mean degree of heterozygosity ($\overline{H}$) at a locus in a subdivided population, i.e. the proportion of diploids that are heterozygous at that locus across all the sub-populations, and the allele frequencies in the total population. From the allele frequencies an expected

heterozygosity value (H_{exp}) can be produced assuming that the population is large and panmictic. The value of G_{ST} is then given by:

$$G_{ST} = \frac{H_{exp} - \overline{H}}{H_{exp}} \qquad \textbf{(2.2.1.1)}$$

The G_{ST} value at the immuno-allelic locus is be described by $\mathbf{G_{ST(d)}}$ and at the neutral locus by $\mathbf{G_{ST(n)}}$.

As stated above, G_{ST} is normally calculated from the heterozygosities at the diploid stage. However, in this model the parasite is haploid for most of its life cycle. Depending on the rate of transmission between hosts, the numbers of diploid stages at any particular time step are expected to be quite small, and thus there is the potential for sampling to cause wide fluctuations in the G_{ST} value calculated.

The diploid stages do, however, reflect the haploid genotype frequencies in each host, albeit weighted by the total parasite density in each host (as the parasite densities in a host correlate with its infectivity). The diploid stage can thus be seen as a sample of the potential diploid stages generated from the parasites in those hosts. Therefore, one can use the allele frequencies in each host, to generate *predicted* heterozygosity values assuming that a large number of vectors bite each host. The predicted homozygosity for a locus for a given host, can be produced by summing the square of the allele frequencies at that locus in that host. The predicted heterozygosity at that locus, in that host, is made by subtracting the homozygosity at the locus from one. Thus the heterozygosity in host k for the immuno-allelic locus ($H_{(d)k}$) can be calculated:

$$H_{(d)k} = 1 - \sum^{i=D} (P_{i1k} + P_{i2k})^2 \qquad \textbf{(2.2.1.2)}$$

The heterozygosity at the neutral locus ($H_{(n)k}$) can be calculated in a similar manner.

To produce a mean heterozygosity across all hosts for a locus, the heterozygosity for each host (H_k) is weighted by multiplying it by the predicted mean number of zygotes each host will produce per time step. Thus the mean heterozygosity for the immune locus can be described:

$$\overline{H}_{(d)} = \frac{\sum^{k=N} H_{(d)k}\, cV_k}{\sum^{k=N} cVk} \qquad \textbf{(2.2.1.3)}$$

Expected heterozygosities can also be calculated assuming that the population is not subdivided; this is done by calculating the total allele frequencies for the population. These total allele frequencies must also be weighted by the infectiousness of each host to mosquitoes. Thus the weighted allele frequency for immuno-allelic allele i ($P_{(w)i}$) is:

$$P_{(w)i} = \frac{\sum^{k=N} cVk \sum^{j=2} P_{ijk}}{\sum^{k=N} cVk} \qquad \textbf{(2.2.1.4)}$$

A similar expression can be derived for the weighted allele frequency of the neutral allele j ($P_{(w)j}$).

The expected heterozygosity for the immuno-allelic locus ($H_{exp(d)}$) is:

$$H_{\exp(d)} = 1 - \sum^{i=D} {P_{(w)i}}^2 \qquad \textbf{(2.2.1.5)}$$

and for the neutral locus:

$$H_{\exp(n)} = 1 - \sum^{j=2} {P_{(w)j}}^2 \qquad \textbf{(2.2.1.6)}$$

G_{ST} can be calculated for each locus by comparing the mean heterozygosity of the locus in each infectious host with the expected heterozygosity if the parasite population is not subdivided.

This is calculated separately for each locus:

$$G_{ST(d)} = \frac{H_{\exp(d)} - \overline{H}_{(d)}}{H_{\exp(d)}} \qquad G_{ST(n)} = \frac{H_{\exp(n)} - \overline{H}_{(n)}}{H_{\exp(n)}} \qquad \textbf{(2.2.1.7)}$$

In the work to be described in this thesis, an important objective is to examine explicitly the relative magnitude of $G_{ST(d)}$ and $G_{ST(n)}$. The statistic ***t*** is used to describe the ratio $G_{ST(d)}/G_{ST(n)}$. Thus if t is greater than one then, the value of $G_{ST(d)}$ is larger than that of $G_{ST(n)}$. This would imply that the apparent level of genetic substructuring (or inbreeding) of the parasite population is greater when an immuno-allelic locus is examined compared to when a neutral locus is studied

The G_{ST} values for both loci change with time. However, their values are of most interest once a stationary distribution has been reached, as once this has occurred, the values should be independent of the starting conditions. Using the Standard Parameter Values, an apparent stationary distribution arises within the first 1500 days (see Chapter 3). It is thus believed to be prudent to start observing the distribution after 4000 days to ensure that a stationary distribution truly has been reached.

Once a stationary distribution has been produced, there is still a lot of variation with time in the G_{ST} values for the population, especially at the neutral locus (see Chapter 3). Thus it seemed sensible to obtain a representative G_{ST} value for each locus by measuring G_{ST} at two-day intervals after day 4000 for 1000 days and taking the mean of the observations. This procedure appears to be sufficient in light of the small

standard errors found when the trials were repeated. The representative value for the statistics $G_{ST(d)}$, $G_{ST(n)}$ and t collected in this manner is described by $\overline{G}_{ST(d)}$, $\overline{G}_{ST(n)}$, $\bar{t}$, respectively.

Each set of parameter values examined with the model was repeated with two different starting conditions. The two sets of starting conditions were:

Initial Condition A) 20% of the host population initially infected, each host equally infected with each possible genotype such that V_k for each infected host = 1.0. In this situation the starting G_{ST} values for both loci = 0.0.

Initial Condition B) 20% of the host population initially infected, each host infected with only 1 parasite genotype, but equal numbers of hosts infected with each possible genotype. In this situation the starting G_{ST} values for both loci = 1.0.

Due to the constraints on computer time, most simulations with each set of parameter values were repeated only once, so that there was one run with each set of starting conditions. The mean values for $\overline{G}_{ST(d)}$, $\overline{G}_{ST(n)}$ and $\bar{t}$ and their standard errors were then calculated with each pair of runs. The mean values for these parameters, when taken across more than one simulation, are referred to by $\overline{\overline{G}}_{ST(d)}$, $\overline{\overline{G}}_{ST(n)}$ and $\overline{\bar{t}}$.

It is possible that, with certain parameter values, a stationary distribution is not reached within the first 4000 days. However, in the studies presented here, the standard errors for $\overline{\overline{G}}_{ST(d)}$, $\overline{\overline{G}}_{ST(n)}$, and $\overline{\bar{t}}$ are very small in comparison to their respective mean values. Thus it seems that either a stationary distribution has been reached, or is very close to being

reached, by day 4000, for each statistic, with each set of parameter values examined. A similar finding is made when examining the mean values after day 4000 of the other statistics collected from the model (described below).

2.2.2) Other Statistics Collected

Several other statistics are also collected when running the simulations. These are collected to help understand better the dynamics of the simulation and explain the effects of different parameters on the G_{ST} values at the two loci studied. There are many other statistics that could be collected, but they were not felt to be of importance in relation to the questions being asked. The other statistics collected are:

2.2.2.1): The proportion of the population of hosts that are infected by parasites.

This statistic ($\mathbf{P_{inf}}$) is calculated by dividing the number of hosts that contain a total parasite density of greater than 1.0×10^{-9} in a particular time step by the total number of hosts (N). A mean value ($\overline{\mathbf{P}}_{\mathbf{inf}}$) for P_{inf} during the stationary distribution is calculated by taking the average value for P_{inf} with time from day 4000 until day 5000 as described above. $\overline{\overline{\mathbf{P}}}_{\mathbf{Inf}}$ is the mean value of $\overline{P}_{inf}$ unless otherwise stated for two simulation runs, one started with Initial Condition A, the other with Initial Condition B.

2.2.2.2): The mean number of immuno-allelic alleles hosts are immune to.

During each time step for which data is collected, the number of alleles to which each host is immune is calculated. The mean value across

all hosts is referred to as **W**. A mean value ($\overline{\mathbf{W}}$) for W during the stationary distribution is calculated by taking the average value for W with time from day 4000 until day 5000 as described above. $\overline{\overline{\mathbf{W}}}$ is the mean value of $\overline{W}$ for two simulation runs (unless otherwise stated), one started with Initial Condition A, the other with Initial condition B.

2.2.2.3): The proportion of double bites, biting from an infected host, that transmit parasites.

During each time step for which data was collected, the proportion of double bites from infected hosts, which transmit parasites (i.e. Z>0) is calculated. This statistic is referred to as $\mathbf{P_v}$. A mean value ($\overline{\mathbf{P}}_\mathbf{v}$) for P_v during the stationary distribution is calculated by taking the average value for P_v with time from day 4000 until day 5000 as described above. $\overline{\overline{\mathbf{P}}}_\mathbf{v}$ is the mean value of $\overline{P}_v$ for two simulation runs (unless otherwise stated), one started with Initial Condition A, the other with Initial Condition B.

2.2.2.4): The mean number of zygotes that are formed in the infective bites.

During each time step for which data was collected, the number of zygotes in each double bite which transmits parasites (i.e. when Z>0) is calculated. The resulting values are then averaged over the total number of infected double bites that occurred in that time step. This statistic is referred to as $\mathbf{Z_{inf}}$. A mean value ($\overline{\mathbf{Z}}_\mathbf{inf}$) for Z_{inf} during the stationary distribution is calculated by taking the average value for Z_{inf} with time from day 4000 until day 5000 as described above. $\overline{\overline{\mathbf{Z}}}_\mathbf{inf}$ is the mean value of $\overline{Z}_{inf}$ for two

simulation runs (unless otherwise stated), one started with Initial Condition A, the other with Initial Condition B.

2.3) Design of the Model

2.3.1) General considerations

The model described above is designed to investigate several specific questions underlying the population genetics of the malaria parasite. Its specific purpose is to investigate the consequences of, and opportunities provided by, strain-specific immunity, assuming that this is haplotype-specific and determined by alleles at a single locus. It is designed to examine these questions in the context of the epidemiology of the parasite assuming a short length of immune memory in the host.

The model is also applicable to situations where haplotype-specific immunity occurs, and is not caused by a single locus, but by two or more loci, as long as these loci are linked. Linkage can either be physical on a chromosome, or epidemiological (Gupta *et al.* 1996). In either case, each locus involved will act in the same manner as the single immuno-allelic locus presented here.

The model is a complex one, with 11 different parameters and many dynamic variables. This complexity is important to allow the population genetics of the parasite to be understood in a biologically and epidemiologically realistic context. However, the added complexity in the

model makes it analytically intractable, and thus a computer simulation has been used.

There are many arguments that can be made in favour of both analytic and computer simulation models. Analytic models produce more precise, repeatable results and the dynamic relationships are often more apparent than is the case in computer simulation models. However, to be solvable, analytic models must in general be quite simple, which, when trying to describe a complex biological system may require many unrealistic assumptions. On the other hand, computer simulations allow much more complex models to be created. Computer simulations also facilitate the study of stochastic effects. However, the system becomes more of a "black box" with the reasons behind any particular result becoming much more inscrutable.

Although each malaria parasite has distinct phases in its life history, these phases are not synchronous between parasites. Thus there are no distinct points in time at which the population of parasites should be examined, and instead they should be examined continuously. One of the problems with computer simulation is the difficulty of constructing a model that acts in continuous time. Thus, a discontinuous time model has been constructed here, where all the calculations of the model are made at discrete points in time. This has, in principle, the potential to lead to unrealistic effects. However, these effects can be minimised by using small time-step sizes. To check that the errors created are small, the results of the simulations can be compared when different step sizes are used (see Appendix 1).

For a general discussion of the advantages of complex and simple mathematical models of biological systems and the use of discrete vs. continuous time in such models, see Crow and Kimura 1970, chapter 1. In general, the situation required here (continuous time and overlapping generations) becomes mathematically intractable except for relatively simple cases (e.g. Charlesworth 1994), and thus computer simulation is necessary.

2.3.2) The Details of the model

2.3.2.1) Parasite growth

In the model presented here, a modified logistic growth equation (Begon *et al* 1990) was chosen to calculate parasite growth in the hosts (see Equation 2.1.3.1 in Section 2.1.3). The difference from the equation's usual form is that, although it is used to describe the growth of individual parasite genotypes, unlike the normal logistic growth equation, the density-dependent effects are based on all the parasite genotypes present (i.e. the total parasite density). The equation is therefore equivalent to the Lotka-Voltera equation (Begon, *et al* 1990) describing population growth when there is complete resource competition between species. The equation is used as it allows for a parasite density-based limitation on growth and is of a simple form.

Although the model is designed to examine the effects of stochastic processes at the level of the population of hosts (e.g. the number of vectors biting a particular host in a particular time step), growth of parasites inside each host is assumed to be deterministic. Although the initial number of

sporozoites infecting a host may be very small, the number of blood stages these sporozoites lead to is very large (in the region of 10^5 or 10^6), and then grow up to much greater levels (Lines and Armstrong 1992). If one assumes that density-dependent effects on growth largely occur in the blood stages, these large numbers of parasites justify a deterministic growth equation.

Despite its simplicity, the logistic growth equation is a much more complex form of parasite growth than is found in previous epidemiological or population genetic models of malaria populations (e.g. Bailey 1982; Gupta *et al.* 1994a; Hastings 1996b; Molineaux and Gramiccia 1980). It is important to take into account parasite growth in the host in this model for two reasons. Firstly, the level of infectiousness of a host to vectors is likely to be correlated to the density of parasites in that host. Secondly, and perhaps more importantly, the dynamics of parasite growth can have important effects on the relative frequencies of different parasite genotypes invading the same host.

There is evidence that the degree of infectivity of hosts to mosquitoes, increases with the density of gametocytes present (Taylor and Read 1996; Tchuinkam *et al.* 1993). It is assumed in the model presented here that infectiousness of a host correlates with parasite density. This is equivalent to assuming, in the real world, that the number of infectious gametocytes present is proportional to the density of asexual parasites in that host. Some correlation between their densities seems very likely, because gametocytes can only be generated from asexual parasites; thus, the more asexual parasites present, the more gametocytes can be generated. There is no clear evidence of a correlation between asexual parasite densities and gametocyte densities in real infections, but in the

absence of data to the contrary it seems to be a sensible working hypothesis, and one used by Hellriegel in her model (Hellriegel 1992).

The degree to which parasite genotypes are represented in a host depends on how parasites grow in that host. This is especially the case where non-specific density-dependent growth limitation is involved, as appears to be the case in malaria (Kwiatkowski 1995). To understand the importance of this, consider a case where a parasite genotype is introduced to a host at a low frequency, and then a few days later a second parasite genotype is introduced at the same low frequency. Assume that both parasite genotypes grow at the same basic rate and that no haplotype specific immunity is acting against either parasite type. At the time the second parasite genotype enters the host, the first parasite has grown up to a density at which non-specific density dependent effects are preventing its further growth. In this case the second parasite genotype would also be prevented from growing and be maintained near its original, low level. Thus despite being introduced at the same low level and having the same basal growth rate, the parasite introduced first into a host is represented at a much greater frequency in that host than one introduced later. The degree to which this effect is evident will depend on the basal growth rate of the parasites and the time between parasite introductions.

In many epidemiological or population genetic models of malaria, the effects of parasite growth in relation to transmission of parasites between hosts is ignored (e.g. Bailey 1982; Gupta *et al.* 1994a; Hastings 1996b). In these models, hosts are deemed either infected or uninfected, and it is assumed that the rate of transmission from infected hosts is uniform. Thus, in these models, the importance of defining a mechanism of parasite growth to determine transmission rates is negligible.

Although many previous models define hosts as either infected and equally infectious, or uninfected and uninfectious, there are several exceptions (e.g. Molineaux and Gramiccia 1980; Saul 1996). In these models, hosts are seen to undergo a latent phase at the start of an infection where they are infected but not infectious. This can be seen as some acknowledgement of parasite growth in the host, with the parasites initially at too low a density, or in the wrong developmental stage, to pass on the infection.

The effects of parasite growth on the co-infection of different parasite genotypes in the same host is not addressed in other genetic or epidemiological models of malaria (although it is addressed in the context of immunity by Hellriegel (1992)). In many epidemiological models no distinct parasite types are examined, and thus this question does not arise (e.g. Bailey 1982; Molineaux and Gramiccia 1980). In others, where different parasite genotypes are addressed it is assumed that no co-infection occurs (Gupta *et al* 1994a) or that parasite types exist in a host totally independently of each other (Gupta and Day 1994b). In Hastings's analytic, genetic model (Hastings 1996) the issue is avoided as the parasites are assumed to undergo discrete generations. Therefore, co-infecting parasite genotypes enter a host at exactly the same time. Thus, most models of parasite growth would have the relative frequencies of the parasites maintained through an infection (in the absence of immunity) in the same proportions as they were at the start of that infection.

2.3.2.2) The immune response

In the model presented here, the immune system is allele specific in both recognition and action, with no cross immunity between alleles. The immune system starts to kill parasites containing a particular immuno-allelic allele after the parasites with that allele have reached a threshold density (M_{min}) and then only after a lag period (mean length of lag=$1/\rho$).

Although this threshold and lag have no obvious specific biological justification, they are designed as a simple tractable form of density-dependent activation of an immune response with a delay in its action. A more complex form of this occurs if immune effector cells are stimulated to increase in numbers by contact with parasites, and this effector cell growth takes time (Anderson and May 1992; Hellriegel 1992). A similar simplified system to that used here (with a threshold and lag) is used in a model by Saul (1994). The time course of parasite densities within hosts produced by Saul's model appears to be a reasonable approximation to the time-course found in real malaria patients (Saul 1994).

Other models of the immune system's interaction with malaria parasites have been developed (e.g. Anderson *et al.* 1989; Hellriegel 1992). These models have tried to be more realistic in terms of the biological processes involved in immune activation and killing. Consequently, these models are highly complex, with many parameters and many interacting variables. This was acceptable when studying a single host, as was done in these models. However, the model presented here is intended to study a population of infected hosts. In increasing the numbers of hosts analysed there is a practical necessity to reduce the complexity of the immune system in each host. This both increases the

ease of interpretation of the results produced, and reduces the computer power needed for the study. Also, despite the more realistic bases for these models, the course of the modelled infections is unlike that found in real infections (Gravenor *et al.* 1995).

In the model presented here, loss of immunity to parasites with a particular immuno-allelic allele is a step-wise function; a host can change from a state where it is immune to a particular allele to a state where it is totally susceptible from one time-step to the next. However, in the natural world, immunity is likely to be lost gradually, with a host becoming gradually more susceptible with time. It was felt that a very simple 'all or nothing' loss of immunity had many advantages over a more complex, 'dwindling' loss of immunity. Firstly, it helps keeps the number of parameters in an already complex model as few as possible. Secondly, the parameter β is easily translatable into 'the mean period of immune memory' ($1/\beta$), which allows for an easy grasp of this aspect of the model, and also allows easy comparisons to other epidemiological and population genetics models of malaria populations which include all-or-nothing loss of immunity (e.g. Hastings and Wedgwood-Oppenheim 1997; Gupta *et al* 1994a). It should be noted that although there is evidence that antibody titres, and T cell responses dwindle with time after exposure to malaria (e.g. Ballet *et al.* 1985), these responses can be very complex (Taylor *et al.* 1996) and the knowledge of how these biochemical and cellular indicators of immunity change parasite growth rate, parasite carrying capacity, and the period of infectiousness in subsequent infections remains small. Thus, if one incorporated dwindling immunity into the model, it has the potential to be less realistic than the model described.

It should also be realised that in simple epidemiological terms, an 'all or nothing response' with a fixed probability of occuring per time-step has similar population effects to dwindling immunity with a predefined time-scale. If one asumes that the lower the level of immunity, the greater the probability of reinfection, then in both cases, the probability of a host being reinfected after a new parasite challenge increases with time since the onset of immunity. Thus there maybe very little difference in measures of simple epidemiological parameters (such as parasite prevelance) between the two systems.

2.3.2.3) Transmission

Malaria is transmitted between mammalian hosts by mosquitoes. While in the mosquito, the parasite undergoes several developmental changes, and many cell divisions. This all takes time. The period of time between an infectious blood meal of *P. falciparum* being taken and that mosquito being able to pass on the parasites to another host is in the order of 10 to 22 days, depending, at least in part, on temperature (Bailey 1982). This lag is known as the "extrinsic incubation period" (Bailey 1982). Clearly the transmission of parasites from one host to another via a vector, all occurring in a single time step, as occurs in this model, cannot be justified on biological grounds. The reasons why this is done is to save on computer memory requirements and simulation time. To have the infections from a couple of weeks earlier contributing to infections in a particular time step, would require keeping a track of the infections in a very large number of vectors. This was felt to be unwieldy in the simulation.

There have been previous epidemiological models which have not taken into account the extrinsic incubation period (Gupta *et al* 1994a). This approach has recently been criticised by Saul (1996) as being potentially unrealistic for examining the dynamics of malaria infections.

As stated earlier, the main interest in the model is to examine the values of $G_{ST(d)}$ and $G_{ST(n)}$ and certain epidemiological parameters when a stationary distribution has been reached. At such a time, whether an infection is transmitted during the same time step, or takes two weeks to occur, should be of no consequence, as both the rate of infection and the composition of the infections should be similar. However, it may effect the dynamics of the epidemic phase of the infection, with a larger epidemic rise and crash in the proportion of hosts infected (P_{inf}) when the extrinsic incubation period is taken into account than when it is not.

The number of double bites that occur from a particular host, during a time step is determined by a Poisson Distribution (**PD**). The PD is a discrete distribution that describes rare and random events. It is assumed that all hosts are identical in their probability of being bitten, and being bitten by one vector in a time step does not preclude being bitten by another. Under these assumptions the use of the PD is appropriate. There is evidence of host heterogeneity in the mosquito biting rates (Knols *et al.* 1995) but this was felt to be an unnecessary complication to the model, and thus the use of a PD was felt appropriate.

While the malaria parasite is developing in the mosquito, sexual recombination occurs (Sinden and Hartley 1985). The malaria life history is particularly unusual in that, in many cases, only one, or very few, zygotes develops in a mosquito (Pichon *et al.* 1996), and thus the exact nature of

meiosis is important. This has not been taken into account by previous population genetic models of malaria (e.g. Curtis and Otoo 1986; Hastings 1996). Assuming that meiosis occurs in the same manner as it does in other eukaryotic organisms that have been examined, a maximum of four genetically distinct haploid progeny should be produced (Alberts *et al.* 1983). In some cases only one of these four progeny develop (e.g. in the development of the vertebrate egg (Alberts *et al.* 1983)). How many of these progeny develop in malaria is not known. However studies on the oocyst stage of the malaria parasite show that two alleles of the same gene can be found in an oocyst (Ranford-Cartwright *et al.* 1993), and thus at least two of the haploid progeny of meiosis do develop.

Detailed analysis of the products of a single meiosis has been made only in fungi (Fincham 1983) and *Drosophila* (Ashburner 1989), by a process known as Tetrad Analysis (in *Drosophila* as half-tetrad analysis). These studies show that, when the four haploid progeny of a single meiosis are examined at two polymorphic loci, all four potential haploid progeny genotypes are not invariably present. This is described in more detail in the model section, and in Appendix 2. In the work described here, the frequency of each haplotype produced from a single meiosis in the parasite is based on studies of fungi and *Drosophila*. Not only is it assumed that four products are formed by each meiosis, and that the frequency of each two locus haplotype is determined in the same manner as it is in fungi, but it is also assumed that each of the four potential haplotypes is passed onto the next host in equal frequency. This is also assumed in other models (e.g. Dye and Godfray 1993; Hastings 1996). However some models assume that only one haplotype is passed on by each infecting vector (e.g. Hastings 1997; Hill and Babiker 1995; Hill *et al.* 1995).

The number of sporozoites passed on to a new host in a single bite has been studied (Beier 1993; Ponnudurai *et al.* 1991; Rosenberg *et al.* 1990). There are large differences in the numbers of sporozoites injected by individual mosquitoes, but estimates of the mean number of sporozoites have been calculated as between 15 and 25. These numbers are clearly small, and thus stochastic effects are likely to be important. Considering the small number of zygotes generally formed in a vector, it is likely that most, though not necessarily all, haplotypes formed in a vector are transmitted to the next host. It is, however, unlikely that these haplotypes will exist in equal frequencies. Thus, in models where only one haplotype is transmitted, there is likely to be an underestimate of the degree of mixing of haplotypes in a host (in this model measured by $G_{ST(n)}$). In models (including this one) where all haplotypes are assumed to be transmitted in equal frequencies there is likely to be an over estimate of the degree of mixing.

The number of zygotes formed in a vector is determined, in this model, by a PD with its mean proportional to the density of parasites in the host. The reasons for assuming a correlation between the density of parasites present in a host, and the infectivity of the host to the vector have been explained above.

Field caught and laboratory fed mosquitoes have been examined from the point of view of oocyst distributions. These studies have all found a Negative Binomial Distribution (**NBD**) of oocysts in mosquitoes (Billingsley *et al.* 1994; Medley *et al.* 1993; Pichon *et al.* 1996). NBDs are found frequently in nature when the spatial distribution of organisms is examined (Crofton 1971; Rutledge *et al.* 1973). One of the bases for an NBD is when many Poisson Distributions, with different means, are summed together (Hill

and Babiker 1995). This is likely to be the case when any particular event is both random and rare (e.g. the formation of a viable zygote in a blood meal), but where the probability of such an event depends on aspects of its local environment (e.g. the density of gametocytes in the blood meal and the size of the blood meal).

Thus, the Poisson Distribution seemed a plausible basis for the NBD found in real infections and is therefore used as the basis for the calculation of the number of zygotes in each double bite in this study. It should also be realised that the use of a NBD as the basis for each double bite, again with the means proportional to the parasite densities, is not likely to produce a NBD across a population of double bites (Grafen and Woolhouse 1993), and thus has no advantage over the system used. Use of an NBD does, however, have two disadvantages: the increase in number parameters of the model and the lack of logical explanation for its use at the individual vector level.

2.3.3) **The potential problems**

The complexity of the model is expected to lead to some difficulty in its analysis. Firstly, as has already been stated, the complexity makes it difficult to understand the precise basis for any particular result. However general trends can be examined and speculation about more specific results given. This can lead to further manipulation of the model to test hypotheses as to the model's nature. As the aim of this study is to gain biological rather than mathematical insights into the nature of malaria populations, this was felt to be a sensible method of analysis.

Secondly, the many parameters involved in the model mean that any parameter that is varied could be examined with a large variety of different values for the other parameters. Constraints on computer time prevented examination of more than a small part of this 'parameter space'. It was decided to use a carefully chosen set of parameter values that, where possible, were chosen to reflect the current level of understanding of the nature of malaria parasite's life history. Each parameter was then varied independently against this background.

This method has the potential to miss important trends that are not apparent in the part of parameter space examined. It is also possible that trends that are felt to be important are only specific to the parameter values examined and not, as might be felt, universal to the model. Some preliminary studies were done to examine other parts of the parameter space, and these agreed with the trends reported here.

Of course, the most important problem with the model is that it must be biologically unrealistic. The life history of the malaria parasite involves interactions with two very different organisms - mosquitoes and man. Such an interaction between three biological systems is likely to be highly complex. Thus there are many simplifications in the model. These occur in many instances to allow the model to be more tractable, and in other cases, simply because that aspect of the biology of the parasite is not known. This is true for all simulation studies. This model was designed, not to describe every detail of the biology of malaria, but to capture the essence of its population dynamics.

It is hoped that these simplifications do not have major qualitative effects on the outcome of the model, although clearly one would expect

quantitative differences between the predictions of the model and malaria in the natural world. There is, however, a risk in all modelling of complex systems that simplifications will also lead to qualitative differences.

2.3.4) Discussion

As has been stated, the complexity of the model has led to its analysis through computer simulation. However, despite its complexity, the model can be seen as having several, interacting aspects each of which are, individually, quite simple. This in turn led to the construction of the simulation programs in a modular form. This is important for two reasons. Firstly, as the biology of malaria parasites becomes better known, aspects of the model may be shown to be inappropriate. Only trivial alterations to the computer program would be necessary to see what effects a more accurate version has on the output statistics.

Secondly, it should be realised that although the model has been designed to answer specific questions about malaria parasites, its modularity allows alterations to be made that could make it easily applicable to other vector born pathogens or even disease agents with other modes of transmission. Thus, the computer programs developed can be seen as a tool allowing the population genetics of many organisms to be examined in the context of their own epidemiologies.

One restriction of the model presented here is that imposed by the availability of computer power and time. If the current trend in increasing computer power with decreasing cost continues, then the modularity of this

model may well allow its expansion to take into account complex aspects of parasite biology not yet addressed.

2.4) Choice of standard parameter values.

In Chapter 4 the parameters of the model described above are examined in detail. This is done by varying the value of each parameter, while keeping the other parameter values constant. Thus, a series of 'Standard Parameter Values' are used. These are the values at which each parameter is set while another parameter is being examined.

Wherever possible, the standard parameter values used reflect the current understanding of the malaria-human-mosquito system. However, for some parameters, very little is known and an arbitrary choice of value was made. In Chapter 4, where the effects of different parameter values are examined, some indication of the effects of such arbitrary choices can be gained.

The standard value for the number of immuno-allelic alleles present in the model (D) was given a value of ten. The number of alleles at the immuno-allelic locus defines the number of immunologically distinct 'strains' that are present. It is not known how many such strains are present in real malaria populations. Thus any value of D used must be seen as arbitrary. In the model by Hastings (Hastings 1996) between 2 and 20 immuno-allelic alleles are used while Gupta used 2 strains in her model (where multiple immunologically distinct strains were present) (Gupta *et al* 1994a). In a different paper, Gupta assumed 5 strains of *P. falciparum* to be

present when analysing data from the Madang region of Papua New Guinea (Gupta *et al.* 1994b).

A standard host population size (N) of 6000 was used. It is obvious that population sizes for hosts of the malaria parasite will vary greatly from locality to locality. It would also be true to say that a useful definition of an epidemiologically distinct host population in the real world may well be difficult to produce. Human populations are often aggregated into villages and towns, with both human and mosquito migrations between them. Clearly in most cases the effective population size for a malarious population will be larger than the few hundred or less found in any village examined, but how much larger is difficult to estimate. Many epidemiological and genetic models of malarious populations assume an infinite population size (e.g. (Gupta *et al* 1994a; Hastings 1996). The standard population size here of 6000 would correspond to a small town or a collection of villages, between which reasonably frequent travel occurs. This is obviously a great simplification .

The host death rate is set at 5×10^{-5} per day. This would correspond to a mean life expectancy of 54.8 years which corresponds reasonably well with that used in other models (Anderson *et al* 1989). It also corresponds reasonably well to an estimated Crude Death Rate measured in a study in Nigeria (approximately 1×10^{-4} per day (Molineaux and Gramiccia 1980)). The host death rate in this model is constant, i.e. it is not age related. This is not the case in real human populations in malarious regions where the majority of deaths occur in the very young and very old (Anderson and May 1992; Molineaux and Gramiccia 1980). For instance, in the Nigerian study the death rate for children of ages one to four years is four to five times greater than the Crude Death Rate measured (Molineaux and Gramiccia

1980). There is also no malaria induced mortality in this model, whereas it has been estimated that, for instance, in a study in Zaire, approximately 14% of children dying between the ages of one and two years die from malaria (Wernsdorfer and Wernsdorfer 1988). Ignoring both age-specific and malaria-specific mortality is a great simplification, but one that was felt would not greatly bias the results of the study while removing some added complexity of the model. Many other models of malaria populations have also assumed a constant, non-malaria induced death rate (e.g. Bailey 1982; Gupta, Swinton and Anderson 1994a; Hastings 1996).

The threshold for immunity, M_{min}, was given a standard value of 0.05. In this model, if the value is above 1/D then, when all immunologically distinct parasite types are present in a host, at equal densities, all parasites would be below the threshold value and thus, no parasites would be cleared from the host. Thus when D has a standard value of 10, a value for M_{min} of between 0.0 and 0.1 is necessary, and a standard value of 0.05 has been chosen arbitrarily.

The probability of an immune response being mounted against a particular allele in any time step (ρ) was given a standard value of 0.025 per day. This corresponds to a mean time between a parasite genotype growing above M_{min} and the start of an immune response of 40 days.

The intensity of the immune response is controlled by s, and this is given a standard value of 0.2 per day^2. With this value, if parasites with only one particular immuno-allelic allele are present in a host, then the mean period of infection, after it has grown to a level above M_{min}, is approximately 53 days.

There is no simple consensus on the duration of natural infections or, more importantly, infectiousness (in this model the duration of infection and infectiousness are completely correlated). Gupta and co-workers (Gupta *et al* 1994a) suggest that infectiousness may well only last about two weeks, and a mean period of between 5 and 18 days was used in the model they presented. Carter, on the other hand (Carter and Gwadz 1980), quotes work by Jeffrey and Eyles, showing that infectiousness lasted in a group of non-immune hosts for at least one month, and for 20% of the participants, for as long as one year. Thus, in this model a compromise position of a mean period of infection of approximately two months seemed appropriate.

The probability of loss of immunity per time step, β was given a standard value of 2.74×10^{-4} per time step. This corresponds approximately to a period of immune memory of one year. This is a little understood topic, there being estimates for periods of immunity of anything from a few days up to nearly lifelong immunity (Deleron and Chougnet 1992; Gupta *et al.* 1994b; Powell *et al* 1972; Saul 1996). Often, the differences in length depends on the differences in the definition of 'immunity'. In this model, sterile immunity is the only concern (anti-parasite or anti-transmission immunity) and there is evidence that this is short lived, perhaps lasting less than a year (Mendis and Carter 1992; Powell *et al.* 1972; Saul 1996; also see Section 1.3.5).

The parasite growth rate, b_0, was given a standard value of 12.3 per 2 days (and therefore a value of $(12.3)^{1/20}$ per time step). In other models it has been assumed to have a value of 8 per day (Anderson *et al* 1989; Hellriegel 1992). Gravenor *et al* (1995) compiled values from earlier workers for *P. falciparum* and found values ranging from 9 to 20.6 with an average of 12.3 per 2 days.

The Double Bite Rate, T, was given a standard value of 0.1 double bites per host per day. Realistically, the Double Bite Rate is likely to vary from place to place, depending on the respective densities of mosquitoes and humans, and the biting habits of the mosquitoes.

The parameter T is proportional to, though less than the epidemiological statistic, the Vectorial Capacity. The Vectorial Capacity is a measure of the mean numbers of human hosts to which an infectious host would pass on an infection per unit time given a fully susceptible population (Bailey 1982). In the model presented here, not all double bites from an infected host will transmit parasites. If however this is the case, T and the Vectorial Capacity would be equal.

Vectorial Capacities of malaria have been measured in many parts of the world at different times. The values have generally been in the region of 0.2 to 30 per day (Graves *et al.* 1990; Molineaux and Gramiccia 1980; Rubio-Palis 1994). These values are higher than the value of 0.1 per day used here. However it has been pointed out by Bailey (Bailey 1982), that these measures generally assume that all infectious mosquitoes which bite a host transmit parasites to that host. This is not necessarily the case. It has been commented by Beier (Beier 1993) that, for non-immune volunteers, five infected mosquito bites are needed to pass on a *P. falciparum* infection reliably. Pull and Grab have estimated that the probability of a *P. falciparum* infected vector infecting a non-immune host is in the region of 0.026 (Bailey 1982; Pull and Grab 1974). Thus it may well be that many of the estimates of Vectorial Capacity are 4-40 fold too high. It should also be noted that Vectorial Capacities are much easier to measure in areas which have high, rather than low levels of transmission , and so the

measured range shown above, is likely to be biased to the high end of the distribution.

The parameter c, determines the infectivity of parasites to the vectors. It determines both the proportion of vectors biting infected hosts that actually transmit parasites and the numbers of zygotes formed in such transmission events. A standard value of c = 1.0 is used in these simulations. With this value for c, if all infected hosts have parasite densities equal to their Carrying Capacities (K), then the mean number of oocysts formed in double bites from infected hosts would be 1.0. The number of studies examining oocyst (the visible product of a zygote) numbers in field caught mosquitoes is small (Billingsley *et al.* 1994). However, two recent studies have examined field data and, using the assumption that there is a Negative Binomial Distribution of oocyst numbers, estimates of the mean numbers of oocysts found in potentially infected vectors have been calculated. These estimates are 1.34 and 2.37 (Billingsley *et al.* 1994; Pichon *et al.* 1996 respectively). Thus, the choice of a value for c of 1.0 is low, however it was not felt to be inappropriate.

The density of parasites that are introduced into a host through an infectious bite, θ, is given a standard value of $2x10^{-5}$. The numbers of sporozoites entering hosts has been shown to be small (Beier 1993; Ponnudurai *et al.* 1991; Rosenberg *et al.* 1990). However it seems likely that there is little density-dependent growth limitation until the parasites have left the liver. The number of parasites at this stage is considerably more, probably between 40,000 and 800,000 although occasionally more may occur (Lines and Armstrong 1992). Parasite numbers in febrile schoolchildren are in the region of 2×10^{10}. Thus, if we assume that the

Carrying Capacity for *P. falciparum* is in the region of 10^{10} or 10^{11}, a value of θ of 2×10^{-5} would appear reasonable.

In conclusion, in this study it has been intended to use realistic parameter values wherever possible. This has not always been the case, and in hindsight some values may seem slightly higher or lower than is appropriate. However, these values are probably in the correct region of the parameter space. In Chapter 4 the parameter values are manipulated, and thus the effects of having inappropriate values for any of the parameters can be evaluated.

3) The Dynamics of the Model

3.1) Introduction

In this chapter the hypothetical parasite population described in the model section is examined with time. It is important to understand the dynamics of the population for several reasons. Firstly, in the next chapter the effects of the different parameters on the outcome statistics of the model are examined, and values of the various statistics are measured once these statistics appear to be in a stationary distribution. An examination of how these statistics change with time helps to justify the choice of time steps used for examining the stationary distribution.

Secondly, while this model incorporates aspects of the parasite population which show epidemiological as well as population genetic features (e.g. the proportion of hosts infected with parasites). Thus to some extent this model can be compared on purely epidemiological grounds to epidemiological models in the literature.

Finally, one of the reasons for studying this hypothetical population is to make predictions as to what is occurring with malaria parasites in the field. In many cases, malaria populations will not be in a stationary distribution but in a state of flux. Whether due to the introduction of control programmes, or seasonal changes in mosquito numbers, or the introduction of new immunological strains (through mutation or migration), or etc., the dynamics, and therefore the genetics, of a malaria population may well be perturbed. It is therefore important to examine the hypothetical population presented here when it is not as well as when it is at a stationary distribution.

3.2) Results

This section are presents the results of simulation runs where the standard parameters were used. The time courses of two specific simulations are examined in detail, and these show features typical to the other simulation runs. One of these, to be referred to in this section as '**Simulation Run 1**', is run with Initial Conditions A, as described in Section 2.2.1, with G_{ST}=0 for both loci. The other simulation will be called '**Simulation Run 2**', and is run under Initial Conditions B, which are also described in Section 2.2.1. Run 2 is initiated with G_{ST}=1 for both loci. In both cases the simulations are sampled every 4 days to produce the time course data. The statistics examined can be broadly separated into three groups: host statistics, vector statistics and parasite genetics.

Unless otherwise stated, when the mean of a statistic is calculated over time, it is calculated between day 4000 and 5000, as described in the Section 2.2.

Mean Values at the Stationary Distribution of the Output Statistics

	Initial Condition A	Initial Condition B
$\bar{G}_{ST(d)}$	0.940142 0.939454 0.939672 0.939737	0.939312 0.940027 0.939991 0.940706
$\bar{G}_{ST(n)}$	0.819915 0.827076 0.816303 0.814390	0.814303 0.821827 0.816029 0.828402
$\bar{t}$	1.146634 1.135874 1.151131 1.153916	1.153516 1.143826 1.151909 1.135567
$\bar{P}_{inf}$	0.662414 0.672420 0.667284 0.672902	0.673311 0.665819 0.666921 0.664893
$\bar{Z}_{inf}$	1.534064 1.543168 1.549157 1.535698	1.549564 1.543041 1.541526 1.541291
$\bar{P}_{v}$	0.483969 0.485909 0.486881 0.486717	0.488509 0.485599 0.488663 0.486874
$\bar{W}$	5.182491 5.174132 5.166388 5.182063	5.189862 5.159174 5.178418 5.148859

Table 3.2a:
The mean statistics presented are calculated from sample values taken every 2 days from day 4000 to day 5000, as described in Section 2.2. Values in the first column are calculated from simulations started with Initial Condition A and values in the second column are from simulations started with Initial Conditions B. The Standard Parameter Values are used.

A Comparison of the Effects of the Two Different Starting Conditions on the Output Statistics

	Initial Condition A	Initial Condition B
$\bar{\bar{G}}_{ST(d)}$	(0.9400327) **0.93975125** (0.9394697)	(0.9405669) **0.940009** (0.9394511)
$\bar{\bar{G}}_{ST(n)}$	(0.8249031) **0.81942101** (0.8139390)	(0.8263917) **0.82014024** (0.8138888)
$\bar{\bar{t}}$	(1.154662) **1.1468888** (1.139115)	(1.1543019) **1.1462045** (1.138107)
$\bar{\bar{P}}_{inf}$	(0.673588) **0.66875501** (0.663922)	(0.671465) **0.66773599** (0.664006)
$\bar{\bar{Z}}_{inf}$	(1.547371) **1.5405218** (1.533673)	(1.54765) **1.5438555** (1.540053)
$\bar{\bar{P}}_{v}$	(0.4871718) **0.48586899** (0.484559)	(0.4888377) **0.48741125** (0.485987)
$\bar{\bar{W}}$	(5.183736) **5.1762685** (5.168801)	(5.187202) **5.1690782** (5.150954)

Table 3.2b:
Each statistic is calculated across 4 runs with each starting condition, these values are presented in bold. The numbers in parentheses above and below the mean values are the upper and lower 95% confidence limits respectively. The simulations were executed using the Standard Parameter Values.

3.2.1) Host statistics

There are two host statistics examined: the proportion of hosts that are infected (P_{inf}) and the mean number of immuno-allelic alleles to which members of the host population are immune (W). Figure 3.2.1a examines how these two statistics change with time in Simulation Run 1. The upper of the two lines in Figure 3.2.1a shows how the proportion of hosts infected changes with time. Recall that the simulations are initiated with 20% of the hosts infected. In the data from the simulation run shown in Figure 3.2.1a, the fraction of infected hosts rises steeply from 20% to about 90% after about 90 days, and then falls to about 60%. The fraction of hosts infected then rises again to about 65%, and vacillates to either side of this value, having apparently achieved a stationary distribution. The mean proportion of hosts infected (averaging over time) was calculated using the procedures described in Section 2.2. As can be seen from the graph, the stationary distribution seems to be well established by day 4,000, and thus the decision to calculate the average from day 4,000 until day 5,000 appears to be justified. The average proportion of hosts infected, between day 4000 and day 5000 ($\overline{P}_{inf}$), for this run is 0.662, with a variance of 1.24×10^{-4}.

During each time step for which data were collected, the number of alleles to which each host was immune were calculated, and the resulting values were averaged over hosts (W). The results of this calculation are shown, for Run 1, by the lower line in Figure 3.2.1a. As can be seen from the figure, W starts at zero, rises to a maximum of about 5.60 on around day 430, and then falls to about 5.2, wavering around this value for the remainder of the run. The variation in the average number of alleles to which individuals were immune was substantially less than the variation

over time in the proportion of hosts infected. The value of $\overline{W}$ was 5.182, with a variance of 1.17×10^{-3}.

Simulations under Initial Condition A were repeated four times. In each case $\overline{P}_{inf}$ and $\overline{W}$ were calculated. The value for each repeat for both statistics can be found in the column labelled "Initial Condition A" in the Table 3.2a. It can be seen for both statistics that the values for each run are similar, and that, in Table 3.2b, the confidence limits for the means of these runs is relatively small in comparison to the mean.

Figure 3.2.1b shows the data from Simulation Run 2 ($G_{ST}=1$). The shape of this graph for both P_{inf} and W can be seen to be very similar to that found in Figure 3.2.1a for Simulation Run 1. The differences to be noted are that the maximum value for P_{inf} is both larger (about 93%) and occurs later (around day 230) in Simulation Run 2. This is also true for W, which reaches a maximum of about 5.79 around day 520 in Simulation Run 2.

The appearance and position of the stationary distributions achieved do, however, appear to be very similar. Tables 3.2a and 3.2b show that $\overline{P}_{inf}$ and $\overline{W}$ values for four different runs under Initial Condition B are very similar to those run under Initial Condition A. The mean values for these runs across each condition have overlapping 95% confidence limits.

The Proportion of Hosts Infected (P_{inf}) and the Mean Level of Immunity (W) versus Time during Simulation Run 1

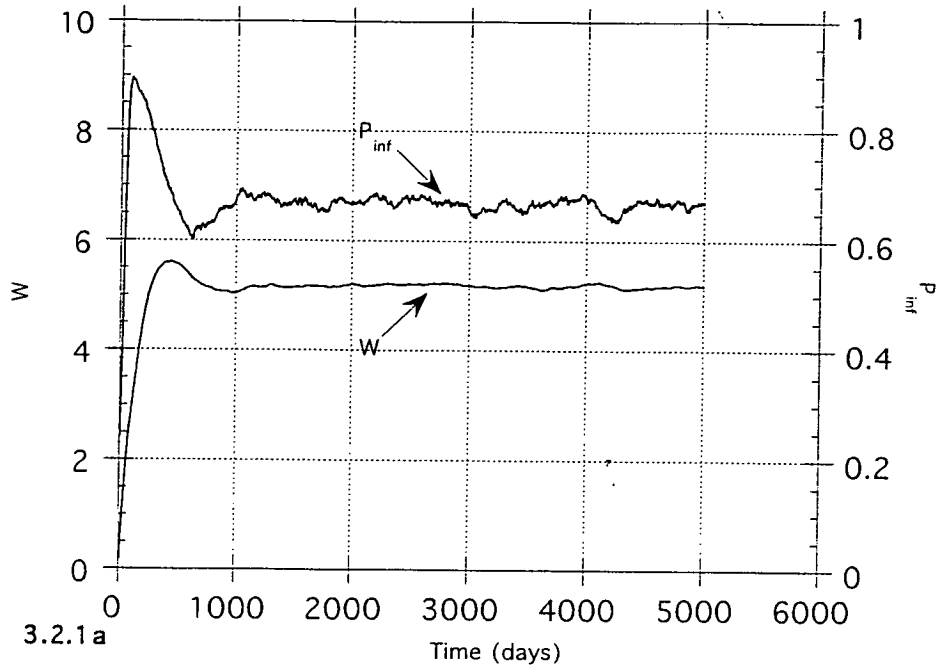


Fig 3.2.1a

The Proportion of Hosts Infected (P_{inf}) and the Mean Level of Immunity (W) versus Time during Simulation Run 2

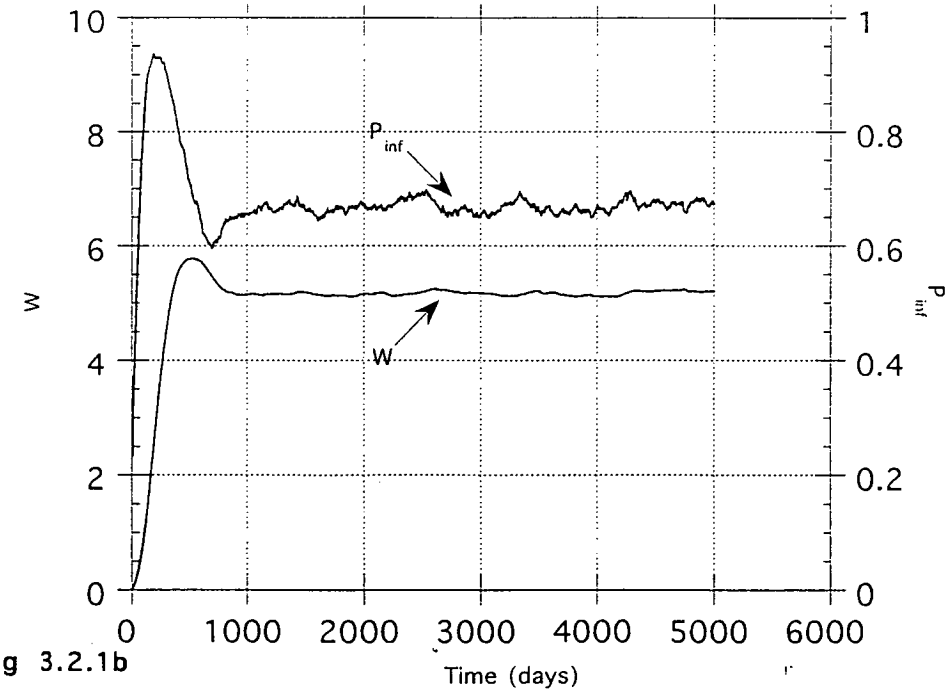


Fig 3.2.1b

Figures 3.2.1a and b:
Values are calculated and plotted every 4 days over the 5,000 days examined. Simulation Run 1 was started with Initial Conditions A and Simulation Run 2 was started with Initial Conditions B. The Standard Parameter Values were used.

