Breeding for Disease Resistance:

Uniting Genetics and Epidemiology

Prof. Steve Bishop

Roslin Institute (Edinburgh), Midlothian, UK

Iowa State University,
Ames,
Iowa

May 19-21, 2003

Acknowledgements

These course notes arose out of a summer course in disease genetics, presented in Montpellier, France, August 12-14, 2002. This course was offered by the Animal Breeding and Genetics team of INA Paris-Grignon jointly with the Animal Genetics department of INRA. It was organised in collaboration with the *ABIES* (Agriculture and food, Biology, Environment, Health) Doctoral School and with the material support of the *Cours Supérieur d'Amélioration Génétique des Animaux Domestiques* (CSAGAD). I thank Prof. Etienne Verrier and Dr. Rachel Rupp for the invitation to present this course, which gave the impetus for the preparation of these notes. I also thank Prof. Verrier and Dr. Rupp for permission to reproduce and update these notes for this course.

This current course arose from an invitation from Prof. Sue Lamont, to whom I am indebted. The following units of Iowa State University are thanked for funding that enabled this course: the Animal Breeding and Genetics Group, the Department of Animal Science, the Center for Integrated Animal Genomics and the Biotechnology Council.

I would like to thank many people who have assisted directly in the research programs that led to my approach to disease genetics, and also in the preparation of the notes themselves. First, I would like to thank all my colleagues who have contributed to the research described in these notes, especially Mike Stear, Katrin MacKenzie, Liesbeth van der Waaij, Dimitrios Vagenas, Frank Jackson, Pieter Knap, Miriam Nagel, Anthea Springbett, Liz Glass and Mary Clapperton. Secondly, I would like to thank colleagues at Roslin Institute for their continued input and encouragement, especially John Woolliams and Chris Haley. Also, many thanks to Mintu Nath for reading and commenting upon earlier versions of these notes, correcting a number of errors, and to Mart de Jong, Douglas Gray and Keith Hammond for inspirational discussions. I would also like to personally thank Prof. Max Murray (Glasgow University) and Prof. Bob Webb (Nottingham University) for their recognition the importance of this work.

Lastly, I wish to acknowledge the many funding bodies that have contributed to my research, both experimental and theoretical, including Defra, SEERAD, MLC, BBSRC, BWMB, DFID and the EC.

Steve Bishop
Roslin Institute
May 2003

Contents

1. Introduction to Disease Biology and Disease Genetics

1.1 Background

Infectious disease may be defined as diseases caused by organisms that colonise a host, such as viruses, bacteria, protozoa, helminths, ectoparasites, etc. These organisms are referred to generally as **parasites**. Infectious disease has major adverse effects on livestock production and animal welfare, worldwide. In market-oriented terms, the costs of disease are estimated as 10-20% of turnover within the livestock sector in the developed world (£1.7 billion in the UK) and 35-50% in the developing world. All animal production systems are subjected to challenges by infectious disease and all animal production systems are liable to large production losses from disease. Disease also poses threats across species barriers as diseases in one species may act as reservoirs for infections in other species. In particular, animal infections pose zoonotic threats to human health.

The control or management of infectious disease is currently achieved by a number of mechanisms, including (i) chemical intervention, notable examples being anthelmintics for nematode parasite control, acaricides for tick control and antibiotics for the control of many bacterial diseases; (ii) vaccination, (iii) sanitation and disinfection, and (iv) culling, isolation and control of the movements of animal and/or animal products. Disease control or management using genetics, i.e. exploiting genetic variation amongst hosts in disease resistance (in the most general sense) is increasingly looked upon as an additional way of controlling many diseases, complementing or sometimes replacing existing strategies. Opportunities for utilising genetics are introduced in section 1.4.

Breeders and agricultural industries now face considerable pressures to select animals for enhanced disease resistance. Over and above the desire of breeders to enhance profitability and seek a competitive edge, the foremost pressure to breed for disease resistance is the evolution of resistance in parasites to chemical or vaccine control measures. Important examples include:

- The evolution of resistance to anthelmintics by nematodes in all major sheep producing countries. This threatens sustainable sheep production throughout the world. Resistance has developed to all classes of anthelmintic, and no new classes are being produced by pharmaceutical companies.
- The evolution of resistance to antibiotics by bacteria. This threatens the efficacy of antibiotics in many production systems. This is especially problematic in intensive production systems where antibiotics are used to control unknown and sometimes sub-clinical disease problems.
- The evolution of resistance to anti-protozoals. For example, there is widespread resistance to drugs used to control animal trypanosomosis.
- The evolution of acaricide resistance by populations of ticks, leading to greater problems of tick infestation and tick-borne diseases.
- The evolution of vaccine resistance. For example, each new generation of vaccines against Marek's disease has apparently been followed by strains of ever more virulent viruses.

A recent phenomenon is that of Governments dictating breeding strategies to farmers. An example is the drive to remove scrapie from sheep flocks in Western Europe, using selection on PrP genotype. Additionally, legislation limiting antibiotic usage may push breeders to select for goals such as increased generalised immunity.

Predicting the consequences of breeding animals for disease resistance is qualitatively different from predicting the consequences of breeding animals for a production trait. A safe assumption with most production traits is that animals express their trait independently of other animals in the herd or flock. This is not always the case with resistance to infectious diseases, where animals may infect each other either directly or indirectly. Hence, the expression of infection or disease status between individual animals within the same population is not independent of each other, depending upon the dynamics or epidemiology of the disease. Therefore, the consequences of genetic improvement will depend upon both the genetic makeup of the host population and the epidemiology of the disease.

Therefore, animal breeders need to widen their perspective to include the dynamics of the disease and ask the question: **what impact will changing host genotype have upon disease dynamics?**

1.2 Basic Definitions and Concepts for Disease Biology and Epidemiology

Following are definitions and concepts to give background information necessary to understand the principals of disease genetics and epidemiology. Many of the basic definitions come from Grenfell and Dobson (1995), however additional context has been added.

Infection vs. Disease
Disease is often used to cover two distinct concepts: **infection** and **disease**. **Infection** is the colonization of a host animal by a parasite. In some but not all diseases, multiplication of the parasite within the host is part of the infection process. Parasite is a general term describing an organism with a dependence upon a host. Parasites include many organisms, including viruses, bacteria, protozoa, helminths, flies and ticks. **Disease** describes the side effects of infection by a parasite. Disease may take several forms including acute, sub-acute, chronic and sub-clinical, which may or may not be debilitating.

Resistance vs. Tolerance
Resistance is the ability of the individual host to resist infection or control the parasite lifecycle. **Tolerance** is the ability of a host to tolerate infection and show little or no measurable detriment, i.e. minimal effects of disease.

Disease control is often used ambiguously to include both the control of the transmission of infection, i.e. resistance, and improvement of tolerance. The control of infection aims to reduce or eliminate the transmission of infection through a population of host animals. An outcome of successful infection management will be a subsequent reduction in the incidence or severity of disease in the host population and also in other populations in contact with the host population. At best, genetic improvement of a population in terms of its resistance to infection may lead to eradication of that disease. This is currently being attempted with breeding programs for scrapie resistance in sheep. The improvement of tolerance will alleviate suffering in infected individuals. However, if the transmission of infection is not simultaneously reduced, it will not reduce the incidence or severity of disease in other populations in contact. In some circumstances the management of tolerance is an effective means of managing a disease problem, especially when there is no means of reducing the infection pressure faced by the population. However, it may not be appropriate when a requirement is to manage transmission of infection in addition to alleviating animal suffering.

At worst, management of the symptoms of disease alone, rather than the underlying cause of disease, may mask problems and lead to greater future disease problems.

Horizontal vs. Vertical Transmission

Horizontal transmission describes the transmission of infection between infected and susceptible individuals, or between disease vectors (an animal or object that transmits parasites) and susceptible individuals. **Vertical transmission** is from parent to unborn offspring. Sources of horizontal transmission include ingestions of contaminated food or drink, inhalation of contaminated air droplets, direct contact, injection via an animal's saliva or bite, invasion via open wounds or penetration of host by actively searching parasite.

Epidemic vs. Endemic

An **epidemic** is a sudden, rapid spread or increase in the prevalence or intensity of a parasite or disease. An epidemic is often the result of a change in circumstances that favour parasite transmission such as a rapid increase in host population density, climatic changes that favour the parasite, the introduction of a new parasite or strain of parasite, or the introduction of a new and susceptible host population. A widely distributed epidemic is known as a **pandemic**. An **endemic** describes a parasite or disease that does not exhibit wide fluctuations through time in a defined location.

Microparasites vs. Macroparasites

Microparasites are parasites that undergo direct multiplication within their hosts (e.g. viruses, rickettsia, bacteria, fungi and protozoa). Microparasites are characterised by small size, short generation intervals and a tendency to induce immunity to infection in those hosts that survive the primary infection (i.e. the first infection of a given parasite in an individual host). Duration of infection is usually short in relation to the expected life span of the host (although there are exceptions, e.g. the slow viruses). **Macroparasites** are parasites such as helminths and arthropods that in general do not multiply within their host, but instead produce transmission stages (eggs and larvae) that pass into the external environment or to vectors (an animal or object that transmits parasites). The immune response elicited by these metazoans (multi-cellular animals) generally depends upon the number of parasites present in the given host and tends to be of a relatively transient nature. An operational definition of a macroparasite is an organism whose population biology requires a full description of the distribution of parasites among hosts (Anderson and May, 1992). For microparasites, such a description is seldom necessary or indeed possible, and simple compartmental models such as 'infected or not' are adequate.

Antibody vs. Antigen

These are commonly used terms used when describing parasites and host immune response (see below). An **antigen** is a substance, generally foreign, that is capable of inducing an antibody response. An **antibody** is a protein produced in the blood of vertebrates in response to an antigen. The antibody produced is able to bind specifically to that antigen, and plays a role in its inactivation or removal by the immune system.

Immunity and herd immunity

Immunity is the ability of the individual host to combat infection or disease due to the presence of antibodies or activated cells. It may be divided into three types: (a) acquired immunity which is conferred on an individual after recovery from a disease, (b) natural or innate immunity which is inherited from parents, or in some cases antibodies may be passed across the placenta and therefore present in the blood at birth, (c) artificial immunity which may be induced by the injection of a vaccine, denatured antigens of a parasite, or antiserum

which contains antibodies and thus may be used when the individual host is already infected. Immunity is often not complete and may vary between individual animals within a population. One of the mechanisms underlying genetic differences in resistance or tolerance is an appropriately targeted immune response: genetic differences in immune response and hence the degree of immunity developed by an individual may be associated with genetic differences in disease resistance or tolerance. **Herd immunity** describes the immunological status of a population of hosts with respect to a given parasite. The level of herd immunity is determined by the net rates at which individual animals acquire and lose their immunity. A feature of herd immunity is that it is sometimes not necessary for all animals in the population to be completely immune for the population as a whole to be free of infection. This is the basis of public health vaccination programs where (for example) not all children are vaccinated against a particular disease, but the mass of vaccinated children effectively protect the unvaccinated children.

Immune responses are very complex and flexixible. The combination of gene segment rearrangement, point mutation (occurring at about 10^7 times the normal rate), and gene conversion generate a very large repertoire of T cell and B cell receptors (approximately 10^{10} for each) with which to sample the antigenic environment. Without these mechanisms, such a repertoire would require more DNA than is present in the entire mammalian genome. An equally important genetic component of the immune system is the major histocompatibility complex (MHC), which derives its complexity from the large number of allelic variants at the more than 200 loci. Significant differences exist in genetic mechanisms for acquired immunity amongst mammalian species and between birds and mammals, which have implications for disease resistance.

Infectious period, latent period and carrier state

The **infectious period** is the period during which an infected animal is able to transmit an infection to a susceptible host or to a vector. The infectious period may or may not coincide with the disease. The **latent period** is the period during which the individual is infected but is not yet capable of transmitting the infection. The **carrier state** is a state in which an infected individual shows no symptoms but is capable of transmitting the infection. Carrier state occurs with many bacterial infections such as tuberculosis. These distinctions are not required when developing the general principals of genetic-epidemiological modelling, but they are important when applying models to some specific diseases and when interpreting field data describing these diseases.

Basic Reproductive Ratio R_0

R_0 is a dimensionless parameter that encapsulates the biological details of different transmission mechanisms. In essence it is the number of successful offspring that a parasite is capable of producing. For **microparasites**, the number of parasites is seldom counted and parasite number is often considered in terms of the number of infected hosts. Therefore, for microparasites, $\mathbf{R_0}$ is defined as the average number of secondary cases in infection resulting from one primary case introduced into a population of susceptible individuals. In this case, susceptible refers to the immune status of the animal, i.e. the animal doesn't have immunity resulting from being challenged by the parasite, rather than the degree of genetic resistance or susceptibility. For example, if the primary animal infects 5 other animals, then R_0 is 5. For **macroparasites**, $\mathbf{R_0}$ is the average number of female offspring (or total offspring for hermaphroditic species) produced by a female mature parasite, which achieve reproductive maturity, in the absence of density-dependent constraints in parasite establishment, survival or

reproduction. This is directly equivalent to Fisher's definition of net reproductive rate for free-living species (Fisher, 1930).

If the host population consists of a mixture of susceptible animals and animals that have previously faced a challenge, or if there are density-dependent constraints limiting parasite population growth, then the concept of R_0 is no longer valid. In this case, the Effective Reproductive Ratio (**R**) is used. However its interpretation is still the same, *viz.* the number of secondary cases (for microparasites) or female offspring (for macroparasites) produced in the host population. Under conditions of stable endemic infection, R=1. In other words, the parasite population is neither growing nor declining.

Estimated R_0 values tend to be variable and dependent upon the nature of the host population. For humans, typical values are 10-14 for measles, 7-12 for chickenpox and 5-7 for polio (Anderson and May, 1992). For domestic animals, R_0 for transmissible gastroenteritis in pigs has been estimated at 2.0 for a breeding farm and 4.0 for a finishing farm (Hone, 1994). Scrapie probably has an R_0 only a little above 1.0, whereas foot and mouth disease almost certainly has a high R_0, well in excess of 10.

Density Dependence

Density Dependent effects are population mechanisms or effects whose intensity of action depends upon the density of the parasite population. When they reduce fecundity or increase mortality (e.g. parasite-induced host mortality), these effects tend to regulate net population density. Density-dependent effects tend to be non-linear. An example is for nematode parasites (worms) in the gut of ruminants: when there is an excessive population of such parasites, the fecundity of these parasites decreases and the establishment of future parasites is inhibited.

Aggregation

Organisms show an **aggregated** (or **over-dispersed**) distribution when they co-occur (e.g. are clumped within certain hosts but not in others) significantly more than would be expected from a completely random Poisson process. This clumping is reflected in a variance/mean ratio significantly greater than unity. Macroparasites are almost always aggregated in their host population. This leads to the well-known 80:20 rule – 80% of parasites being found within 20% of hosts. Pareto (1906) observed that the same rule-of-thumb applies to the distribution of wealth in human populations, although aggregation of wealth is likely to have increased significantly during the last 100 years.

Aggregated distributions are often well described empirically by the negative binomial distribution; the degree of aggregation is inversely proportional to the negative binomial parameter k. Aggregation generally arises from some source of heterogeneity in the host or in the parasite population.

Virulence

Formally, **virulence** is the case mortality rate of a parasite. It is a measure of the damage caused by the parasite to the host, i.e. how pathogenic it is. For the parasite, excessive virulence will be counterproductive as it will result in host death, hence block parasite transmission.

Stochastic vs. Deterministic models

A **deterministic model** is a mathematical model that assumes that all parameters and variables are not subject to random variation, and that the state variables are continuous quantities. The predictions of such models are repeatable point estimates. A **stochastic model** is a mathematical that takes into consideration the effects of random variation in one or more of its parameters or variables. The predictions of such models therefore are not single points but probability distributions.

1.3 Parasitic Organisms and Diseases

1.3.1 Parasitic organisms

Viruses

Viral infections are a fact of life for all organisms, including bacteria and protozoa. Viruses are tiny obligate intracellular parasites. They are incomplete organisms and are dependent on the metabolic and replicative machinery of host cells. Consequently, they can only replicate in host cells that have the necessary cell-surface receptors, and possess the enzymes, co-factors and components that they need for replication. Furthermore, to get established and persist, a virus must evade the innate and adaptive defence mechanisms of the host.

To complete its life cycle a virus must (from Schat and Davies, 2000): (1) gain entry to a host, (2) attach to host tissues, (3) if the site of replication differs from the site of entry travel to the target organ(s), (4) bind to target cells, (5) enter target cells, (6) uncoat its genome, (7) replicate its genome, (8) make structural proteins, (9) assemble capsids, (10) exit the cells, and (11) exit the host and transfer to a new host. Because viruses need to be genetically compatible with their hosts, viruses have restricted host ranges. Furthermore, since they must get to the target organ, must have an appropriate ligand for binding to the target cells, and for replication, require host-cell constituents that may not be present in all types of cells or at all times in the cell cycle, viruses have tissue specificity. A minor mutation in a virus can change either its host range or tissue specificity.