3.2.2) Vector Statistics

Figure 3.2.2a presents data on aspects of vector transmission during Simulation Run 1. The lower of the 2 lines shows the proportion of double bites that actually transmit parasites (P_v). It can be seen that there is a great deal of variation in this value over the time steps sampled, with the value occasionally going above 0.6 and occasionally below 0.4, but generally remaining within these bounds and vacillating around 0.5. The mean value of P_v (over time) ($\overline{P}_v$) for the stationary distribution for Run 1 is 0.484.

The upper line in Figure 3.2.2a shows the mean number of zygotes found in each infected vector in Run 1 (Z_{inf}). There is a very substantial amount of variation in this value between time steps sampled in this simulation, with values as high as 2.0 and as low as 1.2. The mean value between day 4000 and day 5000 ($\overline{Z}_{inf}$) is 1.534.

In Figure 3.2.2b, the same statistics (P_v, and Z_{inf}) are shown for Simulation Run 2. The shapes of the graphs for both statistics appear very similar to those in Figure 3.2.2a, with the values of $\overline{P}_v$ and $\overline{Z}_{inf}$ being 0.489 and 1.550 respectively.

The comparison between the statistics $\overline{P}_v$ and $\overline{Z}_{inf}$ for the two sets of starting conditions can be seen in Tables 3.2a and 3.2b. As for the host statistics described above, the means across four runs for these statistics are very similar under Initial Condition A and Initial Condition B, with overlapping 95% confidence limits.

To study further any underlying trend that is occurring with either P_v or Z_{inf} during the period of the simulations, the values of each were replotted for Run 1, with each point plotted being the mean of the previous

20 time steps (4 days). This can be seen in Figure 3.2.2c. As can be seen there is still a lot of variation between points on the graph, but much less than when each point represented a statistic at a single time step. In this plot one can see more clearly that there appears to be no underlying change in the values of Z_{inf} and that it appears to arrive at a stationary distribution immediately.

With P_v, the lower line in Figure 3.2.2c, there is an initial rise in value to the region of 0.6, it then returns to a lower level within the first 500 days and remains in the region of 0.5 for the duration of the simulation, with the appearance of a stationary distribution.

The Proportion of Double Bites that Transmit Parasites (P_v) and the Mean Number of Zygotes per Transmitting Double Bite (Z_{inf}) versus Time during Simulation Run 1

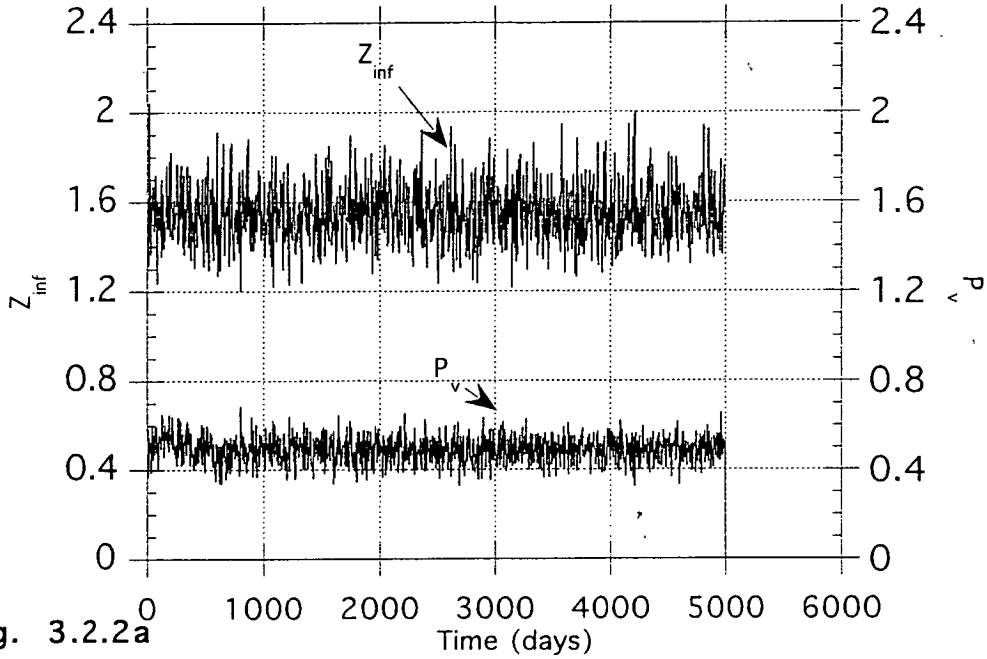


Fig. 3.2.2a

The Proportion of Double Bites that Transmit Parasites (P_v) and the Mean Number of Zygotes per Transmitting Double Bite (Z_{inf}) versus Time during Simulation Run 2

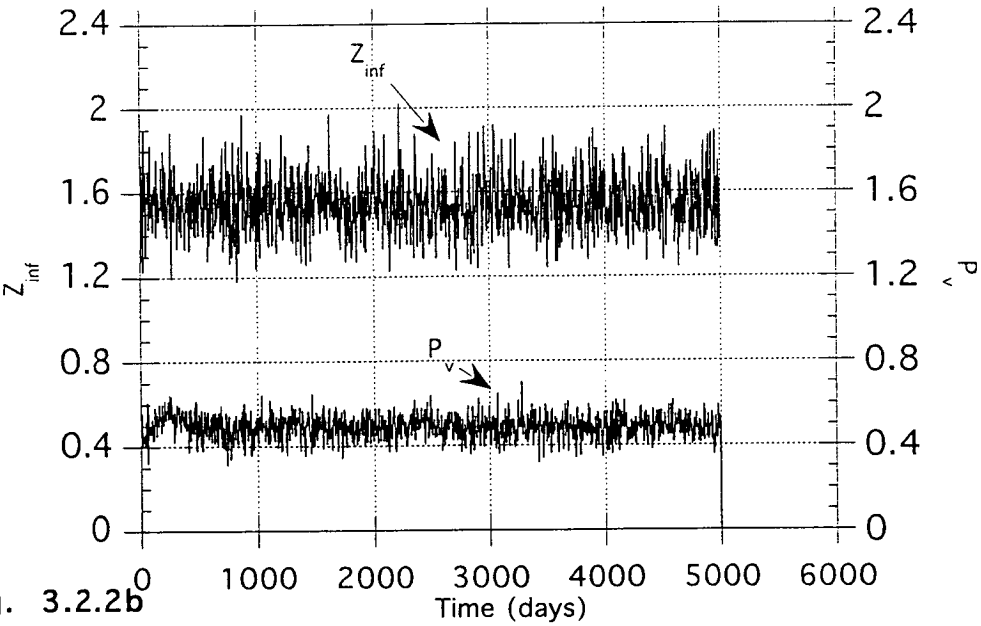


Fig. 3.2.2b

Figures 3.2.2a-b:
Values are calculated and plotted every 4 days over the 5,000 days examined. Simulation Run 1 was started with Initial Conditions A and Simulation Run 2 was started with Initial Conditions B. The Standard Parameter Values were used.

The Proportion of Double Bites that Transmit Parasites (P_v) and the Mean Number of Zygotes per Transmitting Double Bite (Z_{inf}), averaged over time, versus Time during Simulation Run 1

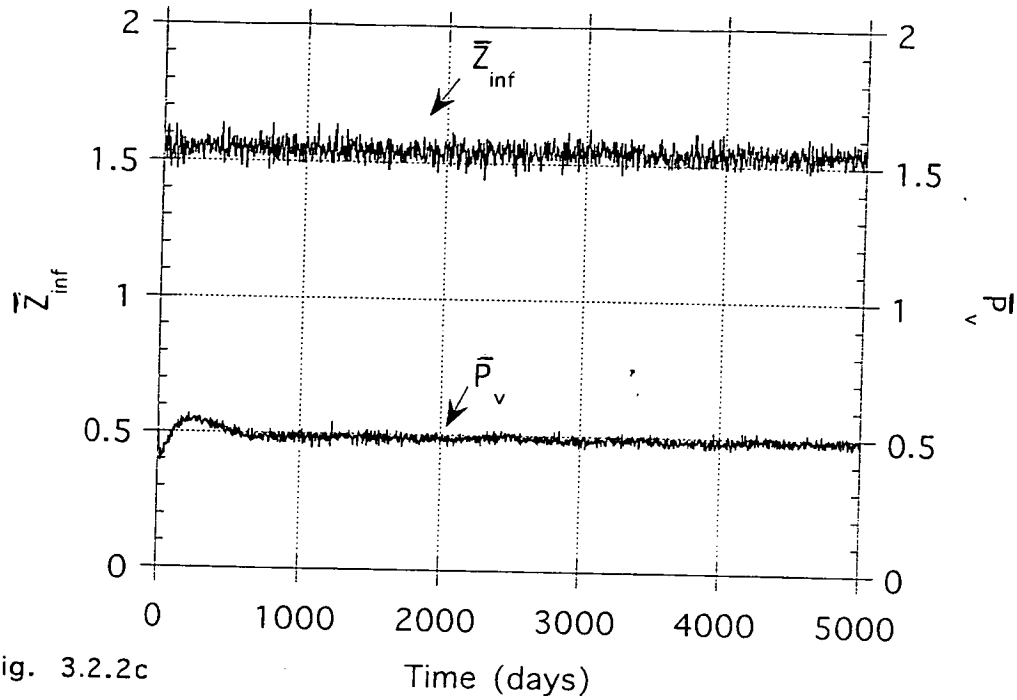


Fig. 3.2.2c

Figures 3.2.2c:
In this case, mean values of each statistic are plotted every 4 days, having been calculated from the measured values at each time step over the 40 previous time steps (4 days). Simulation Run 1 was started with Initial Conditions A. The Standard Parameter Values were used.

3.2.3) Parasite Genetics

Figures 3.2.3a and 3.2.3b show how $G_{ST(d)}$ and $G_{ST(n)}$ change with time during Run 1 and Run 2 (Initiated with $G_{ST}=0$ and $G_{ST}=1$ respectively). First, the value of $G_{ST(d)}$ will be examined, as represented in Figure 3.2.3a by the upper line. It can be seen that the value for $G_{ST(d)}$ rises steeply from 0.0 over the first few days until it reaches a value of about 0.85 on around day 130. There is then a short plateau before continuing to rise more slowly to what appears to be a stationary distribution vacillating around 0.95. The value for $\bar{G}_{ST(d)}$ for this run is found to be 0.940 with a variance of 1.92×10^{-5}.

If we now turn to $G_{ST(n)}$ (the lower line) in Figure 3.2.3a we can see that the value starts at 0.0 at the start of the simulation rising very steeply to a value of about 0.50. After this point the gradient becomes less steep, but continues to rise to a value around 0.82. The value then varies around this point with the appearance of a stationary distribution. The variation in $G_{ST(n)}$, once a stationary distribution has been reached, appears to be greater than it is for $G_{ST(d)}$. The value for $\bar{G}_{ST(n)}$ for this run is found to be 0.820 with a variance of 1.07×10^{-4}.

Figure 3.2.3b shows the values of $G_{ST(d)}$ and $G_{ST(n)}$ with time in Simulation Run 2. The upper of the 2 lines represents the value $G_{ST(d)}$. It can be seen that $G_{ST(d)}$ descends from an initial value of 1.0 to a minimum value of around 0.83 on around day 200 then rises again to around 0.95. It then varies around this value with the appearance of a stationary distribution. The value for $\bar{G}_{ST(d)}$ in Run 2 was found to be 0.939.

The lower of the two lines represents the value of $G_{ST(n)}$. This value starts initially at 1.0 then descends steeply to a minimum of about 0.69 on

around day 340 then rises again to a value around 0.82, it then vacillates around this value with the appearance of a stationary distribution. The value for $\overline{G}_{ST(n)}$ was found to be 0.814.

Though the initial shapes of Figures 3.2.3a and 3.2.3b are very different, it can be seen that the apparent stationary distributions produced for both statistics when compared between the two runs appear very similar. This is confirmed by examining the values of $\overline{G}_{ST(d)}$ and $\overline{G}_{ST(n)}$ over several runs with Initial Condition A and Initial Condition B (see Tables 3.2a and 3.2b).

Figure 3.2.4 shows how the value of *t* changes with time in both Simulation Run 1 and Simulation Run 2. The shape of the curves in the two runs are similar. In both cases, the value of *t* starts at 1.0, rises to a maximum and then descends to a value of around 1.2 and then remains around this value, vacillating on either side with the appearance of a stationary distribution. For Simulation Run 1, the initial rise starts immediately increasing very steeply to a maximum value of around 1.64 around day 80. For Simulation Run 2, however, there is a lag before the value starts to rise it then increases to a maximum of about 1.28 around day 350.

The values of $G_{ST(n)}$ and $G_{ST(d)}$ versus Time during Simulation Run 1

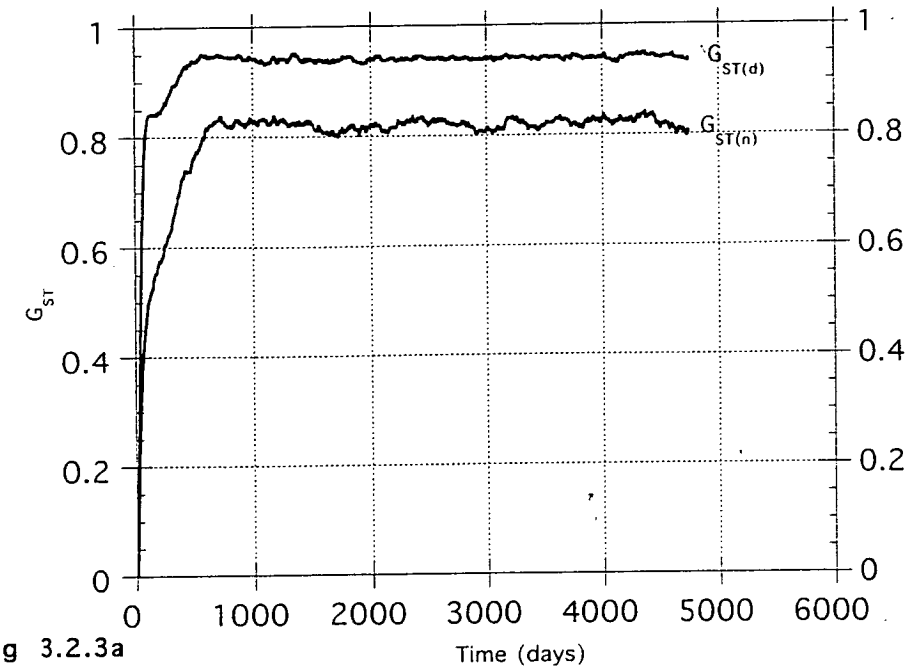


Fig 3.2.3a

The values of $G_{ST(n)}$ and $G_{ST(d)}$ versus Time during Simulation Run 2

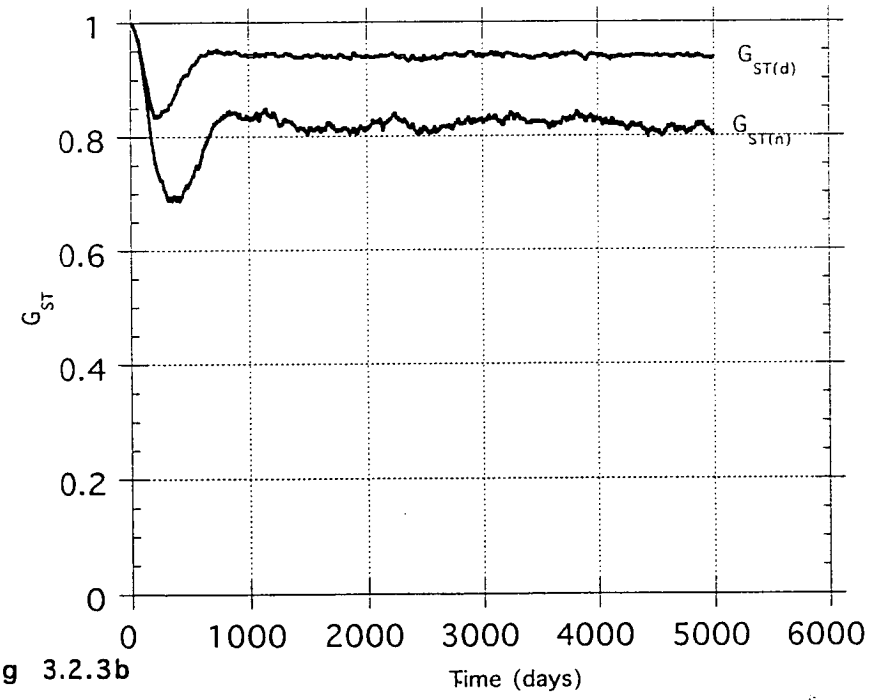


Fig 3.2.3b

Figures 3.2.3a and b:
Values are calculated and plotted every 4 days over the 5,000 days examined. Simulation Run 1 was started with Initial Conditions A and Simulation Run 2 was started with Initial Conditions B. The Standard Parameter Values were used.

The value of *t* versus Time during Simulation Runs 1 and 2

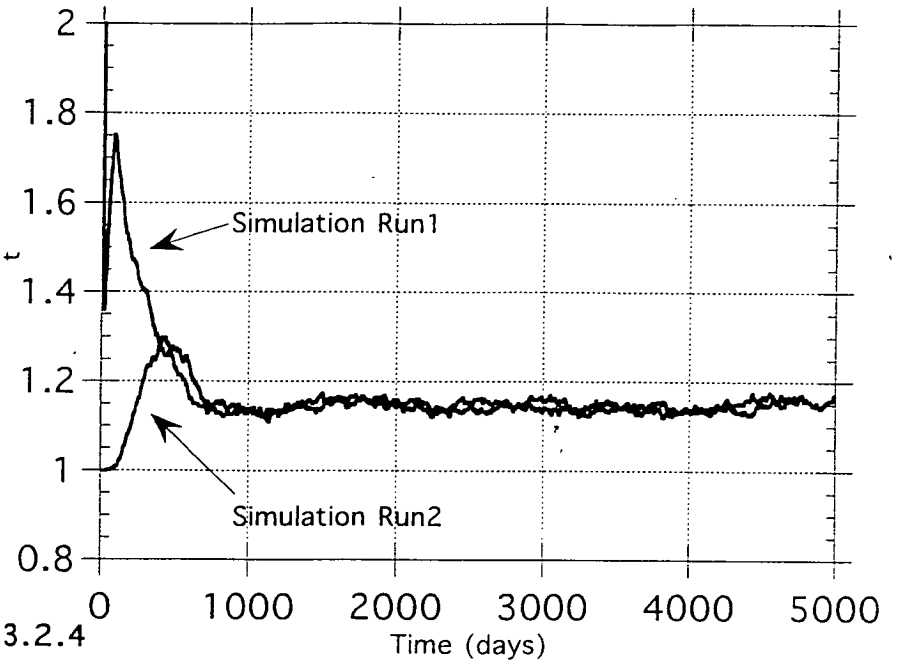


Fig. 3.2.4

Figure 3.2.4:
Values are calculated and plotted every 4 days over the 5,000 days examined. Simulation Run 1 was started with Initial Conditions A and Simulation Run 2 was started with Initial Conditions B. The Standard Parameter Values were used.

3.3) Discussion

In this chapter I have examined how various statistics from the hypothetical population change with time, given a set of standard parameter values. This was intended, in part, to justify the approach used in the next chapter, where the effects of manipulating the parameters of the model are examined. It is also useful for understanding the dynamics and epidemiology of the model.

Let us first turn our attention to P_{inf} and W. When the simulations began, no hosts were immune to any immuno-allelic alleles, and 20% of the hosts were infected. Once the simulations started, an 'epidemic' swept through much of the population in both simulations, causing a peak with a large number of hosts infected. The numbers infected then descended to a low value before approaching a stationary distribution. This "boom and bust" cycle is found in other epidemiological models of the spread of a disease through a susceptible population (e.g. Anderson and May 1992; Gupta *et al.* 1994a). It is likely that the reason for the epidemic rise and crash in P_{inf} is similar to the causes in these other models. Namely, as the infection spreads through the susceptible population, the numbers of hosts susceptible to new infection (those that are uninfected and not immune to parasite infection) become smaller. At some point the number of hosts infected cannot be sustained, as the rate of generation of new hosts is lower than the rate of generation of susceptible hosts (through loss of immune memory, or through birth of new hosts). At this point P_{inf} decreases until the numbers of susceptible hosts are sufficiently replenished for another (smaller) epidemic to occur.

In other models there are often several damped oscillations before an equilibrium state is achieved (e.g. Anderson and May 1992; Gupta *et al.* 1994a). In the equilibrium state there are enough susceptible hosts for any loss of infections through host death and immune clearance to be compensated for by the generation of new infections. Likewise, the loss of susceptible hosts through infection and immunity is compensated for by the birth of new hosts and the loss of immunity.

There are a number of factors that might explain the fact that, in the present model, a long series of cycles in the P_{inf} and W values do not occur, as they do in other, related models. Perhaps the most obvious factor is that, in the present study, the length of immune memory is short in comparison to the other models cited above. It has been shown that a short immune memory can suppress cycling (Gupta *et al*1994a). The present model differs from previous work in a number of other ways, and these other differences may also suppress cycling. For example, in the present study, a larger number of immuno-allelic alleles were used than is the case in previous work. Preliminary explorations of a modified model showed more cycling when the number of alleles was reduced.

It has been pointed out in Section 2.3 that the model employs a simplification of vector transmission, in that transmission occurs in a single time step, whereas in a real population, there would be several days between a mosquito gaining an infection from a host and transmitting it to another host (the "Extrinsic Incubation Period"). This is unlikely to affect the values of statistics at the stationary distribution, but may affect the dynamic epidemiology of the model. Although the effects of this lag have not been examined, it is possible to speculate as to its effects on the results. A lag between the initiation of a transmission event and its conclusion is likely to

increase the sizes of the epidemic boom and bust cycle in the epidemic phase of the infection, due to the lag in the negative-feedback loops inherent in the model. Thus, in a model that took into account the Extrinsic Incubation Period, one might expect to find a larger initial epidemic rise and crash in P_{inf}, and more cycles before a stationary distribution is reached than is found in the model presented here.

The proportion of hosts infected in this model stabilised at around 65%. There is a wide variety of estimates of prevalence of *P. falciparum* infections in different studies and in different parts of the world. Much of this may be due to differences in transmission rates in the populations. This can be shown by examining the seasonal variation in prevalence in a region. In the Garki project in Nigeria (Molineaux and Gramiccia 1980), prevalence levels were found to vary from approximately 45% to around 60% from the wet season to the dry season. Unfortunately, estimates of prevalence in may studies are likely to be low. This has been shown by the recent use of PCR (polymerase chain-reaction) technology where it has been found that many low parasitaemias in hosts go undetected when the more traditional, microscopical methods are used (e.g. Felger *et al.* 1995; Roper *et al.* 1996). Thus, it is only possible to say that the prevalence rates in this study are in the right order of magnitude for many field studies.

It should also be noted that, in this model, all malaria infections are infectious to the vector (though to differing levels depending on the parasite density). This does not appear to be the case in malaria populations in the field, where hosts may be infected with malaria parasites, but are not infectious to mosquitoes. Gupta and co-workers (Gupta*et al*, Swinton and Anderson 1994a) estimated that only one quarter of malaria infections are in fact infectious. Thus, if the value of P_{inf} is compared purely to the

prevalence of infectious hosts in a natural population, it is likely to be too high.

The large degree of variation between the time points sampled for both Z_{inf} and P_v can be explained by the small number of transmission events that occur in any one time step. Only about 65% of the 6000 hosts were infected for most of the simulation run and on average, 0.01 potential transmission events per infected host may occur during a single transmission event. Thus, there would be approximately 36 potential transmission events in any time step sampled. With such a small number of events one would expect a wide variation in the means of P_v and Z_{inf}.

When each point plotted is an average over a large number of time steps, the variance is, as one would expect, much less. There seems to have been very little variation through time in the level of either P_v or Z_{inf} when averaged over time. P_v did show a small rise and fall at the beginning of the run, and then remained roughly constant for the rest of the time course. As these two statistics depend on the degree of infection in each infected host, and not the number of hosts infected, it would appear that, though there were quite wide initial fluctuations in P_{inf}, there was very little change in the distribution of parasite densities among the infected hosts through the simulation run.

It is, unfortunately, very difficult to compare the value of P_v to that found in the field. The statistic, P_v measures the proportion of vectors that, having bitten infected hosts, actually acquire an infection. This is exceedingly difficult to measure in a field situation. It has, however, been studied by feeding laboratory reared mosquitoes on naturally infected hosts in Papua New Guinea. In this case, P_v was found to have a value of 0.38

(Graves *et al.* 1988), as opposed to the value at the stationary distribution in this study of approximately 0.48.

The mean number of zygotes found per infected mosquito (Z_{inf}) was found to have a value of approximately 1.54 at the stationary distribution. This is very close to the maximum, sustainable value for Z_{inf} of approximately 1.58 ($1/(1-e)^{-c}$, with c=1.0), which would only occur if all infected hosts had parasite densities equal to the Carrying Capacity (K). Thus it must be the case that, with the Standard Parameter Values, most infected hosts do indeed have very high parasite densities, either at or near the Carrying Capacity.

Oocyst counts from natural mosquito populations have only rarely been examined (Billingsley *et al.* 1994). However, from the data of Billingsley (Billingsley *et al.* 1994) and Pringle (quoted in Pichon *et al.* 1996) estimated values for Z_{inf} of 7.22 and 13.8, respectively, can be made. These values are much higher than that measured in this study. There appear to be two reasons for this. Firstly, as stated in Section 2.4, the value of c chosen for the Standard Parameter values is too low, and therefore Z_{inf} could not be as high as that found in natural mosquito populations. Secondly, the basis for the distribution of zygotes amongst vectors in this model is a Poisson Distribution (PD), whereas the distributions of oocysts found in these studies appear to be Negative Binomial Distributions (NBDs). NBDs are likely to produce more mosquitoes with a large number of oocysts than a PD is. This is likely to increase the mean value of oocysts in infected mosquitoes compared to a PD. Thus, even if the same mean number of oocysts (across all vectors) is found, a population with an NBD would have a higher value for Z_{inf} than one with a PD.

It is important to note that, despite very different starting values for G_{ST} at the two loci and in the two simulation runs, in both cases the G_{ST} values appear to achieve stationary distributions in less than 1,000 days. It should also be noted that Hastings's prediction that $G_{ST(d)}$ would be greater than $G_{ST(n)}$ (Hastings 1996) holds up under this particular set of parameters. This is the case, not only when a stationary distribution has been reached, but during the initial changes in values too.

The value of $G_{ST(d)}$ in real malaria parasites is, of course, not known, as it is only postulated that an immuno-allelic locus truly exists. However, several estimates of parameters closely related to $G_{ST(n)}$ have been made in several parts of the world, suggesting $G_{ST(n)}$ values from approximately 0.3 in Tanzania (e.g.Babiker *et al.* 1994; Hill *et al.* 1995) to approximately 0.9 in Papua New Guinea (Paul *et al.* 1995). The value at the stationary distribution in these simulations of $G_{ST(n)} = 0.82$, is clearly in this range, and closer to that found in Papua New Guinea than Tanzania.

The value of t in both simulations is always found to be above 1.0, and produced a peak value well above that found at the stationary distribution. This is important to note, for it suggests that if one plans to look for an immuno-allelic locus in a malaria population in the field, even if that population appears to be far from a stationary distribution, then the difference between $G_{ST(d)}$ and $G_{ST(n)}$ may well still be detectable. In fact, it may even be greater at the time studied than when the stationary distribution is reached. In both simulations, the stationary distribution was achieved within 1000 days.

It should also be noted that for each statistic examined in Simulation Run 1 and Simulation Run 2, a stationary distribution appears to have been

realised well before day 4000 when observations started to be collected to produce a mean over time. Furthermore, even in statistics which showed great variation from one time step to the next, such as the mean number of zygotes transmitted by infected vectors (Z_{inf}), the mean with time showed very little variation between repeated simulations. It should also be noted that no significant differences between the two starting conditions are found in the mean values of any of the statistics collected. This helps support the view that a stationary distribution has indeed been reached by day 4000.

4) The Stationary Distributions of the Model

4.1) Introduction

In this chapter, the parameters of the model are manipulated, and the effect of this on the values of the output statistics at the stationary distribution are examined. It is important to examine the effects of parameter manipulations for a number of reasons:

1) Without examining the effects of manipulations of parameters, it is very difficult to gain any understanding as to the underlying processes acting in the model, and the relative magnitude of their effects.

2) The choice of standard parameter values, though intended to be as close to the true nature of malarial biology as possible, may be unrealistic, or may vary from place to place in the a natural population (e.g the host population size and the Double Bite Rate). To understand the effects an inappropriate value for a parameter may have, it is important to alter its value and observe what effects (if any) occur. The model may be very sensitive to the precise values of some parameters, but relatively insensitive to the values of others.

3) It is also the case that different anti-malaria intervention strategies are likely to have different effects on various aspects of the parasite's ecology. Anti-mosquito strategies are likely to reduce the Vectorial Capacity in a population (in this model, this is proportional to the Double Bite Rate, see Section 2.4). However, a vaccine may reduce the size of the susceptible population (equivalent to reducing the size of N), or reduce the

period an infection may last (equivalent to increasing the rate of immune recognition (ρ)), or even reduce the numbers of immunologically distinct strains in a population (equivalent to reducing the value of D in this model). Thus, to understand the possible effects of malaria control strategies on the population genetics of parasites, one needs to understand the effects of different parameter values on the outcomes of this model.

As has been stated in Section 2.3.3, restrictions on computer time prevented a thorough examination of the parameter space of the model. Thus it was decided to alter each parameter independently against a background of standard values for the other parameters. For further discussion see Section 2.3.3.

4.2) Results

4.2.1) Host population size (N)

A range of host population sizes from 2000 hosts to 10,000 was examined. As can be seen from Table 4.2.1a, there are no discernible effects of population size on any of the output statistics examined. To confirm this, a larger number of simulations were performed with N=2,000 and N=10,000 so that a more powerful analysis could be done. A total of ten runs were done with each of the two population sizes, and their means, variances, and standard errors are shown in Table 4.2.1b.

To compare the mean values between these two population sizes for each output statistic, Welch's Approximate t-Test (Sokal and Rohlf 1995) was used. The results of this test for each output statistic is presented in Table 4.2.1b along with information on the power of the test. It can be seen that there is no significant difference in the mean value for any of the output statistics with the two different population sizes. The power of the test shows that, in most cases, less than a 1% difference in mean value of these statistics would be significant at the p=0.05 level. Therefore, it is very unlikely that a substantial difference in the output statistics would have been missed.

Table 4.2.1b also presents the results of the F_{max} Test (Sokal and Rohlf 1995) for each output statistic. This test compares the variances of different samples. The table shows that, for all of the output statistics, there is a significantly larger variance among the mean values of each statistic for each simulation run when the simulations are done with the small

population size (N=2000) compared to when the larger population size (N=10,000) is used.

The statistical tests described above assume a Normal Distribution of the data analysed. Although there is no strong evidence for normality, it is a useful working hypothesis, especially as the tests used are based on F statistics which are robust to deviation from the Normal Distribution. Only if data is highly asymmetric (i.e. highly skewed) is there a much increased chance of error (J. Haig, University of Sussex, pers. com.).

In Figures 4.2.1a-g, a graphical representation of the data used for each analysis is shown. It can be seen that the data are not highly asymmetrical and thus, the use of F statistics seems to be appropriate. It is assumed that the data presented in this chapter are drawn from the same family of distributions, and thus the application of F statistics is appropriate in all cases.

The Effects of Population Size (N) on the Output Statistics of the Model.

N	$\bar{\bar{G}}_{ST(d)}$	$\bar{\bar{G}}_{ST(n)}$	$\bar{\bar{t}}$	$\bar{\bar{P}}_{inf}$	$\bar{\bar{Z}}_{inf}$	$\bar{\bar{P}}_{V}$	$\bar{\bar{W}}$
2,000	**0.9416505**	**0.834956**	**1.1278095**	**0.662226**	**1.541459**	**0.4867245**	**5.135239**
	$7.82\ 10^{-4}$	$4.20\ 10^{-3}$	$4.74\ 10^{-3}$	$1.14\ 10^{-3}$	$4.38\ 10^{-3}$	$4.35\ 10^{-5}$	0.0203
4,000	**0.94065**	**0.824996**	**1.140189**	**0.664849**	**1.546047**	**0.487698**	**5.170577**
	$4.87\ 10^{-4}$	$1.14\ 10^{-3}$	$9.79\ 10^{-4}$	$2.07\ 10^{-3}$	$3.09\ 10^{-4}$	$1.63\ 10^{-3}$	0.017
6,000	**0.939880**	**0.819780**	**1.146547**	**0.6682456**	**1.542189**	**0.4866401**	**5.172673**
	$1.55\ 10^{-4}$	$5.57\ 10^{-3}$	$7.51\ 10^{-3}$	$4.12\ 10^{-3}$	$5.53\ 10^{-3}$	$1.53\ 10^{-3}$	0.0136
10,000	**0.9404155**	**0.8209145**	**1.145572**	**0.666747**	**1.545596**	**0.487579**	**5.166396**
	$1.52\ 10^{-4}$	$2.95\ 10^{-3}$	$1.31\ 10^{-3}$	$2.48\ 10^{-4}$	$7.53\ 10^{-4}$	$1.20\ 10^{-3}$	0.013

Table 4.2.1a:
In most cases, the means and standard errors of each statistic are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for N=6000, in this case, eight runs were done, four with Initial Condition A and four with Initial Condition B. The mean values of each statistic are presented in bold, the standard errors are in parentheses underneath their respective means. Apart from the parameter N, the Standard Parameter Values were used.

The Effects of Population Size (N) on the Output Statistics of the Model

N	$\bar{\bar{G}}_{ST(d)}$	$\bar{\bar{G}}_{ST(n)}$	$\bar{\bar{t}}$	$\bar{\bar{P}}_{inf}$	$\bar{\bar{Z}}_{inf}$	$\bar{\bar{P}}_{V}$	$\bar{\bar{W}}$
2000	**0.9403**	**0.8249**	**1.1399**	**0.6665**	**1.5474**	**0.4880**	**5.1549**
	[2.21 x 10^{-6}]	[7.17 x 10^{-5}]	[1.118 x 10^{-4}]	[3.086 10^{-5}]	[9.788 x 10^{-5}]	[1.708 x 10^{-5}]	[1.000 10^{-3}]
	(4.700 x 10^{-4})	(2.67 x 10^{-3})	(3.34 x 10^{-3})	(1.757 x 10^{-3})	(3.129 x 10^{-3})	(1.307 x 10^{-3})	(0.0100)
10,000	**0.9402**	**0.8211**	**1.14507**	**0.6682**	**1.5444**	**0.48726**	**5.1693**
	[3.467 x 10^{-7}]	[8.770 x 10^{-6}]	[1.722 x 10^{-5}]	[3.9102 x 10^{-6}]	[2.725 x 10^{-5}]	[4.781 x 10^{-6}]	[1.383 x 10^{-4}]
	(1.862 x 10^{-4})	(9.64 x 10^{-4})	(1.312 x 10^{-3})	(6.253 x10^{-4})	(1.651x 10^{-3})	(6.914 x 10^{-4})	(3.720 x 10^{-3})
t'_S	0.198 ns	1.340 ns	1.439 ns	0.911 ns	0.848 ns	0.500 ns	1.350 ns
Power	1.14 x 10^{-3}	6.39 x 10^{-3}	8.08 x 10^{-3}	4.21 x 10^{-3}	7.98 x 10^{-3}	3.34 x 10^{-3}	0.0241
F_{max}	6.37*	8.18**	6.49*	7.89**	3.59*	3.57*	7.23**

Table 4.2.1b:
The mean, variance, and standard error for each statistic are calculated from 10 simulation runs, five with Initial Condition A, the other five with Initial Condition B. The mean values of each statistic are presented in bold, the variances are presented underneath their associated means in square brackets and the standard errors are under those in parentheses. Apart from the parameter N, the Standard Parameter Values were used. The row labelled " t'_S " shows values of t'_S, calculated for Welche's Approximate t Test (Sokal and Rohlf 1995) for each output statistic. The letters "ns" beside a value of t'_S signifies this value was not found to be significant at the p=0.05 level. The Row labelled "Power" gives values for the minimum difference between means which would be statistically significant ($p<0.05$) using Welche's Approximate t Test (The calculation of this value is explained in Appendix 3). The row labelled "F_{max}" shows values of the statistic F_{max} calculated for the F_{max} test (Sokal and Rohlf 1995). The symbol "**" beside an F_{max} value signifies that the difference in variances was found to be significant at the p=0.01 level, whereas the letters "ns" signify that the statistic was not found to be significant at the p=0.05 level.

The Effects of Population Size (N) on the Output Statistics of the Model.

Fig. 4.2.1a

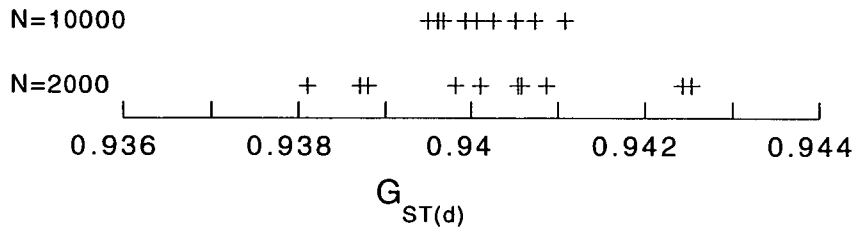


Fig. 4.2.1b

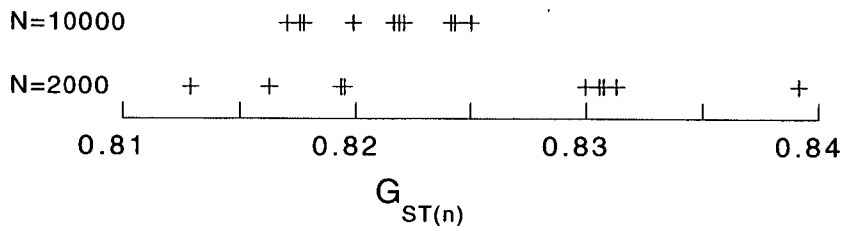


Fig. 4.2.1c

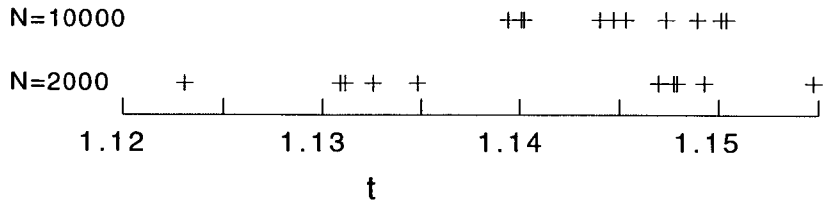
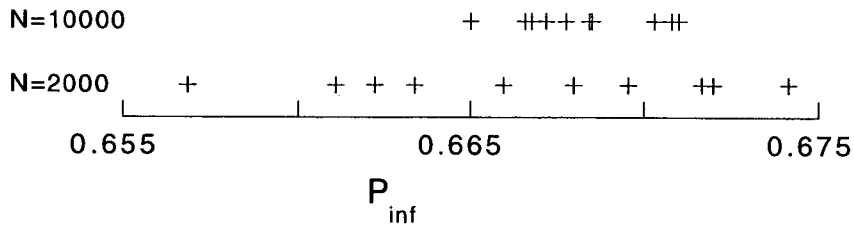


Fig. 4.2.1d



Figures 4.2.1a-g (Figures 4.2.1e-g overleaf):
The values of the output statistics are presented for the 10 simulation runs (five with Initial Condition A, the other five with Initial Condition B) with two different values for N. Apart from the parameter N, the Standard Parameter Values were used.

The Effects of Population Size (N) on the Output Statistics of the Model (continued).

Fig. 4.2.1e

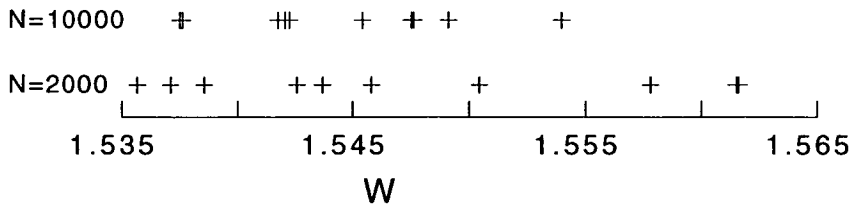


Fig. 4.2.1f

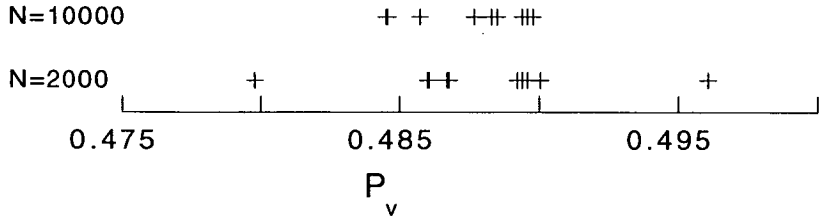
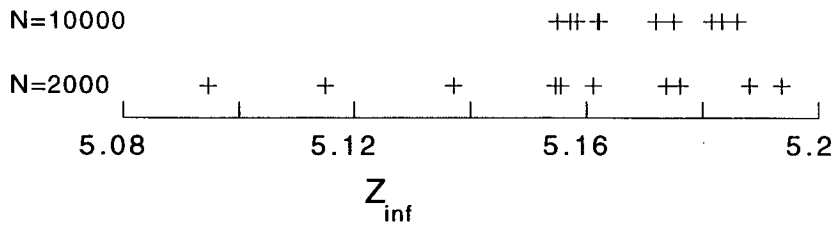


Fig. 4.2.1g



4.2.2) Aspects of transmission

4.2.2.1) Double Bite Rate (T)

The rate at which potential transmission events occur can vary from place to place in the real world, and will depend on the densities of both mosquitoes and humans. A range of Double Bite Rates from 0.05 per day to 2.0 per day was examined.

In Figure 4.2.2.1a, the upper, dashed line represents the values of $\overline{\overline{G}}_{ST(d)}$ as a function of T, the Double Bite Rate. It can be seen that, as T increases, $\overline{\overline{G}}_{ST(d)}$ appears to fall towards an asymptotic value of around 0.9. The value of $\overline{\overline{G}}_{ST(d)}$ decreases from 0.982 to 0.940 (an average rate of change of 0.70 per unit of T) when T is increased from 0.05 to 0.1 per day. However it decreases from 0.900 to 0.898 (an average rate of change of 0.02 per unit of T) with a change in T from 1.0 to 2.0 per day. Thus, over the range of T examined the value of $\overline{\overline{G}}_{ST(d)}$ drops slightly below 0.9 and appears to be approaching or have arrived at an asymptotic value at the higher levels of T.

The lower, solid line in Figure 4.2.2.1a represents the values of $\overline{\overline{G}}_{ST(n)}$. The effects of T on $\overline{\overline{G}}_{ST(n)}$ is much greater than that on $\overline{\overline{G}}_{ST(d)}$. As with $\overline{\overline{G}}_{ST(d)}$, as T increases, the value of $\overline{\overline{G}}_{ST(n)}$ goes down. With $\overline{\overline{G}}_{ST(n)}$ the value decreases from 0.953 to 0.823 (an average rate of change of 2.6 per unit of T) with the increase of T from 0.05 to 0.1 per day, and from 0.437 to 0.340 (an average rate of change of 0.097 per unit of T) with the increase of T from 1.0 to 2 per day. Thus, although the rate of change in $\overline{\overline{G}}_{ST(n)}$ with T is greater than the rate of change in $\overline{\overline{G}}_{ST(d)}$ for any range of values if T examined, in both cases, the rate of change with T also declines as T

increases. No asymptotic value for $\overline{\overline{G}}_{ST(n)}$ is found over the range of values of T examined.

Figure 4.2.2.1a clearly shows that, as T increases, not only do the values for $\overline{\overline{G}}_{ST}$ at the two loci decrease, but the difference between $\overline{\overline{G}}_{ST(d)}$ and $\overline{\overline{G}}_{ST(n)}$ increases. This is shown in Figure 4.2.2.1b, where the effect of T on $\overline{\overline{t}}$ is shown. The value of $\overline{\overline{t}}$ rises from just above one (1.03) to above two (2.64) over the 40 fold range T examined.

In Figure 4.2.2.1c, the effect of T on $\overline{\overline{W}}$, the mean level of host immunity, is represented by the solid line. The larger the value of T, the larger the value of $\overline{\overline{W}}$, although the rate of increase with T of $\overline{\overline{W}}$ declines as T increases. Thus, the value of $\overline{\overline{W}}$ is very low (2.47) when T=0.05 per day, rising steeply to a value of 5.17 when T=0.1 per day. However when T was increased from 1.0 per day to 2.0 per day, the value of $\overline{\overline{W}}$ was found to increase only from 7.97 to 8.04. At this point $\overline{\overline{W}}$ appears to be close to reaching an asymptotic value.

Figure 4.2.2.1c also shows the effect of T on the statistic $\overline{\overline{P}}_{inf}$, the mean proportion of hosts infected. This is represented by the broken line in the figure. As with the value of $\overline{\overline{W}}$, as the value of T increases, so does the value of $\overline{\overline{P}}_{inf}$ and the rate of increase with T declines. The value of $\overline{\overline{P}}_{inf}$ appears to approach an asymptote, in this case, with a value of around 0.88 at the larger values of T.

The changes in the value of $\overline{\overline{P}}_{v}$ with T are shown in Table 4.2.2.1a. These changes are small, and thus statistical analysis to see whether they were explicable by chance alone seemed appropriate. Examination of the estimated variances for each value of $\overline{P}_{v}$ suggests that the variance may be varying greatly between the treatments, and therefore a simple ANOVA

analysis is inappropriate. This is because, when levels of the independent variable lead to different levels of within-group variance, it is not valid simply to average the estimates of within-group variance and then use this average in the ANOVA. To deal with this problem, standard ANOVAs are carried out, but for the estimate of within-group variance (the MS_{within}) the largest sample variance found in any of the treatment groups is used (the degrees of freedom are unchanged from a standard ANOVA). Thus, in the present instance, the sample variance of 7.37×10^{-6} is used (see Table 4.2.2.1b). This procedure involves a loss of statistical power, but it ensures that the probability of type-1 errors is less than the stated p values. In other words, the procedure is crude, but rigorous.

Using this method, despite the loss of statistical power, a significant difference (at the p=0.05 level) was found in the values of $\overline{\overline{P}}_v$ with different values of T. This result is presented in Table 4.2.2.1b.

The effects of T on $\overline{\overline{Z}}_{inf}$ is also shown in Table 4.2.2.1a. There is little change in the values of $\overline{\overline{Z}}_{inf}$ with each value of T. The Modified Anova analysis of the values of $\overline{\overline{Z}}_{inf}$ with T is presented in Table 4.2.2.1c. No significant difference in the values of $\overline{\overline{Z}}_{inf}$ with different values of T was found at the p = 0.05 level.

The changes observed in both $\overline{\overline{P}}_v$ and $\overline{\overline{Z}}_{inf}$ are small. Thus even if these differences are not caused by chance, but by some underlying trend, these differences seem unlikely to have a major bearing on the main statistics of interest (namely $\overline{\overline{G}}_{ST(d)}$, $\overline{\overline{G}}_{ST(n)}$ and $\overline{\overline{t}}$).

As described in the methods section, only a small number of simulations are used to calculate the values of each output statistic for each set of parameter values used. Thus accurate measures of variance are

difficult to obtain. During the rest of this chapter, when Anova analysis is needed, the Modified Anova Analysis described above is used.

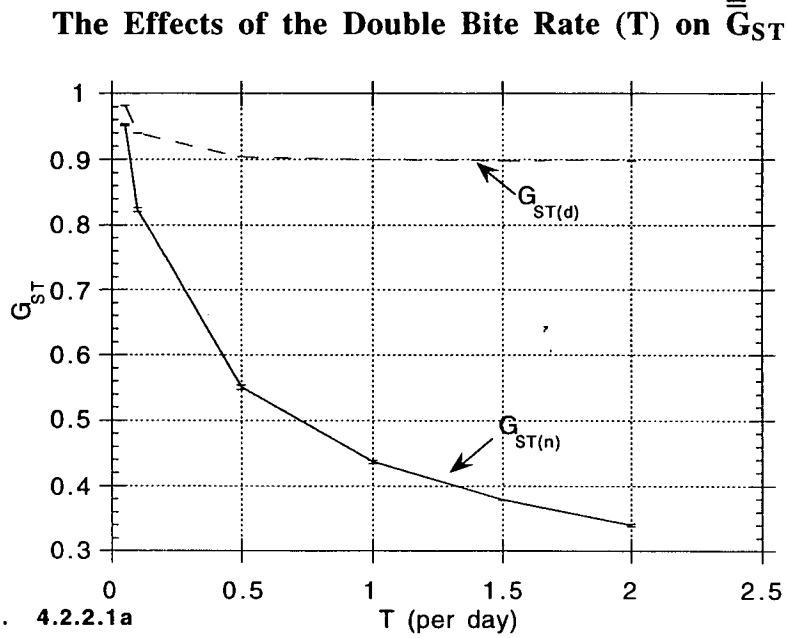


Fig. 4.2.2.1a

Figures 4.2.2.1a - c (Figures 4.2.2.1b-c overleaf):
Mean values are plotted with their standard errors represented as error bars. In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for T=0.1 per day, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Apart from the parameter T, the Standard Parameter Values were used.

The Effects of the Double Bite Rate (T) on $\bar{\bar{t}}$

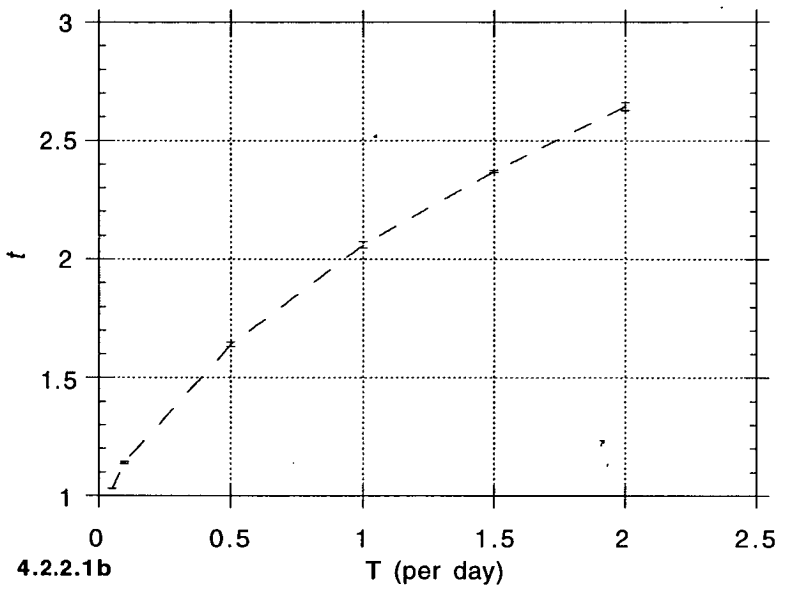


Fig. 4.2.2.1b

The Effects of the Double Bite Rate (T) on $\bar{\bar{P}}_{inf}$ and $\bar{\bar{W}}$

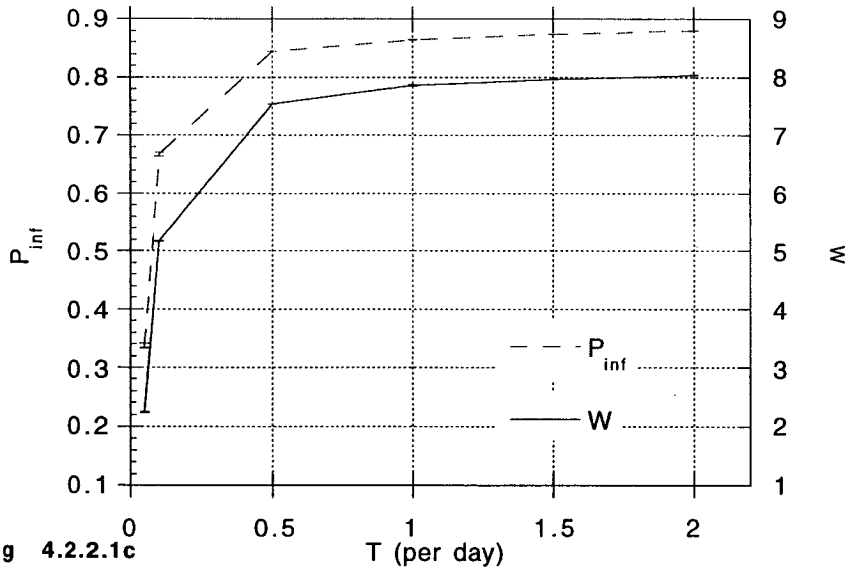


Fig 4.2.2.1c

The Effects of Double Bite Rate (T) on $\bar{\bar{P}}_v$ and $\bar{\bar{Z}}_{inf}$

T (per day)	$\bar{\bar{P}}_v$	$\bar{\bar{Z}}_{inf}$
0.05	**0.44893**	**1.5472**
	[9.94 x 10^{-8}]	[1.10 x 10^{-6}]
	(2.23 x 10^{-4})	(7.42x 10^{-4})
0.1	**0.486640**	**1.54219**
	[2.35 x 10^{-6}]	[3.06 x 10^{-4}]
	(5.42 x 10^{-4})	(1.95 x 10^{-3})
0.5	**0.51278**	**1.54609**
	[7.37 x 10^{-6}]	[1.16 x 10^{-5}]
	(1.92 x 10^{-3})	(2.41 x 10^{-3})
1.0	**0.51270**	**1.5453**
	[8.82 x 10^{-10}]	[3.98 x 10^{-7}]
	(2.1 x 10^{-5})	(4.46 x 10^{-4})
1.5	**0.51095**	**1.5456**
	[1.25 x 10^{-8}]	[1.79 x 10^{-6}]
	(7.9 x 10^{-5})	(9.45 x 10^{-4})
2.0	**0.50793**	**1.54578**
	[1.19 x 10^{-6}]	[4.51 x 10^{-9}]
	(7.73 x 10^{-4})	(4.75 x 10^{-5})

Table 4.2.2.1a:
In most cases, the means, variances and standard errors of $\bar{P}_v$ and $\bar{Z}_{inf}$ are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for T= 0.1 per day, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. The mean value for each statistic is presented in bold, while the variances are presented underneath their associated means in square brackets and the standard errors are under those in parentheses Apart from the parameter T, the Standard Parameter Values were used.

Modified Anova Analysis of $\bar{P}_v$ with different values for T.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	6.855×10^{-3}	1.371×10^{-3}	186**
Within Treatments	12	(1.66×10^{-5})	7.37×10^{-6} †	
Total	17	6.872×10^{-3}		

Table 4.2.2.1b

Modified Anova Analysis of $\bar{Z}_{inf}$ with different values for T.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	6.87×10^{-5}	1.37×10^{-4}	0.447 ns
Within Treatments	12	(2.29×10^{-4})	3.06×10^{-4} †	
Total	17	2.98×10^{-4}		

Table 4.2.2.1c

Tables 4.2.2.1b and c:
SS signifies the "Sum of Squares", while MS signifies the "Mean Square". The SS value for the Within Treatment row is in brackets to show that this value was not used to calculate the Within Treatment MS. Instead the largest Within Treatment Variance was used (see Section 4.2.2.1). This largest Within Treatment Variance is labelled with the symbol "†". The symbol "**" beside a value of Fs signifies that the statistic was found to be significant at the p=0.01 level, whereas the letters "ns" signify that the statistic was not found to be significant at the p=0.05 level.

4.2.2.2) Coefficient of vector infectivity (c)

The parameter c denotes the degree of infectiousness of the parasites within hosts to the vectors. The larger the value of c, both the greater the probability that double bites from an infected host actually transmit parasites to another host, and also, the greater the number of zygotes formed in such bites (see Section 2.1).

In Figure 4.2.2.2a, the effects of c on the $\overline{\overline{G}}_{ST}$ values at the two loci under study can be seen. The upper, broken line represents the effect of c on $\overline{\overline{G}}_{ST(d)}$. It can be seen that the value of $\overline{\overline{G}}_{ST(d)}$ appears to descend asymptotically from a value of around 0.97 when c=0.5 towards a value of around 0.93 when c has a value of 2.5.

The lower, solid line in Figure 4.2.2.2a represents the values of $\overline{\overline{G}}_{ST(n)}$ with different values of c. As with $\overline{\overline{G}}_{ST(d)}$, as c is increased, the value of $\overline{\overline{G}}_{ST(n)}$ decreases. The rate of decrease with c declines as c gets larger. Unlike the value of $\overline{\overline{G}}_{ST(d)}$, that of $\overline{\overline{G}}_{ST(n)}$ does not appear to be approaching an asymptote in the range of values examined.

In Figure 4.2.2.2b, the value of $\overline{\overline{t}}$ is shown in relation to c. It can be seen that as c increases, so does $\overline{\overline{t}}$, and that this rise appears to be roughly linear over the range of values examined. It is found that $\overline{\overline{t}}$ has a value of 1.05 when c=0.5, and a value of 1.38 when c=2.5.

In Figure 4.2.2.2c, the upper, broken line represents the value of $\overline{\overline{P}}_{inf}$ (the mean proportion of hosts infected) with different values of c. The value of $\overline{\overline{P}}_{inf}$ increases with increased values of c. For example, when c=0.5, $\overline{\overline{P}}_{inf}$ = 0.48 and when c=2.5, $\overline{\overline{P}}_{inf}$ = 0.75. Although $\overline{\overline{P}}_{inf}$ increases with increasing values of c, the rate of increase declines as c gets larger.

The lower, solid line in Figure 4.2.2.2c represents the values of $\overline{\overline{W}}$ with different values of c. As c is increased, $\overline{\overline{W}}$ is found to rise too. As with $\overline{\overline{P}}_{inf,}$ the rate of increase of $\overline{\overline{W}}$ with c declines as c increases.

The upper, solid line in Figure 4.2.2.2d represents the value of $\overline{\overline{P}}_v$. As c is increased, the value of $\overline{\overline{P}}_v$ increases. Once again, the rate of increase in $\overline{\overline{P}}_v$ decreases with increasing values of c.

The lower, broken line in Figure 4.2.2.2d represents the values $\overline{\overline{Z}}_{inf}$. The value of $\overline{\overline{Z}}_{inf}$ increases from 1.26 to 2.58 over the range of c examined.

The Effects of the Coefficient of Vector Infectivity (c) on $\bar{\bar{G}}_{ST}$

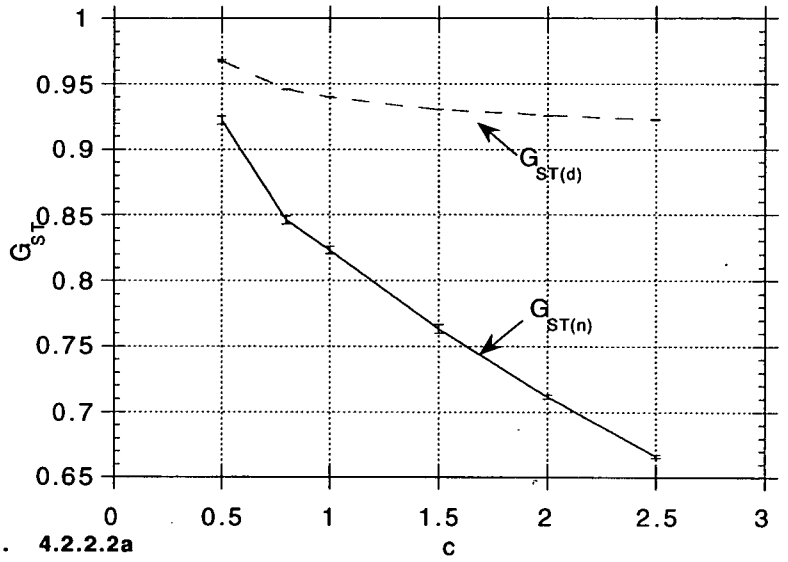


Fig. 4.2.2.2a

The Effects of the Coefficient of Vector Infectivity (c) on $\bar{\bar{t}}$

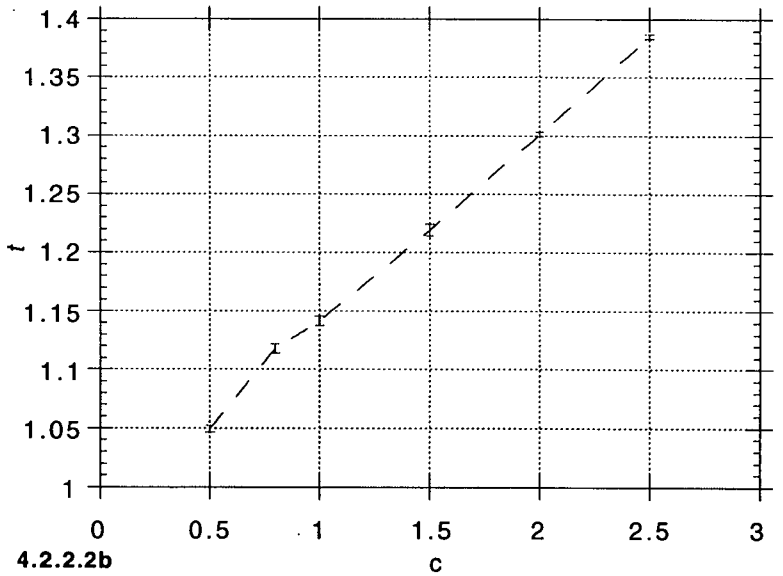


Fig. 4.2.2.2b

Figures 4.2.2.2a - d (Figures 4.2.2.2c - d on following pages): Mean values are plotted with their standard errors represented as error bars. In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for c=1.0, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Apart from the parameter c, the Standard Parameter Values were used.

The Effects of the Coefficient of Vector Infectivity (c) on $\bar{\bar{P}}_{inf}$ and $\bar{\bar{W}}$

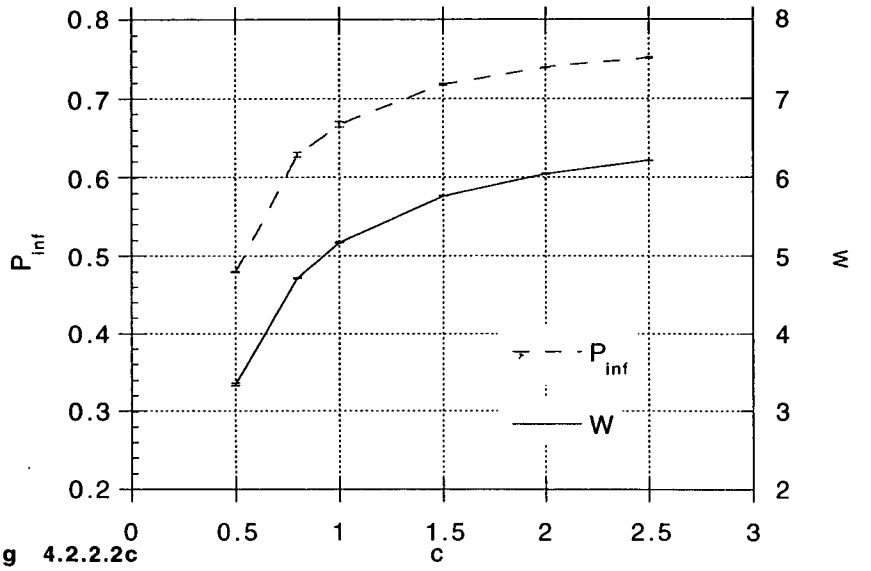


Fig 4.2.2.2c

The Effects of the Coefficient of Vector Infectivity (c) on $\bar{\bar{P}}_{v}$ and $\bar{\bar{Z}}_{inf}$

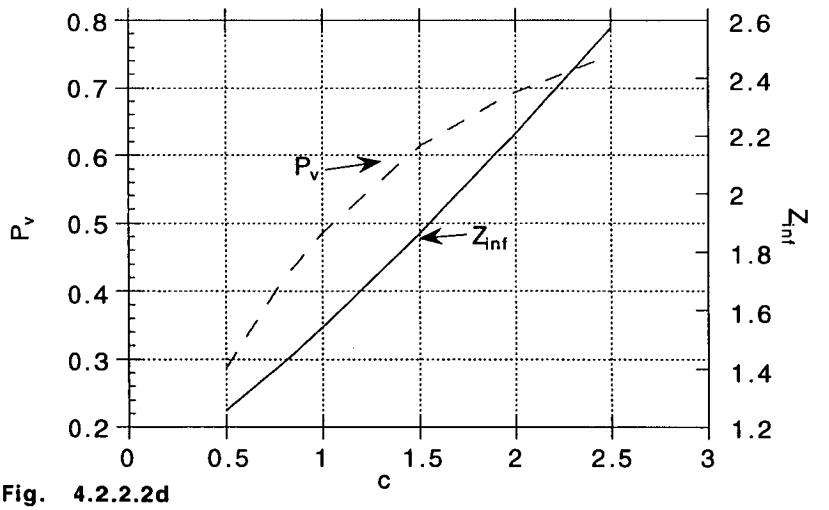


Fig. 4.2.2.2d

4.2.2.2.1) Modifying the model to hold the numbers of zygotes constant

As has been stated above, the value of c has two direct effects in the model. Firstly, c controls the numbers of zygotes transmitted by an infected vector. Secondly, the value of c also effects the proportion of vectors that transmit parasites from one host to another. To separate these effects, simulations were run with different values of c, but with the numbers of zygotes that any infected mosquito transmitted artificially fixed at either one, two, or three, depending on the specific simulation run. Thus the effects of c could, at least in part, be distinguished.

In Figure 4.2.2.2.1a, the effects of c can be seen on the values of $\overline{\overline{G}}_{ST}$ at the two loci. The upper group of lines, shows the effects of c on $\overline{\overline{G}}_{ST(d)}$ under the conditions of having the number of zygotes fixed at either one, two or three zygotes. These lines are very difficult to distinguish by eye, with the greatest difference between any two values being less than 0.5%. However, the differences between these three conditions is found to be statistically significant (see Table 4.2.2.2.1a for analysis). The shape and position of this group of lines is very similar to the line formed when $\overline{\overline{G}}_{ST(d)}$ was plotted against c with the number of zygotes not fixed (see Figure 4.2.2.2a).

The lower group of lines in Figure 4.2.2.2.1a represent the values of $\overline{\overline{G}}_{ST(n)}$ when the number of zygotes are fixed. For each value for the number of zygotes, as c increases, $\overline{\overline{G}}_{ST(n)}$ declines. This is also the case when the number of zygotes is not fixed (see Figure 4.2.2.2a). The rate at which $\overline{\overline{G}}_{ST(n)}$ decreases with increasing c is, however, less when the number of zygotes is fixed than when it is not. It is also evident that for any particular

value of c, the value of $\bar{\bar{G}}_{ST(n)}$ decreases with the numbers of zygotes present.

Figure 4.2.2.2.1b shows the effect of c on the value of $\bar{\bar{t}}$. It can be seen that whether the number of zygotes are fixed to one, two or three, $\bar{\bar{t}}$ increases with increasing values of c The rate of increase in $\bar{\bar{t}}$ with c becomes greater as the numbers of zygotes are increased.

If we now examine Figure 4.2.2.2.1c, we can look at the effects of c, with a fixed numbers of zygotes, on the value of $\bar{\bar{P}}_{inf}$. The lines representing this effect under the conditions of one, two or three zygotes are tightly bunched together. Statistical analysis using the Modified Anova (see Table 4.2.2.2.1b) showed that any differences in value of $\bar{\bar{P}}_{inf}$ between the three different conditions can be explained by chance alone. If this group of lines are compared to the line representing the value of $\bar{\bar{P}}_{inf}$ with different values of c when the numbers of zygotes was not fixed (see Figure 4.2.2.2.1c), then it can be seen that the shape and position of these line are very similar to those when the number of zygotes was not fixed.

The other group of lines in Figure 4.2.2.2.1c represents the mean proportion of double bites, biting from infected hosts, that transmit parasites ($\bar{\bar{P}}_v$), as c is altered, under the three different conditions. Once again, these lines are tightly bunched together, and no significant effect is found when the numbers of zygotes are altered on the values of $\bar{\bar{P}}_v$ (see Table 4.2.2.2.1c).

Figure 4.2.2.2.1d shows the effects of c on the mean level of host immunity ($\bar{\bar{W}}$). Under all three conditions, when the value of c is increased, the value of $\bar{\bar{W}}$ increases as well. However, the greater the number of zygotes, the greater the value of $\bar{\bar{W}}$ for any particular value of c. Though this difference is slight, it is significant (see Table 4.2.2.2.1d).

The Effects of the Coefficient of Vector Infectivity (c) on $\bar{\bar{G}}_{ST}$ with Fixed Numbers of Zygotes.

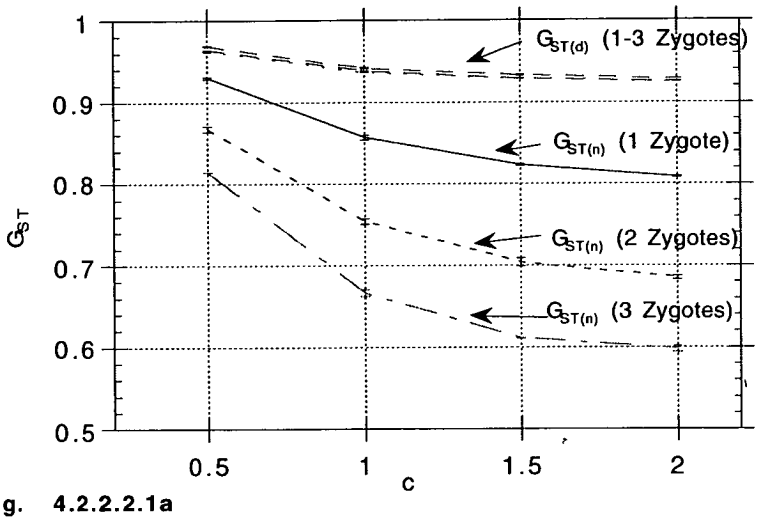


Fig. 4.2.2.2.1a

The Effects of the Coefficient of Vector Infectivity (c) on $\bar{\bar{t}}$ with Fixed Numbers of Zygotes.

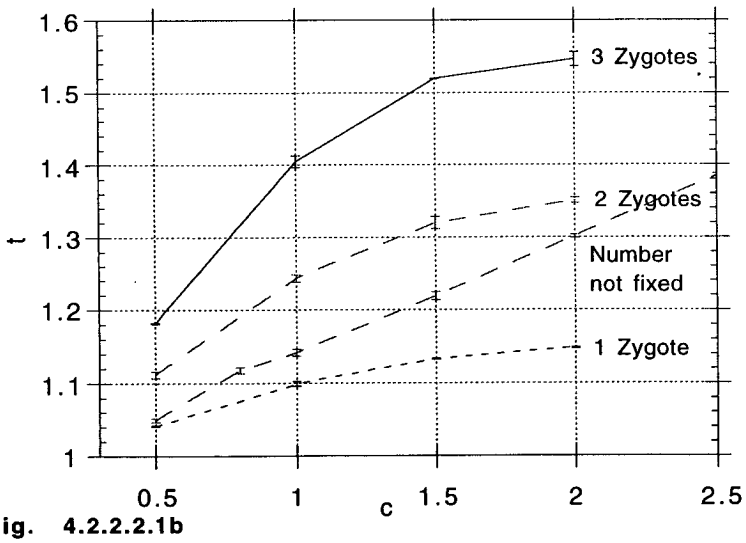


Fig. 4.2.2.2.1b

Figures 4.2.2.2.1a - d (Figures 4.2.2.2.1c - d on overleaf):
Mean values are plotted with their standard errors represented as error bars. In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for c=1.0, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Apart from the parameter c, the Standard Parameter Values were used.

The Effects of the Coefficient of Vector Infectivity (c) on $\bar{\bar{P}}_{inf}$ and $\bar{\bar{P}}_v$ with Fixed Numbers of Zygotes.

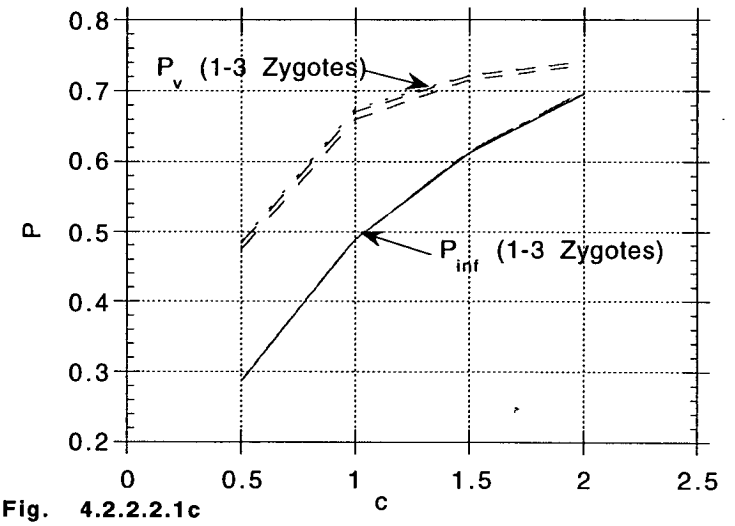


Fig. 4.2.2.2.1c

The Effects of the Coefficient of Vector Infectivity (c) on $\bar{\bar{W}}$ with Fixed Numbers of Zygotes.

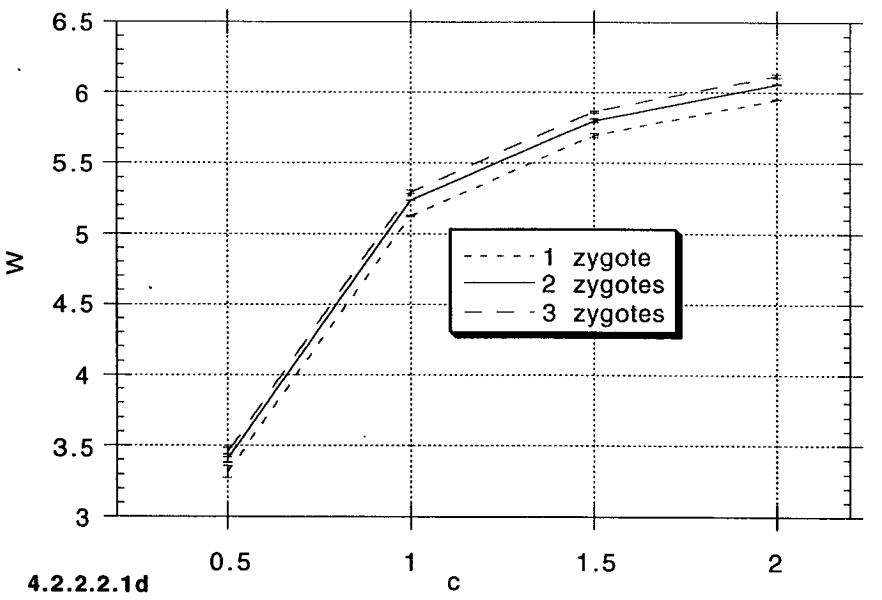


Fig. 4.2.2.2.1d

Two Way Modified Anova Analysis of $\bar{G}_{ST(d)}$ with different values for c and different numbers of zygotes.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Due to c	3	5.62×10^{-3}	1.87×10^{-3}	1341**
Due to Number of zygotes	2	1.289×10^{-4}	6.45×10^{-5}	46.3**
Interaction	6	3.9×10^{-6}	7×10^{-7}	0.50 ns
Error	12	(1.33×10^{-5})	1.39×10^{-6} †	
Total	23	5.77×10^{-3}		

Table 4.2.2.2.1a

Two Way Modified Anova Analysis of $\bar{P}_{inf}$ with different values for c and different numbers of zygotes.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Due to c	3	0.15884	0.05295	12.6**
Due to Number of zygotes	2	0.01004	5.02×10^{-3}	1.19 ns
Interaction	6	0.02343	3.9×10^{-3}	0.93 ns
Error	12	(0.0504)	4.2×10^{-3}†	
Total	23	0.24272		

Table 4.2.2.2.1b

Tables 4.2.2.2.1a - d (4.2.2.2.1c - d overleaf):
SS signifies the "Sum of Squares", while MS the "Mean Square". The SS value for the Error row is in brackets to show that this value was not used to calculate the Error MS. Instead the largest Within Treatment Variance was used (see Section 4.2.2.1). This largest Within Treatment Variance is labelled with the symbol "†". The symbol "**" beside a value of Fs signifies that the statistic was found to be significant at the p=0.01 level, whereas the letters "ns" signify that the statistic was not found to be significant at the p=0.05 level.

Two Way Modified Anova Analysis of $\bar{P}_v$ with different values for c and different numbers of zygotes.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Due to c	3	0.5731	0.1910514	2,233**
Due to Number of zygotes	2	1.17×10^{-5}	5.8×10^{-6}	0.07 ns
Interaction	6	1.85×10^{-5}	3.1×10^{-6}	0.04 ns
Error	12	(1.43×10^{-4})	8.55×10^{-5}†	
Total	23	0.57332		

Table 4.2.2.2.1c

Two Way Modified Anova Analysis of $\bar{W}$ with different values for c and different numbers of zygotes.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Due to c	3	25.688	8.5629	2,414**
Due to Number of zygotes	2	0.1074	0.05368	15.1**
Interaction	6	6.3×10^{-4}	1.1×10^{-4}	0.03 ns
Error	12	(8.39×10^{-3})	3.33×10^{-3}†	
Total	23	25.805		

Table 4.2.2.2.1d

4.2.2.3) Inoculum density(Θ)

Five different inoculum densities were examined over a range from $2x10^{-7}$ to $2x10^{-3}$ (four orders of magnitude). The results from these simulations are presented in Table 4.2.2.3a. It can be seen that the differences in these statistics under each condition are very small, with the largest change being approximately 8% between the largest and smallest value. Statistical analyses (see Tables 4.2.2.3b-h) show that the changes in the values of $\overline{\overline{G}}_{ST(d)}$, $\overline{\overline{P}}_{inf}$, $\overline{\overline{P}}_{v}$, and $\overline{\overline{W}}$ with different values of θ, are found to be significant at the p=0.05 level, while changes in value of $\overline{\overline{G}}_{ST(n)}$, $\overline{\overline{t}}$ and $\overline{\overline{Z}}_{inf}$ are not found to be significant at this level.

The values of $\overline{\overline{G}}_{ST(d)}$, and $\overline{\overline{P}}_{inf}$ descended slightly with increasing values of θ. While the values of $\overline{\overline{P}}_{v}$ and $\overline{\overline{W}}$ rose slightly with increasing values of θ.

The Effects of Innoculum Size (θ) on the Output Statistics of the Model.

θ	$\bar{\bar{G}}_{ST(d)}$	$\bar{\bar{G}}_{ST(n)}$	$\bar{\bar{t}}$	$\bar{\bar{P}}_{inf}$	$\bar{\bar{Z}}_{inf}$	$\bar{\bar{P}}_{V}$	$\bar{\bar{W}}$
2×10^{-7}	**0.9417075**	**0.8238365**	**1.143085**	**0.6735025**	**1.5478525**	**0.467016**	**5.086451**
	(8.05×10^{-5})	(2.26×10^{-3})	(3.24×10^{-3})	(7.39×10^{-4})	(6.03×10^{-4})	(9.72×10^{-4})	(3.93×10^{-3})
2×10^{-6}	**0.941311**	**0.821336**	**1.1460755**	**0.6685195**	**1.535857**	**0.478109**	**5.131025**
	(5.9×10^{-5})	(1.289×10^{-3})	(1.87×10^{-3})	(1.14×10^{-3})	(8.74×10^{-4})	(3.24×10^{-3})	(0.0133)
2×10^{-5}	**0.9398625**	**0.8232315**	**1.141689**	**0.6673505**	**1.542321**	**0.484411**	**5.1716535**
	(1.53×10^{-4})	(2.89×10^{-3})	(4.19×10^{-3})	(3.26×10^{-3})	(3.49×10^{-4})	(7.7×10^{-5})	(6.79×10^{-3})
2×10^{-4}	**0.937443**	**0.814269**	**1.15127**	**0.658829**	**1.538553**	**0.499507**	**5.211783**
	(5.25×10^{-5})	(4.415×10^{-4})	(6.89×10^{-4})	(2.909×10^{-3})	(4.11×10^{-3})	(1.83×10^{-3})	(0.01256)
2×10^{-3}	**0.9327415**	**0.809024**	**1.153085**	**0.640943**	**1.5442915**	**0.508598**	**5.237739**
	(4.0×10^{-5})	(5.04×10^{-3})	(3.24×10^{-3})	(1.00×10^{-3})	(2.73×10^{-3})	(1.68×10^{-3})	(1.57×10^{-3})

Table 4.2.2.3a:
In most cases, the means and standard errors of each statistic are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for $\theta=2 \times 10^{-5}$, in this case, eight runs were done, four with Initial Condition A and four with Initial Condition B. The mean values of each statistic are presented in bold, the standard errors are in parentheses underneath their respective means. Apart from the parameter θ, the Standard Parameter Values were used.

Modified Anova Analysis of $\bar{G}_{ST(d)}$ with different values for θ.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	1.15×10^{-4}	2.86×10^{-5}	147.7**
Within Treatments	11	(1.4×10^{-6})	1.9×10^{-7} †	
Total	15	1.16×10^{-4}		

Table 4.2.2.3b

Modified Anova Analysis of $\bar{G}_{ST(n)}$ with different values for θ.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	3.02×10^{-4}	7.54×10^{-5}	1.48ns
Within Treatments	11	(2.82×10^{-4})	5.08×10^{-5} †	
Total	15	5.83×10^{-4}		

Table 4.2.2.3c

Modified Anova Analysis of $\bar{t}$ with different values for θ.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	1.39×10^{-4}	3.46×10^{-5}	0.34ns
Within Treatments	11	(5.25×10^{-4})	1.02×10^{-4} †	
Total	15	6.64×10^{-4}		

Table 4.2.2.3d

Tables 4.2.2.3a - h (Tables 4.2.2.3e-h overleaf):
SS signifies the "Sum of Squares", while MS the "Mean Square". The SS value for the Within Treatment row is in brackets to show that this value was not used to calculate the Within Treatment MS. Instead the largest Within Treatment Variance was used (see Section 4.2.2.1). This largest Within Treatment Variance is labelled with the symbol "†". The symbol "*" beside a value of Fs signifies that the statistic was found to be significant at the p=0.05 level, while the symbol "**" signifies that the statistic was found to be significant at the p=0.01 level. The letters "ns" beside a value of Fs signify that the statistic was not found to be significant at the p=0.05 level.

Modified Anova Analysis of $\bar{P}_{inf}$ with different values for θ.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	1.48×10^{-3}	3.70×10^{-4}	21.8**
Within Treatments	11	(1.41×10^{-4})	1.70×10^{-5} †	
Total	15	1.62×10^{-3}		

Table 4.2.2.3e

Modified Anova Analysis of $\bar{W}$ with different values for θ.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	0.0298	7.44×10^{-3}	21.1**
Within Treatments	11	(2.20×10^{-3})	3.53×10^{-4} †	
Total	15	0.0320		

Table 4.2.2.3f

Modified Anova Analysis of $\bar{P}_v$ with different values for θ.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	2.20×10^{-3}	5.5×10^{-4}	26.22**
Within Treatments	11	(5.17×10^{-5})	2.10×10^{-5} †	
Total	15	2.25×10^{-3}		

Table 4.2.2.3g

Modified Anova Analysis of $\bar{Z}_{inf}$ with different values for θ.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	1.78×10^{-4}	4.46×10^{-5}	1.32 ns
Within Treatments	11	(2.65×10^{-4})	3.38×10^{-5} †	
Total	15	4.43×10^{-4}		

Table 4.2.2.3h

4.2.3) Aspects of immunity and infection

4.2.3.1) Probability of Loss of Immunity (β)

Several different values of β are examined ranging from $\beta=1.37 \times 10^{-3}$ to $\beta = 10$ per day. For convenience, these rates are described by the reciprocal $1/\beta$, which corresponds to the mean period of time a host is immune to a particular immuno-allelic allele after parasites with that particular allele have been cleared from the host. Thus the range of values of $1/\beta$ examined are from 0.1 days to 720 days.

In Figure 4.2.3.1a, the upper, broken line represents the values of $\bar{\bar{G}}_{ST(d)}$ with different values of $1/\beta$. As the mean period of immune memory is increased, $\bar{\bar{G}}_{ST(d)}$ is found to rise. The value of $\bar{\bar{G}}_{ST(d)}$ was found to rise from a value of 0.732 to a value of 0.973 over the range of values of $1/\beta$ used. The rate at which $\bar{\bar{G}}_{ST(d)}$ increases with $1/\beta$ decreased as the value of $1/\beta$ is increased.

The lower, solid line in Figure 4.2.3.1a represents the values of $\bar{\bar{G}}_{ST(n)}$ with different values of $1/\beta$. As with $\bar{\bar{G}}_{ST(d)}$, it can be seen that as $1/\beta$ is increased, the value of $\bar{\bar{G}}_{ST(n)}$ also increases. The value of $\bar{\bar{G}}_{ST(n)}$ rises from 0.492 to 0.921 over the range of values of $1/\beta$ examined. This rate of increase in $\bar{\bar{G}}_{ST(n)}$ with the mean period of immunity declined as the period of immune memory was increased.

Figure 4.2.3.1b represents the values of $\bar{\bar{t}}$ with different values for $1/\beta$. As the value of $1/\beta$ is increased, $\bar{\bar{t}}$ decreases. The value of $\bar{\bar{t}}$ decreased from approximately 1.5 to just over 1.0 over the range studied.

The rate at which $\bar{\bar{t}}$ decreases with 1/β, declines as the value of 1/β increases.

In Figure 4.2.3.1c, the broken line represents the values of $\bar{\bar{P}}_{inf}$ with different levels of immune memory. As the value of 1/β is increased, the value of $\bar{\bar{P}}_{inf}$ decreases greatly from a value of 0.991 to a value of 0.446 over the range of values of 1/β studied.

The solid line in Figure 4.2.3.1c represents the values of $\bar{\bar{W}}$ with different values for the mean period of immune memory. It can be seen that as the value of 1/β is increased, the value of $\bar{\bar{W}}$ rises greatly from a value of 0.53 when the mean period of immune memory is 0.1 days to a maximum value of 5.78 the value of 1/β is 720 days.

The values of $\bar{\bar{P}}_v$ with different values for the mean period of immune memory can be seen in Table 4.2.3.1a. It can be seen that as the value of 1/β is increased, $\bar{\bar{P}}_v$ decreases from 0.59 to 0.46 over the range studied.

The values of $\bar{\bar{Z}}_{inf}$, with changing values of 1/β, is also shown in Table 4.2.3.1a. There is, however, no significant change in value found (see Table 4.2.3.1b).

The Effects of the Mean Period of Immune Memory (1/β) on the values of $\bar{\bar{G}}_{ST}$

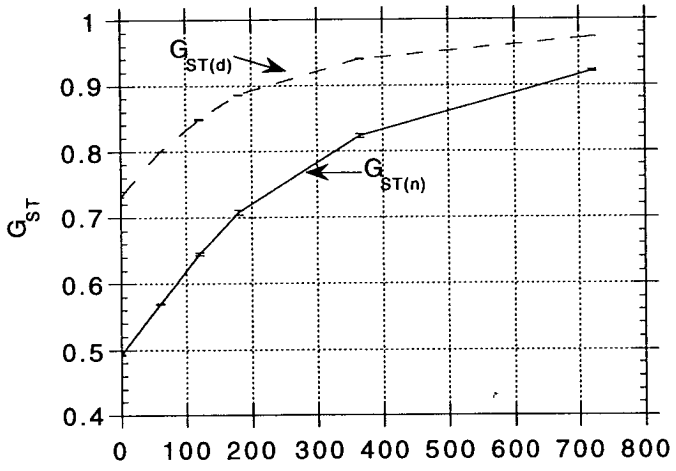


Fig. 4.2.3.1a

The Effects of the Mean Period of Immune Memory (1/β) on the values of $\bar{\bar{t}}$

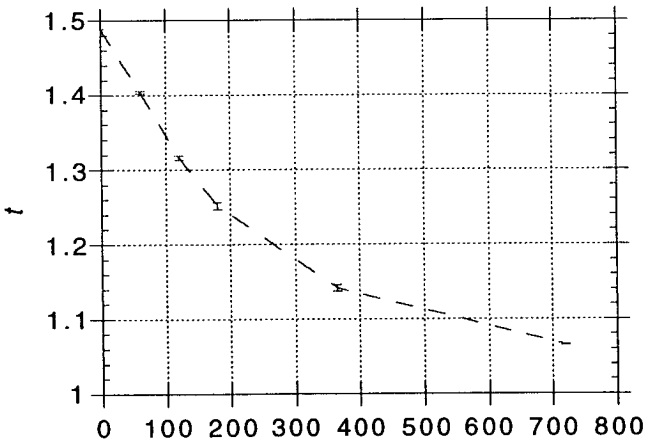


Fig. 4.2.3.1b

Figures 4.2.3.1a - c (Figure 4.2.3.1c overleaf):
Mean values are plotted with their standard errors represented as error bars. In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for 1/β = 365 days, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Apart from for the parameter β, the Standard Parameter Values were used.

The Effects of the Mean Period of Immune Memory (1/β) on the values of P_{inf} and W.

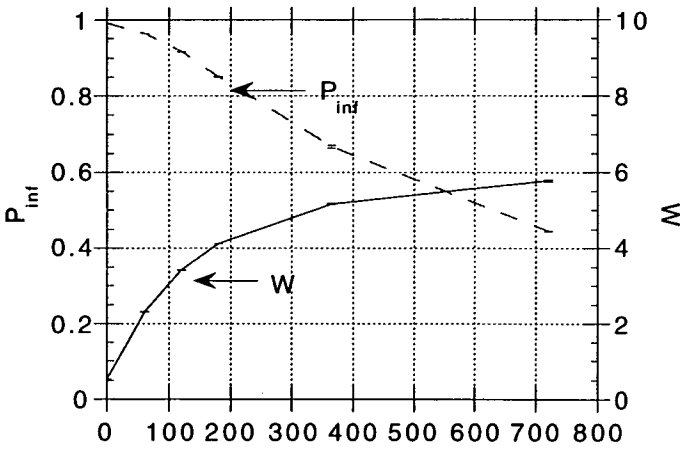


Fig. 4.2.3.1c

The Effects of the Mean Level of Immune Memory (1/β) on $\bar{\bar{P}}_v$ and $\bar{\bar{Z}}_{inf}$

1/β (days)	$\bar{\bar{P}}_v$	$\bar{\bar{Z}}_{inf}$
0.1	**0.586021**	**1.550816**
	(2.3×10^{-5})	(1.31×10^{-3})
60	**0.563163**	**1.542665**
	(1.33×10^{-3})	(4.79×10^{-3})
120	**0.541161**	**1.5470835**
	(5.80×10^{-4})	(2.31×10^{-3})
180	**0.523497**	**1.54090**
	(2.05×10^{-5})	(2.37×10^{-3})
365	**0.484411**	**1.542321**
	(7.7×10^{-5})	(3.49×10^{-4})
720	**0.455592**	**1.548509**
	(4.31×10^{-4})	(1.01×10^{-3})

Table 4.2.3.1a:
In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for 1/β = 365 days, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. The mean value for each statistic is presented in bold with the standard errors presented in parentheses under their associated means. Apart from for the parameter β, the Standard Parameter Values were used.