Illnesses produced by viral infections vary from acute (including death) to subclinical. Recurrent infections characterised by repeated growth and decay of the virus population within the individual may also occur. Immune responses vary from non-existent to life-long immunity. Typically there are three phases in the infection: the latent period, the infectious period and the recovery period when typically viral abundance decays to very low levels and antibody titres (the quantities of antibodies specific to viral antigens) rise to high levels. The time between initial infection and the appearance of disease symptoms is called the incubation period.

Bacteria

Whilst bacteria are also single celled organisms, they are somewhat more complex than viruses and invoke a wider range of immune responses. Bacteria are also clearly more complex antigenically than viruses, yet acquired immunity following infection is neither as complete nor as long lasting as it is for viruses.

The host usually employs three phases of defence to overcome bacterial infections (Adams and Templeton, 1998). The first defence mechanism encountered by the pathogen is the epithelial barrier. This provides a physical barrier, but also provides cell surface receptors for attachment of some bacteria, which is the first step in the establishment of many bacterial

infections. Absence of receptor will prevent colonisation of the host. The second phase of defence is non-specific, or innate, immune mechanisms including cells such as neutrophils, macrophages and NK cells that lyse infected cells. Humoral mediators include the complement system and the interferon system. The third phase of host defence is specific or adaptive immunity, which requires activation, proliferation and differentiation of antigen-specific T and B lymphocytes.

Protozoa

Protozoa are much larger than bacteria and viruses and they present many more antigens, both in size and kind. Protozoan populations maintain genetic variation and hence antigenic variation by means by sexual reproduction. Protozoan infections tend to be more persistent than viruses, or even bacteria, and are often chronic in nature. Certain protozoan species persist within their hosts via successive waves of parasitaemias, as is the case for trypanosomes causing trypanosomosis, induced by immunologically (i.e. genetically) distinct populations of parasites. Acquired immunity is rarely fully protective and its efficacy depends upon the duration and intensity of past exposures.

Protozoa may live in the blood stream (malaria and trypanosomes), in tissues (Leishmania) or inside cells. Antibody is the most important defence against protozoa within the blood stream. Cell mediated immunity is active against tissue-living protozoa. For example, antibody can damage parasites directly, enhance their clearance by phagocytosis, activate complement or block their entry into host cells. Inside cells, the parasite is safe from antibody but oxygen metabolites or other cytotoxic factors may play a role in killing the parasite.

Helminths

Helminths are complex multi-cellular macroparasitic worms, including monogenea, digenia, cestodes, nematodes and acanthocephalans. Of these categories, nematodes cause the major disease problems to livestock and include gastrointestinal nematode parasites (i.e. worms). Other well-known helminth parasites include flukes and tapeworms.

Helminths have generation intervals much longer than microparasites. Sexual reproduction occurs within the animal host but this process involves the production of transmission stages, i.e. eggs and larvae, which must leave the host to complete development. Helminth infections are generally chronic in nature and result in morbidity (i.e. loss of health and performance) rather than mortality. Their large size entails the existence of many antigens. Some of these antigens are even specific to a particular stage of development. Host responses are complex, involving many immune mechanisms and hence genes. Typically immunoglobulin levels are high in the blood, including IgE, IgA and IgG, although these levels are transitory and decrease quickly after removal of the parasite.

Arthropods

By nearly any measure, the most successful animals on the planet are the arthropods. They have conquered land, sea and air, and make up over three-fourths of all currently known living and fossil organisms, or over one million species in all. Since many arthropod species remain undocumented or undiscovered, especially in tropical rain forests, the true number of living arthropod species is probably in the tens of millions. Arthropods include insects, and from a livestock disease viewpoint the most important arthropods are flies and ticks.

Flies and ticks are major direct problems of livestock, causing such problems as tick worry, flystrike and disturbing problems such as nasal botfly where the fly lays eggs in the nasal

cavity and the larvae proceed to burrow upwards. However, probably of even greater is importance is the role of flies and ticks as vector of other diseases, such as trypanosomosis, East Coast fever and babesia.

Prions

Prions are the supposed infectious agent for transmissible spongiform encephalopathies (TSEs). A prion is a wrongly folded specimen of a protein called PrP that is normally produced in the correct folding in cells of animals. This wrongly folded PrP results in newly produced PrP in the cells of animals also being wrongly folded. This is a problem as the wrong folding makes the PrP resistant to enzymes that normally breakdown proteins. Hence the prion protein accumulates as clots in the brain. These clots damage and kill cells. These damages lead to serious clinical symptoms and eventually the death of the affected individual. Known TSE's included scrapie in sheep, Creutzfeldt Jakob's Disease (CJD) in humans and bovine spongiform encephalopathy (BSE) in cattle. BSE has apparently entered the human food chain and has caused cases of CJD that were distinct from previously reported cases. This was remarkable as scrapie (the TSE from sheep) is assumed after epidemiological research to be not transmissible to humans. Because of its human health implications, BSE in cattle is rigorously eradicated to minimise further human cases of the new variant CJD.

1.3.2 Diseases of major importance

The nature of prevalent diseases varies with production systems. For example, macroparasitic infections are generally absent from intensive production systems. Intensive production systems, where animals are housed, generally suffer from bacterial and viral infections, some of which may manifest themselves as subclinical infections. Extensive production systems suffer from macroparasitic and microparasitic infections. However it is the macroparasitic infections, e.g. gastrointestinal nematode infections, tick and fly infestations, that tend to be endemic.

Numerous assessments of important diseases have been undertaken. For example, an assessment of the economic consequences of a number of endemic diseases affecting domestic livestock in the UK is presented at http://www.rdg.ac.uk/livestockdisea/. Curiously, this website does not contain details on the economic costs of nematode infections in sheep or cattle. Apart from the exceptional circumstances caused by foot and mouth disease in Western Europe, in general the infectious disease with the greatest economic consequence in developed countries is mastitis. The relative importance of various dairy health conditions to the US dairy industry has been subjectively ranked by Wells *et al.* (1998). This ranking of the infectious diseases that were considered is presented in Table 1.1. This table also indicates the multitude of ways that the importance of diseases can be ranked.

Table 1.1 *Relative subjective ranking of infectious diseases to the US dairy industry.*

Disease or Pathogen	Production Loss	Zoonotic Potential	International Trade	Animal Welfare
Mastitis	+++		+	
Salmonella spp.	+	++	+	
Johne's disease	++	?	+	+
Bovine viral diarrhoea	++		+	+
Cryptosporidium parvum	+	++		
Campylobacter jejuni		++		
Mycobacterium bovis		+	+	
Bovine leucosis virus	+		+	
Brucella abortus		+	+	
BSE		A	+	+
Blutongue			+	
Vesicular stomatitis	+		+	

A: Curiously, BSE was not listed as a zoonotic threat in this paper in 1998.

Diseases of importance to developing countries differ markedly from those of importance to developed countries. Perry *et al.* (2002) ranked livestock disease in terms or their impact upon the poor in developing countries. The top 'diseases' (with relative rankings) were: helminthosis (100), neonatal mortality (75), foot and mouth disease (71), ectoparasites (61), liver fluke (60), reproductive disorders (58), Newcastle disease (52), anthrax (52), *Toxocora vitulorum* (52), nutrition/micronutrient deficiency (51), haemorhaggic septicaemia (46), peste-des-petits ruminants (43), *Brucella abortus* (43), haemonchosis (43), respiratory complexes (38), trypanosomosis (37), mastitis (26) and coccidiosis (26). The top ten diseases for each of the major livestock species are given in Table 1.2. Alarmingly, the top health problems for donkeys are wounds and injuries, presumably inflicted by disenchanted owners.

Table 1.2 *Top-ranked diseases according to their impact on the poor, by species* (from Perry *et al.*, 2002).

	Cattle	Sheep/Goats	Poultry	Pigs
1	Foot & Mouth dis.	Helminthosis	Newcastle disease	Ecto-parasites
2	Nutritional deficiency	PPR[2]	Helminthosis	Helminthosis
3	Reproductive disorders	Haemonchosis	Coccidiosis	Swine fever
4	Haemorrhagic septicemia	Neonatal mortality	Ecto-parasites	Neonatal mortality
5	Brucella abortus	Respiratory complexes	Neonatal mortality	Foot & Mouth disease
6	Trypanosomosis	Sheep & goat pox	Fowl cholera	African swine fever
7	Liver fluke	Ecto-parasites	Infectious coryza	Cystcercosis
8	Anthrax	Anthrax	Fowl pox	Brucella suis
9	CBPP[1]	Liver fluke	DVE[3]	Trypanosomosis
10	Toxocara vitulorum	Heartwater	Nutritional defiency	Japanese B encepthalitis

1. Contagious bovine pleural pneumonia. 2. Peste-des-petits ruminants. 3. Duck virus enteritis

For many of these important diseases there is genetic variation in either host resistance or host tolerance, as described in section 1.4

1.4 Introduction to Disease Genetics

1.4.1 The breadth of disease genetics

For almost every disease that has been intensively and carefully investigated, evidence for host genetic variation in either resistance or tolerance has been found. However, often it is not clear whether the observed genetic variation is for resistance, tolerance, or a combination of both. Almost certainly there will be genetic variation in resistance or tolerance for a wide variety of additional diseases.

A partial summary of diseases for which there is documented or strong anecdotal evidence of genetic variation in host resistance or tolerance for the major domestic livestock species is given in table 1.3 For brevity, genetic variation in resistance to several genera of nematode parasites from independent populations and studies are collapsed into a single category, *viz.* helminthosis. This category also includes haemonchosis which is listed as a separate disease in Tables 1.2 and 1.3, although strictly speaking it is also a helminthosis.

1.4.2 Mechanisms of disease resistance

Many processes may control tolerance or resistance. For example:

- The animal may have an appropriately targeted immune response against the pathogen. This may enable the animal to successfully combat the infection or avoid pathogenic effects of disease
- The animal may have non-immune response genes that preclude infection, or limit infection in target organs. Examples include the binding protein genes that allow specific strains of *E. coli*, e.g. those with K88 or F18 fimbriae, to attach to the gut of the pig resulting in pre-weaning or post-weaning diarrhoea. Absence of appropriate binding proteins results in resistance to infection.
- The animal may have physical attributes which make infection by the pathogen difficult. An example is the role of skin thickness in conferring resistance to ticks.
- The animal may have behavioural attributes which enables it to avoid infection. An example is the hygienic behaviour of honeybees. This is their dominant natural defence against brood diseases such as American Foulbrood, Chalkbrood and, possibly, the Varroa mite.

Genetic resistance to infection or tolerance of infection may, in some cases, be due predominantly to a single gene. Examples include resistance to *E. coli* diarrhoea in pigs (described above) and the PrP gene which is associated with resistance to scrapie in sheep. In other cases, resistance or tolerance is due to the combined effects of several or many genes. Examples of this include nematode resistance in ruminants, mastitis resistance in dairy cattle and sheep, and trypanotolerance in locally adapted breeds of cattle.

Although resistance and tolerance are sometimes qualitative phenomena, as in the case of *E. coli* diarrhoea in pigs, more often they are quantitative phenomena. That is, they show continuous variation from one extreme to the other. Continuous variation is seen when resistance or tolerance is due to the combined effects of several genes.

Disease resistance tends to be specific to particular diseases rather than general to all diseases. This is because genetic resistance to different diseases is controlled at least to a major degree by different mechanisms. Thus, different genes will generally influence resistance to unrelated diseases. However, similar genes may influence resistance to closely related infections. An example is the often-observed strong genetic correlation between resistance to different species of gastrointestinal nematode parasites.

Table 1.3. *Diseases for which there is documented evidence or strong anecdotal evidence of genetic variation in host resistance or host tolerance.*

Infectious agent	Disease	Comment
Chickens		
virus	Marek's disease	MHC resistance gene used in breeding programs
virus	Infectious laryngotracheitis	Inbred lines differ in resistance. MHC genes involved
virus	Avian leukosis	Single dominant gene affects resistance
virus	Infectious bursal disease	Breed differences in resistance
virus	Avian infectious bronchitis	Inbred lines differ in resistance
virus	Rous sarcoma	Selection for resistance successful. MHC genes involved
virus	Newcastle disease	Heritable differences in resistance long established
bacteria	E. coli	Lines selected for vaccine response
bacteria	Pullorum	Inbred lines differ in resistance
bacteria	Fowl typhoid	Inbred lines differ in resistance
bacteria	Salmonellosis	Well established resistance effects
bacteria	Campylobacter	Line differences observed
protozoa	Coccidiosis	Lines differ in resistance. MHC genes involved
nematode	Ascaridia galli	Differences between commercial lines in resistance
Pigs		
virus	African swine fever	Individual animals survive epidemics
virus	Foot and mouth disease	Anecdotal breed effects. Possibly tolerance
virus	Atrophic rhinitis	Within and between breed variation in resistance
virus	Pseudorabies	Significant resistance QTL detected
bacteria	Neonatal diarrhoea	Major gene conveys complete resistance
bacteria	Postweaning diarrhoea	Major gene conveys complete resistance
Cattle		
virus	Foot and mouth disease	Anecdotal breed effects. Possibly tolerance
virus	Bovine leukaemia	Strong evidence for MHC effects
bacteria	Paratuberculosis (Johnes)	Within-breed resistance variation. Nramp is candidate
bacteria	Mastitis	Genetic effects subtle but well established
bacteria	Tuberculosis	Within & between-breed variation in resistance. Nramp?
bacteria	Brucellosis	Within & between-breed variation in resistance.
bacteria	Salmonellosis	In vitro evidence from Brucellosis resistant cattle
bacteria	Dermatophilosis	Verified breed differences
bacteria	Cowdriosis (heartwater)	Historical field observations only
protozoa	Trypanosomosis	Tolerant breeds exist
protozoa	Theileria (*Theileria annulata*)	Breed differences, possibly tolerance
protozoa	Theileria (*T. sergenti*)	Combination of resistance and tolerance
protozoa	East Coast Fever (*T. parva*)	Breed differences, possibly tolerance
protozoa	Babesia	Breed differences, tolerance
nematodes[1]	Helminthosis	Combination of resistance and tolerance
ticks	Ticks	Combination of resistance and tolerance
Sheep and Goats		
prion	Scrapie	Variation in resistance dominated by a known gene
bacteria	Footrot	Variation in resistance & vaccine response
bacteria	Mastitis	Genetic effects subtle but well established
bacteria	Paratuberculosis (Johnes disease)	Anecdotal evidence of within breed variation
bacteria	Dermatophilosis	Selection for resistance successful in Merinos
bacteria	Salmonellosis	Heritable within-breed resistance. Nramp implicated.
bacteria	Cowdriosis (heartwater)	Breed tolerance
protozoa	Trypanosomosis	Tolerant breeds exist
nematodes[1]	Helminthosis	Combination of resistance and tolerance
plathyhelminths	Liver fluke (*Fasciola hepatica*)	Combination of resistance and tolerance
flies	Cutaneous myiasis (flystroke)	Selection for resistance successful in Merinos

1. Resistance to many species of nematodes has been described.

1.4.3 Exploiting genetic variation in disease resistance

Exploiting genetic differences in resistance, i.e. using host genetic variation to reduce the transmission of infection has the benefit of reducing the incidence of infection in the target population. This reduces the effects of disease and may also reduce transmission of infection to other populations. Quantifying this phenomenon is a major focus of these notes.

Using genetics to assist in the management of the symptoms of disease, i.e. genetically increasing the **tolerance** of a population to infection, will be effective in reducing the incidence or the effect of disease in the target population. However, it may not decrease the prevalence of the parasite. Hence, the disease incidence in other populations in the same environment will not be affected. In worst-case scenarios, the presence of infection in the environment may be masked. For example, in the case of zoonoses, using genetics to manage tolerance of infection would be unwise as it may lead to human disease threats being unrecognised.

A critical issue is the need for disease resistance genes or QTL to assist in selection for disease resistance. This issue will be returned to later in these notes, however phenotypic measurement of animals' responses to an infectious challenge can suffice under many circumstances. Such measurements must be informative as (i) whether or not an individual animal has been challenged, (ii) the degree to which it has been challenged, (iii) the response of the animal to the challenge and (iv) the degree of disease suffered by the animal.

Clearly, actual disease resistance genes or QTL have the benefit of enabling the selection of animals without exposing them to infection. For many diseases molecular techniques may be the only way of enabling effective selection for disease resistance. For others, such as mastitis or nematode parasites, a combination of phenotypic measurement and genetic markers may be the optimal approach.

1.5 Synthesis

A brief summary of concepts necessary to understand disease genetics has been presented. The skills required to understand disease genetics and implement programs to genetically improve disease resistance are multi-disciplinary. Both the genetics of the host animal and the nature of the disease must be understood.

Implementing breeding programs requires decision-making at the outset: will the benefits of breeding for resistance to a particular disease be worthwhile? This may be broken down into components:

- Can the genetic improvement program be combined with other disease control measures?
- Will the incidence and severity of the disease be reduced sufficiently to make the breeding program worthwhile?
- Will animal suffering be alleviated, with consequent improvements in performance, output and profitability?

Once the decision to breed for resistance to a particular disease has been made, the breeding program has to be implemented and monitored, again taking account of the dynamics of the disease.

All these steps require integration of animal genetics and disease dynamics, hence require knowledge and skills over and above those normally required in genetic improvement programs. In particular, knowledge is required of the disease epidemiology and correct decisions often depend upon taking joint account of host animal genetics and disease epidemiology.

These notes describe how genetics and epidemiology may be combined to enable a rigorous approach to disease genetics. In particular, emphasis will be given to genetic-epidemiological modelling as an aid to decision making. Parasite co-evolution will also be considered in order to quantify potential pitfalls of disease genetics. The aim of these notes is to give the reader an appreciation of powerful approaches to disease genetics that will enable better assessment of disease genetics opportunities, effective design of research programs and effective implementation of breeding programs.

2. Introduction to Epidemiological Modelling

2.1 Aims of Epidemiology

Epidemiology is the study of disease in populations and the study of the factors that determine its occurrence. There are five objectives of epidemiology:

- Determination of the origin of as disease whose cause is known
- Investigation and control of a disease whose cause is either unknown or poorly understood
- Acquisition of information on the ecology and biology of a disease
- Assessment of the economic effects of a disease and analysis of the costs and economic benefits of alternative control programs
- Planning and monitoring of disease control programs

The last three aims are of direct relevance to animal genetics and the objective of breeding animals for enhanced disease resistance.

2.2 The Need for Epidemiological Modelling

Quantifying and describing the processes and impacts of infectious diseases in domestic farmed animals presents a considerable challenge. Complex interactions occur between the host animal population and the pathogen, and these interactions will vary temporally, spatially and with the host genotype. Moreover, the infection and subsequent host immune-response processes are multifactorial. In order to predict the effect of any factor, including treatment strategies or genetic selection of resistant hosts, on an infectious disease at the population level, an epidemiological model is required. This model should capture and quantify the dynamics of the disease and allow the impact of various factors affecting the disease severity and epidemiology to be predicted.

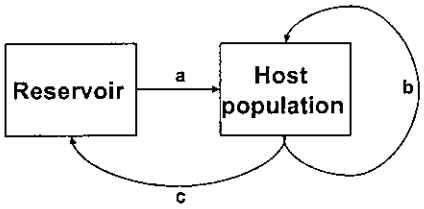
Major uses of epidemiological models have traditionally included (i) evaluating the effects of various control strategies, e.g. vaccination or management options, on the disease severity, (ii) evaluating the impact of local weather conditions on the disease severity and (iii) evaluating the impact of various chemotherapy strategies on likely evolution of drug resistance in the pathogen.

The application of mathematics to the study of infections diseases was pioneered by David Bernoulli in 1760. He evaluated the effectiveness of variolation against smallpox, hoping to influence public health policy. In 1840 William Farr fitted a normal curve to smoothed quarterly data on smallpox deaths. Hamer (1906) postulated that the course of epidemics depends on the contact rate between susceptible and infected individuals, one of the most important concepts in epidemiology. This is the 'mass action' principle where the net rate of spread of infection is assumed to be proportional to the density of susceptibles multiplied by the density of infectives. Ronald Ross (1908) translated the problem from a discrete-time model to a continuous-time framework in his work on the dynamics of malaria. Subsequently much of the theory used to develop epidemic models has been based on differential calculus used to describe **rates** of transmission or processes, or on stochastic approximations to such equations.

2.3 Modelling Approaches: General Principles

Epidemic models must quantify, in general terms, the pathways of transmission shown in simplified form in Figure 2.1. Not all pathways are relevant to all diseases, and some of the pathways ways may be considerably more complex than shown here. For example, pathway **c** may involve intermediate hosts.

Figure 2.1. *Summary of pathways of infection for diseases in domestic livestock.*



In principal, all infections must initially arise from some reservoir of infection outwith the host population. The transmission of infection from the reservoir may be 'continuous' or sporadic. Once infection is in the host population it may follow several transmission pathways. Typically, but not exclusively, microparasitic infections will be transmitted by direct animal-to-animal contact along pathway **b**, whereas macroparasitic infections will be transmitted via some external host, vector or reservoir. There are many diseases where pathway **a** is important and continual, and pathways **b** and **c** are non-existent or trivial. Examples include trypanosomosis and mastitis caused by environmental contamination. Subclinical disease challenges in intensive production systems, addressed using the concept of generalised immunity, may also fit into this framework. Diseases where pathway **b** is critical, with sporadic infection from the reservoir (pathway **a**) include viral diseases. Diseases where pathways **a** and **c** are critical, i.e. a continuous flow of infection between the host population and the reservoir, include nematode infections in ruminants, where the reservoir is the pasture, and some bacterial diseases.

Approaches to epidemic modelling vary dramatically according to the disease scenario. Models take many forms, with the main aim being to describe the disease patterns in relation to events that may change either deterministically or stochastically. Inputs into the models may be explanatory variables describing the infection process or simply empirical relationships between observed phenomena – sometimes with no strong theoretical justification. Classic applications of models using empirical relationships are the weather-based models describing helminth infection in cattle and sheep. In these models, the parasite epidemiology is calculated as a function of the prevailing weather conditions in any particular season, and inputs to the model include a database of meteorological records over many years.

A critical issue when considering epidemic modelling is whether to take a deterministic or stochastic approach. Within animal breeding contexts both approaches generally yield the same expected outcome, although the stochastic approach can yield a greater richness of output – sometimes for less stringent understanding of the underlying processes. In epidemiological modelling, outputs from stochastic and deterministic models can differ dramatically quantitatively and qualitatively. For example, minor epidemics that arise and die out without intervention are predicted by stochastic models but not deterministic models (see section 4). Both approaches have a role to play: deterministic models may give initial insight

into a problem, however stochastic models are required to represent the diversity of outcomes possible from epidemics, and quantify epidemic risks under different circumstances.

In sections 2.4 and 2.5 epidemic modelling approaches for microparasites and macroparasites are presented. These differ somewhat in their approach, due to the different host-parasite interactions and infection dynamics for the two types of parasites. It should be noted that much of the theory presented as describing epidemics as actually specific to microparasites, although this is often not stated.

2.4 Modelling Microparasitic Infections

Readers are referred to Anderson and May (1992) for a full exposition of the principles and applications of epidemic theory for microparasitic infections, and to Renshaw (1991) for the stochastic implementation of this theory. The theory concerns transmission along pathway **b** in Figure 2.1.

2.4.1 Basic principles

Most models of microparasitic infection use the so-called compartmental models, where individuals are categorised into discrete categories such as susceptible, infected, or recovered or removed. This in fact leads to the well-known Kermack-McKendrick SIR model that is described below. The two fundamental parameters of such a model are the (i) **transmission coefficient** β, which describes the expected number of new infections per infectious individual per susceptible individual per unit time (e.g. days), and (ii) the **recovery rate** γ, which is the inverse of the period for which an animal remains infectious, and is defined as the expected number of recoveries per unit time. Other forms of this type of model include SE(xposed)IR, where there is a category of latently infected animal, and SIS, where animals lose their immunity.

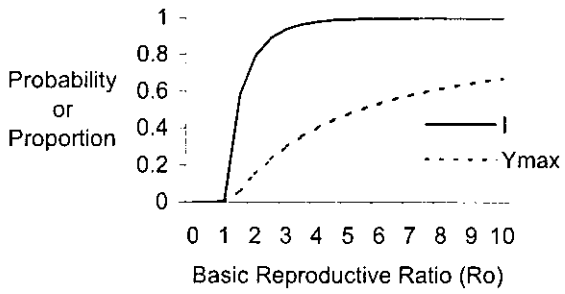
Fundamental to epidemiology is the threshold theory of Kermack and McKendrick (1927), which states that for an epidemic to occur the contact rate between susceptibles and infectives has to be above a certain threshold. This threshold depends upon the density of the population, the infectivity of the parasite and the susceptibility of the host. The threshold theory is encapsulated in the quantity R, the effective reproductive ratio of the microparasite. At stable endemic equilibrium, R=1 and each infectious individual passes on the infection to one susceptible individual. If we assume that the population is homogeneously mixed, then the number of secondary cases is linearly proportional to the probability that any one random contact is with a susceptible individual. Remember from section 1.2 that R_0 is the expected number of secondary infections caused by the introduction of a single infected animal into a population of susceptible individuals. When only a proportion (X) of the population is susceptible then the effective reproductive ratio, R, is defined as $R= R_0x$. If X^* is the proportion of the population that is susceptible at equilibrium, then $R_0X^* = R = 1$. The normal threshold principles apply: if $R<1$, the expectation is that an epidemic will not occur; if $R> 1$, then infection is expected to spread and an epidemic will occur.

In the simple SIR model in a finite population, simple relationships between parameters exist such that $\beta=R_0\gamma/S$, where S is the number of initial susceptible animals (i.e. the initial) population size (Anderson and May, 1992), giving $R_0 = S\beta/\gamma$. Two further functions associated with R_0 are the proportion of animals infected in an epidemic with no intervention (I) and the maximum proportion of animals infected at any one time (y_{max}). Under strict

conditions of homogeneous mixing amongst animals and equal susceptibility, these formulae are given below. The relationship between I, y_{max} and R_0 are shown in Figure 2.2. Essentially, for diseases with R_0 above 4, all individuals in contact with infectious animals can expect to become infected.

$$I = 1-\exp(-R_0 I)$$
$$y_{max} = 1-(1+ln(R_0))/ R_0$$

Figure 2.2. *The relationship of the probability of an animal being infected during an epidemic (I) and the maximum proportion of animals infected at any given time (y_{max}) with R_0*



A widely accepted model for the transmission of infection is the Kermack-McKendrick model, i.e. the SIR model, in which the population of constant size N comprises susceptible (S), infective (I) and removed (R) individuals. In this model animals are removed from the infectious class by recovering or dying, at a rate γ and susceptible animals are infected at a rate βSI. Because rates of change are described by differential equations, this leads to a deterministic models defined by the following differential equations for the three classes:

$$dS/dt = -\beta SI$$
$$dI/dt = \beta SI - \gamma I$$
$$dR/dt = \gamma I$$

These equations can be simply explained. β describes the number of new infections per infectious individual per susceptible individual per unit time, thus βSI describes the total number of new infections per unit time. Every new infected animal implies one less susceptible animal, thus $dS/dt = -\beta SI$. The number of recovered animals is simply the product of the recovery rate and the number of infected animals, hence $dR/dt = \gamma I$. The net change in the number of infected animals is the difference between the number of newly infected animals and the recovered animals. Note that because the population size is assumed to be constant, the sum of the differential equations is zero.

Examining the rates of change of the number of infected animals in the population, it can be seen that $dI/dt > 0$ if and only if the number of susceptible animals at time t, $S_{(t)} > \gamma/\beta$. Under the assumptions of this models where no new individuals are introduced into the population, if $S_{(0)} < \gamma/\beta$, where $S_{(0)}$ is the number of susceptible animals in the population at time 0, then $dI/dt < 0$ for all t and there will not be an epidemic.

Following simple arguments when N is large, R_0 is defined when a single infective animal is introduced into a population of S susceptible animals. R_0 will be the average number of effective contacts of the infectious animal with any other animal, i.e. $R_0 = (\beta S_{(0)} I_{(0)})/\gamma = \beta S/\gamma$ as described in 2.4.1. The condition above where $dI/dt < 0$ is simply the case where $R_0 < 1$.

The SIR model can be extended to include varying degrees of realism by, for example, introducing heterogeneity in parameters for different classes of animals, or by introducing additional parameters describing latent period, maternal antibody protection, or carrier state (as defined in 1.2). For example extending the SIR model to include age specific transmission and recovery rates will lead to the following partial differential equations: $dS/dt + dS/da = -\beta(a)S(a,t)I(a,t)$; $dI/dt + dI/da = \beta(a)S(a,t)I(a,t) - \gamma(a)I(a,t)$; $dR/dt + dR/da = \gamma(a)I(a,t)$. Although complex, these extensions are invariably needed to produce models that represent disease dynamics better. For example, young animals are often more susceptible to infections than older animals. Furthermore, extensions such as 'immigration' (e.g. young animals being born) and disease-dependent and disease-independent mortality are usually required. This can lead to the breakdown of simple results, e.g. it may no longer be true that is $S(0) < \gamma/\beta$ there will be no epidemic. In general, with more complex models, R_0 expressed as a function of the model parameters will simply be an equation where parameters describing the arrival of newly infected animals are on the numerator, and the sum of parameters describing the loss of infected animals is on the denominator (see section 7.3 for an example with recovery and different sources of host mortality).

Under simple circumstances, deterministic models of the transmission of infection can be informative. The biological processes can be written down in relatively intuitive mathematical terms. However, deterministic models can be complex to solve and they only provide point estimates of outcomes. Stochastic models can have a number of advantages insofar as they give probability density functions for various outcomes, e.g. for the numbers of animals in each class or for the probability of an epidemic.

The SIR model can be reformulated in a stochastic setting (Renshaw, 1991). Let X(t) = number of susceptible animals and Y(t) = number of infective animals at time t. In the time interval (t, t+h), for small h, we have the following transition probabilities:

For infection: $\Pr[(X,Y) \rightarrow (X-1, Y+1)] = \beta XYh$
For recovery: $\Pr[(X,Y) \rightarrow (X, Y-1)] = \gamma Yh$

Defining $p_{ij}(t)$ as the probability that there are i susceptible animals and j infective animals at time t, defining $p_{ij}(t+h)$ from the equations above, dividing both sides by h and letting h tend to zero, a differential equation $d(p_{ij}(t))/dt$ can be obtained[1]. Unfortunately, this equation is intractable, however insight may be gained by treating infection transmission as a 'birth and death' process equating infection to birth and recovery to death. With a single infective animal at time t=0 (as is assumed in the definition of R_0), then the transition rate βXY is close to $N\beta Y$, where N is the population size as usual. Setting $\gamma/\beta = \rho$, then if $N \leq \rho$ a major outbreak occurs with probability ρ/N that it is minor and dies out quickly, and $1-\rho/N$ that it is major. Immediately it can be seen that the stochastic formulation leads to results qualitatively differ from the deterministic formulation, insofar as different classes of epidemic, i.e. major and minor, are predicted. This is illustrated below.

[1] $d(p_{ij}(t))/dt = \beta(i+1)(j-1)\, p_{i+1,j-1}(t) - (\beta i + \gamma)j\, p_{ij}(t) + \gamma(j+1)\, p_{i,j+1}(t)$

Whilst the stochastic formulation of SIR and more complex models are intractable analytically, the stochastic formulation lends itself to simulation. Consider a homogeneous population of animals in random contact with each other. Random variation occurs in the time to the next 'event' and what that event actually is. In this simple case, an event will either be an animal becoming infected or an animal recovering. Thus, there are two components to a simple stochastic epidemiological model: the time until the next event and the event type.

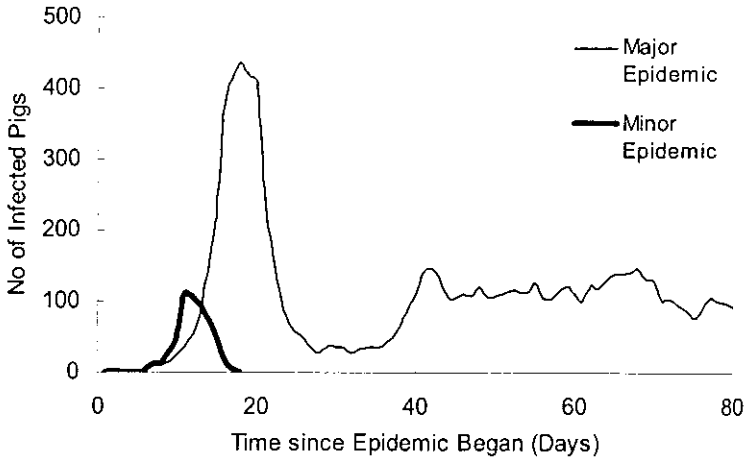
Consider firstly the inter-event time on a farm with Y infected individuals, X susceptible individuals, β and γ. The inter-event time is simply the time until the next infection. Consider an infection from which no animal ever recovers (γ=0). β is defined as above, therefore the mean number of new infections per day is βXY. If we assume that these infections occur at evenly spaced intervals throughout the day then the mean time until the next infection is 1/(mean number of infections per day) or mean inter-event time = $1/\beta XY$. However, the events will not occur evenly spaced, they will be distributed exponentially, ranging from inter-event times of zero to infinity, with a mean of $1/\beta XY$. Therefore, the inter-event time should be sampled from an exponential distribution $-ln(r)/\beta XY$, where r is a random number in [0,1]. It may be verified that this gives the correct mean inter-event time by calculating the expected value of $-ln(r)$, noting that $\int ln(r)dr = r ln(r) - r$.

Now, when γ is not zero, the total number of possible events is $\beta XY+\gamma Y$. Furthermore, with non-random mixing between animals, as we will see on most farms, the actual contact rates between animals must be defined. For example, the contact rate c_{xy} will be the probability that an infectious animal will have contact with a susceptible animal. Therefore, in this case the inter-event time is drawn from an exponential distribution as $-ln(r)/ (Y(\beta Xc_{xy} + \gamma)$, where r is a random number in [0,1].