Modified Anova Analysis of $\bar{Z}_{inf}$ with different values for $1/\beta$.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	1.99×10^{-4}	3.99×10^{-5}	0.87 ns
Within Treatments	12	(2.87×10^{-4})	4.59×10^{-5} †	
Total	17			

Table 4.2.3.1b

Table 4.2.3.1b:
SS signifies the "Sum of Squares", while MS the "Mean Square". The SS value for the Within Treatment row is in brackets to show that this value was not used to calculate the Within Treatment MS. Instead the largest Within Treatment Variance was used (see Section 4.2.2.1). This largest Within Treatment Variance is labelled with the symbol "†". The symbol "**" beside a value of Fs signifies that the statistic was found to be significant at the p=0.01 level, whereas the letters "ns" signify that the statistic was not found to be significant at the p=0.05 level.

4.2.3.2) Immune Threshold (M_{min})

The immune threshold (M_{min}) determines a critical density in the host for parasites with any specific immuno-allelic allele. Above this density there is a chance that an immune response is mounted against that particular immuno-allelic allele. Below this density there is no probability that such an immune response will be mounted against that allele.

A range of values from 0.0 to 0.09 is examined. If a value of 1/D or greater (with the Standard Parameter Values, D=10) is used for the value of M_{min}, then there is the potential for several genotypes to remain in a host without each individually reaching a density equal to M_{min}. Thus, no immune response would ever be mounted against these parasites. For this reason no values for M_{min} greater equal or greater than 0.1 is used.

In Figure 4.2.3.2a the upper, broken line represents the effects of M_{min} on the value of $\overline{\overline{G}}_{ST(d)}$. It can be seen that as the value of M_{min} is increased, the value of $\overline{\overline{G}}_{ST(d)}$ quickly descends from approximately 0.99 when M_{min} has a value of 0.0 to a value of around 0.945 when M_{min} has a value of 0.01. Above this value of M_{min}, the rate of change in $\overline{\overline{G}}_{ST(d)}$ is greatly reduced with $\overline{\overline{G}}_{ST(d)}$ dropping to a value of 0.936 when M_{min} has a value of 0.09

The lower, solid line in Figure 4.2.3.2a represents the values of $\overline{\overline{G}}_{ST(n)}$ as the value of M_{min} is altered. As the value of M_{min} is increased, $\overline{\overline{G}}_{ST(n)}$ descends from around 0.96 to a value of approximately 0.817 over the range of values of M_{min} studied.

Figure 4.2.3.2b represents the effects of M_{min} on $\bar{\bar{t}}$. It can be seen that, as the value of M_{min} is increased, $\bar{\bar{t}}$ initially rises quickly and reaches a value of 1.15 when M_{min}=0.09.

In Figure 4.2.3.2c, the upper, broken line represents the effects of M_{min} on the value of $\bar{\bar{P}}_{inf}$. As M_{min} is increased, $\bar{\bar{P}}_{inf}$ rises towards a value of around 0.67.

The lower, solid line in Figure 4.2.3.2c represents the effects of M_{min} on the value of $\bar{\bar{W}}$. As M_{min} is increased, $\bar{\bar{W}}$ also rises, in this case, reaching a value of approximately 5.2.

The effects of M_{min} on $\bar{\bar{P}}_v$ are shown in Table 4.2.3.2a. The value of $\bar{\bar{P}}_v$ rises from a value of around 0.429 to a value of around 0.487 over the range studied. As with the previous statistics, as M_{min} is increased the rate in change of $\bar{\bar{P}}_v$ with M_{min} is greatly reduced. The differences in $\bar{\bar{P}}_v$ with changes in M_{min} are much smaller than the changes in the output statistics presented above, however, statistical analysis shows that there is a significant difference between the values of $\bar{\bar{P}}_v$ at the 0.05 level (Table 4.2.3.2b).

The apparent effects of M_{min} on $\bar{\bar{Z}}_{inf}$ are also very small. The values of $\bar{\bar{Z}}_{inf}$ with different values of M_{min} can be seen in Table 4.2.3.2a. However, unlike with $\bar{\bar{P}}_v$, when statistical analysis was done on these values, no significant difference was found between the means at the 0.05 level (see Table 4.2.3.2c).

The Effects of the Immune Threshold value (M_{min}) on the values of $\overline{\overline{G}}_{ST}$

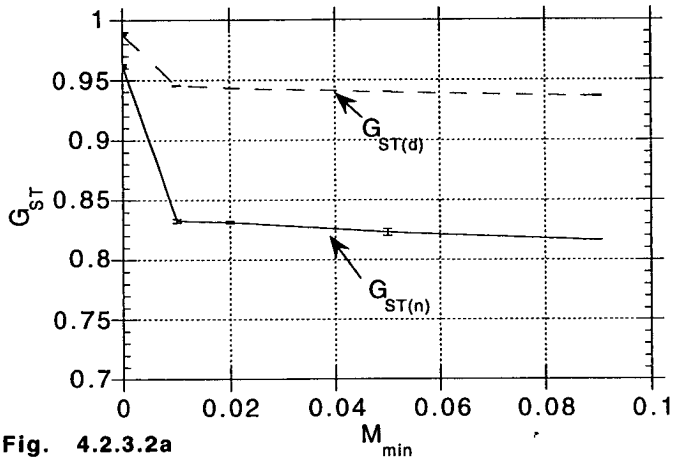


Fig. 4.2.3.2a

The Effects of the Immune Threshold value (M_{min}) on the values of $\overline{\overline{t}}$.

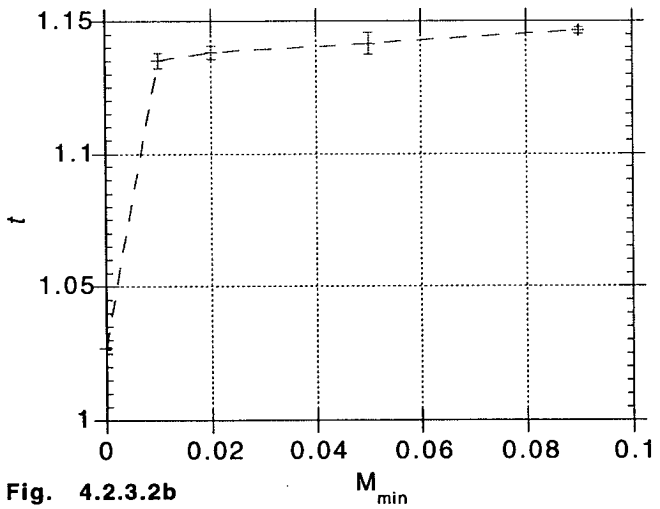


Fig. 4.2.3.2b

Figures 4.2.3.2a - c (Figure 4.2.3.2c overleaf):
Mean values are plotted with their standard errors represented as error bars. In most cases, the means and standard errors are calculated from 2 simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for M_{min}=0.05, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Apart from for the parameter M_{min}, the Standard Parameter Values were used.

The Effects of the Immune Threshold value (M_{min}) on the values of $\bar{\bar{P}}_{inf}$ and $\bar{\bar{W}}$.

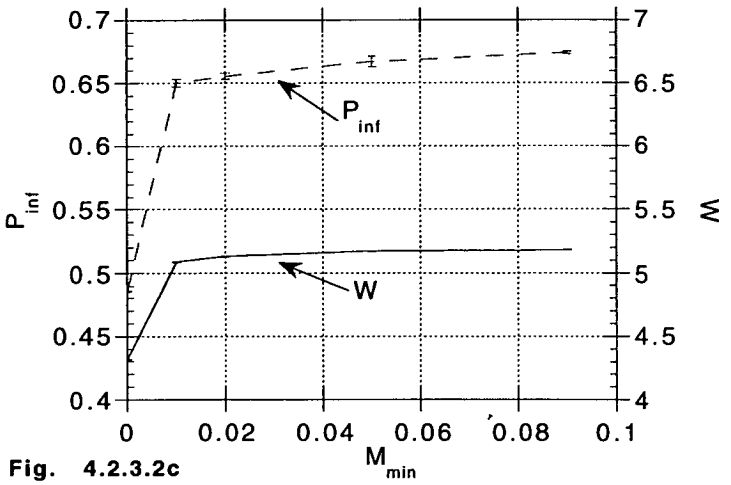


Fig. 4.2.3.2c

The Effects of the Immune Threshold (M_{min}) on $\bar{\bar{P}}_v$ and $\bar{\bar{Z}}_{inf}$

M_{min}	$\bar{\bar{P}}_v$	$\bar{\bar{Z}}_{inf}$
0.0	**0.429221**	**1.535821**
	(1.44×10^{-3})	(0.0118)
0.01	**0.481805**	**1.541502**
	(2.70×10^{-3})	(0.0104)
0.02	**0.481537**	**1.545055**
	(1.63×10^{-3})	(7.61×10^{-3})
0.05	**0.484411**	**1.542321**
	(7.7×10^{-5})	(3.49×10^{-4})
0.09	**0.487503**	**1.541894**
	(3.64×10^{-3})	(1.5×10^{-4})

Table 4.2.3.2a:
In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for M_{min}=0.05, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. The mean value for each statistic is presented in bold with the standard errors presented in parentheses under their associated means. Apart from for the parameter M_{min}, the Standard Parameter Values were used.

Modified Anova Analysis of $\bar{P}_v$ with different values for M_{min}.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	$5.59 \times 10^{--3}$	1.39×10^{-3}	52.4**
Within Treatments	11	(6.69×10^{-5})	2.65×10^{-5} †	
Total	15	5.66×10^{-3}		

Table 4.2.3.2b

Modified Anova Analysis of $\bar{Z}_{inf}$ with different values for M_{min}.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	9.36×10^{-5}	2.34×10^{-5}	0.084 ns
Within Treatments	11	(8.25×10^{-4})	2.78×10^{-4} †	
Total	15	9.18×10^{-4}		

Table 4.2.3.2c

Tables 4.2.3.2b and c:
SS signifies the "Sum of Squares", while MS the "Mean Square". The SS value for the Within Treatment row is in brackets to show that this value was not used to calculate the Within Treatment MS. Instead the largest Within Treatment Variance was used (see Section 4.2.2.1). This largest Within Treatment Variance is labelled with the symbol "†". The symbol "**" beside a value of Fs signifies that the statistic was found to be significant at the p=0.01 level, whereas the letters "ns" signify that the statistic was not found to be significant at the p=0.05 level.

4.2.3.3) The probability of an immune response being mounted (ρ)

The probability of an immune response being mounted (ρ) controls the mean period of time that parasites, with any particular immuno-allelic allele, remain in a host after their density has grown above M_{min} and before an immune response to that allele is mounted. Thus the larger the value of ρ, the shorter the time that a host will remain infected with parasites with any particular immuno-allelic allele.

A range of values of ρ was examined from a value of 0.005 to 0.05 per day. This is equivalent to having the mean period between a parasite genotype increasing above the M_{min}, and the start of immune killing, of between 200 and 20 days respectively.

In Figure 4.2.3.3a, the upper, broken line represents the values of $\overline{\overline{G}}_{ST(d)}$ with different values of ρ. It can be seen that as ρ is increased (i.e. as the mean period of infection decreases), $\overline{\overline{G}}_{ST(d)}$ rises asymptotically towards a value close to 1.0. When ρ is given a value of 0.005, $\overline{\overline{G}}_{ST(d)}$ has a value of 0.708, but when the value of ρ is increased to 0.05, $\overline{\overline{G}}_{ST(d)}$ increases in value to 0.986.

The lower, solid line in Figure 4.2.3.3a represents the values of $\overline{\overline{G}}_{ST(n)}$ with different values of ρ. It can be seen that as ρ is increased, the value of $\overline{\overline{G}}_{ST(n)}$ rises greatly. Thus, when ρ is given a value of 0.005, $\overline{\overline{G}}_{ST(n)}$ had a value of 0.269. However, when the value of ρ is increased to a value of 0.05, $\overline{\overline{G}}_{ST(n)}$ increases to a value of 0.966. The rate of increase in $\overline{\overline{G}}_{ST(n)}$ with ρ declines as ρ increases.

In Figure 4.2.3.3b, the values of $\overline{\overline{t}}$ are represented with different values of ρ. It can be seen that as ρ is increased, $\overline{\overline{t}}$ declines steeply from a

value of approximately 2.63 when ρ was given a value of 0.005 and appears to be very close to 1.0 when ρ=0.05.

In Figure 4.2.3.3c, the broken line represents the values of $\overline{\overline{P}}_{inf}$ with different values of ρ. As ρ is increased, $\overline{\overline{P}}_{inf}$ decreases from a value of 0.996 to a value of 0.307 over the range of values for ρ studied.

The solid line in Figure 4.2.3.3c represents the values of $\overline{\overline{W}}$ with different values of ρ. It can be seen that as ρ is increased, $\overline{\overline{W}}$ does not change monotonically. Instead, $\overline{\overline{W}}$ rises from a value of 3.51 when ρ = 0.005 to a maximum value of 5.42 when ρ = 0.015 before falling to a value of 2.98 when ρ = 0.005.

The value of $\overline{\overline{P}}_v$, with different values of ρ, can be seen in Table 4.2.3.3a. It can be seen that as ρ decreases, so does the value of $\overline{\overline{P}}_v$, decreasing from 0.623 to 0.457 over the range of values for ρ studied.

Table 4.2.3.3a also shows the values of $\overline{\overline{Z}}_{inf}$ with different values of ρ. The value of $\overline{\overline{Z}}_{inf}$ declines from 1.578 to 1.498 with increasing ρ, this decrease is significant at the 0.01 level (see Table 4.2.3.3c).

The Effects of the Rate of Immune Recognition (ρ) on the values of $\bar{\bar{G}}_{ST}$

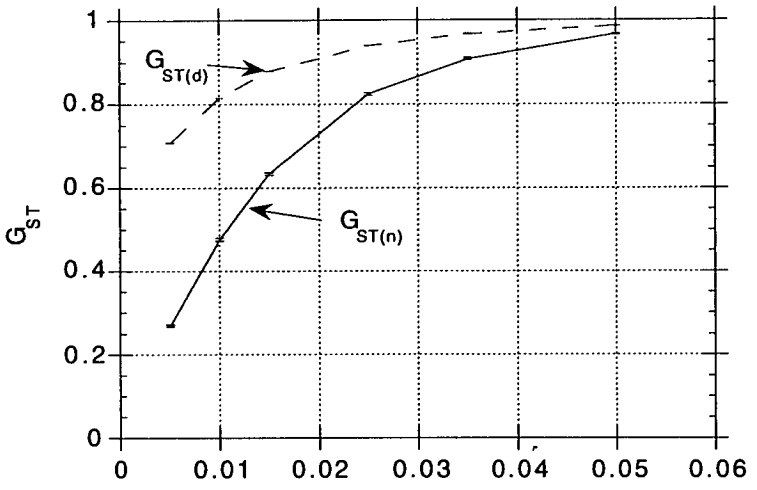


Fig 4.2.3.3a

The Effects of the Rate of Immune Recognition (ρ) on the values of $\bar{\bar{t}}$.

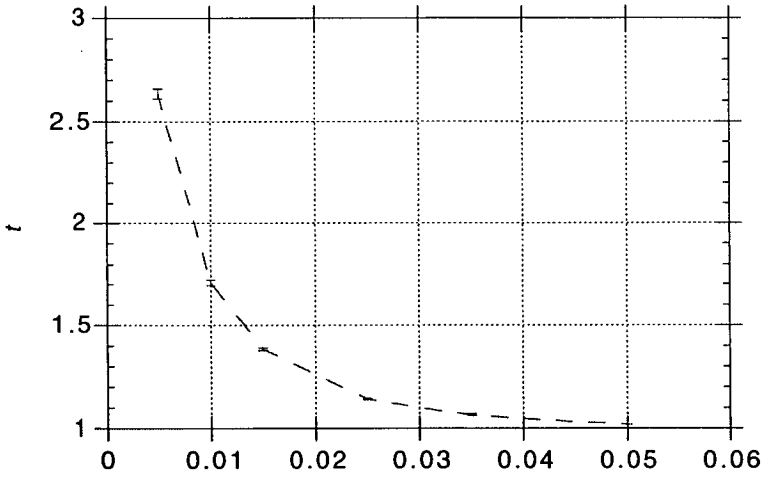


Fig 4.2.3.3b.

Figures 4.2.3.3a - c (Figure 4.2.3.3c overleaf):
Mean values are plotted with their standard errors represented as error bars. In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for ρ=0.025 per day, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Apart from for the parameter ρ, the Standard Parameter Values were used.

The Effects of the Rate of Immune Recognition (ρ) on the values of $\bar{\bar{P}}_{inf}$ and $\bar{\bar{W}}$

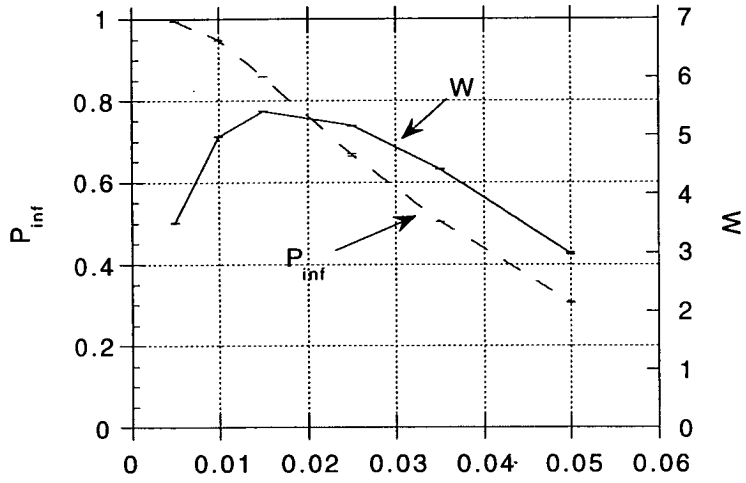


Fig 4.2.3.3c.

The Effects of the Rate of Immune Recognition (ρ) on $\bar{\bar{P}}_v$ and $\bar{\bar{W}}$

ρ (per day)	$\bar{\bar{P}}_v$	$\bar{\bar{Z}}_{inf}$
0.005	**0.623652**	**1.578097**
	(1.35×10^{-3})	(1.87×10^{-3})
0.01	**0.596574**	**1.567082**
	(7.63×10^{-4})	(2.87×10^{-3})
0.015	**0.561215**	**1.562384**
	(2.25×10^{-3})	(2.68×10^{-3})
0.025	**0.484411**	**1.542321**
	(7.7×10^{-5})	(3.49×10^{-4})
0.035	**0.48929**	**1.5286455**
	(6.29×10^{-4})	(2.69×10^{-3})
0.05	**0.357202**	**1.498418**
	(1.96×10^{-3})	(3.73×10^{-3})

Table 4.2.3.3a:
In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for ρ=0.025 per day, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. The mean value for each statistic is presented in bold with the standard errors presented in parentheses under their associated means. Apart from for the parameter ρ, the Standard Parameter Values were used.

Modified Anova Analysis of $\bar{P}_v$ with different values for ρ.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	0.1090	0.0218	2,156**
Within Treatments	12	($3.98\ 10^{-5}$)	$1.01\ 10^{-5}$ †	
Total	17	0.1090		

Table 4.2.3.3b

Modified Anova Analysis of $\bar{Z}_{inf}$ with different values for ρ.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	8.71×10^{-3}	1.74×10^{-3}	56.9**
Within Treatments	12	(2.94×10^{-4})	3.06×10^{-5}†	
Total	17	9.00×10^{-3}		

Table 4.2.3.3c

Tables 4.2.3.3b and c:
SS signifies the "Sum of Squares", while MS the "Mean Square". The SS value for the Within Treatment row is in brackets to show that this value was not used to calculate the Within Treatment MS. Instead the largest Within Treatment Variance was used (see Section 4.2.2.1). This largest Within Treatment Variance is labelled with the symbol "†". The symbol "**" beside a value of Fs signifies that the statistic was found to be significant at the p=0.01 level, whereas the letters "ns" signify that the statistic was not found to be significant at the p=0.05 level.

4.2.3.4) The Rate of Increase in Immune Killing (s)

The Rate of Increase in Immune Killing (s) determines the length of time between an immuno-allelic allele being recognised by the immune system, and parasites with that allele in that host being cleared from the host. Thus the higher the rate of increase in immune killing, the shorter the period that parasites of a particular genotype will remain in a host. A range of values of s from 0.05 to 0.5 per day^2 were examined.

The upper, broken line of Figure 4.2.3.4a represents the values of $\overline{\overline{G}}_{ST(d)}$ with different values of s. It can be seen that as s increases, the value of $\overline{\overline{G}}_{ST(d)}$ increases too. This increase is slight, with the value of $\overline{\overline{G}}_{ST(d)}$ rising from approximately 0.92 to, what seems to be, an asymptotic value of around 0.95 over the range of values of s studied. Despite the small size of this change, the effects of s on $\overline{\overline{G}}_{ST(d)}$ were found to be significant at the p=0.05 level (see Table 4.2.3.4b)

The lower, solid line in Figure 4.2.3.4a represents the values of $\overline{\overline{G}}_{ST(n)}$ with different values of s. As with $\overline{\overline{G}}_{ST(d)}$, the value of $\overline{\overline{G}}_{ST(n)}$ increases with increasing values of s and appears to approach an asymptotic value. The effects of s on $\overline{\overline{G}}_{ST(n)}$ appear to be much larger than they are for $\overline{\overline{G}}_{ST(d)}$.

In Figure 4.2.3.4b the effects of s on the ratio of $\overline{\overline{G}}_{ST(d)}$ to $\overline{\overline{G}}_{ST(n)}$ ($\overline{\overline{t}}$) can be seen. As s is increased, the value of $\overline{\overline{t}}$ decreases. This decrease appears to approach an asymptotic value in the region of 1.13.

The upper broken line in Figure 4.2.3.4c represents the effects of s on the mean proportion of hosts infected ($\overline{\overline{P}}_{inf}$). It can be seen that the value of $\overline{\overline{P}}_{inf}$ decreases with increasing values of s but the rate of this decrease becomes less as s increases.

The lower, solid line in Figure 4.2.3.4c represents the values for the mean level of host immunity ($\bar{\bar{W}}$) with different values of s. The value of $\bar{\bar{W}}$ decreases with increasing values for s, however an asymptotic value for $\bar{\bar{W}}$ of around 5.0 appears to be being reached with the higher values of s.

The value of $\bar{\bar{P}}_v$, with different values of s, can be seen in Table 4.2.3.4a. It can be seen that as s increases, so does the value of $\bar{\bar{P}}_v$, increasing from a value of around 0.53 to a value of around 0.58 over the range of values for s examined. Though this change is small, it is found to be significant at the p=0.05 level (see Table 4.2.3.4c).

Table 4.2.3.4a also shows the values of $\bar{\bar{Z}}_{inf}$ with different values of s. The value of $\bar{\bar{Z}}_{inf}$ also increases with increasing values of s, and this increase is found to be significant at the 0.05 level (see Table 4.2.3.4d).

The Effects of the Rate of Increase in Immune Killing (s) on the values of $\bar{\bar{G}}_{ST}$

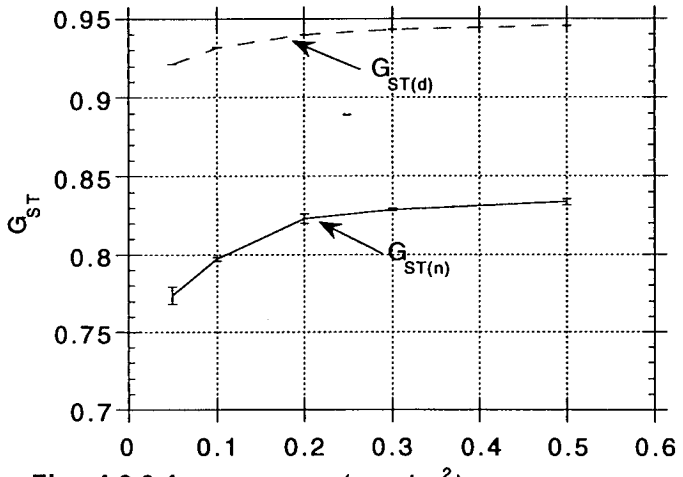


Fig 4.2.3.4a

The Effects of the Rate of Increase in Immune Killing (s) on the values of $\bar{\bar{t}}$.

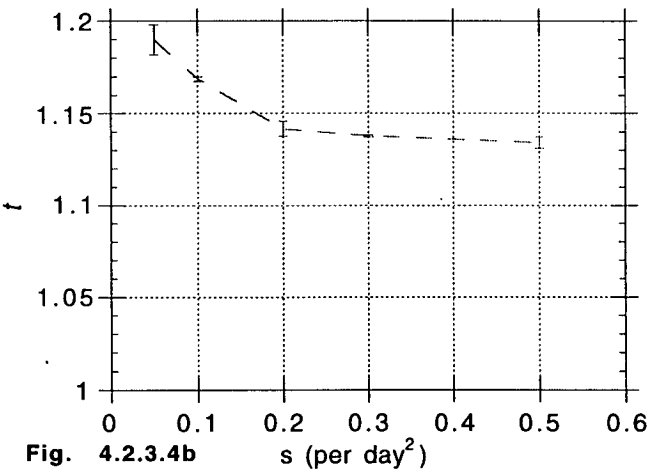


Fig. 4.2.3.4b

Figures 4.2.3.4a - c (Figure 4.2.3.4c overleaf):
Mean values are plotted with their standard errors represented as error bars. In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for s=0.2 per day^2, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Apart from for the parameter s, the Standard Parameter Values were used.

The Effects of the Rate of Increases in Immune Killing (s) on the values of $\bar{\bar{P}}_{inf}$ and $\bar{\bar{W}}$

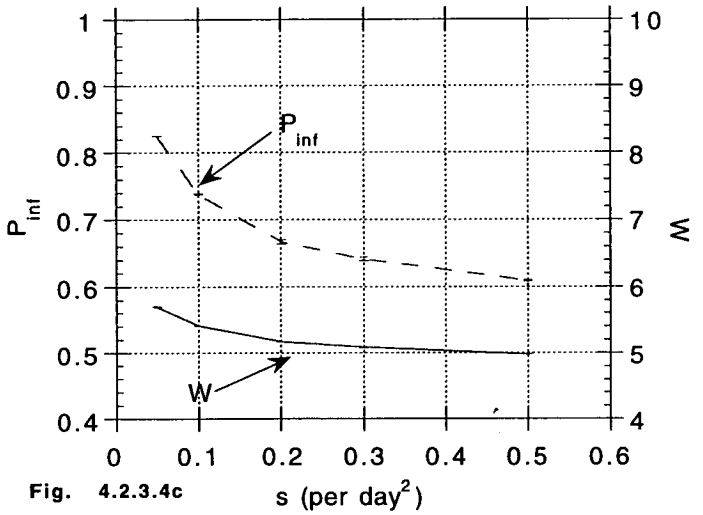


Fig. 4.2.3.4c

The Effects of the Rate of Increase in Immune Killing (s) on $\bar{\bar{P}}_v$ and $\bar{\bar{Z}}_{inf}$

s	$\bar{\bar{P}}_v$	$\bar{\bar{Z}}_{inf}$
0.05	**0.482326**	**1.512025**
	(7.62×10^{-4})	(6.16×10^{-3})
0.1	**0.4811375**	**1.534556**
	(3.447×10^{-3})	(4.68×10^{-3})
0.2	**0.484411**	**1.542321**
	(7.7×10^{-5})	(3.49×10^{-4})
0.3	**0.497147**	**1.5500745**
	(9.48×10^{-4})	(2.72×10^{-3})
0.5	**0.498919**	**1.554081**
	(3.11×10^{-3})	(1.99×10^{-3})

Table 4.2.3.4a:
In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for s=0.2 per day^2, in this case, eight runs were done, four with Initial Condition A and four with Initial Condition B. The mean value for each statistic is presented in bold with the standard errors presented in parentheses under their associated means. Apart from for the parameter s, the Standard Parameter Values were used.

Modified Anova Analysis of $\bar{G}_{ST(d)}$ with different values for s.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	8.68×10^{-4}	2.17×10^{-4}	522**
Within Treatments	11	(2.3×10^{-6})	4.16×10^{-7} †	
Total	15	8.70×10^{-4}		

Table 4.2.3.4b

Modified Anova Analysis of $\bar{P}_{v_r}$ with different values for s.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	4	5.78×10^{-4}	1.45×10^{-4}	6.11*
Within Treatments	11	(6.25×10^{-5})	2.37×10^{-5} †	
Total	15	6.41×10^{-4}		

Table 4.2.3.4c

Modified Anova Analysis of $\bar{Z}_{inf}$ with different values for s.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	6	2.26×10^{-3}	5.66×10^{-4}	7.46*
Within Treatments	13	(3.57×10^{-4})	7.59×10^{-5} †	
Total	19	2.62×10^{-3}		

Table 4.2.3.4d

Tables 4.2.3.4b-8d:
SS signifies the "Sum of Squares", while MS the "Mean Square". The SS value for the Within Treatment row is in brackets to show that this value was not used to calculate the Within Treatment MS. Instead the largest Within Treatment Variance was used (see Section 4.2.2.1). This largest Within Treatment Variance is labelled with the symbol "†". The symbol "**" beside a value of Fs signifies that the statistic was found to be significant at the p=0.01 level, whereas the letters "ns" signify that the statistic was not found to be significant at the p=0.05 level.

4.2.3.5) The Number of Immuno-Allelic Alleles (D)

The number of alleles at the immuno-allelic locus (D) defines the number of immunologically distinct "strains" present in this hypothetical population. A range from D=2 to D=20 is examined.

The upper, broken line of Figure 4.2.3.5a represents the values of $\overline{\overline{G}}_{ST(d)}$ with different values of D. It can be seen that as D increases, the value of $\overline{\overline{G}}_{ST(d)}$ descends. Thus $\overline{\overline{G}}_{ST(d)}$ has a value of very nearly one (0.997) when D=2, and a value of 0.858 when D=20.

The lower, solid line in Figure 4.2.3.5a represents the values of $\overline{\overline{G}}_{ST(n)}$ with different values of D. As with $\overline{\overline{G}}_{ST(d)}$, as D increases, the value of $\overline{\overline{G}}_{ST(n)}$ decreases. The rate of change of $\overline{\overline{G}}_{ST(n)}$ with D is greater than that of $\overline{\overline{G}}_{ST(d)}$, with a value for $\overline{\overline{G}}_{ST(n)}$ of 0.975 when D=2 and a value of 0.690 when D=20.

It can be seen in Figure 4.2.3.5b that as D is increased, so $\overline{\overline{t}}$ increases. Thus, a value for $\overline{\overline{t}}$ of 1.02 is found when D = 2 and a value of 1.24 when D = 20.

There is a very large effect of D on the mean proportion of hosts infected ($\overline{\overline{P}}_{inf}$). The value of $\overline{\overline{P}}_{inf}$ is represented by the broken line in Figure 4.2.3.5c. It can be seen that as D is increased, the value of $\overline{\overline{P}}_{inf}$ rises quickly from 0.186 when D=2 to 0.885 when D=20.

The use of the statistic $\overline{\overline{W}}$ is not particularly informative when comparing the mean level of immunity in populations with different values of D. This is because the maximum value for $\overline{\overline{W}}$ is larger when more alleles at the immuno-allelic locus are present. Thus, to examine the effects of D on

the level of immunity in the host population, the value of $\overline{\overline{W}}/D$ is calculated. This describes the level of immunity of the host population as a proportion of the maximum it could be with D alleles present. This statistic is represented by the solid line in Figure 4.2.3.5c. It can be seen that as D is increased, the value of $\overline{\overline{W}}/D$ decreases slightly. The value of $\overline{\overline{W}}/D$ is found to be 0.573 when D=2, and 0.457 when D=20.

In Table 4.2.3.5a, the effects of different values of D on $\overline{\overline{P}}_v$ are shown. It can be seen that the value of $\overline{\overline{P}}_v$ increases over the range of values of D examined. This change was found to be significant (see Table 4.2.3.5b) at the p=0.05 level.

The effects of different values of D on $\overline{\overline{Z}}_{inf}$ can also be seen in Table 4.2.3.5a. However, the changes in value of $\overline{\overline{Z}}_{inf}$ are not significant at the p=0.05 level (see Table 4.2.3.5c).

The Effects of the Number of Immuno-allelic Alleles (D) on the values of $\overline{\overline{G}}_{ST}$.

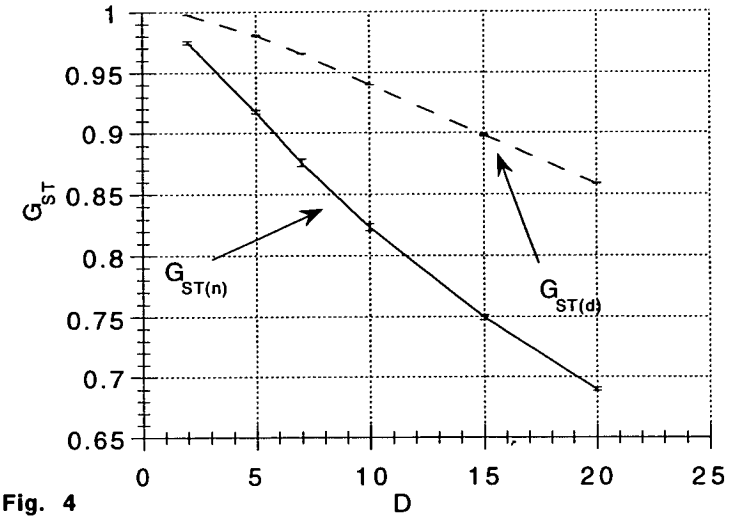


Fig. 4

The Effects of the Number of Immuno-allelic Alleles (D) on the values of $\overline{\overline{t}}$.

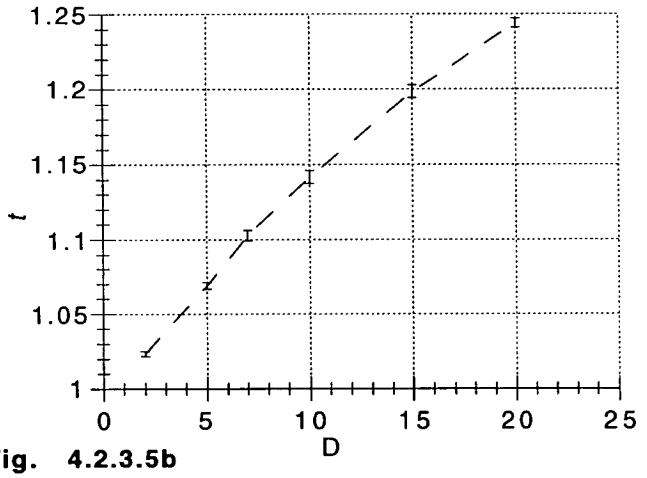


Fig. 4.2.3.5b

Figures 4.2.3.5a - c (Figure 4.2.3.5c overleaf):
Mean values are plotted with their standard errors represented as error bars. In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for D=10, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Apart from for the parameter D, the Standard Parameter Values were used.

The Effects of the Number of Immuno-allelic Alleles (D) on the values of $\bar{\bar{P}}_{inf}$ and $\bar{\bar{W}}/D$

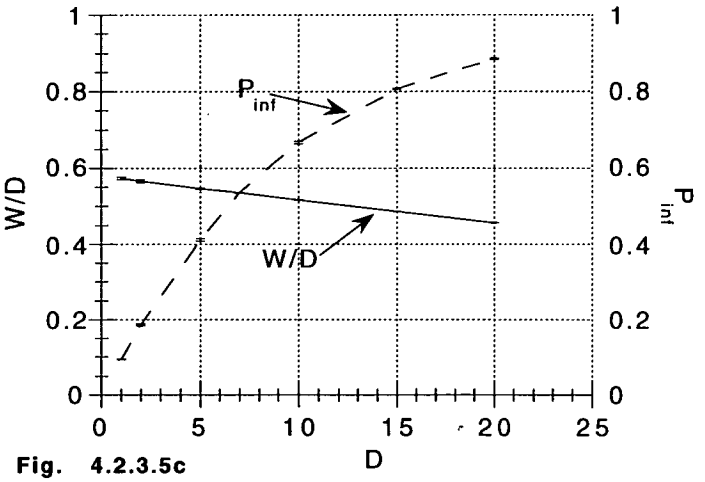


Fig. 4.2.3.5c

The Effects of the Number of Immuno-allelic Alleles (D) on $\bar{\bar{P}}_v$ and $\bar{\bar{Z}}_{inf}$

D	$\bar{\bar{P}}_v$	$\bar{\bar{Z}}_{inf}$
2	**0.4208005**	**1.551969**
	(3.72×10^{-4})	(9.0×10^{-3})
5	**0.4472795**	**1.53334**
	(2.98×10^{-3})	(3.75×10^{-4})
7	**0.46775**	**1.544913**
	(6.7×10^{-5})	(4.86×10^{-4})
10	**0.484411**	**1.542321**
	(7.7×10^{-5})	(3.49×10^{-4})
15	**0.5138585**	**1.546519**
	(1.33×10^{-3})	(5.15×10^{-4})
20	**0.5355935**	**1.5523005**
	(1.57×10^{-3})	(1.03×10^{-3})

Table 4.2.3.5a:
In most cases, the means and standard errors are calculated from 2 simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for D=10, in this case, eight runs were done, four with Initial Condition A and four with Initial Condition B. The mean value for each statistic is presented in bold with the standard errors presented in parentheses under their associated means. Apart from for the parameter D, the Standard Parameter Values were used.

Modified Anova Analysis of $\bar{P}_v$ with different values of D.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	0.01823	3.64×10^{-3}	205.3**
Within Treatments	12	(4.29×10^{-5})	1.77×10^{-5} †	
Total	17	0.01828		

Table 4.2.3.5b

Modified Anova Analysis of $\bar{Z}_{inf}$ with different values of D.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	5.32×10^{-4}	1.06×10^{-4}	0.654 ns
Within Treatments	12	(3.79×10^{-4})	1.62×10^{-4} †	
Total	17	9.12×10^{-4}		

Table 4.2.3.5c

Tables 4.2.3.5b and c:
SS signifies the "Sum of Squares", while MS the "Mean Square". The SS value for the Within Treatment row is in brackets to show that this value was not used to calculate the Within Treatment MS. Instead the largest Within Treatment Variance was used (see Section 4.2.2.1). This largest Within Treatment Variance is labelled with the symbol "†". The symbol "**" beside a value of Fs signifies that the statistic was found to be significant at the p=0.01 level, whereas the letters "ns" signify that the statistic was not found to be significant at the p=0.05 level.

4.2.3.5.1) Comparison of $\overline{\overline{P}}_{inf}$ to the prediction of independent transmission.

In the model examined here, there is potential for competition between parasites with different immuno-allelic alleles and thus 'strains' do not act independently. However, it is possible to produce the predicted effect of D on $\overline{\overline{P}}_{inf}$ if transmission is independent. This predicted value ($\overline{\overline{P}}_{inf[exp]}$) is calculated from the value for $\overline{\overline{P}}_{inf}$ when D=1 ($\overline{\overline{P}}_{inf[1]}$) by the equation:

$$\overline{\overline{P}}_{inf[exp]} = 1 - \left(1 - P_{inf[1]}\right)^{D} \qquad \textbf{(4.2.3.5.1)}$$

In Figure 4.2.3.5.1 the values for P_{inf} with different values of D as found with the model are represented by the dashed line. The solid line represents the predicted values of $\overline{\overline{P}}_{inf}$ ($\overline{\overline{P}}_{inf[exp]}$) under the assumption of independence. $\overline{\overline{P}}_{inf[1]}$ is calculated back from the value of $\overline{\overline{P}}_{inf}$ found with D=20.

A comparison of the effects of D on P_{inf} with and without strain-specific competition

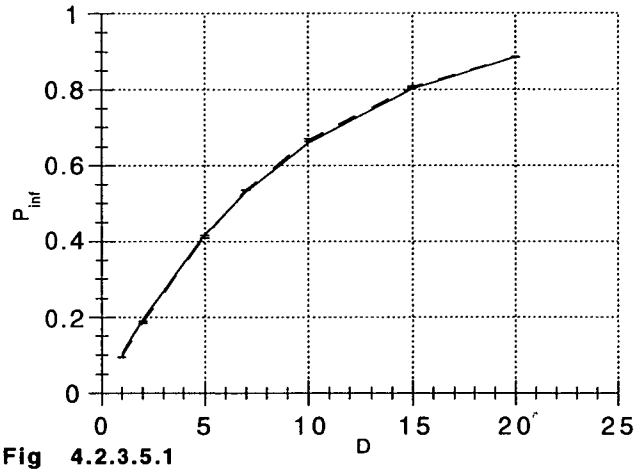


Fig 4.2.3.5.1

Figures 4.2.3.5.1 In the case of the dashed line, mean values are plotted with their standard errors represented as error bars. In most cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for D=10, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Apart from for the parameter D, the Standard Parameter Values were used. For the solid line the values are calculated by the method described in the associated text.

4.2.3.5.2) Examination of the effects of D on the population genetic statistics of the model under constant transmission.

The standard model was modified so that exactly the same number of transmitting double bites are transmitted per time step, irrespective of the numbers of hosts infected, or the level of infection in those hosts. The number of zygotes for each transmitting double bite was fixed at one. This was done to separate the effects of changing the overall rate of transmission in the population (through the changing values of P_{inf}), from other effects that may be caused by altering th value of D.

The effects of changing values of D, in the absence of changes in the overall rates of transmission can be seen in Figures 4.2.3.5.1a and b. In the former, apart from D and the modified transmission system (set to 38 bites per time step) the standard parameter values were used. In the latter, 150 bites per time step were used and the value of M_{min} was given a value of zero.

In Figure 4.3.5.2a it can be seen that as the number of alleles at the immuno -allelic locus (D) increase, the values for both $\overline{\overline{G}}_{ST(d)}$ and $\overline{\overline{G}}_{ST(n)}$ decrease. The value of $\overline{\overline{t}}$ is found to increase with D, though this increases is slight (from 1.06 to 1.11 over the range of D examined), it was found to be significant using the modified Anova with $P<0.01$. The value for W/D was found to decrease greatly with D over the range of values for D examined.

It can be seen in Figure 4.3.5.2b that G_{ST} for both loci tends to descend with increasing values of D. Although $\overline{\overline{t}}$ is always found to have a value greater than one, under the conditions examined, its value is not found to significantly change with different values of D.

The effects of D on the model when modified for constant transmission

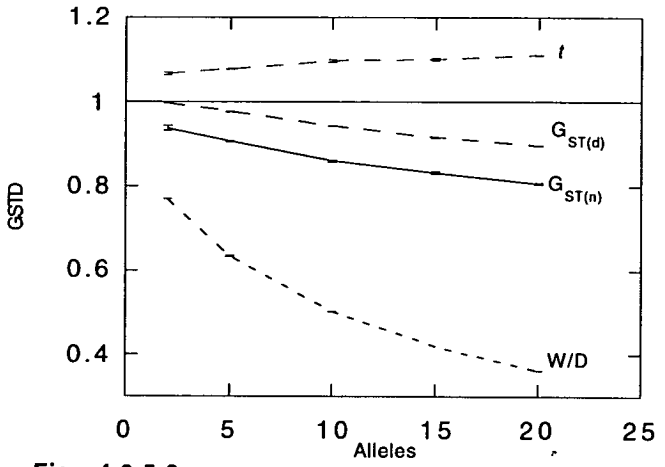


Fig 4.3.5.2a

The effects of D on the model when modified for constant transmission and with M_{min}=0.

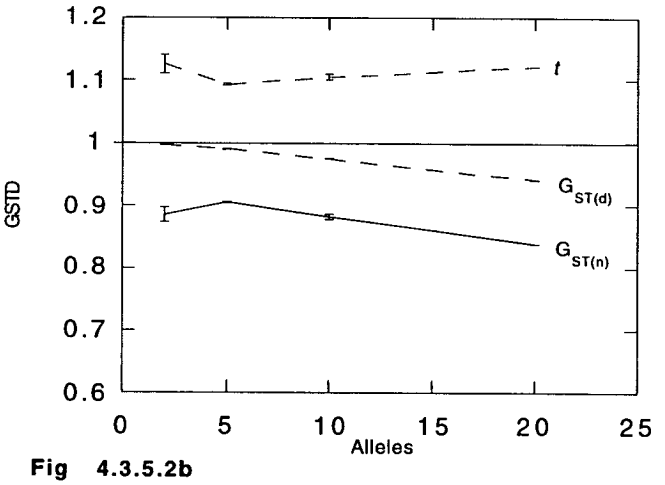


Fig 4.3.5.2b

Figures 4.2.3.5.2a & b:
Mean values are plotted with their standard errors represented as error bars. In all cases, the means and standard errors are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. Note the small size of the standard errors make the error bars difficult to distinguish. Values were calculated using a modified version of the model described in chapter 2, the modifications are explained in the text. Apart from for the parameter D, and in the case of Fig 4.3.5.2b M_{min}(where M_{min}=0), the Standard Parameter Values were used.

4.2.3.6) Parasite Growth Rate (b_0)

The Parasite Growth Rate (b_0) determines the speed at which parasites grow inside a host. This helps determine the rate at which a new parasite genotype entering a host reaches its carrying capacity and, once the immune system is activated, the length of time before that parasite type is cleared. A range of values for b_0 were examined from 6 to 16 per 2 days.

The results from these runs can be seen in Table 4.2.3.6a. It can be seen from examining this table, that the effects of the different values of b_0 examined on the output statistics are very slight. The effects of b_0 on $\overline{\overline{G}}_{ST(d)}$, $\overline{\overline{G}}_{ST(n)}$, $\overline{t}$ and $\overline{\overline{Z}}_{inf}$ were found not to be significant at the p=0.05 level. However the effects of b_0 on the other output statistics were found to be significant at the p=0.05 level or greater. The values of these increased with increasing values of b_0.

Note that the largest effect of b_0 on any of the output statistics was on $\overline{\overline{P}}_v$ where a 6.0% change in value occurred over the range of values for b_0 examined.

The Effects of the Parasite Growth Rate (b_0) on the Output Statistics of the Model.

b_0 (per 2 days)	$\bar{\bar{G}}_{ST(d)}$	$\bar{\bar{G}}_{ST(n)}$	$\bar{\bar{t}}$	$\bar{\bar{P}}_{inf}$	$\bar{\bar{Z}}_{inf}$	$\bar{\bar{P}}_V$	$\bar{\bar{W}}$
6	**0.940261**	**0.825768**	**1.138634**	**0.656497**	**1.538114**	**0.4660435**	**5.020062**
	4.45×10^{-5}	1.97×10^{-3}	2.80×10^{-3}	1.30×10^{-3}	5.56×10^{-4}	1.04×10^{-3}	6.54×10^{-3}
8	**0.9401095**	**0.823595**	**1.141566**	**0.662504**	**1.539547**	**0.478162**	**5.087409**
	1.285×10^{-3}	6.29×10^{-3}	7.11×10^{-3}	5.14×10^{-4}	3.724×10^{-3}	2.43×10^{-3}	3.53×10^{-3}
10	**0.9395215**	**0.821876**	**1.143417**	**0.664263**	**1.544555**	**0.482968**	**5.144009**
	1.28×10^{-4}	1.01×10^{-3}	1.29×10^{-3}	4.34×10^{-3}	2.02×10^{-3}	1.92×10^{-3}	7.12×10^{-3}
12.3	**0.9398625**	**0.8232315**	**1.141689**	**0.6673505**	**1.542321**	**0.484411**	**5.1716535**
	1.53×10^{-4}	2.89×10^{-3}	4.19×10^{-3}	3.26×10^{-3}	3.49×10^{-4}	7.7 x10 -5	6.79×10^{-3}
14	**0.9396115**	**0.819535**	**1.1465205**	**0.671129**	**1.541189**	**0.492433**	**5.197678**
	2.18×10^{-4}	1.02×10^{-3}	1.165×10^{-3}	4.13×10^{-4}	3.59×10^{-3}	3.57×10^{-4}	1.94×10^{-3}
16	**0.939025**	**0.82137**	**1.14326**	**0.671792**	**1.537521**	**0.494503**	**5.22127**
	8.05 $-\times 10^{-4}$	2.23×10^{-3}	2.10×10^{-3}	2.5×10^{-4}	1.21×10^{-3}	1.71×10^{-3}	8.28×10^{-3}

Table 4.2.3.6a.