Now, consider the next event type. In the simple stochastic model considered so far, the next event could be either that the infected animal infects a susceptible animal or that the infected animal recovers. The next event probability is determined utilising the number of possible events. The probability of any particular event is simply the number of events of that type divided by the total number of possible events. Therefore the probability that the next event is an infection is $\beta YX\ c_{xy}/Y(\beta Xc_{xy} + \gamma)$ and a recovery is $\gamma Y/Y(\beta Xc_{xy} + \gamma)$.

This simulation model is very simply implemented. It commences with the introduction a single infected animal. The inter-event time is sampled, the type of the first event is determined (simply by drawing a random number between zero and one and comparing it to the event probabilities) and the epidemic commences. The epidemic runs until either there are no more infected animals on the farm or the epidemic has lasted for some pre-determined time period, e.g. one year. Output describing the course of the epidemic may be collected and summarised across **many** (i.e. 1000s) replicates. Means and probability distributions for any outcome may then computed. An example of the output from an actual stochastic simulation of an epidemic on a pig farm is shown in Figure 2.3 (from MacKenzie and Bishop, 2001a). Shown are the number of infected pigs at any point in time for a minor and a major epidemic. Minor epidemics were simply those that died out quickly. Simulation outputs invariable show a very clear distinction between major and minor epidemics, although the actual maximum length of time of minor epidemics will depend upon the parameters for that particular infection. Note the stable endemic equilibrium that is achieved for the major epidemic.

Figure 2.3. *Example of a minor and a major epidemic, simulated on a structured pig farm.*



The stochastic simulation approach lends itself readily to increasing complexity in terms of disease parameters and heterogeneity. The heterogeneity maybe either spatial, i.e. farm structure, or genetic. More complex applications stochastic of stochastic simulation models are given in section 4.

2.4.3 Modelling Infections, under discrete time

Transition-type matrices are often useful in describing the flow of the disease in the population, especially when dealing with discrete time events. This use of matrices became established when Leslie (1945) published his so-called Leslie matrix. The host population may be divided into n discrete age classes, and the transition rates are defined by probabilities of infective animals of age class j infecting susceptible animals of age class i, at time t. This is sometimes known as the WAIFW ('who acquires infection from whom') matrix (Anderson and May, 1992).

The power of such an approach arises when there is heterogeneity in the host population, e.g. different husbandry classes, geographical differences or genetic differences, and the explicit relationships required for a simulation or mathematical approach are difficult to specify. An excellent example is that given by De Jong *et al.* (1994), describing the dynamics of a viral infection in a pig farm. A transition-type matrix, the next generation operator (M), is specified, describing the probabilities of the infection process throughout the population. It accounts for the age structure of animals in the host population, the susceptibility and infectivity of animals of each class, contact rates amongst animals, possible paths of infection through the population, etc. Of critical importance is the result that R_0 is the dominant eigenvalue of the next generation operator (M) (Diekmann *et al.*, 1990). This result can also be used to summarise outputs from stochastic simulations: in the case of spatial and genetic heterogeneity, the number of secondary infections caused by an animal with a primary infection, within different classes, can be summarised into an M matrix and R_0 then calculated (MacKenzie and Bishop, 2001).

2.5 Modelling Macroparasitic Infections

The modelling of macroparasitic infections generally involves the joint modelling of transmission along pathways **a** and **c** in Figure 2.1. Therefore, the models are generally more complex than those for microparasitic infections, and this complexity is increased by the need to account for (i) the more complex host-parasite interactions and (iii) the distribution of parasites amongst hosts. Nonetheless, the basic principles of modelling parasite dynamics using differential equations, with potential stochastic implementation of these models still hold.

2.5.1 Distribution of macroparasites

Empirically, observed patterns of macroparasite distribution amongst hosts is often described by the negative binomial distribution. This distribution is characterised by two parameters, the mean m and a parameter k that varies inversely with the degree of parasite aggregation. As $k \rightarrow 0$, the parasite population is concentrated on fewer and fewer animals, as $k \rightarrow$ infinity the distribution becomes more random (Bliss and Fisher, 1953). The probability of observing i parasites per animals p(i) is defined as:

$$p(i) = ((k+i-1)!/i!(k-1)!)\ (1+m/k)^{-k-i}(m/k)^{i}$$

Note that as k becomes large, e.g. greater than 5, p(i) converges to resemble a Poisson series. A Poisson series is given by:

$$p(i) \rightarrow m^{i} \exp(-m)/i!$$

The prevalence of infection (the proportion infected) P, calculated from the negative binomial distribution is:

$$P = 1 - (1+m/k)^{-k}$$

2.5.2 Macroparasite dynamics

Consider a simple infection where there is a constant infection rate of the host (Λ), a constant mortality rate of the adult parasite (μ) and the number of adult parasites at time t is M(t) (Anderson and May, 1992). Therefore:

$$dM(t)/dt = \Lambda - \mu M(t)$$

a solution to this differential equation obtained by integration is:

$$M(t) = M^{*}(1-e^{-\mu t})$$

Under this model the worm burden rises monotonically with time until it reaches an equilibrium value $M^{*} = \Lambda/\mu$. These results are instructive. Although the magnitude of the equilibrium M^{*} depends upon the intensity of infection, Λ, the rate of approach towards this equilibrium is dependent solely upon the life expectancy of the parasite ($1/\mu$) and is independent of Λ.

Modelling and describing fully the host-parasite interactions, transmission along pathway **a** and transmission along pathway **c** requires a series of differential equations. For example, in a model of cattle nematode parasites, Grenfell *et al.* (1987) specified no less than 10 equations,

four arguably describing transmission along pathway **c** (from egg production in the gut to the stationing of third-stage larvae on the blade of grass ready to be eaten by the animal) and six describing the various stages from ingestion and establishment of larvae to maturity, i.e. pathway **a**. Moreover, expanding the differential equations from describing the dynamics for a single individual to that of a whole population requires further complexities, as the aggregation of parasites among host complicates the definition of the differential equations.

To simply the task of modelling macroparasitic infections, authors often abandon the theoretical niceties of the differential equation approach and resort to simple procedures such as difference equations (e.g. Bishop and Stear, 1997) or empirical simulations. In the sheep nematode model of Bishop and Stear (described in section 3), the free-living stage of the parasite (parts of pathways **c** and **a**) were condensed into a single parameter, difference equations were solved to define equilibrium conditions and simulations investigated the consequences of perturbing this equilibrium.

The simpler semi-empirical simulation approaches taken for macroparasitic diseases in farm animals, predominantly models of nematode infections in sheep and cattle, may be justified in terms of the aims of the modelling. Often the aim is not to model the transmission dynamics of the disease per se, but to look at the effect of external factors on disease severity. Notably these include the effects of temperature or rainfall, where databases of meteorological records are used as inputs, the effect of anthelmintic treatment upon the evolution of anthelmintic resistance, or the effect of grazing behaviour and pasture management upon the prevalence and severity of infection. Such models often used empirical relationships developed from field data as input parameters. A summary of such models for nematode parasites is given by Bishop (1999).

2.6 Genetic-Epidemiological Models

Of particular interest to animal breeders is how epidemiological modelling and genetics may be combined to enable breeders to predict the consequences of genetically altering the resistance genotype of the host population. Although this is a rich area of innovation that is of practical utility, surprisingly few authors have attempted to do this. Epidemiologists and disease control experts have traditionally been interested in non-genetic means of disease control, whereas animal breeders have probably been unaware of the potential of epidemic modelling. N.B. the use of the term 'genetic-epidemiological models' must be distinguished from the common usage of 'genetic epidemiology' in human medicine where it generally refers to the detection of disease genes, QTL or genetic defects.

Possibly the first author to acknowledge the influence of host genotype for resistance and attempt to quantify the effects of host genotype on disease epidemiology was Barger (1989). Using the model for nematode infection in sheep subsequently published by Barnes and Dobson (1990), he demonstrated that the assumed genetic resistance of the sheep population does indeed regulate the disease epidemiology. Another major application of genetic-epidemiological models has been to scrapie, the sheep TSE. It is the scrapie model of Stringer *et al.* (1998) that is the most complete mathematical model to date explicitly incorporating host genotype for resistance. In this model, specific alleles with Mendelian inheritance are defined to denote resistance or susceptibility, reflecting PrP genotype associations with scrapie resistance/susceptibility. The model then predicts epidemics of several decades duration which, with no intervention, die out through natural selection of resistant genotypes.

Interestingly, the epidemic ends before the resistance allele reaches fixation. Woolhouse *et al.* (1998) then use this model to compare the dynamics of breeding for scrapie resistance with other control strategies – demonstrating the power of the modelling approach.

Subsequently, these notes will detail the application of genetic-epidemiological models to macro- and microparasitic infections, as well as the special (parasite-independent) case where transmission along pathway **a** outweighs all other transmission pathways.

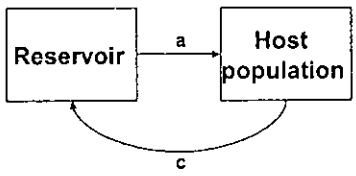
3. Application of Genetic-Epidemiological Models to Macroparasitic Infections

3.1 Introduction

Epidemiological models of macroparasites should be built up from a combination of (i) sound epidemiology theory describing the transmission of infection and the major host-parasite interactions and (ii) field data to parameterise the models. Such models should able to describe the distribution of parasites amongst hosts, the severity of infection and the impact upon the host and, most importantly, the impact of intervention strategies to control the infection. Clearly, genetic epidemiological models should be able to quantify the impact of altering host genotype upon the severity and prevalence of infection. A well-formulated model should also highlight issues requiring further research.

In general, macroparasitic models should take account of transmission along pathways **a** and **c**, as shown in Figure 3.1. Direct host-to-host transmission is generally absent. Additionally, the amount of infection present in the reservoir is often quantifiable. This is the case for nematode infections in ruminants, where the reservoir is simply the pasture that the animal grazes.

Figure 3.1. *Pathways of infection to be modelled for macroparasitic infections.*



3.2 Application to Nematode Parasite Infections of Sheep

Excellent descriptions of non-genetic epidemiological models applied to ruminant nematode infections are provided by Grenfell (1988), Grenfell *et al.* (1987), Grenfell *et al.* (1995) and Smith and Guerrero (1993). In this section a slightly simpler model developed by Bishop and Stear (1997 and 1999) will be described. For full details, refer to these two papers.

3.2.1 Developing the methodology

Defining equilibrium conditions

Recall from section 2.5.2 that a population of parasites within a host will approach an equilibrium population as follows: $M(t) = M^*(1-e^{-\mu t})$. Now consider the true life cycle of the parasite. Three major host-parasite interactions may be defined: (i) *establishment* (E), i.e. the proportion of ingested larvae that survive to become adult parasites (ii) *fecundity* (n), the number of eggs laid per adult worm per day (assuming a constant sex ratio) in the absence of density-dependent constraints; and (iii) *mortality* (μ), the daily death rate of the parasite as defined above.

Under simplifying conditions, the free-living stage of the parasite on pasture may be described by the following parameters: the probability that the egg hatches and develops to

form a third stage larvae and is ingested at time t (y), and the mortality rate of the free-living stage of the parasite (z). An actual epidemic is assumed to be triggered by the ingestion of a high concentration of infective larvae (L_0). Therefore, assuming it takes j time units for ingested and established larvae to reach maturity and assuming continual larval challenge, the parasite population within a single host at time t is:

$$M_{(t)} = M_{(t-1)}(1-\mu) + L_{(t-j)}\,E$$

where m is the parasite mortality rate. In other words the parasite population on day t is the population on day t-1, less the parasites that have died, plus the new parasites that were ingested j days ago, became established and have taken j days to reach maturity.

And the larval challenge at time t (where n is the parasite fecundity) is:

$$L_{(t)} = L_0(1-z)^t + \sum_{i=1}^{t} M_{(i)}\, ny(1-z)^{t-i+1}$$

In other words, larval challenge on day t is the original larval challenge on day zero, less the larvae that have died in the interim, plus the larvae originating from eggs laid by parasites within the host, summed across days.

Over time, the larvae originating from the original larval contamination will tend to zero. At this stage, substituting the larval challenge equation into the worm burden equation, an equilibrium worm burden M* can be defined that holds true whilst (after some algebra):

$$\begin{aligned}\mu &= nEy/z \\ &= nEp\end{aligned}$$

where p is the overall probability of an egg hatching and developing to third stage larvae (the form of larvae that migrates on the grass sward and is eaten by the animal) and being ingested before it dies. It is important to note that this equilibrium condition is not dependent upon the larval challenge level.

If this equilibrium relationship breaks down, then in the absence of density-dependent constraints (described below), the parasite population will approach either extinction or a runaway epidemic. It is assumed that parasite-induced mortality does not exist in this population. Note that the above equations assume a continuous cycle of infection and re-infection. In other words, eggs shed by infected animals will, in time, re-infect the same animals.

Density dependent constraints

The above model is very sensitive to changes in parameter values such that $\mu = nEp$ no longer holds, and also to perturbations in the system, such as treatment of animals with an anthelmintic. In reality, parasite populations are rather stable to perturbations, tending to return towards the pre-perturbed value. One mechanism governing this are the so-called density dependent effects, whereby the value of any of the rate parameters (i.e. E, n or μ) is dependent upon the parasite population already present. Strong density dependent constraints on worm fecundity are demonstrated by Bishop and Stear (2000), in which total egg output increases with increasing worm burden, plateaus and eventually declines. Per capita fecundity rises dramatically as the worm population declines. This may be modelled by redefining fecundity as:

$$n_{(t)} = kM_{(t)}^{b}$$

where 0>b>-1.

Introducing between-animal variation
The theory described so far describes the dynamics for a single animal. The model must describe the dynamics of a population of animals, in order that output variables such as worm burden and faecal egg output show the expected aggregated between-animal distributions. Therefore, within a simulation model representing the processes described above, all time-dependent variables such as larval ingestion, worm burden or egg output at time t may be redefined as vectors of variables. Similarly, parameter values should be redefined as distributions of parameter values. The forms of the distributions for establishment, fecundity and mortality are generally unknown, although information is now available for parasite fecundity (Stear *et al.*, 1997). Assumptions made by Bishop and Stear (1997) were that establishment and mortality were normally distributed and fecundity, being a biological count, had a Poisson distribution (mean = $n_{(t)} = \kappa M_{(t)}^{b}$).

Further between-animal variation exists for daily larval ingestion. Larval ingestion will be proportional to food intake (see below), but will also have a sampling component. Under assumptions of random grazing a Poisson process describes the distribution of larvae between random bites. However, summing over many bites and days of grazing the Central Limit Theorum will dictate that larval intake is normally distributed.

Food intake and growth rate, under conditions of no infection, may be defined as correlated traits of the host in the manner familiar to animal breeders. However, infection has an impact upon the growth rate of the host. Biologically, this impact appears to be a function both of the total mass of worms present in the gut as well as the larval challenge that the animal faces (presumably as a deleterious consequence of immune system activation). The methodology of Leathwick *et al.* (1992) was used to partition reductions in growth rate between larval challenge and worm mass. Following consensus literature values, reductions in growth rate of 25% were chosen for a 'standard' larval challenge, with anthelmintic treatment restoring 33% of this live weight loss.

Partitioning between-animal variation
So far, the model contains no genetic effects. These may be introduced under any chosen genetic model (infinitesimal, finite locus, major gene, etc), by partitioning the between-animal variation into genetic and environmental (temporary and permanent) components. Bishop and Stear (1997 and 1999) chose simple infinitesimal genetic models, with the genetic properties simply defined by the assumed heritabilities of each trait. The model to partition the between animal effects for each trait was:

$$P_{i(t)} = G_{i(t)} + C_{i(t)} + E_{i(t)}$$

Where $P_{i(t)}$ is the time-dependent phenotype, $G_{i(t)}$ is the time-dependent genetic effect, $C_{i(t)}$ is the time-dependent permanent environment effect (defined as a litter effect) and $E_{i(t)}$ is the time-dependent residual term (temporary environmental effect). These effects were sampled from normal distributions defined by trait means, CVs, heritabilities and assumed litter effects (which decline to zero as the lamb ages). For parasite fecundity, individual animal effects were sampled from a Poisson distribution. In the implementation (see below), genotypes were

tracked through the generation according to the infinitesimal model, adding in Mendelian sampling terms.

Heritabilities for live weight and food intake are well known. Less well quantified is the genetic control of the three host-parasite interactions, establishment, fecundity and mortality. However, results from Stear *et al.* (1997) imply that parasite fecundity as a trait of the host is highly heritable at six months of age whereas establishment, as inferred by the number of larvae and mature worms present in the gut, is lowly heritable. Additionally, the results of Bishop *et al.* (1996) suggest that the heritability of faecal egg count rises with age as the hosts' acquired immunity develops. Therefore, the heritability of worm fecundity must also be time-dependent, rising with age.

Defining larval challenge levels

The model is driven by the larval challenge that the population of hosts receives. The initial challenge will determine the equilibrium worm population and hence the overall impact of the infection. Suitable challenge levels can be deduced by parameterising the model such that output traits such as worm burdens, faecal egg counts and growth rates reflect observations from field data.

However, an issue arises relating larval challenge levels at the start of a grazing season with those present at the end of the previous grazing season. For an unselected population of animals this is not a problem, as equilibrium is assumed. However, if genetic progress is introduced then egg output will change, affecting pasture larval contamination and this must be accounted for when defining challenge levels. In general terms, an arbitrary function f(L,w) may be used to define mean pasture larval contamination at any point in time as a function of previous contamination, when there has been a gap in the grazing, e.g. from the end of one season to the start of the next. These 'carry-over' effects are such that if w=0, the initial infection rate at the start of each season or on each new pasture is constant, whereas if w=1, the initial infection rate is directly proportional to (e.g.) the previous season's larval count.

This part of the methodology also allows for input of other factors, such as the impact of weather conditions or grazing management upon larval development on the pasture.

3.2.2 Implementing the methodology

The methodology described above was implemented as a semi-stochastic simulation model. The host-parasite interactions and time course of the infection were simulated deterministically, i.e. each event took place on predictable number of days post infection by larvae. The time-dependent simulation of the host-parasite interactions was simulated on a daily basis, with each animal simulated independently. Initial larval challenges, mean parameter values and coefficients of variation for each parameter distribution were chosen such that output variables matched field observations for unselected animals.

Genetic effects were simulated by imposing a pedigree relationship structure upon the animals, with the population demography of the ewe flock matching those typically seen in UK sheep flocks. This enabled genetic parameters to be estimated for any output traits (flock size of 10000 lambs), for any combination of input parameters. Additionally, the flock genetic structure and demography enabled realistic selection (flock size of 1000 lambs) to be simulated for any trait of interest.

3.2.3 Results

Following are a selection of results taken from Bishop and Stear (1997 and 1999). For unselected animals the benchmark set of parameters resulted in a mean achieved live weight of 33.4 kg at 6 months of age, a corresponding mean live weight loss (compared to the live weight that would have been achieved under conditions of no challenge) of 6.6 kg, and a mean faecal egg count (FEC) of 231 eggs/g. Corresponding heritabilities were 0.17 for achieved live weight, 0.29 for FEC and 0.38 for live weight loss. Apart from the last value, which is unknown, these output parameters reflect known results from field data.

A key issue for animal breeders is the genetic relationship between resistance and performance. This will be returned to more formally in section 5. However, in these simulations the genetic and phenotypic correlations between FEC and live weight were –0.27 and –0.10, respectively. The correlation between achieved live weight and resilience, or live weight loss, places an upper bound on the correlation between FEC and achieved live weight – this is the correlation that would be observed if all variation between animals in live weight gain was a consequence of the infection. This so-called upper bound correlation is therefore the negative value of the correlation between live weight loss and FEC, i.e. -0.54. Genetic and phenotypic correlations between food intake and FEC were –0.07 and –0.04, respectively. Note that no direct feedback mechanisms between food intake and worm burden has been assumed, i.e. disease-induce inappetance has not been assumed.

The influence of pasture larval contamination on achieved live weight and the relationship between productivity and FEC are shown in Table 3.1, for benchmark, low and very high levels of contamination. This gives an indication of the impact of disease epidemiology on genetic relationships. As pasture larval contamination increases, production losses increase and the genetic and phenotypic correlations between achieved live weight and FEC increase. However, the upper limit correlations decrease, as relatively more loss is due to larval intake than worm burden. Even under conditions of heavy infection the phenotypic correlation between FEC and achieved live weight is weak. Increasing the pasture larval contamination has a negligible effect of the heritabilities for FEC and achieved live weight, but it decreases the heritability of resilience.

Table 3.1. *Effect of pasture larval contamination levels on the genetic relationships between productivity, i.e. achieved live weight (kg), and faecal egg count†*

Relative pasture contamination:	Quarter	Benchmark	Quadruple
Mean achieved live weight (kg)	37.3	33.4	22.50
Observed genetic correlation	-0.16	-0.27	-0.40
Observed phenotypic correlation	-0.07	-0.10	-0.14
Upper limit genetic correlation	-0.65	-0.54	-0.41
Upper limit phenotypic correlation	-0.32	-0.24	-0.16
Live weight loss heritability	0.44	0.38	0.34

†Heritabilities for FEC and achieved live weight are 0.29 and 0.17, respectively for all scenarios

The effects of altering the absolute penalties on production losses and relationships between traits are shown in Table 3.2. The production penalties define the mean resilience of the flock. Increasing the production penalties increases the live weight loss, and it also strengthens the correlations between achieved live weight and FEC. The patterns of change are similar to those observed when the larval challenge was altered. However, in this case the upper limit

correlations remain unchanged as there is no intrinsic change in the model, other than a linear transformation of the output values for each animal. Therefore, the heritabilities for FEC and resilience are not affected by varying the absolute penalties, and the heritability for achieved live weight only shows a trivial change.

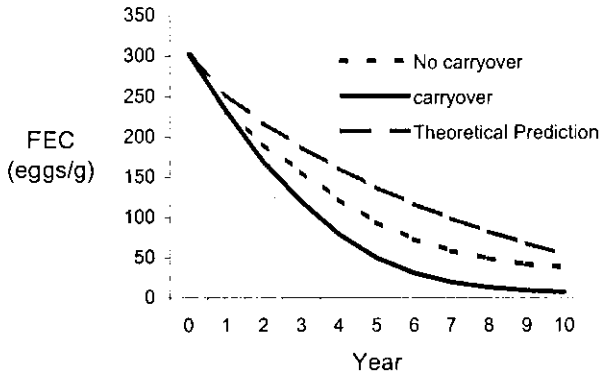
Table 3.2. *Effect of altering the absolute magnitudes of the worm mass and larval intake production penalties on the genetic relationships between productivity and faecal egg count†*

Absolute production penalty:	Half	Benchmark	Double
Mean achieved live weight (kg)	36.7	33.4	26.8
Observed genetic correlation	-0.15	-0.27	-0.45
Observed phenotypic correlation	-0.06	-0.10	-0.16
Upper limit genetic correlation	-0.54	-0.54	-0.54
Upper limit phenotypic correlation	-0.24	-0.24	-0.24
Live weight loss heritability	0.38	0.38	0.38

† Heritabilities for FEC and achieved live weight are 0.29 and 0.17, respectively for all scenarios

Responses to selection for decreased FEC are shown in Figure 3.1. The scenario of 'no carryover effects' assumes that the lambs graze different but equally contaminated fields each year. The 'carryover effects' scenario defines the larval challenge at the start of each season to be proportional to the mean challenge on the field at the end of the previous season. The 'Theoretical Prediction' line represents selection responses expected for an aggregated trait (i.e. back transformed responses on log-transformed values), ignoring disease epidemiology. Selection responses are large and curvilinear, with responses in year 1 for both scenarios being 1.4 times that predicted by quantitative genetic theory. This is due to the positive feedback loop whereby more resistant lambs deposit fewer eggs onto pasture which, in turn, results in a lower subsequent larval challenge to the lambs. Imposing carryover effects increases the selection responses from year two onwards.

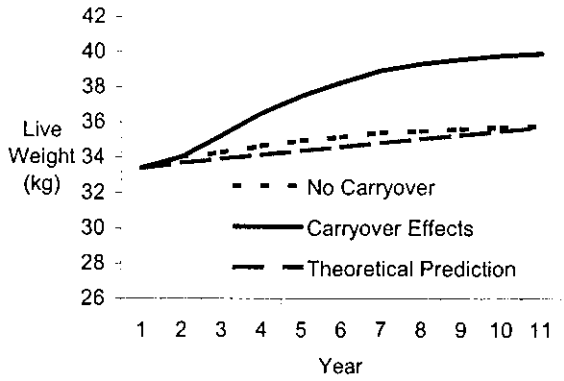
Figure 3.1. *Responses to selection for reduced faecal egg count under various grazing management assumptions ('carryover' and 'no carryover') or ignoring disease epidemiology.*



Correlated responses in live weight gain, when selection is practised on FEC, are shown in Figure 3.2. Selection for reduced FEC results in immediate responses in live weight gain which, again, are larger than predicted by quantitative genetic theory. In year 1, the observed response is 0.655 kg, compared to an expected response of 0.30 kg, i.e. 2.2 times expectation.

If grazing strategies can be devised to exploit the epidemiological benefits of reducing pasture contamination (the 'carryover effects' scenario), then large correlated responses in live weight gain are predicted in subsequent years.

Figure 3.2. *Correlated responses in live weight when selection is for reduced faecal egg count under various grazing management assumptions ('carryover' and 'no carryover') or ignoring disease epidemiology.*



Bishop and Stear (1999) also simulated selection for increased live weight, using the same genetic epidemiological model. In summary, there was a large asymmetry in the correlated responses to selection, depending on whether selection was for reduced FEC or increased live weight. Selection for reduced FEC resulted in large and dramatic epidemiological effects that increased the apparent responses to selection and the correlated responses in achieved live weight. Exploiting the carry-over effects resulted in large increases in achieved live weight. Conversely, selecting for increased live weight resulted in direct and correlated responses which were in line with theoretical expectations. Over time, epidemiological benefits could be exploited, but the effects were much more modest than when selection was for FEC.

A final point to note was that selection also caused the distribution of egg counts to become more aggregated, with a significant decrease in the negative binomial parameter k over time.

3.2.4 Generalisations and Extensions

These results illustrate general features of the impact of host genotype upon the epidemiology of macroparasite epidemiology. In general terms, there is a large impact of host genetics upon disease epidemiology and vice versa. For example, in this situation altering host resistance genotype will change disease epidemiology with a positive feedback, such that apparent extra benefits of selection can arise. (Note that these extra benefits may be invisible insofar as they affect all animals in the flock, not just the selected animals). Conversely, the disease epidemiology, e.g. the level of larval challenge, will affect the genetic relationships between resistance and performance traits.

These observations may be generally applicable to diseases where transmission of infection along both pathway **a** and pathway **c** is important, i.e. there is a positive feedback loop in the infection process. If pathway **a** dominates, i.e. the reservoir of infection is essentially infinite, then the feedback mechanism is no longer important and other approaches (as outlined in section 5) are more appropriate. However, aspects of the approach undertaken here may well

be appropriate for some bacterial diseases where a finite reservoir of environmental infection (e.g. from unsanitary conditions in a pig house) may exist.

Numerous extensions may be made to the methodology outlined here, to include additional environmental information and sources of infection, and also to include other control measures. For example:

Treatment effects
The effect of anthelmintic treatment upon the infection dynamics was described by Bishop and Stear (1997). In general terms, such treatment does not affect the dynamics of selection responses. In principle, models such as this can be used to assess the impact of any form of treatment. Importantly, however, models similar to that described here (minus the genetic component) have been used to model the development of anthelmintic resistance amongst the parasite population. Indeed, this has been one of the major rationales for modelling nematode infections.

Culling of infected animals
Given the aggregated distribution of parasites and egg counts, culling of heavily infected animals seems a simple and attractive option to bring about an immediate reduction in pasture contamination. This has been investigated by Vagenas (2002). The efficacy of culling depends critically upon the repeatability of faecal egg count and the aggregation of the distribution. Unfortunately, for realistically low repeatabilities of egg counts (in the range 0.2 to 0.5), the immediate benefits of culling appear to be somewhat modest.

Nutritional treatments
An alternative means of dealing with parasite infections is through enhanced nutrition of susceptible animals at key times. The impact of such treatment on the interrelationships between host genotype and disease epidemiology has been investigated by Vagenas and Bishop (2002). The outcomes are somewhat similar to the effects of altering the level of larval challenge: increased protein nutritional enhances performance and resistance, and weakens genetic relationships between resistance and performance.

Other host-parasite interactions
In addition to growing lambs, ewes during the late gestation and early lactation (the 'peri-parturient period') are also susceptible to nematode parasites (e.g. Bishop and Stear, 2001). Although, this is not a production problem *per se*, it is a significant source of pasture contamination, i.e. it helps to trigger the lamb epidemic. This phenomenon appears not to have been modelled as yet.

Estimating economic weights
A difficulty often encountered with disease resistance traits is assigning an appropriate economic weight for use in a selection index (including multi-trait BLUP) context. In principle, the epidemiological consequences of increasing resistance to a particular disease serve as a starting point for assigning as economic weight to a resistance trait. This concept will be returned to in section 8.

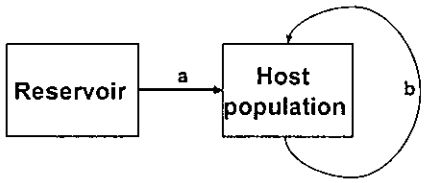
4. Application of Genetic-Epidemiological Models to Microparasitic Infections

4.1 Introduction

As outlined in section 2, epidemic theory and techniques applied to microparasitic infections tends to be more rigorous and less empirical than the theory and techniques applied to macroparasitic infections. However, the parameters required to drive microparasitic models are considerably harder to estimate. In fact actual values of, for example, the transmission coefficient of the recovery rate for some diseases may only be speculated. Additionally, the genetic properties of many of these parameters will probably often have to be guessed at. Experimentation to estimate the parameters and the genetic properties of the parameters will be expensive and imprecise. Nevertheless, considerable insight can be gained from exploration of the impacts of altering each parameter, within a reasonable parameter space.

In general, microparasitic models should take account of transmission along pathway **b**, as shown in Figure 4.1. Pathway **a** can be considered a sporadic one-off event. However, there are circumstances (not covered in these notes) when pathways **a** and **c** may also need to be quantified. Generally, microparasitic models involve SIR models, or extensions of SIR models that take fuller account of the disease biology. Moreover, the aggregation of parasites among hosts is usually ignored.

Figure 4.1. *Pathways of infection to be modelled for microparasitic infections.*



Microparasitic epidemic models have generally been constructed to look at the impact of spatial heterogeneity upon disease transmission, or the impact of various intervention strategies such as vaccination or culling. Incorporating genetics into these same models can be considered to be general extension of such approaches.

4.2 Deterministic Genetic Epidemiological Models for Microparasitic Infections

Apart from recent work by Detilleux on bovine mastitis, there appear to have been two deterministic genetic-epidemiological models published for microparasitic models, the scrapie model of Stringer *et al.* (1998) and the generic discrete-time model for viral infections in pigs of Mackenzie and Bishop (1999). Curiously, these two models were developed simultaneously yet independently by two Scottish-based teams of researchers situated less than two miles apart in the Midlothian countryside. (However, these two groups have since communicated with each other!).

4.2.1 Scrapie genetic-epidemiological model

Full details of the derivation of he scrapie model are given by Stringer *et al.* (1998), and extended interpretation is given by Woolhouse *et al.* (1998). Apparent susceptibility to scrapie

controlled (predominantly) by PrP genotype, with specific combinations of variability at codons 136, 154 and 171 jointly determining the relative susceptibility of sheep to scrapie. Scrapie susceptibility alleles may be defined by the combination of variants at these three codons. Using standard abbreviations for the amino acids coded at each codon, the five known alleles in sheep are: ARR, AHQ, ARH, ARQ and VRQ. Pairwise combinations of these alleles are referred to as the PrP genotype and to date 15 out the possible 25 genotypes have been identified in sheep. Stringer *et al.* (1998) simplified this down to three genotypes: rr, rR and RR. In addition to host genotype, their model incorporated flock demographics, horizontal and vertical transmission and variable initial load of infectious agent. This was incorporated into a series of partial differential equations, which were analysed numerically. Their model predicts scrapie epidemics to be of long duration, and, importantly, natural (or artificial) selection will lead to scrapie eradication before the resistance allele reaches fixation. However, disease eradication takes many years. Interestingly, the Woolhouse *et al.* (1998) extension to the model includes an environmental reservoir of infection, which adds to the duration of the epidemic.

4.2.2 Discrete-time viral model

The first model bringing microparasitic epidemiological concepts to an animal breeding audience was that of MacKenzie and Bishop (1999), in a generic model describing viral transmission in a structured pig farm. This paper developed a discrete-time algorithm first proposed by De Jong *et al.* (1994) for estimating R_0 in a spatially structured farm. In this type of model it is necessary to specify the next-generation operator **M**, i.e. the disease transmission-type matrix as outlined in section 2. The elements of **M** are given by:

$$m_{ij} = f_i g_j \sum_{l=1}^{n} c_{il} e_{lj}$$

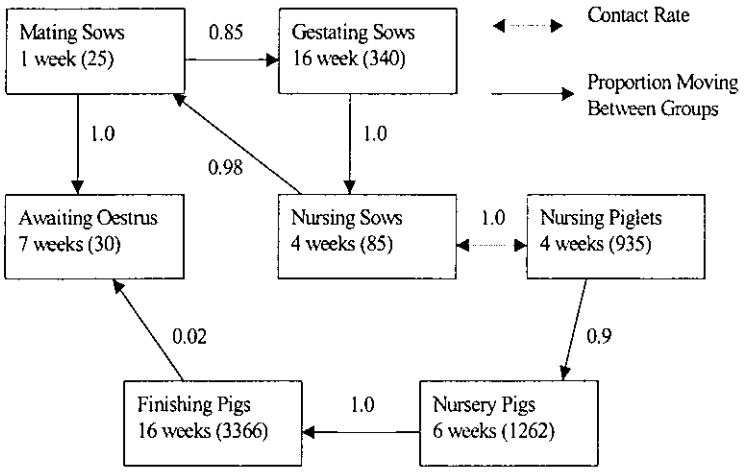
where f_i is the mean infectivity level of animals of type i, g_j is the mean susceptibility level of animals of type j, c_{il} is the contact rate between type i animals and type l animals and e_{lj} is the amount of infection spreading from a type l animal to a type j animal. Susceptibility is a function of the probability that an individual will become infected when exposed to a concentration of infectious material ξ for a period of length Δt. Therefore, integrating across time, the probability that an individual will become infected is: $1 - \exp(-\xi g_i \Delta t)$. Infectivity is similarly defined as the amount of infection spread by an infected animal per unit time. The product of susceptibility and infectivity gives the transmission coefficient (β).