In most cases, the means and standard errors of each statistic are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for b_0=12.3 per 2 days, in this case, eight runs were done, four with Initial Condition A and four with Initial Condition B. The mean values of each statistic are presented in bold, the standard errors are in parentheses underneath their respective means. Apart from the parameter b_0, the Standard Parameter Values were used.

Modified Anova Analysis of $\bar{G}_{ST(d)}$ with different values for b_0.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	1.9×10^{-6}	4×10^{-7}	0.121 ns
Within Treatments	12	(5.6×10^{-6})	$3.30 \times 10^{-6\dagger}$	
Total	17	7.5×10^{-6}		

Table 4.2.3.6b

Modified Anova Analysis of $\bar{G}_{ST(n)}$ with different values for b_0.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	7.59×10^{-5}	1.52×10^{-5}	0.19 ns
Within Treatments	12	(3.18×10^{-4})	$7.90 \times 10^{-5\dagger}$	
Total	17	3.94×10^{-4}		

Table 4.2.3.6c

Modified Anova Analysis of $\bar{t}$ with different values for b_0.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	1.33×10^{-4}	2.66×10^{-5}	0.263 ns
Within Treatments	12	(5.48×10^{-4})	$1.01 \times 10^{-4\dagger}$	
Total	17	6.81×10^{-4}		

Table 4.2.3.6d

Tables 4.2.3.6b-h (Tables 4.2.3.6e-h overleaf):
SS signifies the "Sum of Squares", while MS the "Mean Square". The SS value for the Within Treatment row is in brackets to show that this value was not used to calculate the Within Treatment MS, instead the largest Within Treatment Variance was used (see Section 4.2.2.1). This largest Within Treatment Variance is labelled with the symbol "†". The symbol "*" beside a value of Fs signifies that the statistic was found to be significant at the p=0.05 level, while the symbol "**" signifies that the statistic was found to be significant at the p=0.01 level. The letters "ns" beside a value of Fs signify that the statistic was not found to be significant at the p=0.05 level.

Modified Anova Analysis of $\bar{P}_{inf}$ with different values for b_0.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	3.54×10^{-4}	7.07×10^{-5}	4.17*
Within Treatments	12	$(1.33 \times 10^{-})$	1.70×10^{-5}†	
Total	17	4.86×10^{-4}		

Table 4.2.3.6e

Modified Anova Analysis of $\bar{W}$ with different values for b_0.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	0.06079	0.012	56.3**
Within Treatments	12	(1.85×10^{-3})	2.13×10^{-4}†	
Total	17	0.06264		

Table 4.2.3.6f

Modified Anova Analysis of $\bar{P}_v$ with different values for b_0.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	1.129×10^{-3}	2.26×10^{-4}	19.1**
Within Treatments	12	(4.26×10^{-5})	$1.18x\ 10^{-5}$†	
Total	17	1.171×10^{-3}		

Table 4.2.3.6g

Modified Anova Analysis of $\bar{Z}_{inf}$ with different values for b_0.

Source of Variation	Degrees of Freedom	SS	MS	Fs
Between Treatments	5	8.23×10^{-5}	1.65×10^{-5}	0.539 ns
Within Treatments	12	(2.791×10^{-4})	3.06×10^{-5}†	
Total	17	$3.613\tilde{x}\ 10^{-4}$		

Table 4.2.3.6h

4.3) Discussion

The aim of this chapter was to examine the effects of manipulating the parameters of the model. This was necessary to gain a deeper insight into the dynamics of the model and possibly into the dynamics of human-mosquito-malaria interactions.

4.3.1) Trends in the results

For most points in parameter space, only two simulation runs were examined. This was done to save computer time. However, it has the potential to give an inaccurate value for the mean of each statistic, and to reduce the reliability of any statistical analysis used. The similarity between runs, evident from the small standard errors in the results presented and the use of the Modified Anova Analysis appeared to justify this approach.

There are several important features of the results that should be noted. Firstly, in all cases examined, when a parameter is varied so that the value of $\overline{\overline{G}}_{ST(d)}$ decreases, so does the value of $\overline{\overline{G}}_{ST(n)}$. However, the rate of change of $\overline{\overline{G}}_{ST(d)}$ with the change of parameter value is always found to be slower than the rate of change of $\overline{\overline{G}}_{ST(n)}$. Thus, as the two values decrease with the change in a parameter, the difference between them always increased. Therefore, the value of $\overline{\overline{t}}$ is always found to increase when the value of $\overline{\overline{G}}_{ST(n)}$ and $\overline{\overline{G}}_{ST(d)}$ decrease.

It is generally found that when $\overline{\overline{P}}_{inf}$ (the mean proportion of hosts infected at the stationary distribution) is increased by the change in a parameter value, $\overline{\overline{G}}_{ST(d)}$ and $\overline{\overline{G}}_{ST(n)}$ decrease. However, this is not the case when the value of the inoculum size (θ) is altered. In this case, the values of

$\bar{\bar{P}}_{inf}$, $\bar{\bar{G}}_{ST(d)}$ and $\bar{\bar{G}}_{ST(n)}$ all decrease as the value of θ increases. This will be discussed later.

One of the main aims of designing and examining the model presented here was to compare its results to those of Hastings's analytic model which assumed life-long immune memory (Hastings 1996). Hastings's analytic model examined the ratio of the equilibrium values of $G_{ST(d)}$ and $G_{ST(n)}$ (t) and found it always to be greater than one. Due to the stochastic nature of the simulation study presented here, no equilibrium value of these statistics is ever reached. However the values $\bar{\bar{G}}_{ST(d)}$, $\bar{\bar{G}}_{ST(n)}$ and $\bar{\bar{t}}$ reflect the means of the stationary distributions of these statistics, and thus, are comparable to the equilibrium values of $G_{ST(n)}$, $G_{ST(d)}$ and t in Hastings's model.

In the results presented above, it is found that, under all conditions examined, the ratio of $\bar{\bar{G}}_{ST(d)}$ to $\bar{\bar{G}}_{ST(n)}$ ($\bar{\bar{t}}$) has a value greater than one. This clearly agrees with the general results of Hastings's model. Thus, despite the very different assumptions of the models, this qualitative result seems quite robust.

In Hastings's model, if either the number of independent transmission events leading to an infection, or the number of zygotes formed per vector were increased, then $G_{ST(n)}$ and $G_{ST(d)}$ went down while t was found to increase. In the model presented here, when the Double Bite Rate (T) is increased, the number of independent transmission events increase also. In this case, the values of $\bar{\bar{G}}_{ST(d)}$ and $\bar{\bar{G}}_{ST(n)}$ decrease while $\bar{\bar{t}}$ increases, in accordance with the results of Hastings.

Increasing the value of the coefficient of vector infectivity (c) increases the mean number of zygotes per infected vector (Z_{inf}), however it

also increases the proportion of vectors infected (P_v) and thus the number of independent transmission events. By fixing the numbers of zygotes at either one, two, or three per infected vector, it is possible to separate the effects on the output statistics of these two aspects of c. It is thus possible to show that the greater the number of zygotes per infected vector, the lower the values of $\bar{\bar{G}}_{ST(n)}$ and $\bar{\bar{G}}_{ST(d)}$ and the greater the value of $\bar{\bar{t}}$.

Thus, not only is the general qualitative result of Hastings agreed with in this model ($t>1$), but also these more specific qualitative results. However, one finding of Hastings is not in accord with the findings of this study; this difference is in the effects of the number of immuno-allelic alleles present in the model (D).

In the model by Hastings, as D is increased in value, the equilibrium values of $G_{ST(d)}$ decreases, the equilibrium value of $G_{ST(n)}$ increases, and the value of t decreases. In my model, as the value of D is increased, the value of $\bar{\bar{G}}_{ST(d)}$ decreases (as does $G_{ST(d)}$ in Hastings model). However the value of $\bar{\bar{G}}_{ST(n)}$ also decreases (unlike $G_{ST(n)}$ in Hastings model), and to a greater extent than the value of $\bar{\bar{G}}_{ST(d)}$. Thus the overall effect is that $\bar{\bar{t}}$ increases with increasing values of D in my model, whereas the equilibrium value of t in Hastings's model decreases with increasing values of D. This difference between the two models is examined in detail later in this section.

4.3.2) Discussion of the trends

One problem with computer-simulation studies is that it is often difficult to understand the exact nature of the dynamic processes that produce the results. This is particularly true when different dynamic forces are pushing variables in different directions. However, in the present model, there are a number of cases where the dynamics involved in producing the results seem clear. These results will be discussed in detail, and in some of the other cases, the likely forces involved will be pointed out.

Many of the results produced from the model can, at least in part, be explained by considering the overall rate of transmission in the population, i.e. the rate at which parasites are injected into hosts. This will be referred to as the Inoculation Rate (IR). This Inoculation Rate is the product of three features of the population: (i) the Double Bite Rate (T); (ii) the proportion of vectors, biting infected hosts, that become infected (P_v); and (iii) the proportion of hosts infected (P_{inf}). The Double Bite Rate is a defined parameter of the model, and thus is not affected by any of the other parameters. The value of P_v is controlled by the parameter c and the distribution of parasite densities amongst the hosts. It is clear from the results section above, that the value of P_{inf} is strongly affected by a variety of parameters in the model.

4.3.2.1) The Population Genetic Statistics ($\overline{\overline{G}}_{ST(n)}$, $\overline{\overline{G}}_{ST(d)}$, and $\overline{\overline{t}}$)

The output statistics measuring the population genetic aspects of parasite population (namely $\overline{\overline{G}}_{ST(d)}$,$\overline{\overline{G}}_{ST(n)}$, and $\overline{\overline{t}}$) can largely be explained by considering the Inoculation Rate and the numbers of zygotes per vector. It has already been shown that, in most cases, when these two factors are increased (e.g. through increasing T and c respectively), $\overline{\overline{G}}_{ST(d)}$ and $\overline{\overline{G}}_{ST(n)}$ go down and $\overline{\overline{t}}$ goes up. The reason for this appears reasonably simple.

First, consider the effects of the Inoculation Rate on $\overline{\overline{G}}_{ST(n)}$. The value of $G_{ST(n)}$ describes the amount of genetic diversity at the neutral locus found in each host compared to the amount of genetic diversity at the locus found in the total population. A $G_{ST(n)}$ value of zero would occur if each infected host has the same amount of genetic diversity as the total population. In this instance each infected host would contain parasites with all the alleles at this locus that are found in the total parasite population (amongst hosts) and in the same frequencies as they are found in that population. A $G_{ST(n)}$ value of one would occur if each infected host contained parasites with only a single allele at the neutral locus, while the total population has two or more alleles present.

It is straightforward to see that if a host is infected with parasites with a single allele at the neutral locus it will contribute to a higher $G_{ST(n)}$ value than if it is infected with parasites with multiple alleles at the neutral locus. Thus, the higher the rate of super-infection (infection of a host with parasites derived from different vectors), the greater the chance that a host, infected with parasites containing only one allele, receives parasites introduced with a second allele, which will cause a decrease in the value of $G_{ST(n)}$. Thus, in

almost all cases, the value of $\bar{\bar{G}}_{ST(n)}$ decreases with higher Inoculation Rates. This occurs with increases in the values of T, $\bar{\bar{P}}_v$ and $\bar{\bar{P}}_{inf}$.

The same argument as to why the value of $\bar{\bar{G}}_{ST(n)}$ decreases can be used to explain why the value of $\bar{\bar{G}}_{ST(d)}$ decreases with higher Inoculation Rates. However, the rate of decline is less, and the value of $\bar{\bar{G}}_{ST(d)}$ greater than that of $\bar{\bar{G}}_{ST(n)}$. This difference can be explained by considering the action of the immune system.

Envisage a situation where four parasite haplotypes are transmitted to a host. Let their genotypes be denoted AB, Ab, ab and aB. Letters A and a represent alleles at the immuno-allelic locus and letters B and b represent alleles at the neutral locus. If the host is immune to the immuno-allelic allele A, then only parasite types aB and ab will survive. If the host is immune to allele a, only genotypes Ab and AB will survive. In either case the level of genetic diversity found at the immune locus is less than that found at the neutral locus.

For a difference in genetic diversity to occur between the two loci in a single host, at least two parasite genotypes must infect that host, both bearing the same allele at the immune locus (to which the host is susceptible), but different alleles at the neutral locus. The greater the number of parasite genotypes that are introduced into a host, the greater the probability that such an event occurs. Thus, across a population, one would expect that, the greater the Inoculation Rate, the larger the value for $\bar{\bar{t}}$. Also one would expect that for any value of IR, if the mean number of zygotes in an infected mosquito is increased, a larger value of $\bar{\bar{t}}$ can be expected.

It has been pointed out earlier that the effects of different values of D on the statistics $\bar{\bar{G}}_{ST(d)}$, $\bar{\bar{G}}_{ST(n)}$ and $\bar{\bar{t}}$ is, in this model, very different from the effects found in Hastings's model (Hastings 1996). The two models are far from identical, and one or more of the differences between them must account for the discovery that $\bar{\bar{t}}$ increases with D in the model presented here and *t* decreases withD in the model of Hastings. The effects of two of these differences are explicitly examined in Section 4.2.3.5.2

In Hastings's model, it was assumed that different values of D would not affect any other aspect of the model and thus the genetic consequences of changing the value of D was examined with a fixed Inoculation Rate. It is clear in the model presented here that changing the value of D has a very large effect on $\bar{\bar{P}}_{inf}$ and thus on the IR. It is plausible that this increase in the transmission rate swamps the effects found by Hastings. This hypothesis was tested by examining the effects of D in my model with the transmission rate artificially fixed (see Section 4.2.3.5a).

The results show that the changes in $\bar{\bar{P}}_{inf}$ do have a large effect on the results. In this modified model, the value of $\bar{\bar{t}}$ still increases with increasing values of D, but the increase is much smaller than before. Thus a fixed transmission rate accounts for some but not all of the difference between the two models.

Another difference between the two models which may have an effect on $\bar{\bar{t}}$ is the density dependent initiation of the immune response (controlled by M_{min}) used in the model presented here, but which is not included in the model of Hastings. By setting M_{min}=0, this difference is removed. When this is done in combination with an artificially fixed transmission rate (see Section 4.2.3.5.2b), the difference between the two

models with respect to the effects of D on $\bar{\bar{t}}$ is made even smaller (D is now found to have no significant effect on $\bar{\bar{t}}$) however $\bar{\bar{t}}$ is still not found to decrease with increasing values for D.

The model by Hastings assumes life-long immunity where as the model presented here assume immunity is short lived. As was stated in the introduction, this may have a major effect on the genetics of the model. However, Hastings model has been altered to act as if short term immunity occurred (by the method described in Hastings and Wedgwood-Oppenheim (1997)), and the same qualitative result as Hastings found earlier that t decreased as the number of immuno-allelic alleles increased was found, although the effect appeared to be weaker (Hastings and Wedgwood-Oppenheim, unpublished data).

Whether a short-term or life long immunity model is used in the model by Hastings, it is always assumed that the immunity profile of the population is constant, i.e. the percentage of the population in each immune class is the same. This is not the case here. Even with the transmission rate fixed, the mean level of immunity in the population changes greatly with different values for D (see Figure 4.3.5.2a), and this may also account for some of the difference between the two models. Unfortunately it would be difficult to alter the model to fix the immunity level at a constant value and thus it is only possible to speculate that this, in combination with the features described above may explain this qualitative difference between the two models.

Thus, the results presented do not explain the difference between the two models, although they do point to plausible explanations. It may however be worth pointing out, that with two models so different in many

aspects, it is almost surprising there are not more qualitative differences found.

The effects of different values for the inoculum size (θ) are also surprising. As the value of θ is increased, the values of $\overline{\overline{P}}_{inf}$ and $\overline{\overline{P}}_{v}$ descend, yet $\overline{\overline{t}}$ increases and $\overline{\overline{G}}_{ST(n)}$ and $\overline{\overline{G}}_{ST(d)}$ also increase in value. This is the opposite of the effects when $\overline{\overline{P}}_{inf}$ decreases due to changes in value of any of the other parameters.

It may well be that with many of the parameters, factors other than just the associated changes in Inoculation Rate are affecting the values of G_{ST} at the two loci and $\overline{\overline{t}}$. However, in other cases they are either acting in the same direction as the effects of the IR, or their effects are swamped by the large changes in IR over the range of parameter values examined (as is assumed to be the case when different numbers of alleles are examined at the immuno-allelic locus). Here, although $\overline{\overline{P}}_{inf}$ and $\overline{\overline{P}}_{v}$ do change with changes in the value of θ, those changes are small (2.1% in both cases) Thus, presumably, forces other than the effects of the changes in IR are causing the dominant effects on the genetic output statistics of the model.

4.3.2.2) The epidemiological statistics ($\overline{\overline{P}}_{inf}$, $\overline{\overline{W}}$, $\overline{\overline{P}}_{v}$ and $\overline{\overline{Z}}_{inf}$)

As has been stated above, it seems likely that the major factor affecting G_{ST} at the two loci is the Inoculation Rate. This has three determinants, but only one that varies greatly with the parameters of the model, namely the proportion of hosts that are infected (P_{inf}). Thus, it is important to examine how this statistic changes with the changes in the values of the different parameters of the model.

For a stable level of infection to occur, i.e. for P_{inf} to arrive at a stationary distribution (measured by the value $\overline{\overline{P}}_{inf}$), the number of new infections created must be compensated for by the same number of infections ending. New infections are formed by the inoculation of uninfected hosts with parasites possessing immuno-allelic alleles to which these hosts are susceptible. Infections end through the death of infected hosts or the clearing of parasites from the host via an immune response.

On intuitive grounds, it seems likely that changing a parameter to increase the rate of production of new infections is likely to increase the value of $\overline{\overline{P}}_{inf}$. Changing parameters to increase the rate at which parasite infections end should decrease the value of $\overline{\overline{P}}_{inf}$.

In fact, those parameters that increase the transmission per infected host, namely the Double Bite Rate (T) and the Coefficient of Vector Infectivity (c), do confirm this view. In the case of the parameter c, this is shown to largely be due to the increase in value of $\overline{\overline{P}}_{v}$, not the increase in the number of zygotes per vector that also occurs.

As the mean period of immune memory ($1/\beta$) is increased, $\overline{\overline{P}}_{inf}$ goes down. This can be explained by the fact the proportion of hosts susceptible to any particular immuno-allelic allele decreases with increasing $1/\beta$. Thus the proportion of double bites that successfully pass on an infection decrease with increasing values of $1/\beta$.

The parameters M_{min}, θ, s and ρ affect the mean period that an infection with a particular allele is likely to last. As this period increases, the rate at which hosts clear infections decreases. Thus one expects, and finds, that if these parameters are altered to increase the mean period of infection, then $\overline{\overline{P}}_{inf}$ should also increase.

The parasite basal growth rate (b_0) also effects the mean period of an infection with a particular immuno-allelic allele. However, the manner in which this occurs is more complex and less easy to predict. The exact effect depends on interactions with both M_{min} and s, and are beyond the present scope of this study.

The value of $\overline{\overline{P}}_{inf}$ was found to increase with the number of alleles at the immuno-allelic locus (D). This may initially seem obvious. It seems likely that the more parasite types that are treated differently by the immune system, the larger the value of $\overline{\overline{P}}_{inf}$ that would occur. However, the effects of different values of D are unlikely to be simple. Consider two extreme alternatives:

1) The growth and transmission of parasites are in complete competition (as would happen, for instance, if the immune system did not recognise the immuno-allelic alleles as different). In this case, as the number of alleles is increased, the rate of transmission of each allele would decrease and the value of $\overline{\overline{P}}_{inf}$ would remain constant, irrespective of the value of D.

2) Parasites with different immuno-allelic alleles grow, and are transmitted without competition. This is analogous to the concept of the independent transmission of strains assumed in many models by Gupta (e.g. Gupta and Day 1994b). In this situation, one would expect the value of $\overline{\overline{P}}_{inf}$ to increase with D. The degree to which $\overline{\overline{P}}_{inf}$ increases with D is described by Equation 4.2.3.5.1.

When considering the explicit competition between parasite genotypes (through non-specific density dependence in parasite growth) found in the model, it appears surprising that when the values for $\overline{\overline{P}}_{inf}$ are

compared to that predicted with no competition, i.e. extreme atternative two, (see Figure 4.2.3.5.1) there is so little difference. This result is explored further in Chapter 6 and discussed in Chapter 7.

The other output statistic of the model controlling the Inoculation Rate is the proportion of vectors, having bitten infected hosts, that do transmit parasites (P_v). The estimated stationary distribution value for P_v ($\overline{\overline{P}}_v$) is controlled by two aspects of the model. These controlling factors are the Coefficient of Vector Infectivity (c), and the distribution of densities of parasites in the infected hosts. The distribution of parasite densities is not directly measured in the simulations.

Thus it is expected, and found, that as c is increased, so does the value of $\overline{\overline{P}}_v$. When the number of zygotes are fixed, changing the value of c has the same effect on the value of $\overline{\overline{P}}_v$ as when the number of zygotes are not held constant. This latter result suggests that the number of zygotes per infected vector does not have a discernible effect on $\overline{\overline{P}}_v$.

Though changes in value of some parameters other than c do have a noticeable effect on the value of $\overline{\overline{P}}_v$ (e.g.. s, ρ, $1/\beta$). These effects are unlikely to be caused by the effects of altering the Inoculation Rate. This is suggested by the effects of the Double Bite Rate (T) on $\overline{\overline{P}}_v$. Despite altering the value of IR greatly there is very little effect on the value of $\overline{\overline{P}}_v$. These changes induced in $\overline{\overline{P}}_v$ are not monotonic, suggesting that whatever dynamic processes are involved, they are not simple.

It has been shown that the mean number of zygotes per infected vector at the stationary distribution ($\overline{\overline{Z}}_{inf}$) has an important effect on the values of $\overline{\overline{G}}_{ST}$ at the two loci. The value of $\overline{\overline{Z}}_{inf}$ is controlled by the same factors as is the value of $\overline{\overline{P}}_v$, namely the value of c and the distribution of

parasite densities amongst hosts. Thus one would expect the value of $\bar{\bar{Z}}_{inf}$ to vary with the value of $\bar{\bar{P}}_v$ as the parameters are manipulated. This is not always the case. There are several instances where $\bar{\bar{P}}_v$ shows a significant change in value with the manipulation of a parameter (for example the effects of changing the value of ρ). However no significant change in the value of $\bar{\bar{Z}}_{inf}$ is found. The value of $\bar{\bar{P}}_v$ is calculated from all the double bites from infected hosts that occur in the time steps examined, whereas $\bar{\bar{Z}}_{inf}$ is calculated only from those bites that transmit. Thus, the value of $\bar{\bar{Z}}_{inf}$ is calculated from only about half the number of measurements that $\bar{\bar{P}}_v$ is calculated from. This is likely to explain the smaller number of statistically significant results found. In those cases where changes in the value of $\bar{\bar{Z}}_{inf}$ were found to be significant, it did vary in the same direction as $\bar{\bar{P}}_v$ with the manipulation of a parameter value.

The degree to which hosts are immune to different immuno-allelic alleles is likely to have important effects on the G_{ST} values of the two loci in the model. If hosts are susceptible to all alleles, then no significant difference in value between $\bar{\bar{G}}_{ST(d)}$ and $\bar{\bar{G}}_{ST(n)}$ is likely to be observed. Assuming a constant IR, it would seem likely that the greater the number of alleles that hosts are immune to, the greater the difference between $\bar{\bar{G}}_{ST(d)}$ and $\bar{\bar{G}}_{ST(n)}$, and thus the greater the value of $\bar{\bar{t}}$.

However, the mean level of immunity in the host population at the stationary distribution ($\bar{\bar{W}}$) does not show any clear trend with the value of $\bar{\bar{t}}$. This is probably due to the fact that altering parameter values that change $\bar{\bar{W}}$ also alters the value of IR. As has been previously stated, in most cases the effects of IR on $\bar{\bar{G}}_{ST(d)}$, $\bar{\bar{G}}_{ST(n)}$ and $\bar{\bar{t}}$ appear to be dominant. Nevertheless, it is appropriate to consider the effects different parameter values have on the value of $\bar{\bar{W}}$.

The value of $\bar{\bar{W}}$ can be increased by increasing the rate at which parasites become immune to immuno-allelic alleles. This can occur through two general routes. Firstly, the rate at which hosts are infected by parasites can be increased (this can occur through increases in the inoculation rate or increases in numbers of zygotes per infected vector). This leads to a higher rate at which the immune system of each host experiences new alleles, and thus a greater rate of acquiring immunity to each allele.

Secondly, the rate at which a host, once infected, mounts an immune response to the immuno-allelic alleles in that infection can be increased. This is expected to occur with increasing values of the parameters of M_{min} and ρ (in both cases this should lead to a reduced mean period of parasite infection). Thus, one might assume that increasing the values of these parameters would also increase the value of $\bar{\bar{W}}$. This does not occur. This is because as the mean period of infection decreases, so does the value of $\bar{\bar{P}}_{inf}$. It appears, at least in the part of the parameter space examined, the corresponding decrease in transmission rate has a greater effect in reducing $\bar{\bar{W}}$ than the increase in the rate of development of an immune response has in increasing it.

The effects of altering the value of the mean level of immune memory ($1/\beta$) also produces a somewhat counter intuitive result. As the value of $1/\beta$ increase, the rate of loss of immunity from each host is reduced and thus, one might expect that $\bar{\bar{W}}$ is increased. However, $1/\beta$ decreases with increasing values of $1/\beta$. This result seems likely to have been caused by the large reduction in value of $\bar{\bar{P}}_{inf}$ with increasing values of $1/\beta$, and this has an overriding effect on $\bar{\bar{W}}$.

Chapter 5: Bootstrap Analysis

5.1) Introduction

This chapter is intended to examine the usefulness of the predictions of the model presented in Chapter 4 for finding immuno-allelic loci in real malaria populations. To do this, simulated field data are used and examined to see how likely a field collection of malaria parasites of a particular size is likely to find $G_{ST(d)}$ to be significantly greater than $G_{ST(n)}$ (t>1) with a given set of parameter values.

It is assumed that in such a field study, oocysts would be dissected off the guts of mosquitoes and scored for specific genetic markers. This is a method that has been employed in several field studies (e.g. (Babiker *et al.* 1994)). The oocyst is believed to reflect the genetics of the zygote and thus, in this chapter, zygote genotype data are collected for analysis.

5.2) Methods

For each simulation examined, zygotes are collected and their genotypes recorded every two days, from day 4,000, until 10,000 zygotes have been recorded.

A total of 1000 samples are then drawn (with replacement) from each data set for each sample size. Sample sizes of 500, 200, 100 and 50 zygotes were examined. The bootstrapping method (Sokal and Rohlf 1995) was then employed to examine whether t was significantly greater or less than one for each sample. Each sample was bootstrapped 2,500

times. This is the same method as was employed by Hastings (Hastings 1996), and a program created by Hastings was used to do this analysis.

5.3) Results

5.3.1) Standard Parameter Values

A data set of 10,014 zygotes was collected from a simulation using the Standard Parameter Values and started with Initial Conditions B. In Chapter 3, under these parameter values, values of 0.940, 0.823 and 1.141 were found for $\overline{\overline{G}}_{ST(d)}$, $\overline{\overline{G}}_{ST(n)}$ and $\overline{\overline{t}}$ respectively. The results of the bootstrap analysis can be seen in Table 5.3.1. It can be seen that with only 50 oocysts sampled, *t* is found to be significantly greater than one 43% of the time. However, when 500 oocysts are examined, no non-significant results were found. It should also be noted that in no case was *t* found to be significantly less than one, even with the smaller sample sizes.

5.3.2) Double Bite Rate = 0.05

A data set of 10,004 zygotes was collected from a simulation using the Standard Parameter Values for all parameters except the Double Bite Rate (T) which was given a value of 0.05 per day. In Chapter 4 under these conditions, values of 0.981, 0.952 and 1.031 were found for $\overline{\overline{G}}_{ST(d)}$, $\overline{\overline{G}}_{ST(n)}$ and $\overline{\overline{t}}$ respectively. The simulation was started with Initial Conditions B. The results of the bootstrap analysis can be seen in Table 5.3.2. It can be seen that with only 50 oocysts sampled, *t* is found to be significantly greater than one only 8% of the time. However, when 500 oocysts are examined,

91% of the of samples found t to be significantly greater than one. For a sample size of 100, 0.4% of samples found t to be significantly less than one (i.e. a Type 1 error occurred).

5.3.3) Double Bite Rate = 2.0

A data set of 11,595 zygotes was collected from a simulation using the Standard Parameter Values for all parameters except the Double Bite Rate (T) which was given a value of 2.0 per day. In Chapter 4, under these conditions, values of 0.899, 0.340 and 2.642 were found for $\overline{\overline{G}}_{ST(d)}$, $\overline{\overline{G}}_{ST(n)}$ and $\overline{\overline{t}}$ respectively. The simulation was started with Initial Conditions B. The results of the bootstrap analysis can be seen in Table 5.3.3. It can be seen that with only 50 oocysts sampled, t is found to be significantly greater than one 84% of the time. However, once 200 or more oocysts are examined, 100% of samples found t to be significantly greater than one. Under no sample sizes was t found to have a value significantly less than one.

5.3.4) Rate of Immune Recognition (ρ) = 0.005.

A data set of 10,002 zygotes was collected from a simulation using the Standard Parameter Values for all parameters except for the Rate of Immune Recognition (ρ) which was given a value of 0.005 per day. In Chapter 4 under these conditions, values of 0.708, 0.269 and 2.634 were found for $\overline{\overline{G}}_{ST(d)}$, $\overline{\overline{G}}_{ST(n)}$ and $\overline{\overline{t}}$ respectively. The simulation was started with Initial Conditions B. The results of the bootstrap analysis can be seen in Table 5.3.4. It can be seen that with only 50 oocysts sampled, t is found to

be significantly greater than one only 25% of the time. This is raised to 73% of the time when 200 oocysts are sampled and 94% when 500 are sampled.

Type 1 errors were found to occur in 11 cases (1.1%) with 50 oocysts, 4 cases with 100, 2 cases with 200 and no cases of type 1 errors were found when 500 oocysts were sampled.

Bootstrap analysis with the Standard Parameter Values

Sample Size	p>0.05	p<0.05	P<0.01	p<0.001
50	574	247	124	55
100	229	289	268	214
200	29	96	256	619
500	0	2	8	990

Table 5.3.1

Bootstrap analysis with the T = 0.05

Sample size	p>0.05	p<0.05	P<0.01	p<0.001
50	921	77	2	0
100	755	170	64	7
200	436	284	202	78
500	86	161	295	458

Table 5.3.2

Tables 5.3.1-5.3.4 (Tables 5.3.3 and 5.3.4 overleaf):
The tables presented show the number of times (out of 1000) that bootstrap analysis found a samples of different sizes (shown in the left hand column) to have a value of t significantly greater than 1. The values for p shown in the top row describe the different levels of significance examined. For further explanation, see text.

Bootstrap analysis with the T = 2.0

Sample size	p>0.05	p<0.05	P<0.01	p<0.001
50	155	279	383	183
100	22	77	192	709
200	0	0	12	988
500	0	0	0	1000

Table 5.3.3

Bootstrap analysis with the ρ = 0.005

Sample size	p>0.05	p<0.05	P<0.01	p<0.001
50	749	178	47	15
100	562	244	130	60
200	287	288	237	186
500	43	96	192	669

Table 5.3.4

5.4) Discussion

Collecting oocysts in the field has become a very useful method for population genetic analysis of malaria (Babiker and Walliker 1997). However, it is also a very time consuming and arduous technique and the collections of very large numbers of oocysts is at the moment impractical. It is therefore reassuring that to find a significant difference in G_{ST} values of the scale predicted by the model presented here, hundreds rather than thousands of oocysts are needed. The largest, published collection and genetic analysis of oocysts so far has been 71 (Babiker *et al.* 1994). Thus although the numbers needed are likely to be larger, they are not likely to be unfeasibly so, especially as one collection of oocysts could be seen as a 'reference collection' and used for the study of many different putative immuno-allelic loci. It is also reassuring that, under the wide variety of conditions examined, the probability of Type 1 errors always appeared to be small.

It is somewhat disconcerting to note that the value of t does not appear to be as strong an indicator as to the probability of finding a significant result as had been hoped. This is shown by examining the results of sections 5.3.3 and 5.3.4. With both sets of conditions very similar values for $\bar{\bar{t}}$ had previously been found in Chapter 4 (2.64 and 2.63 respectively) yet the chance of finding a significant result with different numbers of oocysts is found to be very different (100% with 200 oocysts in the former and 73% with 200 in the latter).

If one compares the results for the Standard Parameter Values ($\bar{\bar{t}}$=1.14) to the standard values in the paper by Hastings (t=1.17), despite very similar values for $\bar{\bar{t}}$ and t, very different results of the bootstrap

analysis is found. It was found that 200 or more oocysts was needed for a 90% probability of getting *t* significantly greater than one with the standard parameters for the model presented here, while 500 oocyst were needed in the study by Hastings. The values for $G_{ST(d)}$ and $G_{ST(n)}$ in Hastings's study were much lower than $\overline{\overline{G}}_{ST(d)}$ and $\overline{\overline{G}}_{ST(n)}$ found in this model (0.766 and 0.656 as opposed to 0.940 and 0.823) despite producing a similar value for *t*. When the values of $G_{ST(d)}$ and $G_{ST(n)}$ were artificially raised in Hastings's model to values similar to those in the model here, the bootstrap results became more like those found in this study (Hastings and Wedgwood-Oppenheim, unpublished data).

It would thus appear from both of the above comparisons that the higher the values for G_{ST} at both loci (for the same value of *t*) the greater the ease with which a significant effect can be found in a field study This appears to be in contrast to (Hastings 1996), who suggested that as G_{ST} was reduced, the chance of finding *t* significantly greater than one is increased. However in his study, *t* was assumed to change with G_{ST}. As has been stated above , when G_{ST} is changed but *t* is kept constant, Hastings's model produces very similar results to those found here (e.g. for *t*=1.17, GST(n) = 0.82 and 10 alleles at the immune locus there is a 97% chance of finding *t*>1 with p=0.05 significance).

The precise effects of the values of G_{ST} at each loci and the value of *t* on the probability of finding a *t* to be significantly greater than one has not been fully elucidated and must warrant further study.

6) Strain Competition

6.1) Introduction

In previous models of the effects of strain-specific immunity on malaria infected populations, it has been assumed that each 'strain' acts independently (see Section 1.3.5). In this chapter a series of simple analytical models are studied. First I examine the prevalence of a single parasite strain in a host population. The model is then modified to examine the effects of several independent strains on parasite prevelance. Finally I examine the effects of inter-strain competition on parasite prevelance.

6.2) The Model

6.2.1) A single strain

The models described in this section are modifications of a simple model described by Anderson and May (1991, pgs. 123-5). In their model only a single parasite strain is considered; it is further assumed that there is no geographic or age structuring, no infection-induced mortality, life-long immunity and no heterogeneity in the parasite population. Under these assumptions, and with their model, it can be shown that the equilibrium value for the proportion of hosts susceptible to infection (x*) is described by:

$$x^* = 1/R_0 \qquad \textbf{6.2.1.1}$$

where R_0 is the basic reproductive rate (see section 1.3.5). The equilibrium proportion of hosts infected (P*) is described by

$$P^* = \left(\frac{\mu}{\mu + \nu}\right)\left(1 - \frac{1}{R_0}\right) \qquad \textbf{6.2.1.2}$$

where μ is the *per capita* host death rate (= birth rate = rate of input of new susceptibles into the population), and ν is the *per capita* rate of recovery from an infection (all infections are assumed to be infectious).

Here I consider the situation where life long immunity does not occur, and thus new susceptible hosts do not enter the population just through birth, but also from loss of immunity of previously infected hosts. Equation 6.2.1.2 can then be altered (following Gupta *et al.* 1994a) to take this into account by including the additional term h, which describes the *per capita* rate of loss of immunity.

$$P^* = \left(\frac{\mu + h}{\mu + h + \nu}\right)\left(1 - \frac{1}{R_0}\right) \qquad \textbf{6.2.1.3}$$

If we assume that the birth rate is low compared to the rate of loss of immune memory, then this can be simplified to:

$$P^* = \left(\frac{h}{h + \nu}\right)\left(1 - \frac{1}{R_0}\right) \qquad \textbf{6.2.1.4}$$

It should be recognised that the terms R_0 and ν are not necessarily independent of each other in the first of these two equations. This is because R_0 can be described by:

$$R_0 = bC / \nu \qquad \textbf{6.2.1.5}$$

where C describes the Vectorial Capacity of the vector population (Bailey 1982) (see Section 2.4) and b, the probability that an infectious vector, when biting a susceptible host, transfers that infection (MacDonald 1957).

6.2.2) *Multiple strains with no competition*

Here the models of Anderson and May and Gupta *et al.* described above are modified to consider a situation where there are several (D) strains of malaria parasite which act independently of each other then one can predict the total parasite prevalence (P_T^*) as

$$P_T^* = 1 - \prod_{i=1}^{i=D} \left(1 - P_i^*\right) \qquad \textbf{6.2.2.1}$$

where P_i^* is the prevalence of infection of strain *i* in the population. If each strain has the same values for v, h and R_0, then this can be simplified to

$$P_T^* = 1 - \left(1 - P^*\right)^D \qquad \textbf{6.2.2.2}$$

and is thus equivalent to:

$$P_T^* = 1 - \left(1 - \left(\frac{h\left(1 - 1/R_0\right)}{h + v}\right)\right)^D \qquad \textbf{6.2.2.3}$$

6.2.3) *Multiple strains with competition*

There are many ways in which strains may compete in a population: there may be competition for resources in hosts; for vector transmission between hosts; for new hosts; for a mixture of the above.

In this model it is assumed simply that competition is only for transmission. A simple form of transmission competition is that the vectorial capacity for the host population is constant irrespective of the numbers of parasite strains present.

I assume that each strain is equal in its competitive ability and thus that the Vectorial Capacity for each strain is equal and is reduced by the presence of other strains in the population:

$$C_i = {}^{C_T}/_{D} \qquad \textbf{6.2.3.1}$$

where C_T is the Vectorial Capacity for a strain if it is alone in the host population, and C_i is the Vectorial Capacity for a strain if other strains are present in the population. If it is assumed that there is no competition between strains in each host, then v is not altered by the presence of other strains in the population and thus:

$$R_{0_i} = {}^{R_{0_T}}/_{D} \qquad \textbf{6.2.3.2}$$

where R_{0_T} is the basic reproductive rate for a strain if it is alone in the host population, and R_{0_i} is the basic reproductive rate for a strain if other strains are present in the population. This situation can be seen as analogous to the situation described by Saul (Saul 1996), where the total R_0 for a population, can be seen as the mean of the R_0s for all the immunologically distinct strains present. With this form of competition the total parasite prevalence is:

$$P_T^* = 1 - \left(1 - \left(\frac{h\left(1 - {}^{D}/_{R_{0_T}}\right)}{h+v}\right)\right)^D \qquad \textbf{6.2.3.3}$$

6.3) Results

6.3.1) A single strain.

It is evident from the equation 6.2.1.4 that for a single parasite strain, in the absence of other strains, as the value of R_0 increases, so does the proportion of hosts infected. The maximum value for P^* is determined by the relative values of h and ν, with P^* tending to h/ν when R_0 is very large. This is illustrated graphicaly in Figure 6.3.1. It can be seen that as 1/h is increased (with a fixed value for ν), the prevalance, P^*, for each value of R_0 is less.

The effects of R_0 on parasite prevalence with a singe parasite strain

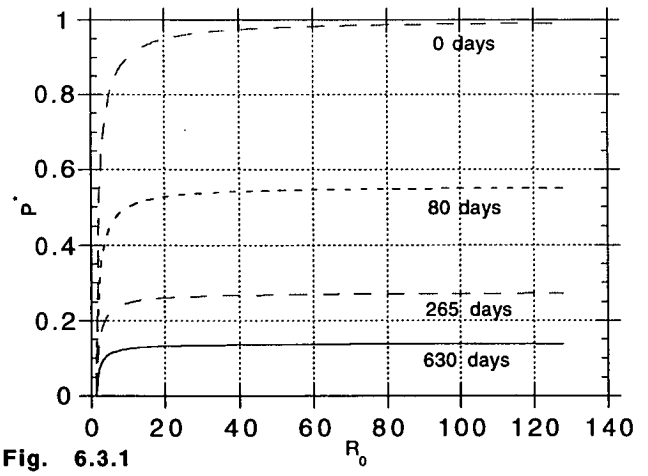


Fig. 6.3.1

Figure 6.3.1: Parasite prevalence is plotted against R_0 using equation 6.2.1.4. 1/ν is set at 100 days, while each line has a different value for h. The values of 1/h is written under each line.

6.3.2) Parasite prevalence with no competition.

In Figure 6.3.2 the effects of the number of strains present on the total parasite prevalence can be seen under the assumption of no competition between strains. As the number of strains present in the population increases, so does the parasite prevalence. The total parasite prevalence appears to be asymptotically approaching a value of one. The rate at which the prevalence approaches one is dependent on the values of ν and h. It is also effected by the value of R_0, especially when the value of R_0 is small (see Equation 6.2.2.3).

The Effects of the Number of Parasite Strains on Parasite Prevalence with no Inter-strain Competition.

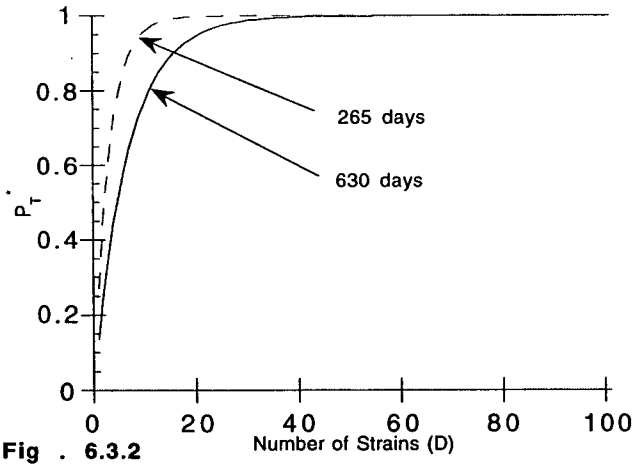


Fig . 6.3.2

Figure 6.3.2: Parasite prevalence is plotted against the number of parasite strains (D) using equation 6.2.2.3. $1/\nu$ is set at 100 days, while each line has a different value for h. The value of 1/h is written under each line. R_0 is given a value of 100.

6.3.3) Parasite prevalence with inter-strain competition.

Figure 6.3.3.1 represents the effects of strain competition on the parasite prevalence (see equation 6.2.3.3) in a population. The upper two lines represent the effects of strain competition on the total parasite prevalence in a population. The lower two lines represent the effects of strain competition on the prevalence of a single strain in the population.

If one compares the effects of number of strains on total parasite prevalence with and without inter-strain competition (see Figures 6.3.2, 6.3.3.1 and 6.3.3.2), it is evident that when the number of strains are large (i.e. close to the R_0 for the population), there is a very large effect. However when D is small in relation to R_0 this difference is quite small.

The Effects of the Number of Parasite Strains on Parasite Prevalence with no Inter-strain Competition.

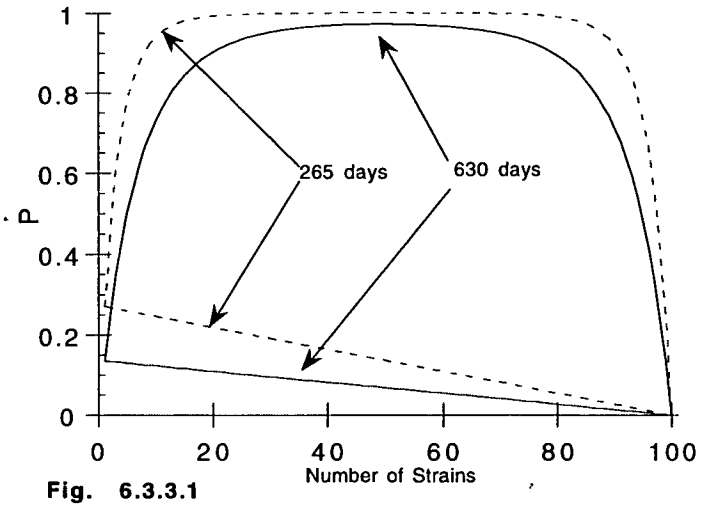


Fig. 6.3.3.1

Figure 6.3.3.1: The upper two lines represent the total parasite prevalence with different numbers of parasite strains (D) with different values for 1/h using equation 6.2.3.3. The lower two lines represent the parasite prevalence of individual strains as the parasite prevalence increases. 1/v is set at 100 days, while each line has a different value for h. The value of 1/h are presented on the graph. R_0 is given a value of 100.

Comparison of Effects of the Number of Parasite Strains on Parasite Prevalence With and Without Inter-strain Competition.

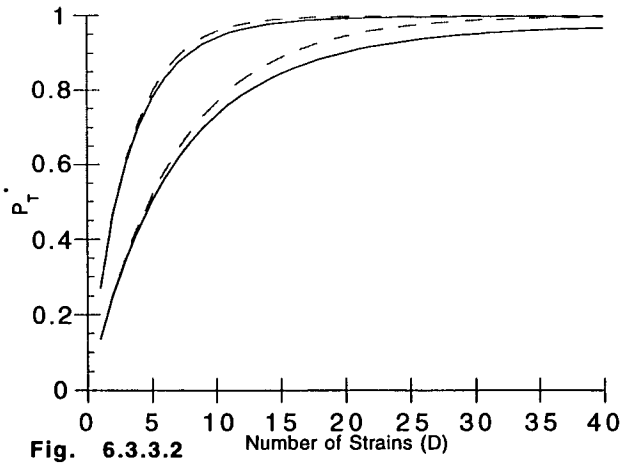


Fig. 6.3.3.2

Figure 6.3.3.2: The dashed lines represent different total parasite prevalence with no inter-strain competition, while the solid lines represent total parasite prevalence with inter strain competition. The upper pair of lines have a value for 1/h of 265 days, while the lower pair have a value of 630 days. The value 1/v is set at 100 days, while each line has a different value for h. R_0 is given a value of 100.

6.4) Discussion

The models presented here show that inter-strain competition can have very profound effects on the epidemiology of malaria population. However these effects are small, and only evident when the numbers of strains is large and approaching the value of R_0 for the population. Most profoundly this study shows that, under the assumptions stated, a population cannot sustain more parasite strains than the value for R_0 of the total population. This result is easily understood by examining equation 6.2.3.2, for as D approaches R_{0T}, R_{0i} approaches one at which point parasites cannot be maintained in the population.

This is an interesting result and may have implications as to the control of parasite populations (see Chapter 7), however that depends on whether the method of inter-strain competition described is applicable to the real world. Clearly more work needs to be done to examine whether these qualitative results are robust to other, possibly more realistic, forms of inter-strain competition (e.g. involving intra-host competition).

7) General Discussion

The aim of this study has been to examine the effects of strain-specific immunity on the population genetics of malaria parasites in the context of their epidemiology. This has been done through the construction of a theoretical model which has been examined through computer simulation. A second mathematical model has also been examined to look just at the effects of strain-specific immunity on the epidemiology of malaria populations

The study has focused on the use of Wright's F statistics (specifically G_{ST}) to examine the degree of genetic substructuring within hosts and potential for outcrossing in malaria parasites, although other aspects of the parasite population have also been examined. The values of G_{ST} for the two loci have been studied, one being immuno-allelic, the other neutral.

7.1) Discussion of the results

The G_{ST} values for both loci were examined, first over time from very different starting conditions, and second at the stationary distributions that were achieved. At all times and under all conditions examined, G_{ST} at the immuno-allelic locus was found to be greater than at the neutral locus (i.e. t was found to be greater than one).

Manipulation of the parameter values in the model showed that, in almost all cases, where a parameter was altered such that the proportion of hosts infected during the stationary distribution ($\bar{\bar{P}}_{inf}$) was increased, $\bar{\bar{t}}$ was

also found to increase, and the $\overline{\overline{G}}_{ST}$ values for the two loci were found to decline.

When the effects of the number of alleles at the immuno-allelic locus (D) were examined it was found that as D was increased, the values of $\overline{\overline{G}}_{ST}$ at the two loci decreased, while $\overline{\overline{t}}$ and $\overline{\overline{P}}_{inf}$ increased.

A detailed discussion of the results has been presented at the end of Chapters 3 to 6. However several important points can be drawn from the results and these are highlighted here.

1) The dynamics of the model examined in Chapter 3 show that the hypothetical parasite population acts in an epidemiologically expected manner. The population shows an initial epidemic "boom and bust" cycle as one would expect from parasites invading a population of susceptible hosts.

2) It has been postulated at the end of Chapter 4 that the population-genetic results presented in that chapter can largely be explained by considering the Inoculation Rate (IR), where the IR is the product of the Double Bite Rate (T), the proportion of hosts infected (Pinf), and the proportion of double bites biting from an infected host that transmit parasites (P_v). When a single parameter was manipulated, the larger the value of IR, in almost all cases, led to lower values of $\overline{\overline{G}}_{ST(n)}$ and $\overline{\overline{G}}_{ST(d)}$, and greater values of $\overline{\overline{t}}$. The likely reason for this was explained at the end of Chapter 4.

The changes in value of IR may also explain the changes in G_{ST} found during the epidemic phase of the infections described in Chapter 3. Unfortunately, these changes do not appear to be simple. There is a lag in

the response of t to P_{inf}, with t rising to a peak shortly after the peak in P_{inf}. This is not surprising as one would expect G_{ST} to take time to equilibrate to a new inoculation rate (Crow 1986), this time being lengthened by the time taken for parasite densities to grow up after their initial inoculation into a host. However the lag in t's increase is very different in the two runs, suggesting that there are other factors involved in determining the dynamic phase of G_{ST} than just the IR. These factors are beyond the scope of this study.

3) It has been stated in Section 1.3.5 that the independent transmission of immunologically distinct parasite strains is unlikely. The model presented is constructed with several features that cause parasites with different immuno-allelic alleles to interact and compete. These features are (i) a single population of vectors; (ii) a small number of zygotes in most transmitting vectors; and (iii) total parasite density dependent growth in the hosts. Surprisingly, in Chapter 4, the values of $\overline{\overline{P}}_{inf}$ found with different values of D suggest that the immuno-allelic alleles appear to act close to independently in terms of overall transmission.

The studies in Chapter 6 suggest that there is a simple explanation for this. Using a different model of malaria epidemiology, it is shown that when the number of immunologically distinct strains present in a population is small compared to basic reproductive rate of the parasites, the proportion of hosts infected is a poor indicator of inter-strain competition.

It is not known what the basic reproductive rate of the parasite population in Chapter 4 is, however it is likely to be larger than the number of alleles at the immuno-allelic locus (D). It is also seems likely that the competition between strains in the simulation model described in Chapter 2 is less stringent than that used in Chapter 6. In combination, these factors

are likely to explain the, initially surprising, effects of D on $\bar{\bar{P}}_{inf}$ found in Chapter 4.

7.2) Potential problems with the study

It is clear from the literature that some form of strain-specific immunity occurs (see Section 1.3.2). However, to what degree it is genetically or epigenetically controlled is not known. Here we have assumed that strain specificity is completely genetically controlled (i.e. it is haplotype specific), and by a single locus. This need not be the case.

There has been much interest recently in antigenic switching, especially since some of the *var* genes have been cloned (Su *et al.* 1995), and these genes may be important in strain-specific immunity (Borst *et al.* 1995). If this is the case, this is likely to be through a mixture of epigenetic (e.g. the switching of antigen expression in a particular clone) and genetic (e.g. different repertoires of *var* genes in each host) factors. This study has not attempted to examine the effects of antigenic variation on the population genetics of malaria parasites. As yet, too little is known about the diversity of alleles amongst hosts, the control of switching and the importance of the *var* gene products in controlling human malaria infections for a population genetic model to be constructed to take antigenic variation in malaria populations into account. Until more is known of the biology of this gene family, it will be difficult to predict the effects of antigenic switching on the population genetics of malaria parasites. When more is known of the action of the *var* genes it should be straightforward to build them into the model. Indeed, antigenic switching in a host can be seen as similar to inoculation

of that host with a new parasite clone with a new immuno-allelic allele, but with the same neutral allele

The *var* genes may define part of the specification of immunologically distinct strains, but other immuno-allelic loci may also be involved. If this is the case, these other loci may still be identified by examining their G_{ST} values with respect to the G_{ST} values at neutral loci. Assuming the immuno-allelic loci are not closely linked to the *var* genes, the presence of the var genes will act as background 'noise' to the measurements of G_{ST}, and should not obscure the difference between $G_{ST(n)}$ and $G_{ST(d)}$.

If immunity is largely haplotype specific, there are several other methods by which it can be defined other than by a single locus with no cross-reactivity between alleles (as is assumed in this study). Namely, there could be multiple loci involved, and/or their alleles could show some degree of cross reactivity. Though the model presented here does not address these alternatives, it is possible to speculate on the population genetic outcomes of these alternatives.

Firstly, the effects of multiple loci on the strain structure of parasites has become a topic of interest recently (Gupta *et al.* 1996; Hastings and Wedgwood-Oppenheim 1997). The work by Gupta and co-workers has proposed that if alleles at multiple loci are responsible for strain-specific immunity, then a strong degree of linkage disequilibrium will develop between alleles at the different loci. This is likely to mean that no matter how many loci are present, the $\overline{\overline{G}}_{ST(d)}$ value for each locus will be equal, and greater than a $\overline{\overline{G}}_{ST(n)}$ value measured. The loci are likely to act epidemiologically as a single locus with the number of alleles equal to the smallest number of alleles that any of the loci have.

Secondly, partial cross-reactivity between alleles at a locus will reduce the average degree of susceptibility of the host population and consequently reduce the value of $\overline{\overline{P}}_{inf}$. Thus in much the same manner as increasing the mean period of immune memory ($1/\beta$), it is likely to increase $\overline{\overline{G}}_{ST(d)}$ and $\overline{\overline{G}}_{ST(n)}$ and decrease $\overline{\overline{t}}$.

The model presented here assumes that immunity to a particular immuno-allelic allele is completely sterile against parasites with that allele. However, as stated in Chapter 1, sterile immunity against malaria is a rare phenomenon in man. In most cases anti-parasite immunity appears as a reduced parasite density and a shorter period of patent infection (e.g. Boyd and Kitchen 1943; Ciuca *et al.* 1934).

It is possible to speculate as to the effects of partial immunity on the outcomes of the model. It seems likely that the less effective the immune response, the greater the proportion of hosts that become infected. This is likely to cause an increase in the Inoculation Rate, which is itself likely to reduce the values of $\overline{\overline{G}}_{ST(n)}$ and $\overline{\overline{G}}_{ST(d)}$ and increase the value of $\overline{\overline{t}}$. This prediction may seem counter-intuitive, however the situation can be seen as analogous to the reduction in the period of immune memory shown in Section 4.2.3.1

7.3) Implications of the study

7.3.1) The usefulness of examining G_{ST} values from field data to look for immuno-allelic loci, and the choice of field site.

The results of the model clearly show that the general prediction of Hastings, namely that G_{ST} values can be used to detect an immuno-allelic locus, is robust to very different assumptions about the biological nature of malaria parasites. More specifically, Hastings's prediction holds up even when a short period of immune memory is assumed, and also when the parasite population is in a state of flux as well as at a stationary distribution.

It has been suggested that, by measuring the G_{ST} values for potentially immunologically important loci in a population, and comparing these values to those measured at neutral loci (e.g. microsatellites), one can help to discover a locus' immunological importance (Hastings 1996b; also see Section 1.4.3).

The results in Chapter 4 show that a wide variety of parameters of the model, have large effects on the magnitude of $\bar{\bar{t}}$. It was also shown in Chapter 5 that the magnitude of $\bar{\bar{t}}$ does not necessarily correlate with the probability of finding $t>1$ in a field study.

When comparing different real malaria populations, most of the parameters of the model are likely to be constant, or at least similar in value (unless comparing different species). The only parameters that are likely to change drastically between potential study sites are the size of the human population and the transmission rate. The size of the human population has been shown to have no discernible effect on $\bar{\bar{t}}$, however, larger the

transmission rates have been shown to both lead to larger values of $\bar{\bar{t}}$ and to increase the chances of finding $t>1$ in a field study. Thus, the results presented here agree with Hastings's view, that if such a technique is to be used in the field, examining parasites in an area of high transmission, rather than low transmission would be better for detecting a difference between the two loci.

The model highlights the importance of examining the level of transmission from the point of view of rate of inoculation *to* the population rather than the rate of transmission *from* infected hosts, i.e. the Inoculation Rate is a more important factor for comparison between areas than is the Double Bite Rate. In epidemiological terms this is equivalent to comparing Entomological Inoculation Rates rather than Vectorial Capacities (Bailey 1982; Molineaux *et al.* 1988).

The accurate measure of vectorial parameters can be difficult (Gupta and Snow 1996), however, increasing transmission rate does appear to correlate (at least in the model presented in Chapter 2) with the reduction in $\bar{\bar{G}}_{ST(n)}$. Thus, perhaps the best way to choose a study site to look for immuno-allelic loci, is to look for ones with low estimates for $G_{ST(n)}$. Thus Tanzania or the Gambia, may well be better than Papua New Guinea for such a study (Babiker *et al.* 1994; Carter and McGregor 1973; Paul *et al.* 1995).

In Chapter 3 it is shown that changes in the value of t lag behind, but appear to mirror, the value of P_{inf} during the epidemic phase. In the two simulation runs examined in detail (with different starting conditions), very different lag periods occurred, from a few days to around half a year. Thus, when a population is in flux, it is not possible to predict when the largest

value of t is likely to occur. In such a situation it would again appear to be most sensible to measure t when $G_{ST(n)}$ is at its lowest value, as this does appear to correlate reasonably well with high values of t during the epidemic phase as well as during the stationary distributions examined.

In the studies presented in Chapters 3 and 4, G_{ST} values are calculated from the parasite genotype densities in each host. According to the assumptions of this model, this should predict the G_{ST} values one would find in the zygotes in the vectors. In Chapter 5 zygote data was explicitly examined. Population genetic studies of real malaria populations have calculated outcrossing rates by looking at genetic polymorphisms in either oocysts or blood stage parasites. If oocysts reflect the genetics of the parasite zygotes, using them is clearly an accurate way of measuring G_{ST}. Examining outcrossing rates by prediction from the blood stages has much more room for error (see Section 1.4.1), although in the one study where the outcrossing rates from oocyst and blood data was examined they two did agree (Hill and Babiker 1995; Hill *et al.* 1995). However, despite the increased risk of error, examining the blood stages has many advantages. For instance, blood samples are much easier to collect than dissecting oocysts from the guts of mosquitoes. Also it is much easier to collect parasites at the blood stage from specific age groups than it is to collect oocysts from mosquitoes that have fed on those age groups.

It would be of great interest to know whether the inherent inaccuracies in calculating G_{ST} from blood collections translate into a large loss of power when testing whether a locus is immuno-allelic. This could be done through the simulation of blood data sets with specific G_{ST} values and the examination of these data sets having simulated the use of different experimental techniques (e.g. isoenzyme polymorphisms as used by Carter

and McGregor (1973) or the polymerase chain-reaction as used to generate the data for the study by Hill and Babiker (1995)). Such simulations could then be used, via bootstrap analysis, to weigh up the relative merits of blood versus oocyst data analysis when using G_{ST} measurements to identify immuno-allelic loci. This is clearly important and warrants study.

However, the results of the bootstrap analysis in Chapter 5 suggest however that only 200 to 500 oocysts are needed for the reliable finding of an immuno-allelic locus. Thus, although it would clearly be more convenient to use blood data, an oocyst based study should be within the bounds of practicality for a field study

7.3.2) Other implications

The dynamic nature of this model may provide a useful tool in the design of control strategies. The value of $G_{ST(n)}$ in a population is a useful measure of the degree of outcrossing in that population. This may have implications on the spread of drug and vaccine resistance and the level of virulence in the population (see Section 1.4.1). The dynamic nature of the model takes into account the effects of parameter values on the epidemiology of the system, and its knock-on effects on the population genetics. Thus, unlike previous population genetic models (specifically (Hastings 1996) the effects of different control strategies on $G_{ST(n)}$ (and thus, the degree of outcrossing) can be predicted.

The effect of the number of immuno-allelic alleles on $\overline{\overline{P}}_{inf}$ may be useful in the design of vaccine strategies. If independently transmitted, immunologically distinct strains exist in a malaria population (e.g. scenario (2) in Section 4.3.2.2), and if a vaccine removes only a few of these strains

from the population, then one would expect a large reduction in the parasite prevalence in the population. If, however, there is complete competition between strains (scenario (1) in Section 4.3.2.2) then such a partial vaccine would have no effect on the total prevalence. In this model, it was found that, despite no independent transmission of immuno-allelic alleles, there was still a sizeable difference in P_{inf} with different values of D, and thus a vaccine against only a few of the alleles at the immuno-allelic locus would still be expected to reduce P_{inf} in the population and thus may be worth implementing.

However, the studies in Chapter 6, suggest that if the number of immuno-allelic alleles are large, i.e. a similar number to the R_0 for the population, then a small reduction in the number of strains present may in fact increase the total parasite prevalence. This result is only preliminary and more work needs to be done on this topic before any serious concerns along these lines should be raised.

7.4) Possible Future Directions

There are many interesting features of the model reflected in the results described and discussed above, and there are clearly many potential follow-on studies. These can generally be described under two headings, though there is much overlap between them.

1) Further examination of the processes underlying the results of the model.

As has been stated previously, the nature of a computer simulation study is such that some of the underlying causes of a result can be intractable. However, deliberate modifications of the model can be used to clarify the various causes to some extent. This was done when examining the coefficient of vector infectivity (c) in Section 4.2.2.2

There are many possible manipulations of the model which may produce illuminating results. For instance, the dominant nature of IR in the study may well be obscuring other interesting interactions in the model. Thus, it is of interest to control deliberately the inoculation rate and to manipulate the parameters against this background. This may be of use in predicting the effects of the same inoculation rate on heterogeneous population (e.g. the effects of age - see below).

2) Examination of the possible application of the model to field studies of malaria populations.

It has been shown in a study by Ntoumi and co-workers (Ntoumi *et al.* 1995) that the number of alleles at the MSP2 locus, found in individuals in an endemic population, decreases with age. This result fits with the general concept of strain-specific immunity. As the population gets older the mean number of immunologically distinct strains each individual has experienced increases, thus the probability of being immune to any new inoculation increases, and thus the probability of a mixed infection is reduced.

In the model presented here, the host population is age structured, and includes births, deaths, and hosts experiencing more alleles at the immuno-allelic locus with increasing age. Thus, one would expect older hosts to have higher levels of immunity than the younger ones. If the older

and younger age groups were separate populations, P_{inf} for the older population would be expected to be smaller than for the younger one, and thus the IR to be smaller, the G_{ST} values to be higher, and the value of *t* to be smaller. However, as the populations are not separate but interconnected, a high IR, due to the high P_{inf} in the younger age groups is, in part, experienced by the older ages. This higher IR coupled with the higher levels of immunity in the upper age groups may well cause an increase in value of *t* due to the greater levels of immunological restriction. If this is the case ,it may be of great importance as the most effective way to examine Hastings's predictions in the field may be to look specifically at one age group. This has not been examined in this study, but should be done.

7.5) In conclusion

The study presented here lends weight to the predictions of Hastings as to a possible method for identifying immunologically important loci. This model shows that these predictions are robust even in the face of more realistic parasite life-history assumptions, including the use of a short period of immune memory. However, if the immune memory to any particular allele at the locus identified is only short-lived, what use is its identification?

Once a locus is identified as being an important target of anti-parasite immunity and that its polymorphism is involved in the evasion of this immunity, it may be possible to construct a vaccine that promotes an immune response against invariant regions of the locus and also aimed at eliciting longer-term immune memory. However, even a vaccine which elicits short term protection may still be useful under certain conditions. For example, it may be useful to protect very young children through their first few years of life where they appear to be at their most vulnerable; it may be useful in the prevention of the spread of an epidemic in a region with sporadic, unstable malaria; and it maybe useful as a "visitors' vaccine" for people visiting an endemic region from a non-malarious one (Saul 1994).

As well as confirming the robustness of Hastings's results, this study provides some insights into the level of interdependence of different immunologically distinct strains, and moreover it highlights the inter-relationships of population genetics and the epidemiology of malaria parasites. The models constructed for these studies may also prove to be a resource in any future examination of the population genetics and epidemiology of malaria parasites and other disease agents.

8) Bibliography

Alberts, B. *et al.*, 1983 *Molecular biology of the cell.* Garland Publishing, Inc., New York.

Anders, R. F. *et al.*, 1988 Antigens with repeated amino acid sequences from asexual blood stages of *Plasmodium falciparum.* Progress in Allergy **41:** 148-172.

Anderson, R. M., and R. M. May, 1992 *Infectious diseases of humans.* Oxford University Press, Oxford.

Anderson, R. M. *et al.*, 1989 Non-linear phenomena in host-parasite interactions. . Parasitology **99:** S59-S79.

Ashburner, M., 1989 *Drosophila: A laboratory handbook.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Babiker, H. A. *et al.*, 1994 Random mating in a natural population of the malaria parasite *Plasmodium falciparum.* .

Babiker, H. A., and D. Walliker, 1997 Current views on the population structure of Plasmodium falciparum: Implications for control. Parasitology Today **13:** 262-267.

Bailey, N. T. J., 1982 *The Biomathematics of Malaria.* Charles Griffin, High Wycombe.

Baird, J. K., 1994 Age and "Strain-transcending" Immunity to *Plasmodium falciparum.* Parasitology Today **10:** 302.

Baird, K. J., 1995 Host age as a determinant of naturally acquired immunity to *Plasmodium falciparum.* Parasitology Today **11:** 105-111.

Ballet, J. J. *et al.*, 1985 Human lymphocyte responses to Plasmodium falciparum merozoite antigens. A functional assay of protective immunity? Transactions of the Royal Socitey of Tropical Medicine and Hygiene **79:** 497-499.

Baruch, D. I. *et al.*, 1995 Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitised human erythrocytes. Cell **82:** 77-87.

Beale, G. H., and D. Walliker, 1988 Genetics of malaria parasites, pp. 379-393 in *Malaria: principles and practice of malariology*, edited by W. H. Wernsdorfer, and I. A. McGregor. Churchill Livingstone, Edinburgh.

Begon, M. *et al.*, 1990 *Ecology: individuals, populations, and communities.* Blackwell Scientific Publications, Oxford.

Beier, J. C., 1993 Malaria sporozoites: survival, transmission and disease control. Parasitology Today **9:** 210-215.

Billingsley, P. F. *et al.*, 1994 Relationship between prevalence and intensity of *Plasmodium falciparum* infection in Natural populations of Anopheles mosquitos. American Journal of Tropical Medicine and Hygiene **51:** 260-270.

Borst, P., 1991 Molecular genetics of antigenic variation. Immunoparasitology Today - A combined issue of Immunology Today and Parasitology Today : A29-A33.

Borst, P. *et al.*, 1995 Antigenic Variation in Malaria. Cell **82:** 1-4.

Boyd, M. F., and S. F. Kitchen, 1943 On attemptes to hyperimmunize convalescents from vivax malaria. American Journal of Tropical Medicine **23:** 209-225.

Boyd, M. F., and C. B. Matthews, 1939 Further observations on the duration of immunity to the homologous strain of *Plasmodium vivax*. American Journal of Tropical Medicine **19:** 63-67.

Brown, N. K., and I. N. Brown, 1965 Immunity to malaria: antigenic variation in c hronic infections of *Plasmodium knowlesi*. Nature **208:** 1286-1288.

Carter, R., and R. W. Gwadz, 1980 Infectiousness and gamete immunization in malaria, pp. 263-297 in *Malaria*. Academic Press Inc.

Carter, R., and I. A. McGregor, 1973 Enzyme variation in *Plasmodium falciparum* in the Gambia. Transactions of the Royal Society of Tropical Medicine and Hygiene **67:** 830-837.

Carter, R., and A. Voller, 1975 The distribution of enzyme variation in populations of *Plasmodium falciparum* in Africa. . Transactions of the Royal Society of Tropical Medicine and Hygiene **69:** 371-376.

Charlesworth, B., 1989 The evolution of sex and recombination. Trends in Ecology and Evolution **4:** 264-267.

Charlesworth, B., 1994 *Evolution in age-structured populations*. Cambridge University Press, Cambridge.

Ciuca, M. *et al.*, 1934 Immunity in malaria. Transactions of the Royal Society of Tropical medicine and Hygiene **27:** 619-622.

Creasey, A. *et al.*, 1990 Genetic diversity of *Plasmodium falciparum* shows geographical variation. . American Journal of Tropical Medicine and Hygiene **42:** 403-413.

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

Crow, J. F., 1986 *Basic concepts in population, quantitative and evolutionary genetics*. W. H. Freeman, San Francisco.

Crow, J. F., 1992 An advantage of sexual reproduction in a rapidly changing environment. Journal of Heredity **83:** 169-173.

Crow, J. F., and M. Kimura, 1970 *An introduction to population genetics theory*. Harper and Row, New York.

Curtis, C. F., and L. N. Otoo, 1986 A Simple Model of the Build-up of Resistance to Mixtures of Ant-malarial Drugs. Transactions of the Royal Society of Tropical Medicine and Hygiene **80:** 889-892.

Day, K. P., and K. Marsh, 1991 Naturally acquired-immunity to *Plasmodium falciparum*. Immunoparasitology Today - a combined issue of Immunology Today and Parasitology Today **:** A68 - A71.

Deleron, P., and C. Chougnet, 1992 Is immunity to malaria really short-lived? Parasitology Today **8:** 375-379.

Dye, C., 1991 Population genetics of nonclonal, nonrandomly mating malaria parasites. Parasitology Today **7:** 236-240.

Dye, C., and H. C. F. Godfray, 1993 On sex ratio and inbreeding in malaria populations. Journal of Theoretical Biology **161:** 131-134.

Felger, I. *et al.*, 1995 The use of the polymerase chain-reaction for more sensitive detection of *Plasmodium falciparum*. Papua New Guinea Medical Journal **1995:** 52-56.

Fincham, J. R. S., 1983 *Genetics*. Wright, Bristol.

Frank, S., 1994 Kin selection and virulence in the evolution of protocells and parasites. Proceedings of the Royal Society, London, series B **258:** 153-161.

Frank, S. A., 1996 Models of parasite virulence. Quaterly Review of Biology **71:** 37-78.

Garnham, P. C. C., 1988 Malaria parasites of man: life-cycles and morphology (excluding ultrastructure), pp. 61-96 in *Malaria: Principles and practice of malariology*, edited by W. Wernsdorfer H., and I. McGregor. Churchill Livingstone, Edinburgh.

Grafen, A., and M. E. J. Woolhouse, 1993 Does the negative binomial distribution add up? Parasitology Today **9:** 475-477.

Gravenor, M. B. *et al.*, 1995 The regulation of malaria parasitaemia: parameter estimates for a population model. Parasitology **110:** 115-122.

Graves, P. M. *et al.*, 1988 Measurement of malarial infectivity of human populations to mosquitoesin the Madang area, Papua New Guinea. Parasitology **96:** 251-263.

Graves, P. M. *et al.*, 1990 Estimation of Anopheline survival rate, vectorial capacity and mosquito infection probability from malaria vector infection rates in villages near Madang, Papua New Guinea. J. App. Eco. **27:** 134-147.

Gupta, S., and K. Day, 1994a Age and 'strain-transcending' immunity to *Plasmodium falciparum* - reply. Parasitology Today **10:** 302.

Gupta, S., and K. P. Day, 1994b A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. Parasite Immunology **16:** 361-370.

Gupta, S. *et al.*, 1996 The maintenance of strain structure in populations of recombining infectious agents. Nature Medicine **2:** 437-442.

Gupta, S., and R. W. Snow, 1996 How do bednets influence the transmissibility of *Plasmodium falciparum*? Parasitology Today **12:** 89-90.

Gupta, S. *et al.*, 1994a Theoretical studies of the effects of heterogeneity in the parasite population on the transmission dynamics of malaria. Proc. Ro. Soc. Lond. B. **256:** 231-238.

Gupta, S. *et al.*, 1994b Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*. Science **263:** 961-963.

Harinasuta, T., and D. Bunnag, 1988 The clinical features of malaria, pp. 709-734 in *Malaria: Principles and practice of malariology*, edited by W. H. Wernsdorfer, and I. McGregor. Churchill Livingstone, Edinburgh.

Hartl, D. L., and A. G. Clark, 1989 *Principles of Population Genetics*. Sinauer Associates, Inc., Sunderland.

Hastings, I. M., 1996 Population genetics and the detection of immunogenic and drug-resistant loci in *Plasmodium*. Parasitology **112:** 155-164.

Hastings, I. M., 1997 A model for the origins and spread of drug-resistant malaria. Parasitology **115:** 113-141.

Hastings, I. M., and B. A. Wedgwood-Oppenheim, 1997 Sex, strains and virulence. Parasitology Today (in press) .

Hellriegel, B., 1992 Modelling the immune response to malaria with ecological concepts: short-term behaviour against long term equilibrium. Proceedings of the Royal Society of London Series B. **250:** 249-256.

Hill, W. G., and H. A. Babiker, 1995 Estimation of numbers of malaria clones in blood samples. Proceedings of the Royal Society, London, Series B **262:** 249-257.

Hill, W. G. *et al.*, 1995 Estimation of Inbreeding Coefficients from Genotypic data on multiple alleles, and Aplicatication to estimation of clonality in Malaria Parasites. Genetical Research **65:** 53-61.

Hughes, A. L., 1992 Positive selection and interallelic recombination at the Merozoite Surface Antigen-1 (MSA-1) Locus of *Plasmodium falciparum*. Molecular and Biological Evolution **9:** 381-393.

Hughes, M. K., and A. L. Hughes, 1995 Natural selection on *Plasmodium* surface proteins. Molecular and Biochemical Parasitology **71:** 99-113.

Jeffery, G. M., 1966 Epidemiological significance of repeated infections with homologous strains and species of *Plasmodium*. Bulletin of the World Health Organisation **35:** 873-882.

Kemp, D. J. *et al.*, 1990 Genetic diversity in *Plasmodium falciparum*. Advances in Parasitology **29:** 75-149.

Knols, B. G. L. *et al.*, 1995 Differential attractiveness of isolated humans to mosquitoes in Tanzania. Transactions of the Royal Society of Tropical Medicine and Hygiene **89:** 604-606.

Kondrashov, A. S., 1993 Classification of hypotheses on the advantage of amphimixis. Journal of Heredity **84:** 372-387.

Kwiatkowski, D., 1991 Cytokines and anti-disease immunity to malaria. Research in Immunology **142:** 611-738.

Kwiatkowski, D., 1995 Malaria toxins and the regulation of parasite density. Parasitology Today **11:** 206-212.

Lines, J., and J. R. M. Armstrong, 1992 For a few parasites more: Inoculum size, vector control and strain-specific immunity to malaria. Parasitology Today **8:** 381-383.

MacDonald, G., 1957 *The epidemiology and control of malaria*. Oxford University Press, Oxford.

Marsh, K., 1992 Malaria - a neglected disease. Parasitology **104:** S53-S69.

Maynard Smith, J., 1982 *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.

McGregor, I. A., 1986 The development and maintenance of immunity to malaria in highly endemic areas. Clinics in Tropical Medicine and Communicable Diseases **1:** 29-53.

Medley, G. F. *et al.*, 1993 Heterogeneity in patterns of malarial oocyst infections in the mosquito vector. Parasitology **106:** 441-449.

Mendis, K. N., and C. Carter, 1992 Transmission blocking immunity may provide clues that antimalarial immunity is largely T-independant. Research in Immunology **142:** 687-690.

Mendis, K. N. *et al.*, 1991 Antigenic polymorphism in malaria: is it an important mechanism for immune evasion? . Immunol Today **12.**

Meyer, T. F. *et al.*, 1990 Variation and control of protein expression in Neisseria. Annual Review of Microbiology **44:** 451-477.

Molineaux, L., and G. Gramiccia, 1980 *The Garki Project: Research on the epidemiology and control of malaria in the Sudan savanna of West Africa*. World Health Organisation, Geneva.

Molineaux, L. *et al.*, 1988 The epidemiology of malaria and its measurement, pp. 999-1089 in *Malaria: principles and practice of malariology*, edited by W. H. Wernsdorfer, and I. McGregor. Churchill Livingstone, Edinburgh.

Ntoumi, F. *et al.*, 1995 Age-dependant carriage of multiple *Plasmodium falciparum* merozoite surface antigen-2 alleles in asymptomatic malaria infections. American Journal of Tropical Medicine and Hygiene **52:** 81-88.

Paul, R. E. L. *et al.*, 1995 Mating patterns in malaria parasite populations of Papua New Guinea. Science **269:** 1709-1711.

Pichon, G. *et al.*, 1996 A quantitative-analysis of the distribution of *Plasmodium falciparum* oocysts in *Anopheles gambiae*. Parasite-Journal De La Societe Francaise De Parasitologie **3:** 161-167.

Ponnudurai, T. *et al.*, 1991 Feeding behaviour and sporozoite ejection by infected *Anopholes stephensi*. Transactions of the Royal Society of Tropical Medicine and Hygiene **85:** 175-180.

Powell, R. D. *et al.*, 1972 Clinical aspects of acquisition of immunity to *falciparum* malaria. Proceedings of the Helminthological Society of Washington **39:** 51-66.

Pull, J. H., and B. Grab, 1974 A simple epidemiological model for evaluating the malaria inoculation rate and the risk of infection in infants. . Bulletin of the World Health Organisation **51:** 507-516.

Ranford-Cartwright, L. C. *et al.*, 1993 Frequency of cross-fertilization in the human malaria parasite *Plasmodium falciparum*. Parasitology **107:** 11-18.

Read, A. F. *et al.*, 1995 Sex allocation and population structure in malaria and related parasitic protozoa. Proceedings of the Royal Society, London, series B **260:** 359-363.

Read, A. F. *et al.*, 1992 Gametocyte sex-ratios as indirect measures of outcrossing rates in malaria. Parasitology **104:** 387-395.

Riley, E. M. *et al.*, 1994 Malaria, pp. in *Parasitic Infections and the Immune System*. Academic Press Inc.

Roberts, D. J. *et al.*, 1994 Age and "Strain-transcending" Immunity to *Plasmodium falciparum* - Reply. Parastology Today **10:** 303.

Roper, C. *et al.*, 1996 Detection of very-low level *Plasmodium falciparum* infections using the nested polymerase chain-reaction and a reassessment of the epidemiology of unstable malaria in Sudan. American Journal of Tropical Medicine and Hygiene **54:** 325-331.

Rosenberg, R. *et al.*, 1990 An estimation of the number of malaria sporozoites ejected by a feeding mosquito. Transactions of the Royal Society of Tropical Medicine and Hygiene **84:** 209-212.

Rubio-Palis, Y., 1994 Variation of the vectorial capacity of some anophelines in western Venezuela. . Am J Trop Med Hyg **50:** 420-4.

Rutledge, L. C. *et al.*, 1973 *Plasmodium* spp.: dispersion of malarial oocyst populations in anopheline and culicine mosquitoes. . Exp Parasitol **34:** 132-141.

Saul, A. J., 1994 Testing the vaccine, pp. 245-259 in *Molecular Immunological Considerations in Malaria Vaccine Development*, edited by M. F. Good, and A. J. Saul. CRC Press, Boca Raton.

Saul, A. J., 1996 Transmission dynamics of *Plasmodium falciparum*. Parasitology Today **12:** 74-79.

Schofield, L., 1991 On the function of repetitive domains in proteins of *Plasmodium* and other eukaryotic parasites. Parasitology Today **7:** 99-105.

Sinden, R. E., and R. H. Hartley, 1985 Identification of the meiotic division of malarial parasites. Journal of Protozoology **32:** 742-744.

Smith, J. D. *et al.*, 1995 Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherance phenotypes of infected erythrocytes. Cell **82:** 101-110.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry*. Freeman, New York.

Su, X.-Z. *et al.*, 1995 The large diverse gene family var encodes proteins involved in cytoadherance and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. Cell **82:** 89-100.

Taylor, L. H., and A. F. Read, 1996 Why so few transmission stages? Reproductive restraint by malaria parasites. Parasitology Today (in press) .

Taylor, R. R. *et al.*, 1996 Selective recognition of malaria antigens by human serum antibodies is not genetically determined but demonstrates some features of clonal imprinting. International Immunology **8:** 905-915.

Tchuinkam, T. *et al.*, 1993 Experimental infections of *Anopheles gambiae* with *Plasmodium falciparum* of naturally infected gametocyte carriers in Cameroon: factors influencing the infectivity to mosquitoes. Trop Med Parasitol **44:** 271-276.

Tibayrenc, M., 1995 Population genetics of parasitic protozoa and other microorganisms. Advances in Parasitology **36:** 47-115.

Tibayrenc, M. *et al.*, 1990 A clonal theory of parasitic protozoa: The population structures of *Entamoeba, Giardia, Naegleria, Plasmodium, Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. Proc. Natl. Acad. Sci. **87:** 2414-2418.

Walliker, D., 1991 Malaria parasites: Randomly interbreeding or 'clonal' populations. Parasitology Today **7:** 232-235.

Wernsdorfer, G., and W. H. Wernsdorfer, 1988 Social and economic aspects of malaria and its control, pp. 1421-1471 in *Malaria: principles and practice of malariology*, edited by W. H. Wernsdorfer, and I. McGregor. Churchill Livingstone, Edinburgh.

White, N. J., and D. A. Warrell, 1988 The management of severe malaria, pp. 865-888 in *Malaria: principles and practce of malariology*, edited by W. H. Wernsdorfer, and I. McGregor. Churchill Livingstone, Edinburgh.

World Health Organisation, 1995a Malaria Control, pp. . World Health Organisation, http://www.who.ch/.

World Health Organisation, 1995b *Tropical disease research: progress 1975-1994: highlights 1993-94: twelth programme report of the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR)*. World Health Organisation, Geneva.

Appendix 1: Choice of time-step size.

As explained in the text (Section 2.1), simulation runs were repeated with different time-step sizes. This was done to check that the step size used in subsequent simulations is appropriate. Step sizes of 0.05 to 0.8 days were examined. The Standard Parameter Values involve a low value for the Double Bite Rate (T=0.1 per day), and it was felt that the effects of an inappropriate time-step size might be greater when T had a larger value. Therefore, the effects of the step size on the output statistics was examined not only with the Standard Parameter Values, but also with T=2.0 per day (with the other parameters at their standard values). The results of these simulations are presented in the Tables A1.1 and A.1.2.

The time-step size of 0.2 days seems to be appropriate as there is very little change in value of any of the output statistics, compared to the values of the statistics when the time-step is reduced to 0.05 days.

The Effects of Time-step Size on the Output Statistics of the Model with T=0.1 per day.

Step size (days)	$\bar{\bar{G}}_{ST(d)}$	$\bar{\bar{G}}_{ST(n)}$	$\bar{\bar{t}}$	$\bar{\bar{P}}_{inf}$	$\bar{\bar{Z}}_{inf}$	$\bar{\bar{P}}_{V}$	$\bar{\bar{W}}$
0.05	**0.940057**	**0.822159**	**1.1434025**	**0.664723**	**1.535148**	**0.48864**	**5.154949**
	1.03×10^{-3}	1.61×10^{-3}	9.92×10^{-4}	2.76×10^{-3}	0.0139	2.27×10^{-3}	7.96×10^{-3}
0.1	**0.940289**	**0.8246105**	**1.140291**	**0.663497**	**1.548212**	**0.487505**	**5.167035**
	2.11×10^{-4}	2.21×10^{-3}	3.32×10^{-3}	4.84×10^{-3}	2.58×10^{-3}	1.87×10^{-3}	0.0145
0.2	**0.9398625**	**0.8232315**	**1.141689**	**0.6673505**	**1.542321**	**0.484411**	**5.1716535**
	1.53×10^{-4}	2.89×10^{-3}	4.19×10^{-3}	3.26×10^{-3}	3.49×10^{-4}	7.7 x10 -5	6.79×10^{-3}
0.5	**0.938933**	**0.8208165**	**1.143906**	**0.675939**	**1.544008**	**0.497081**	**5.205043**
	4.025×10^{-4}	1.45×10^{-3}	2.51×10^{-3}	4.94×10^{-3}	9.15×10^{-4}	4.27×10^{-4}	0.01097
0.8	**0.938849**	**0.817455**	**1.148503**	**0.684751**	**1.537852**	**0.50474**	**5.219799**
	8.53×10^{-4}	7.96×10^{-4}	7.4×10^{-5}	3.12×10^{-3}	1.36×10^{-3}	4.74×10^{-4}	3.08×10^{-3}

The Effects of Time-step Size on the Output Statistics of the Model with T=2.0 per day.

Step size (days)	$\bar{\bar{G}}_{ST(d)}$	$\bar{\bar{G}}_{ST(n)}$	$\bar{\bar{t}}$	$\bar{\bar{P}}_{inf}$	$\bar{\bar{Z}}_{inf}$	$\bar{\bar{P}}_{V}$	$\bar{\bar{W}}$
0.05	**0.898872**	**0.339997**	**2.64381**	**0.863082**	**1.545411**	**0.5030455**	0.8041268
	3.21×10^{-4}	1.464×10^{-3}	0.0104	2.28×10^{-4}	6.47×10^{-4}	4.55×10^{-4}	4.03×10^{-3}
0.1	**0.897076**	**0.339904**	**2.639301**	**0.8694975**	**1.545332**	**0.507025**	**8.032764**
	1.87×10^{-4}	2.02×10^{-3}	0.0162	1.45×10^{-4}	4.84×10^{-4}	1.96×10^{-4}	6.0×10^{-3}
0.2	**0.8987325**	**0.3401095**	**2.6426**	**0.880011**	**1.5457805**	**0.50793**	**8.036869**
	3.78×10^{-4}	2.36×10^{-3}	0.017	5.39×10^{-4}	4.75×10^{-5}	7.73×10^{-4}	4.49×10^{-3}
0.5	**0.900913**	**0.345707**	**2.606082**	**0.9034355**	**1.544930**	**0.5035715**	**8.073682**
	1.71×10^{-4}	1.864×10^{-3}	0.0145	6.79×10^{-4}	5.38×10^{-4}	4.48×10^{-4}	1.625×10^{-3}
0.8	**0.904817**	**0.351966**	**2.570768**	**0.92349**	**1.544153**	**0.494937**	**8.12116**
	3.35×10^{-5}	9.27×10^{-4}	6.86×10^{-3}	4.3×10^{-5}	3.67×10^{-4}	2.85×10^{-4}	3.75×10^{-3}

Tables A1.1 and A1.2. In most cases, the means and standard errors of each statistic are calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is with T=0.1 per day and the Time-step Size=0.1 days, in this case, eight runs were done, four with Initial Condition A and four with Initial Condition B. The mean values of each statistic are presented in bold, the standard errors are in parentheses underneath their respective means. Apart from the parameter T, the Standard Parameter Values were used.

Appendix 2: Meiosis

The genetic consequences of individual meiotic events have been studied extensively in fungi. This is possible because of the biology of many fungal species, where the products of individual meiotic events are found together in what are known as 'tetrads'. Genetic analysis of these tetrads has greatly increased the understanding of meiosis.

Let us consider an organism with two unlinked loci, with two alleles at each locus. The diploid stage is heterozygous for both loci. There are four combinations of alleles possible for the haploid progeny following meiosis. As meiosis produces four haploid progeny, it is an easy misconception, that, if the loci are unlinked, all four combinations will be produced from each meiosis. However examination of fungal tetrads has shown that this is not necessarily the case. In general either two or four different combinations of alleles are produced in each meiosis. The reasons for this and the relative frequencies of the different outcomes can be explained by considering the mechanisms by which meiosis occurs.

Figure A2 shows a schematic diagram representing meiosis with 2 unlinked loci and 2 alleles at each locus. It shows two of the several different ways crossover events and segregation can occur in meiosis, and demonstrates two of the possible outcomes. In one case (Alternative A), this leads to four genetically distinct progeny being produced, in the other case shown (Alternative B), only two progeny types are produced.

The relative frequencies at which all four progeny types , or the two alternative pairs of progeny types, are produced from a single meiosis, depends upon the degree of linkage between the two loci and also the

degree of linkage between each locus and its respective chromosomal centromere.

Assuming no interference between recombination events, and the two loci are unlinked with respect to each other, then there is a continuum between two extreme alternatives. These two alternatives are firstly, that the two loci are completely linked to separate centromeres, in which case meiosis can not lead to the production of all 4 progeny types and the two alternative pairs of progeny types are produced in equal frequencies. Secondly, either one, or both loci are so distant from their chromosomal centromeres that there is effectively no linkage between them. In this case, all four progeny types are produced (the tetratype) in 2/3 of cases, and the two alternative pairs of progeny types produced (the parental and non-parental ditypes) each occur in 1/6 of cases.

This topic is discussed further by Fincham (1983).

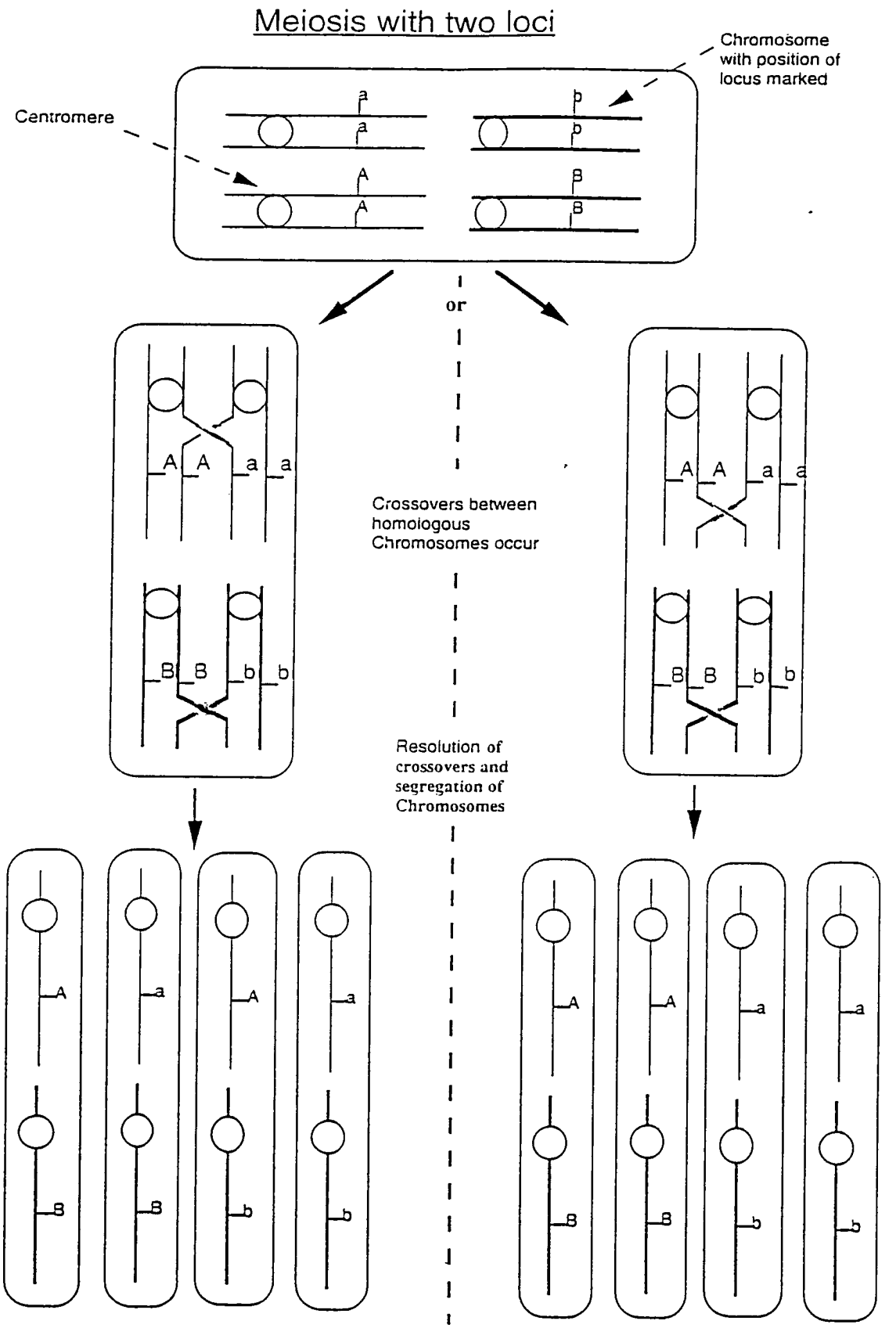

Meiosis with two loci
Chromosome with position of locus marked
Centromere
or
Crossovers between homologous Chromosomes occur
Resolution of crossovers and segregation of Chromosomes

Fig. A2

Appendix 3: Calculating the power of Welch's Approximate t test.