Definitions of animal types and contact rates between animal types depend upon the structure of the farm, but in general terms describe discrete categories of animals such as newborns, growing animals, gestating dams, lactating dams, etc. The definition of animal type will depend upon spatial structure upon the farm, physiological status of animals as is relevant to disease resistance and, in the case of genetic progress, genetically distinct cohorts of animals. An example of different types of animals and the contact structure amongst these types is given in Figure 4.2.

Host genetic effects are now easily incorporated into this model by specifying the product fg (i.e. the transmission coefficient β) as a genetically controlled trait of the host. Because of the heterogeneity in the model, the implementation of the model requires the product $f_i g_j$, i.e. infectivity and susceptibility of different groups of animals. Therefore, it is easier as a first assumption, to specify either f or g as the variable trait, holding the other constant.

Genetics were implemented into this model either by assuming an infinitesimal model with constant genetic progress in susceptibility (g), or by assuming a major recessive gene for g, two copies which confer essentially zero susceptibility. In both cases it was assumed that genetic progress was made in a nucleus selection herd outwith this particular herd. Full details are provided by MacKenzie and Bishop (1999).

Figure 4.2 *Example pig farm structure used in the models of MacKenzie and Bishop (1999, 2001a and 2001b).*

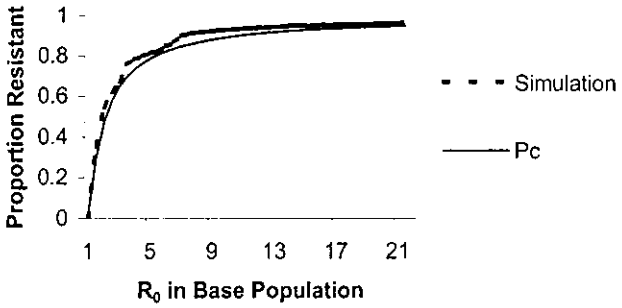


The model was implemented by tracking animals and genotypes through the farm and, at regular intervals, constructing the **M** matrix and calculating R_0. Inferences can then be made from R_0 on the likely severity of epidemics, should they occur at this point in time, using the formulae outlined in section 2. It is important to note that these inferences are only approximate, as they assume both spatial and genetic homogeneity. Additionally, however, inferences can also be made on the effect of changing other aspects such as the spatial structure of the farm.

Apart from illustrating non-linear impacts of selection for disease resistance upon the likely severity of epidemics, the most important output from this was the demonstration that it is not necessary to take resistance genes to fixation in order to protect the population from major epidemics, i.e. to reduce R_0 below 1.0. This finding mirrors that of Stringer *et al.* (1998) from their scrapie model, and is in line with the concept of herd immunity in which not all a population has to be immune to protect the population as a whole. The specific result obtained is shown in Figure 4.3, showing both the result obtained from simulation and an approximation under genetic and spatial homogeneity (Pc, as described below). The required proportion of completely resistant animals is a function of R_0 in the base population, i.e. in a population of genetically susceptible animals. For high values of R_0, essentially all animals must be resistant. But for moderate or low R_0 values, e.g. less than 4, only a proportion of animals need be resistant to protect the population as a whole. Pc shows the line $1-1/R_0$, derived from the relationship in section 2, $R_0X^*=1$. This result has previously been used to justify vaccination programs that fail to reach 100% of the target population. Potentially, the result shown in Figure 4.3 has important implications for the deployment of resistance genes or QTL, and it is developed more formally in section 6.

The strength of the discrete-time approach is that it is very simply implemented and quickly allows a variety of scenarios to be investigated. The limitation of the approach lies firstly in that it is deterministic, i.e. only point estimates of outcomes can be obtained and, secondly, that it only gives a limited range of output parameters – *viz.* parameters associated with the eigenvalue of M (i.e. R_0) and the eigenvectors of M (in theory showing the source of infection). The stochastic approach outlined in the next section has the ability to give a much greater range of information and outcomes.

Figure 4.3 *Proportion of resistant animals necessary to reduce the realised R_0 to below 1.0, showing simulation results and theoretical expectation under complete homogeneity ($Pc=1-1/R_0$)*



4.3 Stochastic Genetic Epidemiological Models for Microparasitic Infections

4.3.1 Developing the Methodology

Definition of parameters

The following methodologies and techniques are summarised from MacKenzie and Bishop (2001a and b) and MacKenzie (2000). As described in section 2, a stochastic model describes the infection transmission process as a series of random events in space and time, governed by parameters describing the disease biology. The two components are the inter-event time and the event type. For an SIR model, from section 2, the inter-event time is thus drawn from an exponential distribution as $-ln(r)/(Y(\gamma + \beta X)$, where r is a random number in [0,1]. Additionally, the probability that the event is an infection is $\beta YX\ c_{xy}/Y(\beta X c_{xy} + \gamma)$ and a recovery is $\gamma Y/Y(\beta X c_{xy} + \gamma)$.

We wish to apply this methodology to a structured pig farm, where there is spatial heterogeneity. The description given above can be extended to allow heterogeneity between pigs as follows. Assume there are n types (classes) of animals. An animal of type i has contact with animals of type i and type j given by a matrix with elements c_{ji}. The inter-event time is now :

$$-\ln(r)/(\gamma\sum_{i=1}^{n}Y_i + \beta\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ji}Y_iX_j)$$

where r is again a random number in [0, 1] and c_{ji} is the contact rate between type j and type i pigs. The recovery rate is γ and β is the transmission coefficient, both of which are constant for all types in this example.

To determine the next event type, the sum $\sum_{i=1}^{n}\sum_{j=1}^{n} Y_j(\beta X_i c_{ji} + \gamma)$ is calculated. If we denote this sum by RATE then the probability that the next event is infection of a pig of type i by a pig of type j is given by

$\beta X_i Y_j c_{ji}$/RATE

for all i, j and the probability that the next event is recovery of a type j animal

$Y_j\gamma$/RATE

for all j.

Epidemic probabilities

Implementing this simple model and recording the number of simulations that result in an epidemic, the probability that an epidemic will occur can be determined. Where n is the number of infected animals, if n=1 at the end of the simulation there is no epidemic; if n>1 but the epidemic dies out within a predetermined time period the epidemic is deemed minor; if n>1 and the epidemic does not die out, the epidemic is major. Illustration of major and minor epidemics was shown in Figure 2.3. The so-called 'predetermined' time will be parameter dependent, however from a small number of simulations it will become apparent that minor epidemics generally die out within a short time (e.g. 2-3 infectious periods), whereas major epidemics may last a very long time.

Estimation of R_0, and properties of the estimator

R_0 may be estimated in the same manner as for the discrete-time model, by creating the next-generation matrix **M** describing the disease transmission probabilities, from the simulation outputs. R_0 will then be the dominant eigenvalue of this matrix. For example, an epidemic may be simulated with each animal in turn being the index case. The number of secondary infections caused in type j pigs when the index case was initially type i is then stored. Thus the element m_{ij} of **M** is the number of secondary infections in type j animals caused by the index case of type i, averaging across simulations. R_0 is then calculated as the dominant eigenvalue of **M**.

Using this method, an estimate of R_0 is obtained. However, estimates of biological parameters are generally estimated with imprecision. This will be the case in an experiment that has limited replication, and the same will be true for estimates obtained from stochastic simulation. Numerical techniques may be used to quantify the variability of the estimator of R_0. To estimate the standard error of the estimate of R_0 from a complete set of simulations, bootstrapping may be applied to the disease transition matrix, **M**. The method involves repeated sampling with replacement of the data used to obtain the parameter under investigation. Thus, the standard error of the estimate of R_0 may be obtained by repeated sampling, with replacement, of the number of secondary infections caused by each type of animal. In the results presented below, one sample was drawn for each animal on the farm. The samples were used to construct a new disease transition matrix from which R_0 was estimated. This process was repeated 1000 times. The distribution of the 1000 bootstraps gives an estimate of the distribution of estimated values of R_0, the mean of the distribution being the mean estimate of R_0, and the standard deviation being the standard error of the estimate of R_0.

To investigate the distribution of estimates of R_0 obtainable from the stochastic model, as opposed to the accuracy of the estimate of the mean R_0, simulation may be performed where **M** is constructed using a single epidemic for each type of pig. Again, a large number of simulations, e.g. 1000, should be performed to provide the estimate for the distribution of the

estimate of R_0. This gives an indication of the possible range of estimated values of R_0 given a particular set of parameters. The two techniques together allow determination of the accuracy of the estimate of R_0 as well as the distribution of estimated values of R_0.

Other outputs
In essence the stochastic simulation provides a rich range of outputs. Almost any descriptive statistic may be gathered, including epidemic severity (i.e. numbers of animals infected at any time or in total), epidemic duration and the impact of epidemics starting on different parts of the farm.

Incorporating genetics
Once again, the genetic model of choice may be used, ranging from infinitesimal to major gene. In the example provided below, both models were implemented, with the transmission coefficient β chosen as the trait under genetic control. However, in principle any parameter describing a host-parasite interaction could be modelled as being under genetic control.

Simulation procedure
Simulation proceeds by incorporating genetic change into the herd, as determined by the genetic model chosen. At chosen time intervals, the herd is subjected to hypothetical epidemics, by introducing infected animals of a chosen type. Stochastic outcomes are very variable and considerable replication is required, both of the genetic structure of the population and of the simulated epidemics, themselves. For example, the results of MacKenzie and Bishop (2001a and b) are based on >5000 replicates per scenario.

4.3.2 Application to stochastic GEM to transmissible gastroenteritis
The pig viral disease Transmissible Gastroenteritis (TGE) provides an excellent example of the application of stochastic GEMs to real diseases. This disease has been previously modelled using a deterministic model (Hone, 1994), thus point estimates of all necessary parameters are available. However, genetic variation in host resistance to TGE has not been demonstrated or even investigated, therefore the genetic control of resistance in this application is an assumption. The details presented here are summarised from MacKenzie and Bishop (2001b).

The SIR model is generally inadequate for most diseases, and additional parameters describing the biology of the target disease are required, although the basic methodology remains the same. For TGE these parameters include the latent period, the disease-dependent mortality (which is age dependent) and the rate of loss of immunity in recovered pigs. Hone (1994) has previously estimated these parameters and these estimates are: transmission coefficient = 0.0007/d, recovery rate = 0.057/d, latent period = 2 d, thus the rate at which latent pigs become infectious = 0.5/d, mortality rate = 0.006/d for pigs over 4 wk of age and 0.1712/d for piglets 4 wk of age or less, and the rate of loss of immunity = 0.0031/d. The model is further extended to allow each type of pig to have a different value for the transmission coefficient, thus allowing incorporation of host genetic effects in this parameter. In this particular model, the recovery rate is assumed to be constant across all types, as are the latent period and rate of loss of immunity, although the disease-dependent mortality varies (non-genetically) according to animal type.

The possible event types for this model are: a susceptible animal becomes latently infected, a latently infected animals becomes infectious, an infected animal recovers or dies as a result of infection, or an animal that has recovered loses immunity.

For a population with Y infected animals, X susceptible animals, L animals in the latent class and Q recovered animals, the inter-event time in a TGE epidemic has a mean:

$$1/(\gamma\sum_{i=1}^{n}Y_i+\sum_{i=1}^{n}Y_i\varepsilon_i+\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_j c_{ji}Y_iX_j+\sigma\sum_{i=1}^{n}L_i+\omega\sum_{i=1}^{n}Q_i)$$

where γ is the recovery rate, β_j is the transmission coefficient for a pig of type j and c_{ji} is the contact rate between type j and type i pigs, ε_i is the disease dependent mortality rate for a pig of type i, ω is the rate of loss of immunity and σ is the rate at which latent pigs become infectious. Thus the inter-event time is drawn from an exponential distribution as $-ln(r)$x(mean inter-event time), where r is a random number in [0,1].

The next event type is calculated as follows. The sum $\sum_{i=1}^{n}\sum_{j=1}^{n}(Y_j(\beta_iX_ic_{ji}+\gamma_j+\varepsilon_j)+\sigma L_i+\omega Q_i)$ is calculated and is denoted by RATE. The probability that the next event is the infection of a pig of type i by a pig of type j, moving that pig to the latent class, is given by $\beta_iX_iY_jc_{ji}$/RATE for all i,j. The probability that it is the movement of a latent pig to the infectious class is given by σL_i/RATE for all i and that the next event is recovery of a type j animal is γY_j/RATE for all j. The probability that the next event is the death of a type j pig is ε_jY_j/RATE for all j and that it is the loss of immunity of a previously infected pig is ωQ_i/RATE for all i.

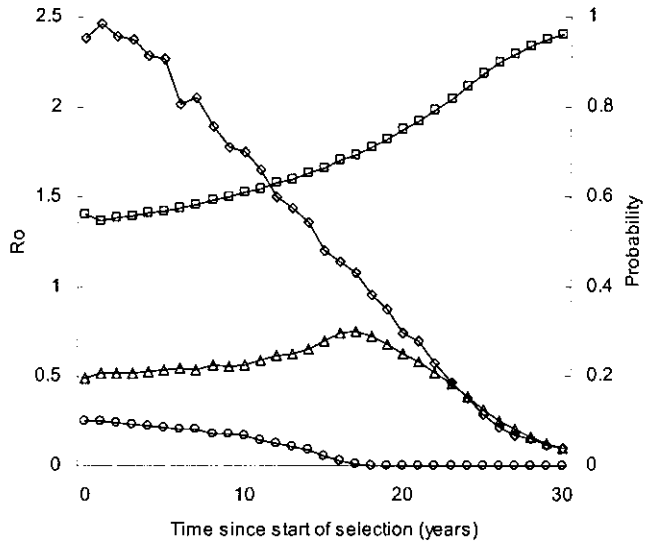
Implementation of this in the base (unselected) population yielded an estimated R_0 for TGE of 2.24 (s.e. = 0.2). Hone (1994) estimated values of 2.0 for a breeding farm and 4.0 for a finishing farm. The farm structure used in this model is a combination of the two farm structures; therefore, the model is reproducing expected results. The total and maximum proportions of pigs infected during an epidemic were 0.85 and 0.19 respectively. Table 4.1 summarizes the results for the base population.

Table 4.1. *Results from stochastic TGE model for the base population*

Parameter	No Epidemic	Minor Epidemic	Major Epidemic
Probability	0.55	0.20	0.25
Probability standard error	0.002	0.004	0.004
Maximum number of Cases	1	4	18110
Total proportion infected during epidemic	0	0.00053	0.65
Maximum proportion infected at one time	0	0.00038	0.13

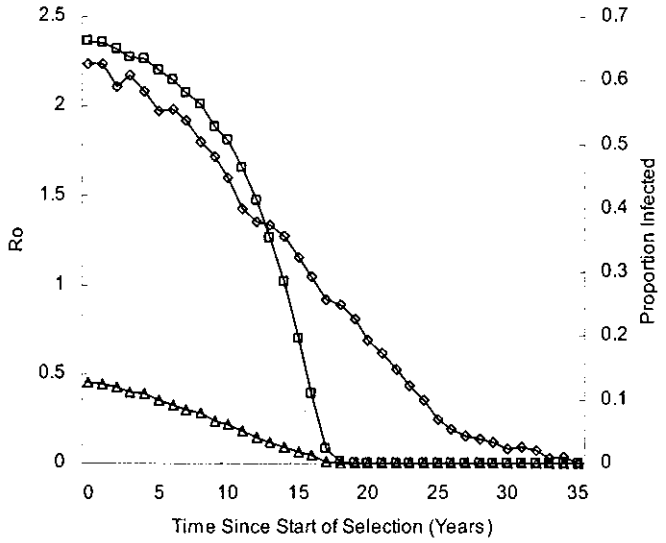
Responses to selection, assuming an arbitrary genetic progress of 4%/year are shown in Figure 4.4. R_0 shows an almost linear decline with time. The probability of a major epidemic declines in a curvilinear fashion, reaching zero as R_0 falls below 1.0. Conversely, the probability of a minor epidemic rises slightly with time, then declines once R_0 falls below 1.0. The interpretation of this result is that epidemics that would have been major can no longer be supported by the population once it becomes more resistant, hence are only minor epidemics. The probability no epidemic, i.e. the population being safe, rises continually with time.

Figure 4.4. *Effect of selection for resistance to TGE on R_0 (◇) and the probability of no epidemic (□), a minor epidemic (Δ) or a major epidemic (○) at each time point, under the continuous selection model.*



The effect of selection upon the severity of major epidemics is shown in Figure 4.5. As selection proceeds, not only does the probability of a major epidemic decrease, but the severity of such epidemics also decreases. Conversely, the severity of minor epidemics actually increases with time and genetic progress (MacKenzie and Bishop, 2001b), before declining as R_0 becomes very small. However, it should be realised that in minor epidemics very few animals are infected compared to major epidemics.

Figure 4.5. *Effect of selection for resistance to Transmissible Gastroenteritis on R_0 (◇), the total proportion (□) and the maximum proportion (Δ) of animals infected during major epidemics at each time point, under the continuous selection model.*



These stochastic GEMs are also informative in terms of determining the impact of the location of the source of infection upon the eventual epidemic outcome. This type of result is very important for overall disease management and is illustrated in Figure 4.6. It can be clearly seen that the location of the outbreak has a profound effect upon the subsequent dynamics of the epidemic. Consider Figure 4.6a, where the probabilities of major epidemics are shown, as a function of where on the farm the initial infection was. Infections starting in areas crowded with pigs in close proximity, i.e. the nursery and finishing pig groups, are likely to lead to a major epidemic. However, infections arising amongst (e.g.) gestating sows are very unlikely to lead to major epidemics. This is a function both of pig density and also of the potential pathways of transmission of infection around the farm.

The ranking of the different types of pigs in terms of their potential danger remains the same for major epidemics, as selection proceeds. However, for minor epidemics the relative risk posed by different parts of the farm actually changes with selection. For example, the riskiest group of pigs in terms of minor epidemics at the outset in this example are the nursing sows – their minor epidemic risk is greater than the major epidemic risk mainly as function of the limited transmission possibilities. However, as selection proceeds, groups such as finishing pigs that were previously a major epidemic risk become a greater minor epidemic risk – this is a function of the reduced transmission coefficient but the much greater possibilities for transmission relative to (e.g.) the nursing sows.

In practical terms, the results in Figure 4.6 show where the greatest efforts in disease prevention must be placed.

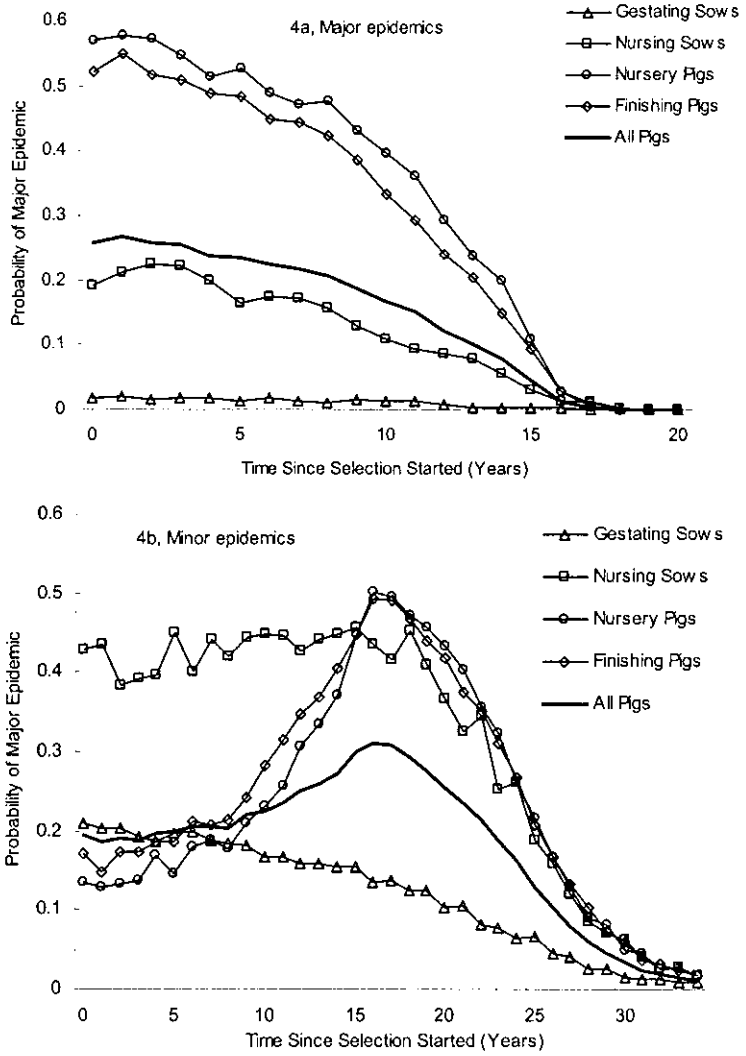
4.3.3 Exploring disease dynamics using GEMS: PRRS example

The ultimate rationale for developing GEMs is for decision-making. For example, will selection for resistance be worthwhile? Will it require major genes or will phenotypic measurements suffice? If the disease is chosen for research, where should the research effort focus?

An example of such an exploration is for the disease: porcine reproductive and respiratory syndrome (PRRS) (MacKenzie, 2000). This is a widespread viral disease of pigs, the consequences of which include high piglet mortality, increased mortality in sows, nursery and finishing pigs, increased premature farrowing, abortion and reduced farrowing rate in infected sows. It should be noted that the effects on productivity are a function of both the primary PRRS infection and also secondary (i.e. non-PRRS) infections.

Extensive literature is available for PRRS describing both case studies and the outcomes of observed epidemics. From this literature it was apparent that the parameters required in a PRRS model were the transmission rate, latent period, infectious period, disease dependent mortality, reduction in reproductive efficiency and the loss of immunity. Additionally, implementation of the model and comparison of its predicted outputs with published results defined reasonable parameters spaces as follows: transmission coefficient 0.0005 to 0.008, latent period 2 days (variation in latent period had little effect), infectious period 20 to 100 days, piglet mortality mean of 70%, sow mortality mean of 2%, growing pig mortality mean of 30%, reduction in farrowing rate 45 to 70%. The transmission rate and the infectious period both have wide parameter spaces, however their effects are correlated and satisfactory models require combinations such as low transmission rate and long infectious period, or high transmission rate and short infectious period.

Figure 4.6. *The probability of (a) major epidemics, and (b) minor epidemics by index case type (heavy black line indicated mean for all pigs), during selection for resistance to transmissible gastroenteritis, under the continuous selection model*



For a wide range of parameter values, the PRRS stochastic epidemic model predicted R_0 values between 9 and 11, i.e. a highly infectious disease. Given the presence of an infected pig on the pig farm, the probability of a major epidemic was estimated at 88%, a minor epidemic 2% and no epidemic 10%. During the course of a major epidemic, it was predicted that 77% of the pigs on the farm would become infected with an overall average mortality rate of 17%.

Whilst actual verification of the results predicted by the model is not possible, the results do provide insight into the disease process and into the utility of different disease control options. For example, with an R_0 as high as inferred from these model explorations, selection for resistance using a phenotypic indicator trait is unlikely to be an effective approach – it would take a very long time to reduce R_0 to a level whereby the disease poses no major problem. However, use of a gene or QTL for resistance would be worthwhile, as it would quickly allow the overall resistance of the herd to be increased.

5. Continuous Challenge, Generalised Immunity and Relationships with Performance

5.1 Introduction

The epidemic theory described so far either assumes some infection equilibrium for macroparasitic infections, or risks of epidemics for specified parasites in the case of microparasitic infections. There are circumstances where neither of these models is appropriate. For example, there are situations where there is essentially continuous challenge such that pathway **a** outweighs pathways **b** and **c**. Intuitively, this would be expected to lead to an epidemic no matter what transmission properties of the parasite. The fact that this is true from an epidemiological perspective is easily seen from the fact that when there are α initial infected animals rather than one, the probability that there is a major epidemic increases in proportion to $1-1/R_0^{\alpha}$ (derived from Renshaw, 1991) (note: this is not the absolute risk of a major epidemic – see section 6). Therefore, when α is large, an epidemic is inescapable. Continuous challenge scenarios are briefly considered in section 5.2.

Other issues that lie outwith the epidemiology theory given so far are concepts of generalised immunity, which is an attempt to increase resistance to many diseases rather than just one target disease, and the more general issue of relationships between performance and resistance. Generalised immunity is often strongly argued for as an alternative to selection for resistance to specific diseases, using arguments covering animal performance, relevance to a wide variety of diseases and reducing parasite coevolution risks (see section 7). As will be argued in section 5.3, there are situations where resistance to specific disease challenges is an appropriate goals and situations where generalised immunity is more appropriate.

Observed relationships between resistance and performance are complex, and predicting or accounting for these relationships requires consideration of disease epidemiology, immune system activation (section 5.4) and resources available to the animal. A conceptual framework for discussing such relationships is given in section 5.5.

5.2 Continuous Challenge Scenarios

Some diseases pose continuous challenges to animals from an essentially infinite reservoir of infection, in which feedback along pathway **c**, if it exists, has a negligible impact upon the size of the reservoir. Typical examples of such diseases include many tropical diseases, where the disease is transmitted from some intermediate host. A good example is trypanosomosis, a protozoan disease transmitted by the tsetse fly. In such diseases the host animal population has often had a long evolutionary relationship with the parasite, and as a result has evolved a varying degree of tolerance towards the parasite. Whether or not it is host resistance to the parasite or tolerance of infection, the question of interest to animal breeders in this scenario is the impact of the infection upon performance. Epidemiological issues relating to the transmission of infection are less relevant because of the dominance of pathway **a**.

An approach towards addressing this problem is given by van der Waaij *et al.* (2000), considering resilience (i.e. tolerance) of animals to continuous challenge. Their model works equally well whether it is resistance or resilience that is modelled. In this model, the level of resistance of an animal to the infection affects the production that can be achieved ("observed production"). Individuals that are less resistant to the infection will be less productive. When

a high infection pressure occurs in combination with a low level of resistance, production can drop to, or even below (e.g., in case of growth) zero. On the other hand, animal performance will not be reduced in the presence of the disease, if the animals have a level of resistance or tolerance beyond a certain threshold.

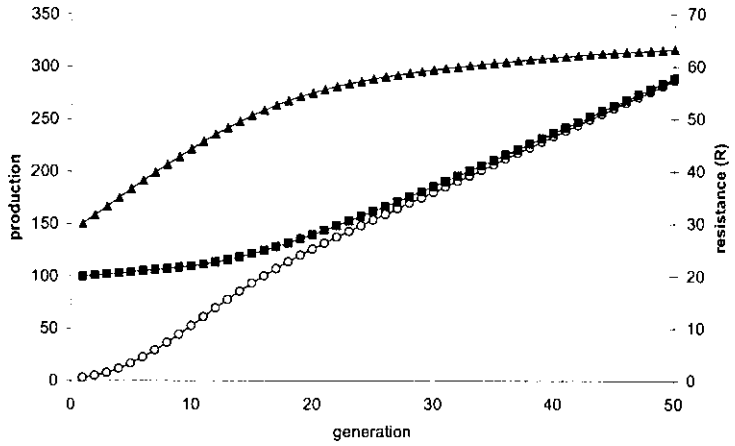
Van der Waaij *et al.* (2000) assumed an infinitesimal model, under which the level of resistance (R) and production potential (Pp) were defined as continuous normally distributed traits. The level of production that would have been achieved if the animals were completely resilient or resistant, or free of infectious challenge, was denoted as Pp. Observed production in the face of infectious challenge (Po) was less than or equal to the underlying production potential (Pp), and influenced by the level of disease resistance (R).

In this model, two thresholds of resistance divide the population into three sub populations; a lower threshold (L) assumed for R, below which production drops to a minimum (Po=0) and an upper threshold (U) for R, above which the animal is assumed to be fully resistant and production equals production potential (i.e. Po=Pp). Between L and U, Po depends on both Pp and R. Thresholds L and U may be considered fixed but they will depend on the type of infection and other environmental factors. These thresholds may be determined from actual disease data, although I know of no cases where this has been done. For example, if animal performance and packed cell volume (PCV) (an indicator of anaemia) is analysed in trypanosome-challenged animals, it may be determined whether or not there are PCV values below which and above which relationships between performance and PCV break down. If so, these would define the L and U thresholds.

In each of the sub populations, the interaction between Pp and disease resistance was defined as $Po=Pp \times f(R)$, where f(R) was dependent on the value of R in relation to the thresholds, and may represent any function, for example a linear, exponential or logistic function. It is assumed that there was no correlation between Pp and R. Therefore, the influence of R on Po differed in each sub population and Po could be considered as consisting of three separate distributions; I: $Po=0$, for $R \leq L$; II: $Po=Pp \times f(R)$, for $L<R<U$; III: $Po=Pp$ for $R \geq U$. As first approximation f(R) was assumed to be a linear function: $f(R)=(R-L)/(U-L)$.

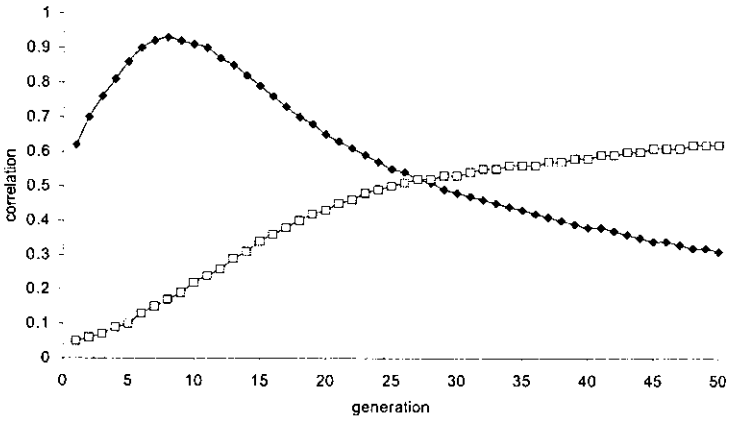
This simple model led to somewhat complex deterministic expectations (for full details see Appendix 1 and 2 in van der Waaij *et al.*, 2000), but some general results of interest emerged. Shown in Figure 5.1 are simulated (rather than deterministic) responses to selection for improved observed production (Po) under this model, starting with a population with an arbitrarily low resistance to the disease. The point to note is that increases in Po under this model in early generations are achieved almost entirely by increases in resistance. Only once resistance reaches an acceptable level does the genotype for underlying performance increase at a rapid rate. Although the absolute results are dependent upon the assumptions in the model, the results do illustrate the general impact of resistance upon performance.

Figure 5.1. *Simulation results for the change in phenotypic population mean for Po (o), R (▲) and Pp (■) during 50 generations of phenotypic selection on Po, where h^2_R = 0.2, distance between L and U = 3.0 s.d.*



This model also results in an interesting pattern of change in the correlation between resistance and observed production as resistance and performance both increase, as shown in Figure 5.2. The correlation between observed performance and resistance is maximised when the overall level of resistance is moderate, then decreases as resistance increases. In contrast, the correlation between potential and observed performance increases steadily as resistance improves, i.e. as differences in performance are less masked by differences in resistance.

Figure 5.2. *Simulation results for the phenotypic correlations between Po and R (♦) and between Po and Pp (□) during 50 generations of phenotypic selection, where h^2_R = 0.2, distance between L and U = 3.0 s.d.*



This model has been extended by van der Waaij *et al.* (2002) to include discontinuities in the infection pressure and also to investigate the role of genetic markers in enabling selection for resistance. In this case, genetic markers were used to enable selection in the absence of infectious challenge, i.e. in the way that is expected to be most beneficial. It was found that genetic markers for resistance were beneficial (i.e. responses to selection were greater than for mass selection on observed production) when a large proportion of the additive genetic variance was explained by the QTL and/or when the heritability for resistance was low (e.g.

$h^2_R = 0.1$). Under constant infection pressure, incorporating QTL information did not increase selection responses in observed production when the QTL effect explained less than 25% of the genetic variance.

5.3 Generalised Immunity

All the arguments so far in these notes have assumed that there is a known disease of overriding importance to address. In many situations, especially in the more extensive production systems, this will be the case and easily identifiable diseases of importance should be obvious. These will include diseases such as nematode infections, ticks and tick borne diseases, flystrike, footrot, mastitis and trypanosomosis. However, it sometimes becomes more difficult to identify overriding diseases of importance in intensive production systems, where most endemic diseases are under control. In this case, a more generalised approach may be desirable, e.g. looking for general resistance to a variety of diseases, or generalised immunity.

Generalised resistance to a wide range of diseases, covering different types of parasites, probably doesn't exist. Mechanisms of disease resistance tend to be specific to individual classes of diseases, and therefore it is unlikely to expect animals or breeds to be resistant to a wide range of diseases, unless they have had a long evolutionary relationship with this disease. However, resistance to diseases caused by related parasites does tend to be correlated. For example, the genetic correlation between the resistance of ruminants to infections caused by difference species/genera of nematode parasites does tend to be correlated (e.g. $rg \sim 0.5$), indicating that many of the mechanisms controlling resistance may be common to different types of nematode parasites. Therefore, there is an opportunity to improve generalised resistance to different classes of diseases. This is most often addressed through the concept of generalised immunity.