When the two sample sizes under consideration are equal, then the test statistic t'_s is calculated using the formula:

$$t'_s = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n}(s_1^2 + s_2^2)}}$$

In this case, the difference between Welch's approximate t test, and Student's t test is the numbers of degrees of freedom involved. In the Student's t test, where the sample variances are equal, the degrees of freedom = 2(n-1). In this case, where the variances are unequal the degrees of freedom = n-1 (Sokal and Rohlf 1995).

To calculate the power of such a test, it is necessary to examine the minimum difference between two sample means that would be found significant. In the case presented in the text, the degrees of freedom are 9, and the level of significance of interest is p=0.05. By examining statistical tables, it is found that the value of $t_{0.05[9]}$= 2.262.

Knowing the sample size (n), the estimated variances of the two samples (s_1^2 and s_2^2), and the critical value of t'_s, it is then possible to calculate the minimum difference between $\bar{Y}_1 - \bar{Y}_2$ that would be found to be significant. (The null hypothesis assumes that $\mu_1 - \mu_2$=0).

Appendix 4: The tests of the computer program

To examine the model described in Chapter Two, it was translated into a computer program (described in Appendix 5). It is possible that the computer program has mistakes in it often referred to as 'bugs'. Most bugs are easily recognised, as they prevent the program working, or it produces clearly anomalous results. However, some can be more subtle, and less evident. There are several ways in which these more subtle bugs can be found. The model can be deliberately altered in ways that strong predictions can be made as to the outcomes of the simulations, and those predictions then tested. This is done in sections A4.1 and A4.2a of this appendix.

It is also possible that with some parameters of the model, predictions can be made as to their 'limit conditions'; i.e. when they are given values at their maximum or minimum extremes. Such conditions are often biologically unfeasable and are thus not examined in Chapter 4 (where the effects of different parameter values are examined). However in Chapter 4, in several cases, trends can be seen as a parameter tends towards its limit condition. Thus, as both the parameters, c (the coefficient of vector infectivity) and T (the double bite rate) are reduced, G_{ST} for both loci tend towards one and $\bar{\bar{t}}$ tends towards one (see Sections 4.2.2.1and 4.2.2.2). This is because in both cases as the parameters tend towards zero, the rate of mixed infections is reduced, and thus the value of G_{ST} at both loci tend towards one. Unfortunately, in both of these cases, although the trends towards the limit can be seen, the limit conditions themselves cannot be examined as the population of parasites will become extinct before the limit is reached. The trend towards parasite extinction with

decreasing values of T and c (as evidenced by Figures 4.2.2.1c and 4.2.2.2c) can also be seen as tests of the model. Reducing the values of these parameters is effectvely reducing the value of R_0 for the parasites. Standard epidemiological theory shows that as R_0 is reduced towards one, the prevelance of parasites in the population will tend to zero (Anderson and May 1992).

Two cases where the limit conditions can be examined and have strong predicted outcomes are examined in section A4.2.2b of this appendix.

A4.1): Drift at the neutral locus:

There is much population genetic theory as to the dynamics of alleles at a neutral locus. It is a general prediction that the probability of any particular neutral allele reaching fixation in a population is directly proportional to its initial allele frequency (Hartl and Clark 1989).

To test whether this occurs in the computer program used, it was necessary to increase the rate at which fixation of alleles occurred, this was done by making the host population size small (N=20). Unfortunately under the Standard Parameter values, with such a small population size, the parasite population quickly becomes extinct. This was countered by altering $1/\beta$ to equal 5 days. To speed up the simulations, D was set to 2 alleles, for all other parameters the standard parameter values were used. The population was initiated with 30% with Allele 1 at the neutral locus and 70% with Allele 2 at the neutral locus. Simulations were run until one allele or the other at the neutral locus was fixed. This was repeated 5000 times.

The probability distribution for fixation of alleles at the neutral locus is expected to follow a binomial function with a mean frequency for Allele 1 of 0.3. The standard deviation for a binomial function is calculated by

$$\sigma = \sqrt{p(1-p)/k} \qquad \textbf{A4.1.1}$$

(Sokal and Rohlf 1995) where p is the frequency at which one of two alternative events (e.g. fixation of Neutral Allele 1) occurs, and k is the number of trials executed (in this case 5000). Thus, in this case the frequency of fixation of neutral Allele 1 has a predicted mean of 0.3 with a standard deviation of 6.48×10^{-3}. When the trials were examined, 1484 were fixed for Allele 1 and 3516 for Allele 2. This frequency (0.2968 for Allele 1) is well within 2 standard deviations of the predicted frequency.

A4.2): Making the immuno-allelic locus neutral.

The general result of the simulations examined in this thesis is that $t>1$, i.e. $G_{ST(d)} > G_{ST(n)}$. The reason for this should be due to the differential response of the immune system to alleles at the immuno-allelic locus. Thus one would expect $G_{ST(n)}$ not to be significantly different to $G_{ST(d)}$ if: (a) all the alleles at the immune locus completely cross-react in their recognition by the immune immune system , and (b), if the immune system has no effect on parasite densities. Both of these predictions are examined.

A4.2a) Immuno-allelic alleles treated the same

In this case D (the number of immuno-allelic alleles) was given a value of 2, however the immune system was then altered so that both alleles were treated the same. I.e. immunity to one allele conferred identical immunity to the other allele. If both alleles are treated identically then no selection can act to increase G_{ST} at the immune locus compared to the neutral locus, and thus one would expect $\bar{\bar{t}}$ not to differ significantly from one. The results of eight such simulations can be seen in the Table A4.2.1. The mean value for $\bar{\bar{t}}$ was found to be $0.998641 \pm 4.36 \times 10^{-3}$. Thus $\bar{\bar{t}}$ was not found to be significantly different from one.

$\bar{G}_{ST(d)}$	$\bar{G}_{ST(n)}$	$\bar{t}$
0.703169	0.706081	0.995877
0.703064	0.706489	0.995152
0.707689	0.701967	1.008152
0.702850	0.700040	1.004015
0.707311	0.706894	1.000590
0.704835	0.714019	0.987138
0.702505	0.702492	1.000017
0.701652	0.702927	0.998187

Table A4.2.1
The mean statistics presented are calculated from sample values taken every 2 days from day 4000 to day 5000, as described in Section 2.2. The upper four values in each column are calculated from simulations started with Initial Condition A and the lower four values in each column are from simulations started with Initial Conditions B. Apart from for the parameters D (D=2), β ($1/\beta$ =60 days) and T (T=1.0 per day), the Standard Parameter Values were used.. The model was modified so that each allele at the immuno-allelic locus was viewed identically by the immune system (see text).

A4.2b) Immune response has no effect

This was examined in 2 different ways. Firstly by examining the output statistics with s=0 (where s is the rate of increase in immune killing) and secondly examining the statistics as ρ=0 (where ρ is the probability of an immune response being mounted). In both cases there is a potential problem of the length of time the simulations take to reach a stationary distribution. To decrease this, the turnover in parasite densities is increased,. This is done by increasing the host death rate (γ) to 1/365 per day.

1) The effect of s=0 on the simulations

Eight runs were repeated with the s=0 and γ = 1/365 per day. The other parameters with their standard parameter values. The results of simulations are presented in the table A4.2.2. The mean value for $\bar{\bar{t}}$ was found to be 0.99975 ± 0.00194. Thus, the 95% confidence limits include the value one. It is therefore possible to conclude that no significant difference between $\bar{\bar{G}}_{ST(d)}$ and $\bar{\bar{G}}_{ST(n)}$ was found.

2) The effects of ρ=0 on the simulations

Eight runs were repeated with the ρ=0 and γ = 1/365 per day and the other parameters with their standard parameter values. The results of simulations are presented in the table A4.2.3. The mean value for $\bar{\bar{t}}$ was found to be 0.99745 ± 0.010. Thus, the 95% confidence limits include the

value one. It is therefore possible to conclude that no significant difference between $\overline{\overline{G}}_{ST(d)}$ and $\overline{\overline{G}}_{ST(n)}$ was found.

It may appear surprising that when $\rho=0$, t is not significantly different from one. In Section 4.2.2.3, $\overline{\overline{t}}$ was shown to get larger as ρ was decreased in value, with the value of $\overline{\overline{t}}$ never being found below one. However, it should be noted that in Section 4.2.3.3 ρ was always much larger in value than γ, the host death rate. Host death is the other method for the host population to lose infections and is not immuno-allelic allele specific. As the value of ρ is reduced to values near, or below, the host death rate (in this section $\gamma=0.027$ per day), the relative importance of allele-specific versus non-specific loss of infection becomes smaller , and the value of $\overline{\overline{t}}$ starts to become smaller. When ρ is much larger than γ (see Section 4.2.3.3), host death has very little effect on the population genetics, and other aspects of the model such as the Inoculation Rate have the dominant effects on the genetics of the system. This is evident in Figure A4.2.

$\overline{\overline{G}}_{ST(d)}$	$\overline{\overline{G}}_{ST(n)}$	$\overline{\overline{t}}$
0.881046	0.879321	1.001962
0.881566	0.879975	1.001808
0.880029	0.881363	0.998486
0.880048	0.886090	0.993182
0.888279	0.891934	0.995903
0.882368	0.885246	0.996748
0.888727	0.889885	0.998698
0.878543	0.868808	1.011205

Table A4.2.2:
The mean statistics presented are calculated from sample values taken every 2 days from day 4000 to day 5000, as described in Section 2.2. The upper four values in each column are calculated from simulations started with Initial Condition A and the lower four values in each column are from simulations started with Initial Conditions B. Apart from for the parameters s (s=0), and γ (γ=0.0027 per day) the Standard Parameter Values were used.

$\overline{G}_{ST(d)}$	$\overline{G}_{ST(n)}$	$\overline{t}$
0.880044	0.881362	0.998505
0.883863	0.881238	1.002978
0.882925	0.889043	0.993118
0.889008	0.887185	1.002055
0.890718	0.894105	0.996212
0.881778	0.879023	1.003134
0.889273	0.894165	0.994529
0.885489	0.895248	0.989099

Table 4.2.3
The mean statistics presented are calculated from sample values taken every 2 days from day 4000 to day 5000, as described in Section 2.2. The upper four values in each column are calculated from simulations started with Initial Condition A and the lower four values in each column are from simulations started with Initial Conditions B. Apart from for the parameters ρ (ρ=0), and γ (γ=0.0027 per day) the Standard Parameter Values were used.

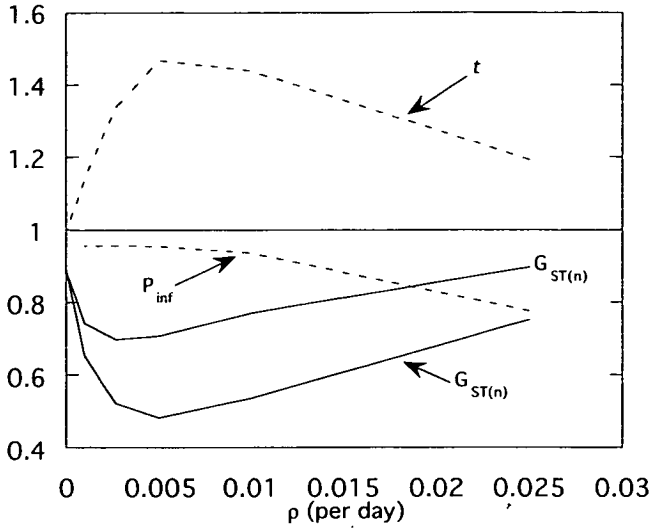


Figure A4.2:
Mean values are plotted. In most cases, the means calculated from two simulation runs, one with Initial Condition A, the other with Initial Condition B. The only exception is for $\rho=0$, in this case, eight runs were done, four with Initial Condition A and 4 with Initial Condition B. Apart from for the parameters ρ, and γ ($\gamma=0.0027$ per day) the Standard Parameter Values were used.

A.4.3) Discussion

It is impossible to rule out any bugs in the computer except by examining the same model by analytic methods (in this case an impractical solution). The results of the tests presented here are consistent with the computer program accurately representing the model. When these tests are considered in addition to the results presented in chapters Three and Four, and the logical explanations for the results in these chapters, it would appear reasonable to believe that the analysis of this thesis is not the embarrassing study of a 'bug'.

Appendix 5: The computer program

Below is a printout of a sample program used to examine the model presented in Section 2.1. The program is written in "C" and was run on UNIX workstations. The file "standardhybrid.h" (not shown) contains routines for the memory management of large arrays, methods for generation of pseudo-random numbers and methods for simple data analyses. These routines are based on those found in (Press *et al.* 1992).

```
#define seed_value -2173
#include "standardhybrid.h"
#define file_name "Output 3"
#define graph1 "mathematica1"
#define popfile "population"
#define start_sampling 4000
#define max_pop_size 6000
#define max_num_allele 10.
#define imm_classes 11 /* 1 more than num alleles*/
#define death_rate 0.00005
#define r 0.2
#define maxbites (max_pop_size * 3)
#define immune_min 0.05
#define kap 10.

/* data file declarations*/
FILE *fp, *fp2, *fp3, *fp4, *fp5, *fp6, *fp7;

/*structure declarations*/
/**********************************/
typedef struct people HOST;
struct people
{       long idx;
        int age;
        double total_para;
        double old_total_para;
        int total_immune;
        int age_to_first;
        HOST *infected_idx;
        HOST *infectous_idx;    };
/*HOST host[pop_size];*/
HOST *first_infected;
HOST *host;
HOST *link1;
HOST *previous;
HOST *first_infectous;
HOST *previous_infectous;
HOST *link2;
/**********************************/
/* Global declarations*/
long background_run, host_dex, death_count,allele_dex,numbers_infectous, time_to_quit,
        time_step, last_tickcount,total_infectous_hosts, numbers_infected, pd, npd, tt, fcount,
        pop_size, zygotes, inf_mosies, immune_cnt, zero_inf_cnt, tot_mos_bite;
```

```
double  total_hetero,tot_neut_hetero, *allele_freq, estim_hetero, neut_estim_hetero, *neutfreq;
double ***frequencies, *hetero, *neut_hetero, meanf, avf, bite_rate, n_meanf, *f, *nf;
double ***allele_pop,***old_allele_pop, avnf, varf, varnf, step_size, immune_memory,
mean_mean_f, mean_mean_nf, square_mean_f, square_mean_nf, variancef, variancenf,
imm_mem_var, coef_zygote;
double pop_weight, mean_infected, mean_zygotes, mean_inf_mos, mean_immunity, imm_recog,
                        hap_inoc, selection;

int **immune, **mosquitoes, mos_idx, num_immune, trial, sample_start, check, num_oocysts,
                **para_arrive, **immune_count, imm_allele, neut_allele, bitten;
float b0, num_allele;
float d0;
float kb;
int alph;
long  y3, y4, j, d;
long jgy = 0;
/* end of declarations*/

/*function prototypes*/
void new_calc_and_print_stats(void);
void duplicate (void);
void debug(void);
void sampling (void);
void average_etc (void);
void summary_stats (void);
void print_to_file(void);
void files (void);
void infectous_link (void);
void time_step_initialise(void);
void set_to_zero (void);
double growthcurve(double, double, double);
void gamete_sampling (void);
void transmission (void);
void mosquitoe_gets_parasites(int);
void immune_system (void);
void create_initial_pop(void);
void initialise(void);
void the_deed_is_done(void);
void wait_a_while(void);
void host_births_and_deaths(void);
void host_stats(void);
void parasite_pop_dynamics(void);
void calc_and_print_stats(void);
void birth (void);
void parasite_intro(void);
void setup_links (void);
void infected_procs (void);
void infectous_procs (void);
void pop_procs(void);
void indiv_hetero (void);
void fcalc (void);
void immune_mem(void);
void new_transmission (void);
void new_mosquito (void);
void new_gamete_grab (void);
void new_recombination (int);
void new_fcalc (void);
void new_infected_procs (void);
/*end of function prototypes*/
```

```
void new_mosquito (void)
{ int x, i;
/*checks poisson distrobution of zygotes.*/
x = poi_val(coef_zygote * link1->old_total_para);
zygotes += x;
if (x >=1)
{ inf_mosies ++;
bitten = (rand_nu * pop_size) + 1;
while (bitten > pop_size) bitten = (rand_nu * pop_size) + 1;
for (i =1; i <= x; i++)
        { new_recombination(x); }

/*For each zygote samples, recombines then sends to same host with
same total frequency*/
} }


void new_recombination (int zygs)
{ int alle1, alle2, neut1, neut2;
        double rnd;
        alle1 = 0;
        alle2 = 0;
        neut1 = 0;
        neut2 = 0;
                new_gamete_grab();
                        alle1 = imm_allele;
                        neut1 = neut_allele;
                new_gamete_grab();
                        alle2 = imm_allele;
                        neut2 = neut_allele;
                if ( (alle1 AND alle2) AND ((alle1 <= num_allele) AND (alle2 <= num_allele)))
                                {
                                if (bitten <= pop_size)
                                { rnd = rand_nu;
                                if (rnd <= 0.166666667)
                                        { pd++;
                                allele_pop[bitten][alle1][neut1] += (2 * hap_inoc)/(double)zygs;
                                allele_pop[bitten][alle2][neut2] += (2 * hap_inoc)/(double)zygs;
                                        }
                                else    {
                                        if (rnd <= 0.33333333)
                                        { npd++;
                                allele_pop[bitten][alle1][neut2] += (2 * hap_inoc)/(double)zygs;
                                allele_pop[bitten][alle2][neut1] += (2 * hap_inoc)/(double)zygs;
                                        }
                                        else
                                        { tt++;
                                allele_pop[bitten][alle1][neut1] += hap_inoc/(double)zygs;
                                allele_pop[bitten][alle2][neut2] += hap_inoc/(double)zygs;
                                allele_pop[bitten][alle1][neut2] += hap_inoc/(double)zygs;
                                allele_pop[bitten][alle2][neut1] += hap_inoc/(double)zygs;
                                        }       }
                                host[bitten].total_para += 4 * hap_inoc/(double)zygs;           }
                        else
                                { fprintf(fp, "-bitten out of range"); } }
                else {
                        if (alle1 > num_allele) fprintf(fp,"alle1 out of range");
                        if (alle2 > num_allele) fprintf(fp,"alle2 out of Rage");
                        } }
```

```
void new_gamete_grab (void)
{       double gamete_grab, gamete_prob, total_inf;
        int loc_allele_dex, loc_allele_dex2, neut;
        int test;
        loc_allele_dex =0;
        loc_allele_dex2 = 0;
        total_inf = 0.;
        gamete_grab = 2;
        test = 0;
        gamete_prob=0;

        /* adds all clones in a host together.*/
        for (loc_allele_dex = 1; loc_allele_dex <= num_allele; loc_allele_dex++)
                { total_inf += (old_allele_pop[link1->idx][loc_allele_dex][1]
                                        +old_allele_pop[link1->idx][loc_allele_dex][2]);
                }
        imm_allele =1;
        neut_allele =1;
        while (test == 0)
        { while (gamete_grab>=1.0 OR gamete_grab<1.0e-10)
                { gamete_grab = rand_nu;
                if (gamete_grab >1.0 OR gamete_grab<1.0e-10)
                        fprintf(fp, "random number out of range");
                }
        /* Checks to see if allele sampled is this one*/
        for (neut =1; neut <= 2 AND (gamete_grab > gamete_prob); neut++)
        { for (loc_allele_dex2 = 1; (loc_allele_dex2 <= num_allele) AND
                (gamete_grab > gamete_prob); loc_allele_dex2++) /* goes through alleles
                                until one is sampled.*/
                {
                gamete_prob += old_allele_pop[link1->idx][loc_allele_dex2][neut] /total_inf;
                /* Check to see if latest allele sampled if so return its value!*/
                if (gamete_grab <= gamete_prob)
                        { imm_allele = loc_allele_dex2;
                        neut_allele = neut;
                        test=1;
                        }       } }
if (test == 0)
{ fprintf(fp, "oops\n");
fprintf(fp, "rand = %f\n", gamete_grab);
fprintf(fp, "total allele Freqs in host num %d, is %f\n", link1->idx, gamete_prob);
fprintf(fp, "total parasites in it are = %f\n", total_inf);
fprintf(fp, "numbers infectous = %d\n", numbers_infectous);
fprintf(fp, "loc_allele_dex 2 = %d\n", loc_allele_dex2);
fprintf(fp, "total_parasites = %f\n", link1->total_para);
exit(1); }      }       }


void new_transmission(void)
{ long x, i;
                        if (sample_start ==1) new_calc_and_print_stats();
                        x = poi_val((bite_rate * step_size));
                        tot_mos_bite += x;
                        for (i = 1; i <=x; i++)
                        {       new_mosquito(); } }


void duplicate (void)
{ int loc_allele_dex;
for (loc_allele_dex = 1; loc_allele_dex<= num_allele; loc_allele_dex++)
```

```
{old_allele_pop[host_dex][loc_allele_dex][1] = allele_pop[host_dex][loc_allele_dex][1];
old_allele_pop[host_dex][loc_allele_dex][2] = allele_pop[host_dex][loc_allele_dex][2];
host[host_dex].old_total_para = host[host_dex].total_para; } }

void pop_procs(void)
{for (host_dex = 1; host_dex <= pop_size; host_dex++)
        { duplicate();
        host_births_and_deaths();
        immune_mem();
        setup_links();
        host[host_dex].age++;
        }}

void immune_mem(void)
{
int loc_allele_dex;
if (immune[host_dex][1] > 0.5) num_immune++;
for (loc_allele_dex = 1; loc_allele_dex<=num_allele; loc_allele_dex++)
        { if (immune[host_dex][loc_allele_dex] > 0.5) immune_cnt++;
        if ((immune[host_dex][loc_allele_dex] > 0.5)
                AND (allele_pop[host_dex][loc_allele_dex][1] +
allele_pop[host_dex][loc_allele_dex][2] < 1e-9))
                {       if (rand_nu < (1/immune_memory) * step_size)
                        { immune[host_dex][loc_allele_dex] = 0;
                        } }     } }


void new_infected_procs (void)
{ while (link1 NOT NULL)
        {       if ((link1->idx) > pop_size OR (link1->idx) < 1)
                { fprintf(stderr, "link1 out of range");
                 fprintf(fp, "link1 out of range");
                exit(1);
                link1 = NULL; }
        host_stats();
        new_transmission();
        link1 = link1->infected_idx;
        }

if (sample_start ==1)
{tot_neut_hetero =tot_neut_hetero/pop_weight;
total_hetero = total_hetero/pop_weight;
new_fcalc ();
}       }

/*works out mean hetrozygosity. and total allele freqs*/
void new_fcalc (void)
{int loc_allele_dex;
double estim_homo, neut_estim_homo;
estim_homo = 0.;
neut_estim_homo =0.;
neutfreq[1] = 0.;
neutfreq[2] =0.;
/*adds up all allele freqeuncies and finds mean*/
for (loc_allele_dex = 1; loc_allele_dex <= num_allele; loc_allele_dex++)
        { allele_freq[loc_allele_dex] = 0.;
        previous->infected_idx = NULL;
        link1 = first_infected;
        while (link1 NOT NULL)
                { if ((link1->idx) > pop_size OR (link1->idx) < 1)
                        { fprintf(stderr, "link1 out of range");
```

```
                    fprintf(fp, "link1 out of range");
                    exit(1);
                    link1 = NULL;
                    }
                allele_freq[loc_allele_dex] += (frequencies[link1->idx][loc_allele_dex][1]
                 + frequencies[link1->idx][loc_allele_dex][2]) * link1->old_total_para;
                neutfreq[1] += frequencies[link1->idx][loc_allele_dex][1] * link1->old_total_para;
                neutfreq[2] += frequencies[link1->idx][loc_allele_dex][2] * link1->old_total_para;
                link1 = link1->infected_idx;
                }
allele_freq[loc_allele_dex] =(allele_freq[loc_allele_dex]/pop_weight);
estim_homo += pow(allele_freq[loc_allele_dex],2);
        }
neutfreq[1] =neutfreq[1]/pop_weight;
neutfreq[2] =neutfreq[2]/pop_weight;
neut_estim_homo =       pow(neutfreq[1], 2) + pow(neutfreq[2],2);
if (estim_homo > 1.)
        {       fprintf(fp, "in gen %d, estimated homozygosity was %f\nand rounded down to 1.\n",
time_step, estim_homo);
        estim_homo = 1.; }
if (neut_estim_homo > 1.)
        { fprintf(fp, "in gen %d, Neut estim homozygosity was %f\nand rounded down to 1.\n",
time_step, neut_estim_homo);
        estim_homo = 1.;  }
estim_hetero = 1 - estim_homo;
neut_estim_hetero = 1 - neut_estim_homo;
n_meanf = (neut_estim_hetero - tot_neut_hetero)/neut_estim_hetero;
if (neut_estim_hetero == tot_neut_hetero) n_meanf = 0;
meanf = (estim_hetero -total_hetero)/estim_hetero;
if (estim_hetero == total_hetero) n_meanf = 0;
sampling();     }


void sampling (void)
{mean_mean_f += meanf;
square_mean_f += pow(meanf, 2);
mean_mean_nf += n_meanf;
square_mean_nf += pow(n_meanf, 2);
mean_infected += ((double)numbers_infected/pop_size);
if (inf_mosies > 0) mean_zygotes += (double)zygotes/inf_mosies;
else zero_inf_cnt ++;
mean_inf_mos += (double)inf_mosies/tot_mos_bite;
mean_immunity +=(double)immune_cnt/pop_size;
fcount++; }

void parasite_intro(void)
{       int loc_intro_dex;
        int loc_allele_dex;
        int intro, xintro;
        for (loc_intro_dex = 1; loc_intro_dex<=(int)(pop_size/5)+1; loc_intro_dex++)
                {if (trial%2 == 0)
        {check++;
                for (loc_allele_dex = 1; loc_allele_dex <= num_allele; loc_allele_dex++)
                        {
                        allele_pop[loc_intro_dex][loc_allele_dex][1] = 1/(2.* num_allele);
                        allele_pop[loc_intro_dex][loc_allele_dex][2] = 1/(2.* num_allele);
                        } }
                else    {
                intro = (loc_intro_dex%(2 * (int)num_allele)) +1;
                xintro = intro - num_allele;
                if (intro <= num_allele)
```

```
                { allele_pop[loc_intro_dex][intro][1] = 1; }
                else
                {allele_pop[loc_intro_dex][xintro][2] = 1;} }              } }


void create_initial_pop(void)
        {double total_pop;
        long loc_host_dex, loc_allele_dex, loc_bite_dex;
        for ( loc_bite_dex=1; loc_bite_dex<= maxbites; loc_bite_dex++)
                { mosquitoes[loc_bite_dex][1]=0;
                mosquitoes[loc_bite_dex][2]=0; }
        for (loc_host_dex = 1; loc_host_dex<= pop_size; loc_host_dex++)
                {
                host[loc_host_dex].idx=loc_host_dex;
                host[loc_host_dex].age=0.;
                host[loc_host_dex].total_para=0.0;
                host[loc_host_dex].total_immune=0.;
                host[loc_host_dex].age_to_first=0.;
                for (loc_allele_dex = 1; loc_allele_dex<= num_allele; loc_allele_dex++)
                        {
                        allele_pop[loc_host_dex][loc_allele_dex][1]=0.;
                        allele_pop[loc_host_dex][loc_allele_dex][2]=0.;
                        immune[loc_host_dex][loc_allele_dex]=0;
                        para_arrive[loc_host_dex][loc_allele_dex]= -5;
                        immune_count[loc_host_dex][loc_allele_dex]= -5;
                        }}
        parasite_intro();
        for(loc_host_dex =1; loc_host_dex<= pop_size; loc_host_dex++)
                {total_pop= 0.;
                for(allele_dex =1; allele_dex<=num_allele; allele_dex++)
                        {
                        total_pop += (allele_pop[loc_host_dex][allele_dex][1]
                                        + allele_pop[loc_host_dex][allele_dex][2]);
                        }
        host[loc_host_dex].total_para = total_pop;
                }
        num_immune = 0;
        }


void initialise(void)
        {host = malloc(sizeof(HOST) * max_pop_size);
        host--;
        check = 0;
        pd=npd=tt=0;
        /* assigments of global declarations */
        f = dvector(1,max_pop_size);
        nf = dvector(1,max_pop_size);
        allele_freq = dvector(1, max_num_allele);
        frequencies = d3tensor(1, max_pop_size, 1, max_num_allele, 1, 2);
        mosquitoes = imatrix(1, maxbites, 1, 4);
        hetero = dvector(1, max_pop_size);
        neut_hetero = dvector(1, max_pop_size);
        neutfreq=dvector(1,2);
        allele_pop=d3tensor(1, max_pop_size, 1, max_num_allele, 1, 2);
        old_allele_pop=d3tensor(1, max_pop_size, 1, max_num_allele, 1, 2);
        immune=imatrix(1, max_pop_size,1, max_num_allele);
        para_arrive=imatrix(1, max_pop_size,1, max_num_allele);
        immune_count=imatrix(1, max_pop_size,1, max_num_allele);
        background_run=1;          }
```

```
void files(void)
{if (trial == 1)
        { fp=fopen("output1", "w");          }

if (trial ==2)
        { fp=fopen("output2", "w");          }

if (trial ==3)
        { fp=fopen("output3", "w");          }

if (trial ==4)
        { fp=fopen("output4", "w");          }
if (trial >=5)
        { fp=fopen("output5", "w");          }          }


void setup_links (void)
        {/* checks through all population to see which ones are infected*/
        if (host[host_dex].total_para > 1.0e-9)
                { if (host[host_dex].total_para > 0.00001)numbers_infected ++;
                if (first_infected == NULL)
                        {
                        first_infected = &host[host_dex];
                        previous = &host[host_dex];
                        }
                else    {
                        previous->infected_idx = &host[host_dex];
                        previous = &host[host_dex];
                        }}          }


/*Logistic growth of parasites*/
double growthcurve (double both, double tot, double self)
        {
        double change_pop;
        double sel;
        sel = (selection * pow(step_size,2));
        change_pop = (b0 - (kb *(both + (alph*(tot-both)))) -
                                (sel * immune[link1->idx][allele_dex])) * self;
        return (change_pop);          }


void birth(void)/*resets host [host_dex]*/
        { int  loc_allele_dex;
        for (loc_allele_dex = 1;loc_allele_dex<=num_allele; loc_allele_dex++)
                { allele_pop[host_dex][loc_allele_dex][1]=0;
                allele_pop[host_dex][loc_allele_dex][2]=0;
                immune[host_dex][loc_allele_dex]=0;
                immune_count[host_dex][loc_allele_dex]= -5;
                para_arrive[host_dex][loc_allele_dex]= -5;
                }
        host[host_dex].age=0.;
        host[host_dex].total_para=0.;
        host[host_dex].total_immune=0.;
        host[host_dex].age_to_first=0.;
        }


void the_deed_is_done(void)
        {fprintf (fp, "number of deaths were %d", death_count);
```

```
        fclose(fp);
        if (trial <3)
        { fclose(fp2);
        fclose(fp3);
        fclose(fp4);
        fclose(fp5);
        }
        }

void host_births_and_deaths(void)
        { if (rand_nu < (death_rate * step_size))
                        { birth();
                        death_count++;
                        }}

void parasite_pop_dynamics (void)
        {
        double change1, change2, together;
        together = old_allele_pop[link1->idx][allele_dex][1] + old_allele_pop[link1-
>idx][allele_dex][2];
        immune_system();
        change1 = growthcurve(together, link1->old_total_para, old_allele_pop[link1-
>idx][allele_dex][1]);
        change2 = growthcurve(together, link1->old_total_para, old_allele_pop[link1-
>idx][allele_dex][2]);
        allele_pop[link1->idx][allele_dex][1] += change1;
        allele_pop[link1->idx][allele_dex][2] += change2;
        if (allele_pop[link1->idx][allele_dex][1] < 1.0e-9)
                allele_pop[link1->idx][allele_dex][1] = 0.;
        if (allele_pop[link1->idx][allele_dex][2] < 1.0e-9)
                allele_pop[link1->idx][allele_dex][2] = 0.;
        }


void immune_system (void)
        {
        if (immune[link1->idx][allele_dex] == 0)
                {
                if ((allele_pop[link1->idx][allele_dex][1] + allele_pop[link1->idx][allele_dex][2])
                                > immune_min)
                        {
                        if (rand_nu < (imm_recog * step_size))
                                immune[link1->idx][allele_dex] = 1;
                                jgy++;
                        }       }
        else immune[link1->idx][allele_dex] ++;
        }


void host_stats(void)
        { double total_pop;
        total_pop = 0;
        pop_weight +=  link1->old_total_para;
        for(allele_dex =1; allele_dex<=num_allele; allele_dex++)
                {
        parasite_pop_dynamics();
        total_pop += (allele_pop[link1->idx][allele_dex][1] + allele_pop[link1->idx][allele_dex][2]);
                                }
        link1->total_para = total_pop;
        if (link1->total_para < 1.0e-9)
                link1->total_para = 0.;
```

```
}


void time_step_initialise (void)
        {
        long loc_bite_dex, loc_allele_dex, loc_host_dex;
        mos_idx=0;
        total_hetero = 0.;
        tot_neut_hetero = 0.;
        total_infectous_hosts=0;
        numbers_infected=0;
        numbers_infectous=0;
        estim_hetero= 0.;
        neut_estim_hetero=0.;
        num_immune = 0;
        pop_weight = 0;
        zygotes = 0;
        inf_mosies = 0;
        immune_cnt =0;
        tot_mos_bite = 0;
        zero_inf_cnt = 0;
        for (loc_host_dex = 1; loc_host_dex<= pop_size; loc_host_dex++)
                {
                for (loc_allele_dex = 1; loc_allele_dex<= num_allele; loc_allele_dex++)
                        {
                        frequencies[loc_host_dex][loc_allele_dex][1] = 0.;
                        frequencies[loc_host_dex][loc_allele_dex][2] = 0.;
                        }       }
        for (loc_bite_dex=1; loc_bite_dex<= maxbites; loc_bite_dex++)
                {
                mosquitoes[loc_bite_dex][1]=0;
                mosquitoes[loc_bite_dex][2]=0;
                }       }

void new_calc_and_print_stats(void)
{
        long loc_allele_dex2;
        double total_inf;
        total_inf =0.;
        /*works out infectous parasite allele frequencies*/
        for (loc_allele_dex2 = 1; loc_allele_dex2 <= num_allele; loc_allele_dex2 ++)
                {
                if (link1->idx > pop_size)
                        { fprintf(fp, "oh bugger, %d %d",link1->idx, pop_size);
                        exit(1);
                        }
                frequencies[link1->idx][loc_allele_dex2][1] =
                                (old_allele_pop[link1->idx][loc_allele_dex2][1]/link1-
>old_total_para);
                frequencies[link1->idx][loc_allele_dex2][2] =
                                (old_allele_pop[link1->idx][loc_allele_dex2][2]/link1-
>old_total_para);
                if (frequencies[link1->idx][loc_allele_dex2][1] < 0)
                                frequencies[link1->idx][loc_allele_dex2][1] = 0.;
                if (frequencies[link1->idx][loc_allele_dex2][2] < 0)
                                frequencies[link1->idx][loc_allele_dex2][2] = 0.;
                }
        indiv_hetero ();
        total_hetero += hetero[link1->idx]*link1->old_total_para;
        tot_neut_hetero += neut_hetero[link1->idx]*link1->old_total_para;
        }
```

```
/*calculates each infectous hosts heterozygosity*/
void indiv_hetero (void)
        {
        double homozyg, neu1, neu2, neut_homo;
        int loc_allele_dex;
        homozyg = 0.;
        neut_homo = 0.;
        neu1 = 0.;
        neu2 = 0.;
        for (loc_allele_dex = 1; loc_allele_dex <= num_allele; loc_allele_dex++)
                {homozyg += pow(
                                        (frequencies[link1->idx][loc_allele_dex][1]+
                                                frequencies[link1->idx][loc_allele_dex][2]), 2);
                                }
        for (loc_allele_dex = 1; loc_allele_dex <= num_allele; loc_allele_dex++)
                {       neu1 += frequencies[link1->idx][loc_allele_dex][1];
                        neu2 += frequencies[link1->idx][loc_allele_dex][2];
                }
        neut_homo = pow(neu1, 2) + pow(neu2,2);
        if (homozyg >1.)
                {if (homozyg > 1.01)
                        { fprintf(fp, "homozygosity in host %d, greater than 1.01 in gen %d\nand
rounded to 1\n", link2->idx, time_step);
                        fprintf(fp, "homozygosity = %f\n", homozyg);
                        }
                if (homozyg > 1.2)
                        { for (loc_allele_dex = 1; loc_allele_dex <= num_allele; loc_allele_dex++)
                                { fprintf(fp, "freqs are: %f\n",
                                        frequencies[link1->idx][loc_allele_dex][1] +
                                                frequencies[link1->idx][loc_allele_dex][2]);  }
                        fprintf(fp, "wibble");
                        exit(1);  }
                homozyg = 1.;
                }
        if (neut_homo > 1.01)
                {
                 fprintf(fp, " neut homozygosity in host %d, greater than 1.01 in gen %d\nand
rounded to 1\n", link2->idx, time_step);
                 neut_homo = 1.;
                 }
        if (neut_homo > 1.) neut_homo = 1.;
        hetero[link1->idx] = 1 - homozyg;
        neut_hetero[link1->idx] = 1 - neut_homo;
        }


 void print_to_file(void)
 { fprintf(fp3, " %f %f\n", time_step * step_size, meanf);
  fprintf(fp5, " %f %f\n", time_step * step_size, n_meanf);
                fprintf(fp4, " %f %f\n", time_step * step_size, ((float)num_immune/pop_size));
                fprintf(fp2, " %f %f\n", time_step * step_size, ((float)numbers_infected/pop_size));
 }


 void average_etc(void)
 { long dx;
 double neu1, neu2;
 int loc_allele_dex;
 dx = 1;
```

```
previous->infected_idx = NULL;
link1 = first_infected;
while (link1 NOT NULL)
        {
        if ((link1->idx) > pop_size OR (link1->idx) < 1)
                {
                fprintf(stderr, "link1 out of range");
                fprintf(fp, "link1 out of range");
                exit(1);
                link1 = NULL;
                }
        f[dx] = (estim_hetero - hetero[link1->idx])/estim_hetero;
        if (estim_hetero <= hetero[link1->idx]) f[dx] =0.0;
        nf[dx] = (neut_estim_hetero - neut_hetero[link1->idx])/neut_estim_hetero;
        if (neut_estim_hetero <= neut_hetero[link1->idx]) nf[dx] =0.0;
        if (nf[dx] <= 0.00000000001)
                {
                for (loc_allele_dex = 1; loc_allele_dex <= num_allele; loc_allele_dex++)
                {
                        neu1 += frequencies[link1->idx][loc_allele_dex][1];
                        neu2 += frequencies[link1->idx][loc_allele_dex][2];

                }                       }

        dx++;
        link1 = link1->infected_idx:
        }
if (dx >1)
{
avf= mean(f, dx);
varf = variance(f, dx);
avnf= mean(nf, dx);
varnf = variance(nf, dx);
varf = (varf *(dx-1))/dx;
varnf = (varnf *(dx-1))/dx;
}
else {
avf = -1;
fprintf(fp, "dx is %d\n", dx);
} }


void summary_stats (void)
{  int immunity[imm_classes];
int loc_allele_dex, loc_host_dex, imm_dex;
average_etc();
mean_mean_f = mean_mean_f/fcount;
mean_mean_nf = mean_mean_nf/fcount;
mean_infected = mean_infected/fcount;
mean_zygotes = mean_zygotes/(fcount - zero_inf_cnt);
mean_inf_mos = mean_inf_mos/fcount;
mean_immunity = mean_immunity/fcount;
variancef = (square_mean_f - (pow(mean_mean_f, 2) * fcount))/(fcount - 1);
variancenf = (square_mean_nf - (pow(mean_mean_nf, 2) * fcount))/(fcount - 1);
for (loc_allele_dex = 0; loc_allele_dex <= num_allele; loc_allele_dex++)
                                        {  immunity[loc_allele_dex] = 0.;
                                        }
fprintf (fp, "jgy is %d\n", jgy);
fprintf (fp, "in gen %d\n", time_step);
fprintf (fp, "meanf was %f\n", meanf);
fprintf (fp, "avf was %f\n", avf);
```

```
fprintf (fp, "avnf was %f\n", avnf);
fprintf (fp, "varf was %f\n", varf);
fprintf (fp, "varnf was %f\n", varnf);
fprintf (fp, "neutral meanf was %f\n", n_meanf);
fprintf (fp, "tot neut Hetero was %f\n", tot_neut_hetero);
fprintf (fp, "neutral estim hetero was %f\n", neut_estim_hetero);
fprintf (fp, "bite rate was  %f\n", bite_rate);
fprintf (fp, "square_mean_f is %f\n", square_mean_f);
fprintf (fp, "mean_mean_f is  %f\n", mean_mean_f);
fprintf (fp, "fcount is  %d\n", fcount);
fprintf (fp, "check is  %d\n", check);

for (loc_host_dex = 1; loc_host_dex <= pop_size; loc_host_dex++)
        {  imm_dex = 0;
        for (loc_allele_dex = 1; loc_allele_dex <= num_allele; loc_allele_dex++)
                                {
                                if (immune[loc_host_dex][loc_allele_dex] > 0.5)
                                        imm_dex ++;  }
immunity[imm_dex] ++;
        }
fprintf (fp, "pop_size is %d\n", pop_size);
for (loc_allele_dex = 0; loc_allele_dex <= num_allele; loc_allele_dex++)
                        {
                        fprintf (fp, "pop_size is %d\n", pop_size);
                        fprintf (fp, "hosts immune to %d alleles are %f, pop_size is %d\n",
loc_allele_dex, immunity[loc_allele_dex], pop_size);
                        fprintf (fp, "pop_size is %d\n", pop_size);
                if (loc_allele_dex > 0)
                        fprintf (fp, "Freq in pop of allele %d is %f\n", loc_allele_dex,
allele_freq[loc_allele_dex]);
                        }
fprintf (fp, "neutral alleles are at freq %f, %f.\n", neutfreq[1], neutfreq[2]);
if (neutfreq[1] <=0.000001) mean_mean_nf = 2000;
if (neutfreq[2] <=0.000001) mean_mean_nf = 1000;
if (trial ==1)
        {
        fp6 = fopen("transrun1","w");
        }
        else fp6 = fopen("transrun1","a");
fprintf (fp6, "%f %d %f %f %f %f %f %f %f %f %f %f %f\n", bite_rate,
                        (int)num_allele, immune_memory, mean_mean_f, variancef,
                                mean_mean_nf, variancenf,
mean_mean_f/mean_mean_nf, mean_infected, mean_zygotes, mean_inf_mos, mean_immunity);
fclose (fp6);
fprintf (fp, "Mean no. of bites = %f in %d trials\n", (float)j/d, d);
fprintf (fp, "numbers sampling 3 is %d 4 is %d",  y3, y4);
fprintf (fp, "numbers of PD, NPD, TT are %d, %d, %d respectively", pd, npd, tt);
}

void trial_initialise (void)
{
        tot_mos_bite = 0;
        zero_inf_cnt = 0;
        mean_infected = 0;
        mean_zygotes = 0;
        mean_inf_mos = 0;
        mean_immunity = 0;
        variancef = 0.;
        variancenf = 0.;
        square_mean_f = 0.;
        square_mean_nf = 0.;
```

```
        mean_mean_f = 0.;
        mean_mean_nf = 0.;
        fcount = 0;
}

main()
{
        initialise();
        selection = 0.2;
        hap_inoc = 0.000005;
        imm_recog = 0.025;
        bite_rate = 0.1;
        num_oocysts = 1;
        coef_zygote = 1;
        immune_memory = 365;
        imm_mem_var= immune_memory * 3.0;
        step_size = 0.2;
        pop_size = 6000;
        alph = 1;
        time_to_quit = 5000;
        for (trial = 1; trial <= 20; trial++)
        {
        trial_initialise();
        pd=npd=tt=0;
        num_allele =10;
        if (trial == 1) bite_rate = 0.05;
        if (trial == 2) bite_rate = 0.05;
        if (trial == 3) bite_rate = 0.1;
        if (trial == 4) bite_rate = 0.1;
        if (trial == 5) bite_rate = 0.5;
        if (trial == 6)       bite_rate = 0.5;
        if (trial == 7) bite_rate = 1.0;
        if (trial == 8) bite_rate = 1.0;
        if (trial == 9) bite_rate = 1.5;
        if (trial == 10) bite_rate = 1.5;
        if (trial == 11) imm_recog = 0.005;
        if (trial == 12) imm_recog = 0.005;
        if (trial == 13) imm_recog = 0.01;
        if (trial == 14) imm_recog = 0.01;
        if (trial == 15) imm_recog = 0.015;
        if (trial == 16) imm_recog = 0.015;
        if (trial == 17) imm_recog = 0.035;
        if (trial == 18) imm_recog = 0.035;
        if (trial == 19) imm_recog = 0.05;
        if (trial == 20) imm_recog = 0.05;
        /* New growth calc*/
        b0 = (pow(12.3, (step_size/2))) -1;
        d0 = 0.;
        death_count = 0.;
        kb = b0 - d0;
        sample_start = 0;
        files();
        create_initial_pop();
        for (time_step=1; time_step <= (int)(time_to_quit * (1/step_size)); time_step++)
                {
                if ((time_step >= start_sampling/step_size)
                        AND (time_step%(int)(2/step_size) == 0)) sample_start = 1;
                else sample_start = 0;
                time_step_initialise();
                first_infected = NULL;
                /*previous->infected_idx = NULL;*/
```

```
                debug();
                pop_procs();
                previous->infected_idx = NULL;
                link1 = first_infected;
                if (time_step == (int)(time_to_quit * (1/step_size))) summary_stats();
                new_infected_procs();
                if (time_step == 2000)
                        fprintf(fp, " In gen 2000, numbers infectected is %d,\n",
                         numbers_infected); }
        fprintf(fp, "Done!\n");
        the_deed_is_done();
        }}

void debug(void)
        {
        static FILE *deb;
        int i;
        if (time_step == 1)
        {
         deb = fopen("growthc", "w");
        for (i = 1; i <=20; i++)
        {
        fprintf(deb, "%f, %f, %f, %f, %f, %f, %f, %f, %f, %f\n", allele_pop[i][1][1],
                                        allele_pop[i][2][1], allele_pop[i][3][1],
allele_pop[i][4][1], allele_pop[i][5][1],
                                        allele_pop[i][6][1], allele_pop[i][7][1],
allele_pop[i][8][1], allele_pop[i][9][1],
                                        allele_pop[i][10][1]);
        fprintf(deb, "%f, %f, %f, %f, %f, %f, %f, %f, %f, %f\n\n", allele_pop[i][1][2],
                                        allele_pop[i][2][2], allele_pop[i][3][2],
allele_pop[i][4][2], allele_pop[i][5][2],
                                        allele_pop[i][6][2], allele_pop[i][7][2],
allele_pop[i][8][2], allele_pop[i][9][2],
                                        allele_pop[i][10][2]);
        }       }
        if (time_step ==2) fclose(deb);
        }
```

Appendix 6: Paper produced in conjunction with Dr. Ian Hastings.

In Press, October 1997 Issue.

Sex, strains and virulence.

Ian M. Hastings[1,2] *and B. Wedgwood-Oppenheim*[3]

[1]UMR CNRS/ORSTOM u° 9926,
Laboratoire Génétique Moléculaire des Parasites et des Vecteurs,
ORSTOM,
911 Av. Agropolis BP5045,
34032 Montpellier cedex 1,
France.

[2]current address:
[3]Institute of Cell, Animal and Population Biology,
University of Edinburgh,
West Mains Road,
Edinburgh EH9 3JT.

Phone: 0131-650-5484
Fax: 0131-650-6564
Email: eang18@tattoo.ed.ac.uk

The last 6 years have seen a long running debate as to the extent to which populations of pathogens can be split into a series of genetically distinct "strains" or clones[e.g.1-11, refs.therein*]. A clonal structure has implications for epidemiology as it may allow the history of infections to be traced, and may form the basis for clinical decisions. If genetically stable strains do exist then genes, including those affecting virulence, will segregate within this strain background and the strains may be associated with clinical or epidemiological differences. Unfortunately, and despite its obvious importance, this debate remains unresolved. The purpose of this brief discussion is not to attempt a definitive resolution, but rather to demonstrate how the application of basic population genetic approaches provides a framework within which we can consider the putative emergence of a strain structure. The discussion will tend to focus on *Plasmodium falciparum* both because of its clinical importance as the most debilitating widespread eukaryotic infection, and because its population genetics are known in some detail thus enabling theory to be grounded in reality. In particular, it will be shown that earlier arguments about the role of recombination may need to be (re)interpreted in the context of immunity.

Sex plays a pivotal role in the process of strain formation and maintenance because it breaks down genetic differentiation between the putative strains. The mode of sexual recombination differs between eukaryotes such as *Plasmodium,* Trypanosomes etc., where sex is presumed to occur via meiosis, and bacteria (and viruses) where sexual recombination appears to take the form of sporadic exchange of small segments of genetic material[12,13,refs therein.] As the following calculations make clear, the consequences of these differing modes of genetic exchange affect the extent to which populations may become "clonal". First, a note about our use of the terms "outcrossing, "sex" and "recombination". The last term is potentially the most confusing as it appears to be used in at least three different ways when eukaryotes are considered. Firstly, in the molecular genetic sense of the breaking and rejoining of DNA during meioses; our subsequent use

* Also the topic of a recent meeting at CDC, Atlanta "International Workshop on molecular epidemiology and evolutionary genetics of pathogenic microorganisms", June 1996.

of the term "recombination" ignores this connotation. Secondly, it may be used according to the older genetic definition of a process leading to the production of non-parental genotypes. In most higher eukaryotes this is equivalent to the frequency of sexual reproduction but this is not the case in "lower" eukaryotes: for example, in *Plasmodium* sex involving gametogenesis, fertilisation and meiosis is mandatory but if between identical genotypes, then no novel genotypes arise. The term "outcrossing" is more useful for this definition of recombination as it describes the situation where sex occurs between non-identical haplotypes, resulting in the re-arrangement of genetic material: thus in *Plasmodium* outcrossing can be estimated as 66%[14,15] or 10%[16]. The third usage is the population genetic definition of the frequency of re-arrangement between two loci; this is the usage employed here in discussion of eukaryotes and has a maximum of 0.5. It can be calculated by estimating the expected recombination rate due to physical linkage (or lack thereof) of two loci and then multiplying by the outcrossing rate to get the true value; for example recombination between unlinked loci is 0.5 but if only 0.3 of reproduction involves outcrossing, than the true average rate of recombination is 0.5x0.3=0.15 per genome per generation. When considering prokaryotes such as bacteria, recombination is clearly defined as the rate of genetic transfer between organisms per gene segment per genome per generation[13]. In these calculations it is assumed that mating is entirely random and consequently that there is no co-transmission of distinct but related genotypes through the same source of infection (such as infected food or mosquito bites). This restriction was assumed both for mathematical tractability and to make the results as general as possible; the same assumption has been made in several other studies[14,15,17]. It may also be worthwhile pointing out for non-specialist readers that "linkage disequilibrium" simply describes the situation where alleles at two loci are found together more frequently than would be expected purely by chance; importantly it does not imply, nor preclude, that the two loci are physically linked on the same chromosome. As an example familiar in this context consider the case of 2 loci, one with alleles *a* and *b*, the other with alleles *x* and *y*: it has been shown[10] that if these encode antigens then under certain circumstances 2 of the 4

possible combinations will dominate the population resulting in large scale linkage disequilibrium.

There appears to be two main underlying mechanisms by which a strain structure can arise in a pathogen population. Critically, both these models represent processes which act in essentially the same manner: they reduce the probability of co-infection of different genotypes within the same host and hence reduce the opportunities (via recombination) for gene flow between putative "strains". This same functional basis allows us to describe these phenomena with the same basic population genetic models. We shall call these the "epidemiological" and "immune mediated" models. The former may occur because of low infection rates, geographical isolation, differing host preferences[18] or an "epidemic" population structure[19]; the latter is included as "epidemiological" even though its ultimate cause may be that a novel gene combination evades immunity. The "immune mediated" model arises from the actions of acquired immunity against either a single immunodominant locus[8,9], or against several immune loci[10].

The most useful population genetic approach to address these models is to regard each strain as a genetic "island" and to estimate genetic differentiation between them. This "Island Model" is attributable to Wright[20] and is a standard tool in population genetics[e.g.21,22]. It is explicitly used to investigate geographic islands each with endogenous populations of, for example, Drosophila, mice or humans. The extent of genetic differentiation between the islands depends on the extent of gene flow through migration of individuals between islands; increased migration obviously increases gene flow and limits, or prevents, the genetic differentiation of populations on separate islands. The same maths apply if we redefine the sub-division of the population into genetic islands or "clones" rather than geographic islands. The barriers in this case are immunogenic or epidemiological rather than geographic but the same maths applies, the only difference being that recombination rate replaces migration rate as a measure of gene flow. Firstly, we investigate the situation where a individual pathogen undergoes large clonal expansion either as a result of encountering a novel new environment, or because a novel gene combination gives it an advantage; this corresponds to the "epidemic" structure of Maynard Smith *et al*[19]. We follow its fate by using a single island model. Conventionally, this corresponds to a small isolated population which has colonised an island off a large continental mainland. This small population is assumed to be genetically distinct because it was derived from a small number, possibly a single, colonist(s) which by chance will contain an unrepresentative gene sample (the so-called "founder effect"). We are interested in calculating how long it takes this genetic differentiation to decay through genetic recombination between the "clone" and "non-clone" individuals (Box Ai). Similarly, we can follow the population genetics of more numerous clones by assuming that the pathogen population is, or can be arbitrarily (e.g. by host species), divided into a number of distinct "strains". We then adapt the conventional "Finite Island" model to investigate how much gene flow (through

recombination) is necessary to prevent genetic differentiation of the clones (Boxes Aii and B).

One of the uses of strains is in epidemiological tracking. Figure 1a shows the expected "lifespan" of a newly emergent strain, that is the time (in generations), before its genetic character has decayed back to 90% that of the parental population. Obviously, the conversion to real time depends on the timescale of the disease. If we regard an infective generation as around 2-4 weeks, there will be roughly 10-20 generation per year. Figure 1a suggests that the clone would be stable for at least 5 to 10 years if recombination rate is 1%, and for 50-100 years if recombination is 0.1%. These figures appear to be much higher than previously supposed and show that an epidemic may well be useful for epidemiological tracking after its initial expansion has taken place. The extent to which a population is likely to be structured into a series of clones as a result of repeated epidemics is shown on Figure 1b. In this analysis s is the probability of an epidemic occurring in a generation, and p is the proportion of the population represented by this epidemic; it is the product of these two factors, sp, which is important. For example, if an epidemic is believed to occur about once every 100 generations (s=0.01, or about every 10 years), and to involve 1% of the populations (p=0.01), then $sp=10^{-4}$ and it can be seen that a significant degree of "clonality" will only exist if recombination rate is less than about 10^{-3}. As the frequency and/or size of epidemics gets larger then the homogenising effects of recombination become swamped, and a clonal structure can become established despite high rates of recombination. In this consideration of strains created as a result of epidemics, there will (by definition) be genetic differences between the strains which may therefore be accompanied by differences in pathology, virulence and response to treatment.

The situation where epidemics are absent, small or rare enough to be negligible can also be investigated (Box Aii). As a working hypothesis, this assumes that putative clones are present, for example as a result of geographical or host-species isolation. The critical question in terms of virulence patterns is not so much whether these strains actually exist, but whether they are likely to differ genetically and hence possibly differ in factors such as virulence or other clinical properties. In essence, the problem is to determine the amount of recombination necessary to prevent wholesale genetic differentiation of the population into distinct "strains". The degree of genetic differentiation is conventionally measures as G_{st}, the subscript st arising because it measure the differentiation of subpopulations or strains, relative to the total population. G_{st} lies between 0 and 1, representing zero and complete differentiation respectively and is plotted as a function of recombination rate on Figure 1c. Again, the effect is non-linear and it appears unlikely that a clonal structure will arise in populations where recombination rates exceed about 10^{-4}.

In protozoan pathogens there are few estimates for recombination rates apart from *Plasmodium*. One problem is that demonstrating that recombination is rare is insufficient. It must be demonstrated that it is exceedingly rare (less than about 0.0001

judging from Fig 1c) before we could accept a clonal population structure as a working hypothesis. For example, although sexual recombination is known to occur in *Trypanosoma brucei*, its effect on population structure and the existence of strains remains obscure[18]. Similarly, recombination can be inferred in *T.cruzi*[23] but at a level insufficient to break down it's predominantly "clonal" population structure[11] . It is difficult to estimate recombination rates per DNA segment per individual per generation in bacteria. In many respects this depends on the level of assay resolution and on the effect of transformations on gene function. For example sequencing the gene would detect far more genetic variation than would be possible by electrophoresis, and small recombinational events may not affect the function basis (including virulence) of the gene product; both these factors decrease the "effective" recombination rate. Estimates for raw recombination rates per nucleotide are generally low[e.g.24] and we conjecture that the effective rate, i.e. those events creating detectable variation or changes in virulence, are also likely to be low and probably less than about 10^{-5} to 10^{-6} per segment per individual. It is clear that this is low enough that a strain structure may result (see figs 1b and 1c), and explains the observation that a clonal population structure may be observed in *E. coli* despite the presence of recombination[24,25].

In contrast to these epidemiological models, it has recently been proposed that a strain structure could result from acquired immunity and be associated with differing clinical patterns[8,9,26], an influential idea which has generated considerable debate in the literature[e.g.27-30]. One problem arises from the qualitative argument that if a strain is defined at a single immunodominant locus, then alleles affecting factors such as virulence and R_0 (an epidemiological measure of the potential rate of pathogen spread) would segregate freely against a "strain" background unless very closely physically linked to, or encoded by, the immunogenic locus; thus strains could be defined immunologically but would not differ in their clinical or epidemiological properties. In the specific case of *Plasmodium falciparum*, an argument can be made that the PIESA (or PfEMP1) locus determines both strain structure and virulence and that strains may therefore differ clinically[8], but this problem of free segregation restricts the general application of the model. Subsequent work investigating the consequences of immunity to several loci[10] has made the process far more plausible (Box B). This occurs because the barrier to gene flow between strains is increased by the effects of meiosis, (the "meiotic barrier" i.e. parameter *a* in Box B) which depends geometrically on the number of genes defining a strain. Under the supposition of 2, 5, or 10 loci, the magnitude of outcrossing necessary to prevent the creation of separate strains becomes $8x10^{-4}$, $6x10^{-3}$ and 0.2 , respectively, assuming lifelong immunity and that wholesale differentiation throughout the genome occurs when recombination is less than 10^{-4} (Figure 1c). In the case of the short term immunity model (where 90% of infections are assumed to be re-infections), these values are increased by a factor of 10 subject, of course, to a maximum value of r=0.5. This clearly shows that increasing the number of loci greatly relaxes the conditions necessary for the emergence of strains but that, at least for P. *falciparum* the conditions must be far more stringent than the 2 locus, lifelong immunity model originally proposed[10] However, it is not clear how many separate strains can co-exist. The maximum number of "discordant" strains[10] (i.e. those not sharing any common alleles) cannot exceed the minimum number of alleles at any one locus. In

other words, if one locus has only 2 alleles then this effectively limits the number of strains to two. This is critically important as when only 2 strains are present, very stringent conditions are necessary for the emergence of strain structures (Figure 1c). Thus increasing the number of loci defining a strain is a two edged sword. It increases the barriers to gene flow (see above) but increases the likelihood of reducing the number of discordant strains. It is known that many antigenic loci appear to encode 2 major allelic types[e.g.31], which would limit the number of discordant strains to two, but whether these major divisions obscure more numerous immunologically distinct variants is unclear. Obviously a final assessment awaits a more detailed understanding of the immunological relationships between pathogen and host.

We are now in a position to quantify the consequences of acquired immunity acting against antigen loci. To investigate the illustrative case of *Plasmodium falciparum* we assume outcrossing rates of 0.1 to 0.66[14-16] and use the methodology of Box B, arbitrarily assuming there are 5 strains defined by either a single locus or by 5 unlinked loci. Intuitively, loci closely physically linked to the immunogenic locus/loci have reduced levels of recombination between strains and hence will become more differentiated between strains. The results are illustrated on Figure 2. As might be expected from Figure 1c, selection at immune loci does not necessarily result in wholesale genetic differentiation of the genome. Rather it occurs in "foci" around the selected locus or loci; the size of these foci being determined by factors which act to reduce recombination. In the case of Figure 2 these are outcrossing rate, number of loci which define a "strain", and the length of immunity. It has been estimated that in *P. falciparum,* one centimorgan corresponds to 15-30 kb of DNA[32], and these calculations suggest that genetic differentiation would only occur upto 60 kb of DNA either side of a single immune locus. It seems improbable that such a small piece of genetic material could encode major differences in both virulence and R_o, so, at least in *Plasmodium,* the available data argue against the differentiation of clinically distinct strains (although one could, of course, increase the size of this putative differentiated region by postulating a larger number of loci defining a strain, and/or by postulating that a higher proportion of infections are re-infections; that is increasing either *a* or *b*, the meitotic and immune "barriers", in the methodology of Box B). Similar calculations can be made for other pathogens as details of their biology become available. It may be that in other eukaryotic pathogens, such as Leishmania, that outcrossing rates are so low that these foci are much wider and/or that the effects of immunity and segregation decrease the rate of recombination between strains to the extent that they become differentiated at all loci, including those unlinked to the immune loci (Figure 1c).

These results show that strains will not be either differentiated or undifferentiated, but these are merely the ends of a continuum. As outcrossing rate is decreased the differentiation around the immune loci expands outwards along the chromosome (Figure 2) until ultimately it falls sufficiently that even loci on separate chromosomes become differentiated (when recombination rate between clones is less than about 10^{-4}, Figure 1c). Figure 2 also suggests another test to identify immune loci. Alleles at the

immunogenic loci defining a strain are subject to "balancing" selection, i.e. they become selectively more advantageous as the strain becomes rarer (since the number of non-immune susceptible hosts increases) and are therefore maintained in the population indefinitely. Variation therefore accumulates around such loci whereas in other regions of the genome variation is periodically lost by selective sweeps or stochastic processes; this difference in the level of variation may be diagnostic of loci undergoing balancing selection[33] and therefore be used to identify immunogenic loci defining a strain. It is dubious whether the map of molecular markers is dense enough at present (re Figure 2) but in principle it constitutes a fourth independent population genetic test for immunogenicity (the others being the rate of molecular evolution[34], differences in genetic variation within hosts[35], and linkage disequilibrium between putative immune loci[10]).

An important point to note in these calculations is the differences between bacteria and eukaryotes in the putative emergence of strains. Since the former only exchange small pieces of DNA, the recombinant offspring will not inherit an (unfavourable) mixture of immune loci but will retain the same "strain" genotype and antigenic profile. In other words, the meiotic barrier is absent in bacteria (a=1 in the terminology of Box B) so there is no segregation load on the population and prokaryotic strain structure will evolve independently of the number of loci defining the strain's antigenic phenotype. The presence of strain structure in prokaryotes acts to decrease recombination between strains solely through the actions of acquired immunity (the "immune barrier") which reduces the probability of co-infection of separate strains (the factor b in the terminology of Box B). The other major consequence is that recombination is unaffected by chromosomal location relative to immune loci, so "foci" of increased genetic differentiation and variation (see Figure 2) will not build up around immune loci; in the case of prokaryotes the results shown on Figure 1c quantify the extent of genetic differentiation as a function of recombination rate.

The calculations based on a model of multi-locus immunity, show that the demonstration of significant outcrossing within a population is insufficient to reject the hypothesis that no strain structure is present. Even the relatively high outcrossing rate of 0.67 obtained for *P. falciparum* in Tanzania could, in principle, represent outcrossing *within* strains while outcrossing and recombination *between* strains could be orders of magnitude lower (see methodological note in Box B) and allow a strain structure to result (Figure 1c). Thus it may be logically more robust to search for linkage disequilibrium rather than outcrossing as a test of strain structure[3,11], although care has to be taken to ensure that spurious positive results do not occur through geographic subdivision[3] or restrictive sampling of a single non-representative epidemic[19]. Paradoxically, genetic differentiation at unlinked loci may result in strains of different clinical characteristics but prevent the identification of immunogenic loci either by the linkage disequilibrium method proposed by Gupta *et al*[10] (because all loci will be linked, not just the immunogenic), or by the G_{st} method proposed by Hastings[35]. Another consequence of short-term immunity restricting outcrossing to predominantly

within strains is that the production of deleterious inter-strain genotypes is drastically reduced. This greatly reduces the genetic load (or, more specifically, the segregational load) within the population which could be used as an argument against multi-locus strain structure[11].

In many infections there may be a positive correlation between short-term transmission rate and virulence. Increased virulence may be favoured since it increases the short-term rate of transmission despite the fact that it may ultimately reduce the total number of transmissions[e.g.36-38]. Thus, the processes of strain formation and virulence levels are correlated and spring from the same epidemiological factor: as mixed infections increase so does expected virulence and recombination rate, and hence strain differentiation decreases (Figure 1b, 1c). Thus we may expect that as the infection rate falls due to epidemiological or immune models described above, recombination rate falls and increasing strain differentiation will result. These strains may differ in virulence due to random genetic differences, but will on average be expected to be less virulent than in areas of intense transmission. Conversely, in areas of intense transmission, recombination rate will increase, strain structure will be less apparent, and overall levels of virulence may be higher.

As stated at the outset, the results described above cannot give a definitive answer to the current debate on "clonality", rather they provide an objective basis within which this debate can be conducted. They provide general pointers, for example to the fact that it is the extent of outcrossing and recombination which is important, rather than its presence or absence. Figures 1a, 1b and 1c all suggest that only at levels of recombination below about 10^{-3} or 10^{-4} are clinical differences in strains likely to emerge. They also show that certain hypothesis are incompatible: for example a model of *Plasmodium* strains defined at two loci and provoking lifelong immunity[10], even if true, is extremely unlikely to result in differing clinical properties unless the actual strain-defining loci (antigens) are themselves responsible for pathology (Figure 2). The calculations also provide a framework for interpretation of field data. For example it has been shown that the genotypic diversity of malarial infection declines with age and in adults is characterised by occasional bouts of infection. This could, in principle, be interpreted as putative support for a strain specific pattern with children becoming immune to strains and adult infections occurring as immunity is lost. However if epidemiological evidence suggests that the latter only accounts for 10% of total transmission, then k=0.1 in the model of short-term immunity of Box B and it is unlikely that these putative "strains" would differ genetically (Figure 2) and hence clinically. Perhaps more importantly the application of an objective body of theory from a different branch of biology provides a fresh perspective on the important parameters in the evolution of virulence; the most critical one in this respect being whether immunity is strain-specific and if so whether it is lifelong or periodically lost.

We thank J. Maynard Smith, M.Tibayrenc, S. Gupta and M. Fisher for comments and/or discussion. This work was supported by the Medical Research Council.

References.

1. Tibayrenc, M., Kjellberg, F. and Ayala, F.J. (1990) **A clonal theory of parasitic protozoa: the population structure of *Entamoeba, Giardia, Leishmania* and *Trypanosomes*, and its medical and taxonomic consequences.** *Proc. Natl. Acad. Sci. USA* 87, 2414-2418.

2. Tibayrenc, M. and Ayala, F.J. (1991) **Towards a population genetics of microorganisms: the clonal theory of parasitic protozoa.** *Parasitol. Today* 7, 228-232.

3. Dye, C. (1991) **Population genetics of nonclonal, nonrandomly mating malarial parasites.** *Parasitol. Today* 7, 236-240.

4. Tibayrenc, M. *et al.* (1991) **Are eucaryotic microorganisms clonal or sexual? A population genetics vantage.** *Proc. Natl. Acad. Sci. USA* 88, 5129-5133.

5. Walliker, D. (1991) **Malaria parasites: randomly interbreeding or "clonal" populations?** *Parasitol. Today* 7, 232-235.

6. Read, A.F. and Day, K.P. (1992) **The genetic structure of malaria parasite populations.** *Parasitol. Today* 8, 239-242.

7. Day, K. *et al.* (1992) **Population genetics and dynamics of *Plasmodium falciparum*: an ecological view.** *Parasitology* 104, S35-S52.

8. Gupta, S. and Day, K.P. (1994) **A strain theory of malaria transmission.** *Parasitol. Today* 10, 476-481.

9. Gupta, S. *et al.* (1994) **Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*.** *Science* 263, 961-963.

10. Gupta, S. *et al.* (1996) **The maintenance of strain structure in populations of recombining infectious agents.** *Nature Medicine* 2, 437-442.

11. Tibayrenc, M. (1995) **Population genetics of parasitic protozoa and other microorganisms.** *Adv. Parasitol* 36, 47-115.

12. Maynard Smith, J., Dowson, C.G. and Spratt, B.G. (1991) **Localised sex in bacteria.** *Nature* 349, 29-31.

13. Cohan, F.M. (1994) **The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes.** *Am. Nat.* 143, 965-986.

14. Hill, W.G. *et al.* (1995) **Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malarial parasites.** *Genet. Res.* 65, 53-61.

15. Hill, W.G. and Babiker, H.A. (1995) **Estimation of numbers of malaria clones in blood samples.** *Proc. Roy. Soc. Lond. B.* 262, 249-257.

16. Paul, R.E.L. *et al.* (1995) **Mating patterns in malarial parasite populations of Papua New Guinea.** *Science* 269, 1709-1711.

17. Hastings, I.M. (1996) **The evolution of drug and vaccine resistance in *Plasmodium.*** *submitted*

18. Hide, G. *et al.* (1996) **The origins, dynamics and generation of *Trypanosoma brucei rhodesiense* epidemics in East Africa.** *Parasitol. Today* 12, 50-55.

19. Maynard Smith, J. *et al.* (1993) **How clonal are bacteria?** *Proc. Natl. Acad. Sci. USA* 90, 4384-4388.

20. Wright S. (1978) *Evolution and the genetics of populations. Vol. 4: variability within and between natural populations.* University of Chigaco Press, Chicago.

21. Hartl D.L. and Clarke A.G. (1989) *Principles of Population Genetics.* 2nd ed. Sinauer Associates,Sunderland, Ma., USA.

22. Crow J.F. (1986) *Basic Concepts in Population, Quantitative and Evolutionary Genetics.* W.H. Freeman, New York.

23. Carrasco, H.J. *et al.* (1996) **Genetic exchange as a possible source of genomoc diversity in sylvatic populations of *Trypanosoma cruzi.*** *Am. J. Trop. Med. Hyg.* 54, 418-424.

24. Guttman, D.S. and Dykhuizen, D.L. (1994) **Clonal divergence in *Escherichia coli* as a result of recombination not mutation.** *Science* 266, 1380-1383.

25. Desjardin, P. *et al.* (1995) **Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from randomly amplified polymorphic DNA and restriction-fragment-length polymorphism.** *J. Mol. Evol.* 41, 440-448.

26. Gupta, S. *et al.* (1994) **Parasite virulence and disease patterns in *Plasmodium falciparum* malaria.** *Proc. Natl. Acad. Sci. USA* 91, 3715-3719.

27. Saul, A. (1996) **Transmission dynamics of *Plasmodium falciparum.*** *Parasitol. Today* 12, 74-79.

28. Gupta, S. and Anderson, R.M. (1996) **Transmission dynamics of *Plasmodium falciparum*: reply.** *Parasitol. Today* 12, 82-83.

29. Dye, C., Lines, J.D. and Curtis, C.F. (1996) **A test of the malaria strain theory.** *Parasitol. Today* 12, 88-89.

30. Gupta, S. and Snow, R.W. (1996) **How do bednets influence the transmissibility of *Plasmodium falciparum*?** *Parasitol. Today* 12, 89-90.

31. Smith, N.H., Maynard Smith, J. and Spratt, B.G. (1995) **Sequence evolution of the *porB* gene of *Neisseria gonnorrhoeae* and *Neisseria meningitidis*: evidence of positive Darwinian selection.** *Mol. Biol. Evol.* 12, 363-370.

32. Walkerjonah, A. *et al.* (1992) **An RFLP map of the *Plasmodium falciparum* genome, recombination rates and favored linkage groups in a genetic cross.** *Mol. Biochem. Parasitol.* 51, 313-320.

33. Norborg, B., Charlesworth, B. and Charlesworth, D. (1996) **Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species.** *Proc. Roy. Soc. Lond. B.* 263, 1033-1039.

34. Hughes, M.K. and Hughes, A.L. (1995) **Natural selection on *Plasmodium* surface proteins.** *Mol. Biochem. Parasitol.* 71, 99-113.

35. Hastings, I.M. (1996) **Population genetics and the detection of immunogenic and drug resistant loci in *Plasmodium.*** *Parasitology* 112, 155-164.

36. Levin, B.R. and Bull, J.J. (1994) **Short-sighted evolution and the virulence of pathogenic microorganisms.** *Trends in Microbiol.* 2, 76-81.

37. Lenski, R.E. and May, R.M. (1994) **The evolution of virulence in parasites and pathogens: reconciliation between two competing hypothesis.** *J. theor. Biol.* 169, 253-265.

38. Frank, S.A. (1996) **Models of parasite virulence.** *Quart. Rev. Biol.* 71, 37-78.

39. Roberts, M.S. and Cohan, F.M. (1993) **The effects of DNA sequence divergence on sexual isolation in Bacillus.** *Genetics* 134, 402-408.

40. Attwood, K.C., Schneider, L.K. and Ryan, F.J. (1951) **Periodic selection in *Escherichia coli.*** *Proc. Natl. Acad. Sci. USA* 37, 146-155.

41. Levin, B.R. (1981) **Periodic selection, infectious gene exchange, and the genetic structure of *E. coli* populations.** *Genetics* 99, 1-23.

42. Maynard Smith, J. (1991) **The population genetics of bacteria.** *Proc. Roy. Soc. Lond. B.* 245, 37-41.

43. Barton, N.H. (1983) **Multilocus clines.** *Evolution* 37, 454-471.

44. Wedgwood-Oppenheim, B. (1996) **Immmunity and the population genetics of malaria.** *Thesis, submitted.*

Box A: epidemiological strains.

(i) population genetics of a new "epidemic" strain

For convenience we follow the derivation and symbolism of Crow[22]. We assume the frequency of a gene in the mainland population is P, in the original epidemic clone it is p_0, and in the epidemic clone after t generations it is p_t. It is easily shown (assuming that the clone is sufficiently small that mating within the clone can be ignored) that

$$\frac{p_t - P}{p_0 - P} = (1 - r)^t \qquad \text{(Crow[22], rearrangement of eqn 3.3)}$$

In other words, the difference in allele frequency decays at a rate of r per generation. Linkage disequilibrium also decays at the same rate, assuming the main part of the population is in equilibrium, so the "lifespan" will be the same irrespective of whether we define a strain by its allele frequencies or by its linkage disequilibrium. For convenience, we assume that the "clone" ceases to be genetically distinct once it has lost 90% of its original differentiation. It is straightforward to calculate the expected "lifespan", l, of the clone in generations as a function of recombination rate. This is shown of Figure 1a and it is obvious that genetically distinct clones can exist on a timescale useful in epidemiological tracking at low recombination rates. In this and subsequent calculations, we assume that there is no barrier to recombination within the "species", although in bacteria it can be shown that recombination (via transformation) decreases as the genetic difference between strains increases[39]. We also assume that an epidemic expansion occurs, and then terminates with no expected difference in fitness relative to other individuals of the same species. Such a situation may occur when a pathogen population is expanded purely by chance (for example, if it contaminates a water supply). The situation where a "clone" expands as a result of a favourable mutation has been considered elsewhere[13]

We can make some quick approximate calculations to determine how this affects the overall "clonality" of the population. Assume that epidemics occur at frequency s and each epidemic spreads to represent a proportion p of the population (e.g. if its eventual size is 5% of the population p=0.05). The chance of an individual *not* being in a clone is the probability that its ancestor in each of the previous l generation had not been incorporated into an epidemic which is $(1-sp)^l$. Clonality is defined as the probability that a randomly selected individual *is* a member of a genetically distinct clone which is therefore $1-(1-sp)^l$. Clonality as a function of recombination rate is shown on Figure 1b for a variety of values of sp. Again, this shows very clearly that a "clonal" population structure can only persist at low recombination rates. Note that this assumes that a epidemic spreads and then becomes selectively neutral, i.e. is no longer favoured by natural selection. When this is not the case, a selectively advantageous allele can sweep through a strain removing genetic diversity at linked loci by hitch-hiking: this reduces diversity within the strain[40-42] and may increase genetic differentiation between them although the dynamics are complicated[13]. "Clonality" measures the probability that an individual will belong to a genetically distinct strain, but does not tell us how many strains will be present. This occurs because s and p are confounded: a high value of s (with low p) will result in more strains being present than the case of low s (with high

p): the exact number expected will also depend on population size and the extent to which the population is sampled.

(ii) population genetics of extant strains

Genetic differentiation between the clones (or islands) is conventionally measured as G_{st} whose scale is from 0 to 1. It can be shown that:

$$G_{st} = \frac{1}{4Nr_s\alpha + 1} \quad \text{(from Crow}^{22}\text{, eqn 3.18)}$$

where $\alpha = (\frac{n}{n-1})^2$, N is the number of individuals per strain, and n is the number of strains. Note that r_s is the effective recombination rate between *strains*; if there are n strains then this is $r_s=r(1-1/n)$ because a proportion 1/n of recombinational events will be within the same strain. The factor r_s therefore replaces migration rate m in the original equation. Total population size ($=Nn$) is assumed to be large (in this case one million). G_{st} as a function of recombination rate is shown on Figure 1c assuming differing number of clones within the population.

Box B: Gene flow in immunity-derived strain structure.

It is assumed that strong immune selection on antigenic loci structures the pathogen population into a series of antigenically distinct "strains". As in Box A we need to calculate an effective recombination rate *between* strains r_s as $r_s=r(1-1/n)ab$ where r is recombination rate in the absence of immunity, and is scaled by a to take account of the barrier to gene flow due to the segregation at meiosis (the "meiotic barrier"), and b to account for the reduction in outcrossing between different strains due to host immunity preventing their coexistence (the "immune barrier"). There is a slight methodological question here as when we measure recombination or outcrossing in the field it is already reduced by any effects of immunity, ultimately it may merely reflect outcrossing rate within clones; thus we obtain underestimates of r_s (i.e. recombination in the absence of immunity) from field data. We could reduce this by investigating immunologically naive or compromised individuals or merely by regarding r_s as a conservative estimate.

If two strains undergo "conventional" sexual recombination (e.g. *Plasmodium)*, then the probability of retaining the original strain structure after meiosis and an unlinked allele "migrating" into the other strain is 0.5^z ,where z is the no. of loci defining the strain. Thus for "conventional" sexual diploids we calculate $a=0.5^z$ but for bacteria it appears that only small, individual pieces of DNA take part in "recombination" so a=1. The calculation for conventional sexual species make the conservative assumption that all intermediate antigenic combinations immediately die and that segregation at meiosis forms a "meiotic barrier" to gene flow between strains. If this is not the case, "stepping stone" models or models of gene flow across a selective barrier of a "hybrid zone"[43] could be employed. The critical points are that a difference exists between the magnitude of a for "conventional" sexual species and bacteria and that the assumption of lethality over-estimates the extent of genetic isolation (and hence differentiation) between strains.

We can make an estimate of b using the existing methods[35](note that in these papers G_{st} are not comparable: in the former[35] individual hosts were analogous to islands whereas in the present case islands correspond to strains). Assume there are n strains and that hosts encounter and then become resistant to them. We can regard this as a flux with hosts classified into, and moving through, immune status groups progressing from resistant to 0, to resistant to 1, to 2 etc. upto resistant to $n-1$ strains (we are not interested in people resistant to all n since they won't get any infections). Assuming immunity is lifelong (and death rate is low relative to infection rate) the reduction in outcrossing rate *between strains, b,* can be calculated as

$$b = \frac{1}{n}\sum_{i=0}^{n-1}\left(1 - \frac{i}{n-1}\right) = \frac{1}{2}$$

The term in brackets is the probability that the second haplotype has survived the effects of acquired immunity. This is an important result as it shows that lifelong, strain-specific immunity cannot reduce outcrossing rate between strains below 50% of the value it would take in a naive population. The weighting factor n arises because immunity is lifelong so there are n classes of immune status each of which have the

same weight. This is not the case when immunity is not lifelong[44]. We assume that immunity is periodically lost so that hosts slip back into the class of being susceptible to only a single strain and that infection rate is so rapid that they become re-infected rather than falling back into immune status group n-2; we call this the "short-term immunity" model. Recombination between strains obviously cannot occur in the class n-1 so if the proportion of infections which occur in this manner is k, then we can recalculate b as

$$b = \frac{(1-k)}{2}$$

Thus under short-term immunity, the reduction in outcrossing rate can greatly exceed the maximum of 50% which occurs when immunity is lifelong. When interpreting the results of these calculation (e.g. Figure 2), it is important to remember that the estimate of a ia an underestimate; thus the general conclusion that it is difficult to produce genetically differentiated strains in eukaryotic species rests on conservative assumptions and is robust. Note that the same procedure can be used to investigate the situation where strains are adapted to different environments (such as hosts species) by allelic differences at z loci which act epistatically. The value of a is unchanged as it measures the barrier to gene flow caused by the breakdown, at meiosis, of favoured allele combinations. The parameter b now becomes a measure of the relative ability of a strain to establish itself in a non-preferred environment or host type.

Figure 1a. The "lifespan" of a clone as a function of recombination rate. The lifespan is arbitrarily taken as the time (in generations) for 90% of its genetic differentiation to disappear. Note that it is assumed that the loci are unlinked in eukaryotes in which case recombination rate is half outcrossing rate.

Figure 1b. The extent of clonality within a population as a result of an epidemic population structure, plotted as a function of recombination rate (see caption to Figure 1a). Clonality is defined as the probability that an individual randomly sampled from the population belongs to a genetically distinct clone.

Figure 1c. The extent of genetic differentiation between "strains" as a function of r_s, the recombination rate between clones (see text for details of how this is calculated). It is assumed that 2,5,10 or 20 strains are present and that total population size of pathogens is 1,000,000.

Figure 2. Genetic differentiation, G_{st}, in the chromosome surrounding an immunogenic locus at position "0"; in *Plasmodium falciparum* 1 centimorgan is equivalent to approximately 15-30 kb[32]. Outcrossing rates were 0.1 or 0.67 corresponding to estimates in *P. falciparum* from Papua New Guinea[16] (PNG) and Tanzania[14,15] (Tanz) respectively. There were assumed to be 5 distinct strains defined either by a single locus or by 5 unlinked loci. Immunity was assumed to be either lifelong or short term (i.e. is periodically lost such that 90% of infections are re-infections of hosts previously exposed to that strain, see Box B for further details). Note that this only applies to pathogens undergoing "standard" eukaryotic meiosis and recombination. In prokaryotes, recombination is independent of chromosome location so no such foci of genetic differentiation would be expected; this is illustrated by the example "bacteria" where recombination is $5x10^{-5}$ and G_{st} remains constant at 0.2 over the whole genome.

Figure 1a

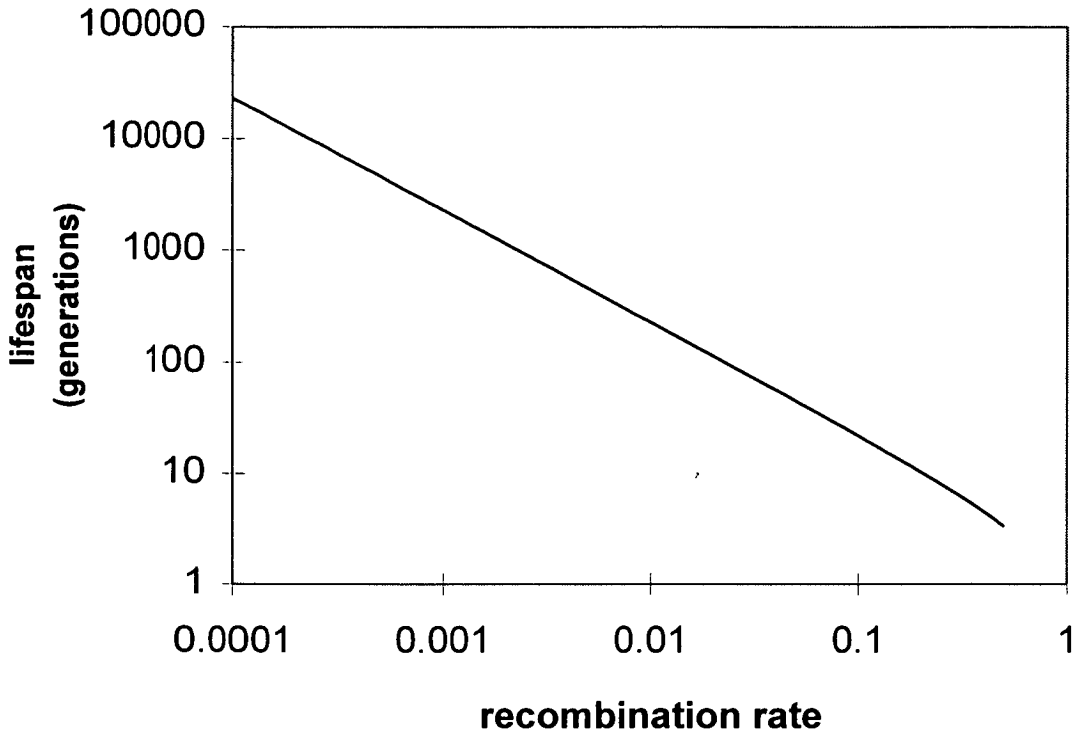

100000
10000
1000
100
10
1
lifespan (generations)
0.0001
0.001
0.01
0.1
1
recombination rate

Figure 1b

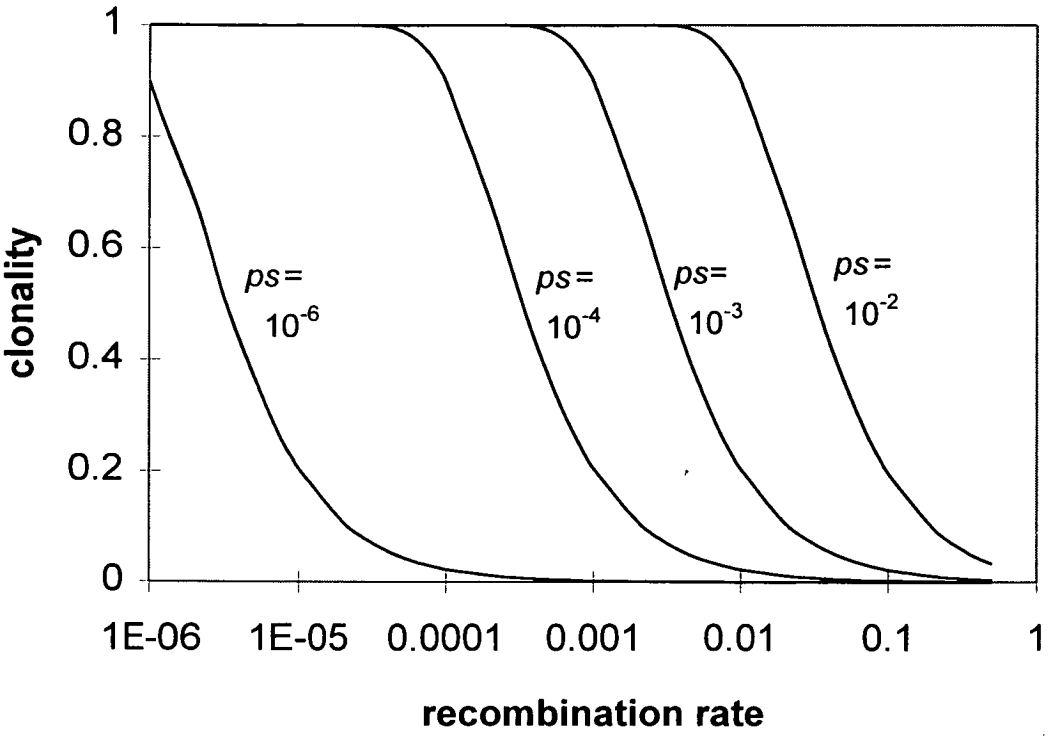

clonality
1
0.8
0.6
0.4
0.2
0
ps= 10-6
ps= 10-4
ps= 10-3
ps= 10-2
1E-06
1E-05
0.0001
0.001
0.01
0.1
1
recombination rate

Figure 1c

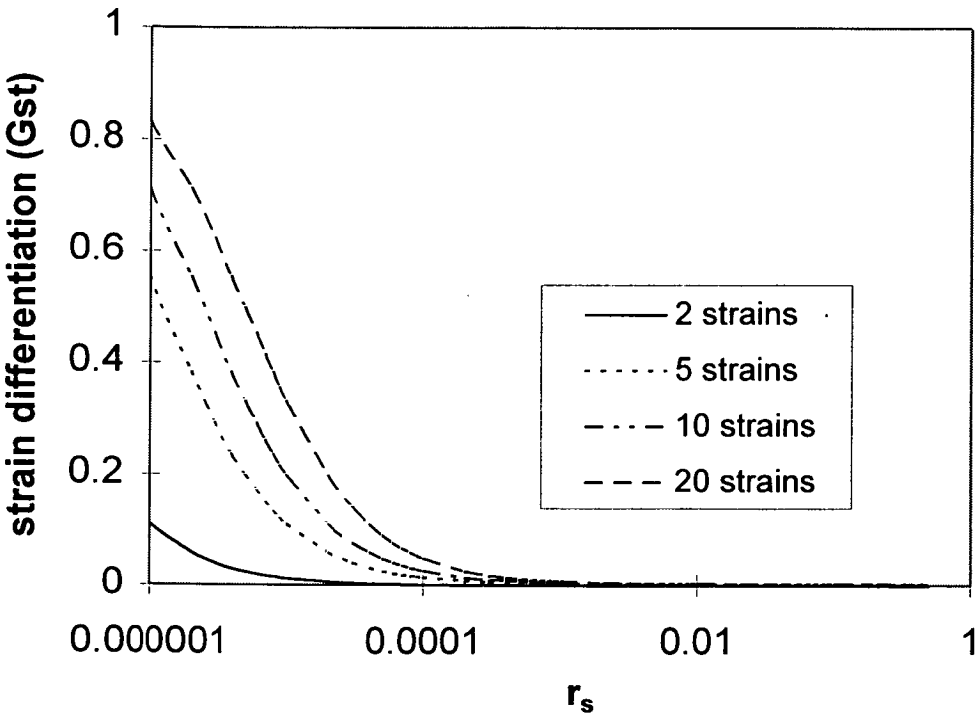
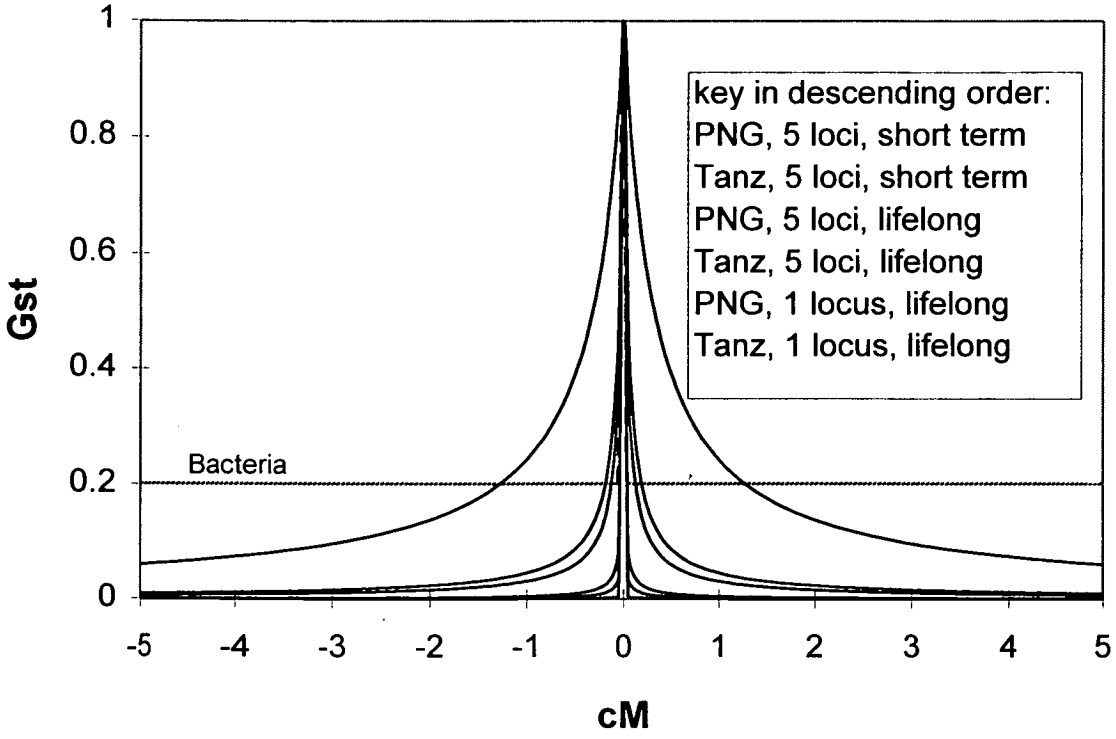

strain differentiation (Gst)
1
0.8
0.6
0.4
0.2
0
0.000001
0.0001
0.01
1
r_s
2 strains
5 strains
10 strains
20 strains

Figure 2


1
0.8
0.6
0.4
0.2
0
Gst
Bacteria
key in descending order:
PNG, 5 loci, short term
Tanz, 5 loci, short term
PNG, 5 loci, lifelong
Tanz, 5 loci, lifelong
PNG, 1 locus, lifelong
Tanz, 1 locus, lifelong
-5
-4
-3
-2
-1
0
1
2
3
4
5
cM