Generalised immunity may be considered as a combination of immune responses to a variety of immune system challenges, the aim being to identify animals that are better able to respond to a variety of challenges. The goals of generalised immunity are (i) to genetically improve the overall health status and performance of animals, in the face infectious unknown challenges and (ii) attempt to provide animals with protection against accidental infection from existing or new diseases. The generalised immunity approach may also be desirable if the breeder does not wish to base a genetic management strategy upon a single mode of resistance, if selection for this mode has been shown to adversely affect resistance to other diseases.

The immune measurements used in generalised immunity may be specific challenges, but more often they are non-specific challenges such as sheep red blood cells, or even simply measures of the innate immune response, e.g. Natural Killer (NK) cells measured in animals known to be facing subclinical challenges. However, it must be realised that at the immunological level the distinction between selecting for resistance to a specific disease (if resistance is immunologically controlled) and generalised immunity may not be clear. For example, allelic variants that affect immunity to many pathogens may be part of the innate immune system and could be selected for both when selecting for specific resistance and also when selecting for generalised immunity.

The feasibility of selecting for generalised immunity and immune responsiveness has been

explored and demonstrated by a number of authors, most notably in pigs by Mallard *et al.* (1998) and in chickens by Pinard *et al.* (1992) and Sarker *et al.* (1999). A feature of all these studies is that the immune response traits tend to be highly heritable, a result verified by the study of immune response traits in sheep challenged by nematode parasites (Strain *et al.*, 2002). These high heritabilities are verified by the relative ease of finding QTL for immune measurements (e.g. Edfors-Lilja *et al.*, 1998 and 2000). In general terms it may be summarised that heritabilities tends to rise as one goes from general disease category to specific disease resistance to specific immune response, with antibody responses sometimes being highly heritable.

Correlated responses in resistance to specific diseases are variable from these generalised studies. For example, pigs selected for increased immune responsiveness by Mallard *et al.* (1998) had significantly less peritonitis and pleuritis, but more severe arthritis. The latter may be due to an excessively active immune system. Also, for the lines of birds described by Pinard *et al.* (1992), whilst the low-selected line was more susceptible to Marek's disease, within-line relationships between antibody response and resistance to Marek's disease could not be established. It is probably unreasonable to expect selection for generalised immunity or immune responsiveness to have major effects on resistance to specific diseases, although it should improve resistance to a variety of diseases. However, to formally test which diseases have, or have not, been affected by such selection would be a large task.

The other rationale for improving generalised immunity is to improve health status and hence performance under environments in which there are a variety of subclinical challenges (the so-called 'dirty' environments, typical of commercial intensive production units). Certainly, Mallard *et al.* (1998) found that the line selected for increased immune responsiveness had enhanced performance – a result that is predictable using the framework described in section 5.5. Currently (i.e. in 2003), the aim of several generalised immunity studies in Europe, mainly in pigs, is to determine whether or not there are immune measurements taken on pigs reared under 'clean' specific-pathogen-free conditions that are predictive of progeny performance and progeny disease status, as measured under commercial ('dirty') conditions. If this proves to be the case, then selection for generalised immunity, especially for components of the innate immune system, may have an important role to play in the breeding of animals for intensive husbandry situations.

From a modelling perspective, generalised immunity should be treated as a continuous challenge scenario. It is reasonable to assume that under 'dirty' conditions there will be some degree of continuous challenge. Models that are appropriate for this situation may be parameterised using considerations of how immune system activation affects performance (section 5.4). The framework for considering relationships between performance and resistance/immunocompetence (section 5.5) gives a formal basis for developing such models.

5.4 Immune System Activation and Performance

Immune system activation following an infectious challenge will have subsequent effects upon animal performance. The immune system can actively regulate the various components of nutrient metabolism in several ways (Klasing *et al.*, 1991): (i) by direct neural connections to the central nervous system, which may trigger behavioural adaptations and/or release of hypothalamic and pituitary hormones; (ii) by release of hormones such as ACTH and thyrotropin by immune cells; (iii) most importantly, through cytokines which trigger not only

anorexia and fever but also processes like the up-regulation of gluconeogenesis from glycogen, fatty acids and amino acids, accompanied by the down-regulation of muscle protein deposition (and/or muscle proteolysis) to support this gluconeogenesis and to support acute-phase glycoprotein synthesis.

As a result, infection and the associated activation of the immune system will lead to a cascade of resource-reallocating processes in the host, most notably the following: (i) the production of acute-phase glycoproteins, immune cells and immunoglobulins, which requires extra protein synthesis; (ii) repair of damaged tissue, which may cause a strong increase in protein turnover rates and hence increase metabolism considerably; (iii) fever, with the same metabolic effect as subcritical ambient temperature; (iv) depression of voluntary feed intake (anorexia), the process with the most dramatic effects on energy metabolism.

The wave of cytokines that induces nausea and anorexia (Spurlock, 1997) is a transitory phenomenon, however one with a marked impact upon performance. Anorexia seems a paradox at a time when animals' metabolic needs for energy and protein are increased, especially to support the extra energy required to mount a fever and to manufacture the proteins required for an effective immune response. Maybe it is an adaptive response to minimise the risk of continue challenge from the same source of infection.

Strong phenotypic evidence of relationships between various aspects of immune system activation and performance is given by Clapperton *et al.* (2003) in a cross-sectional study of performance and immune status in pigs. These authors found consistent phenotypic relationships between performance traits (growth rate, food intake and efficiency) and immune indicator traits (NK cells, B cells and Monocytes), such that elevated levels of these cells was always associate with decreased performance, in pigs kept in the same environment. Additionally, in the context of intensive pig production systems, Stahly (1996) summarised the impact of high and low immune system activation (achieved by conventional weaning and medicated early weaning, respectively; Williams *et al.*, 1997ab). He concluded that "minimizing the activation of the pig's immune system" in those trials resulted in higher *ad libitum* feed intake, growth rate, muscle development, and feed efficiency.

The issue may also be approached from the opposite perspective: can the impacts of infection be minimised by overcoming the major deleterious impact of immune system activation, *viz.* reduced protein availability and utilisation? Certainly, in the case of nematode parasitic infections in sheep this is the case: protein supplementation of infected sheep does enhance the effectiveness of the immune system in coping with nematode infections, enhance the expression of resistance and minimise the impact of the infection (see Kyriazakis *et al.*, 2002 for an overview).

The impact of immune system activation upon performance may seem somewhat peripheral to disease genetics, but potentially it plays a central role in a key issue in disease genetics, *viz.* the genetic relationship between resistance and performance (discussed in section 5.5). Additionally, the impact of immune system activation on performance is still relatively poorly quantified and represents an area with considerable opportunity for modellers aiming to quantify performance and the environmental factors affecting performance.

5.5 Relationships between Performance and Resistance

Relationships between resistance, tolerance and animal performance are often poorly quantified, and even more often misunderstood. For example, a fallacious conclusion sometimes drawn is that resistant animals are less productive animals. This conclusion is often drawn when comparing productive characteristics of (e.g.) hardy locally adapted animals *vs.* high performing susceptible exotic animals. Most likely, all that is being observed in this case is different selection histories for the two populations. Even the relationships between performance described and simulated in sections 3 (resistance to nematode parasites *vs.* growth rate) and 5.2 (resistance to a continuous challenge *vs.* observed performance) are only partial summaries of the true underlying relationships. It is necessary to create a more rigorous framework to consider these relationships.

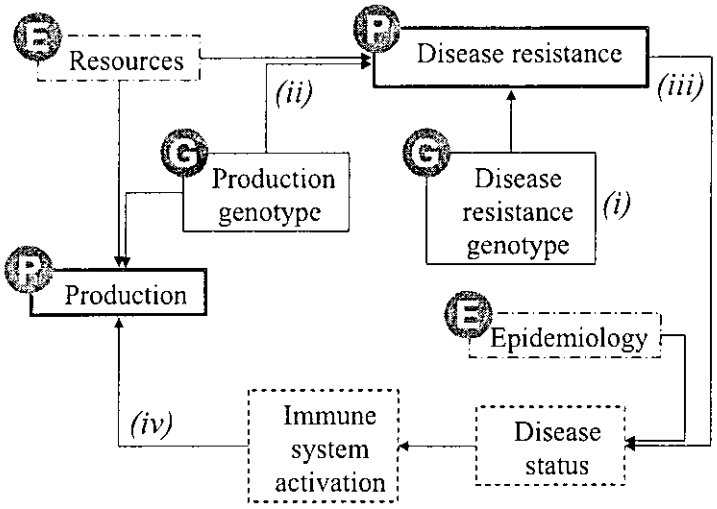
Firstly, several basic concepts must be stated to avoid common misunderstandings:

- Infection compromises performance as described above: it is evident that the growth and performance of animals that are responding to pathogen challenges will usually be inferior to unchallenged animals. This may be due to the pathology of the disease, as well as the costs and the consequences of the immune response.
- Ability of animals to withstand disease is environment dependent: the nutritional and physiological status of an animal may affect its ability to mount an effective response to pathogen(s). For example, nematode resistance in sheep is enhanced by protein supplementation. Additionally, pregnant ewes are more susceptible to nematodes than barren ewes facing the same larval challenge.
- Resistance is not generally associated with performance at the underlying genetic level. Examples exist of both favourable and unfavourable associations between resistance and performance. Moreover, the observed relationship will be dependent upon the environment and the disease challenge, as described below.

In general terms it is difficult to predict genetic correlations between performance and resistance: as described below, such correlations depend upon both the consequences of being infected as well as the costs of mounting (or being able to mount) appropriate immune responses. These effects will influence relationships between resistance and performance in opposite ways, with the actual relationship depending upon the balance between the two. To understand thus, consider the framework proposed by Knap and Bishop (2000). Figure 5.3 shows in a simplistic way both the direct and the indirect relationships between performance and disease resistance.

The underlying genetic relationship between performance and disease resistance is simply pathway (ii) in this Figure. There is no particular reason to expect this to be either positive or negative, and a safe assumption to make is that this correlation is generally neutral. However, this underlying correlation will rarely be directly expressed or even measurable. Firstly, both disease resistance and performance compete for resources in their expression, particularly protein. Unless particular effort is made to supply excess protein, this may induce an apparent trade-off between resistance and performance. This is part of the environmental dependence described above, and may be considered as a cost of being resistant. However, disease resistance has a consequence as well, as described by pathways (iii) and (iv). Disease resistance phenotype along with the disease epidemiology will determine the disease status of the animal or population. Immune system activation follows on directly from disease status, and can lead to reductions in performance.

Figure 5.3. *Schematic diagram of interrelationships between performance and disease resistance.*



Therefore, the observed genetic (and phenotypic) relationships between disease resistance and performance are a balance of (i) the costs of being resistant *vs.* (ii) the beneficial consequences of being resistant. Greater costs of resistance will tend to make the relationship more unfavourable. Greater benefits of being resistant compared to being susceptible will make the relationship more favourable, in the presence of infectious challenge.

From Figure 5.3 two things are immediately apparent. Firstly, genetic relationships between resistance and performance are environment-dependent. The greater the resources (particularly protein) available to the animal, the less is the potential trade-off (in agreement with Kyriazkis *et al.*, 2002). Secondly, genetic relationships between resistance and performance are also dependent upon the disease epidemiology. The more severe the infectious challenge (i.e. the greater the impact of the disease epidemiology), the more favourable the relationships between resistance and performance are likely to be. This was the result obtained by Bishop and Stear (1999), although it must be noted that these authors only modelled the consequences of infection and failed to account for the costs.

Figure 5.3 has further consequences for animal breeders. This Figure shows how apparent genotype by environment (GxE) interactions can arise when the breeder is focussing solely on performance traits. If the same trait is evaluated in environments that differ in their disease challenge (e.g. contaminated *vs.* clean pastures for nematode infections, or tsetse fly infected *vs.* free areas in the context of trypanosome challenge) and this difference in disease challenge is ignore, apparent GxE interactions will arise from the impact of the disease upon performance.

5.6 Summary

In this section, situations have been considered where focus should change from considering the dynamics of a specific disease to considering relationships between disease resistance (or immunocompetence) and performance. The framework described in section 5.5 is

illuminating in accounting for relationships between performance and productivity. Indeed, it can also be used to predict the consequences of selection for performance under certain environmental conditions on disease resistance or immunocompetence or, conversely, the effects of selecting for disease resistance on performance.

Examples may illustrate this.

- Tropical breeds selected in the face of continuous disease challenge and exotic breeds will show differences in productivity that are completely dependent upon the challenge environment. This has been elegantly demonstrated for nematode parasite infections in Red Maasai (resistant and tolerant) *vs.* Dorper sheep (susceptible) (Baker *et al.*, 2002).
- Pigs selected for performance under limited resources in a commercial environment may be expected to allocated additional resources to immune responsiveness, if the benefits of doing so outweigh the costs. Experimental results verify that this is the case.
- Again for nematode parasite infections in sheep, genetic relationships between resistance and various production traits depend upon the relative balance of costs *vs.* consequences, for that trait. For example, for lambs infected by *Teladorsagia circumcincta*, the correlation between nematode egg counts is generally negative (favourable) (Bishop *et al.*, 1996, plus additional unpublished observations), because the benefits of being resistant outweigh the costs. However, in ewes the genetic relationship between lamb output (i.e. her milk production) and egg count is positive (unfavourable) (Bishop and Stear, 2001), since nematode infections probably have only a small impact upon the ewe, but scarce resources during lactation must either be partitioned towards milk production (at her expense) or towards immune response (at the expense of lactation).

6. Advanced Genetic Management Issues: Major Genes and Genetic Diversity

6.1 Introduction

Sections 2, 3 and 4 covered broad issues regarding the impact of genetic change in populations, the use of major genes, and impacts upon disease epidemiology and disease risks. However, these were covered at a somewhat superficial level, and somewhat simplistic assumptions were made regarding the effects of the major genes for disease resistance. For example, in the consideration of major resistance genes, it was implicitly assumed that the resistance allele conferred complete resistance. In reality this might not be the case, i.e. resistance may not be complete. Additionally, there might be several discrete classes of resistance, as opposed to simple a comparison of 'resistant' *vs.* 'susceptible'. This generalisation has yet to be explored.

The major theme in the calculations and arguments so far is that host genotype affects the transmission of infection. From this starting point, arguments relating host population genetic homogeneity to disease transmission may be developed. Also, maintenance of genetic variation to safeguard populations against potential disease risks is a theme that often runs through population genetic and ecological arguments, especially when disease dynamics in natural populations are concerned. However, it remains necessary to show, from an epidemiological viewpoint, that these arguments are valid and to quantify them.

This section shows the theoretical and simulation developments necessary to explore issues relating to the deployment of disease resistance genes and the impact of genetic heterogeneity or biodiversity. These arguments are somewhat conceptual and theoretical, however they give the necessary background for applying theoretical results to specific examples.

6.2 Strategies for Utilising Disease Resistance Genes

It has already been shown that deterministic expectations are such that if R_0 is greater than 1.0 it is expected that a epidemic will occur upon the introduction of an infected animal, but if $R_0 \leq 1.0$ an epidemic is not expected. However, stochastic chance events ensure that even for R_0 values greater than 1.0, there is a probability that there will be no epidemic, or a minor epidemic which dies out without intervention. Likewise, for $R_0 \leq 1.0$, minor epidemics can arise. By definition, a population comprised solely of genetically (completely) resistant animals results in the pathogen having an R_0 of less than 1.0. This is irrespective of R_0 in the population of genetically susceptible wild-type animals. A rational aim of a disease control strategy is to reduce R_0 in the target population to below 1.0. This may be achieved by selecting animals for enhanced resistance or, in the case of discrete levels of resistance, by mixing genetically resistant animals with susceptible wild-type animals. Here, the precise nature of these heterogeneous mixtures are considered, along with the associated epidemic properties, when there is one or a small number of genes conferring resistance, and when resistance may either be complete (i.e. $R_0<1$) or partial (i.e. $R_0>1$) (Bishop and MacKenzie, 2003).

Required proportions of resistant animals

Under assumptions of no spatial heterogeneity and equal contact amongst animals of different genotypes, i.e. a fully mixing genetically heterogeneous population, R_0 for the population as a

whole will be the weighted average of the R_0 values within each subgroup (Dushoff and Levin, 1995). Assume that for a given disease, if the host population comprises animals with a wild-type genotype, then disease transmission is described by R_{01}. Now suppose that a resistance allele, r, is found, such that if the host population comprises animals homozygous for this allele will alter transmission to R_{02}, where $R_{02} < R_{01}$. Our objective is to use judicious selection of animals to construct a mixed population such that $R_0 = (R_{01}\rho_1 + R_{02}(1-\rho_1))$, where ρ_1 is the proportion of wildtype animals. Hereafter this is called the two-genotype model.

The two-genotype model can be extended to the general case of *n* genetic categories. These categories might represent animals that are combinations of, say, homozygous for the susceptibility allele, heterozygous and homozygous for the resistance allele, across several loci. Now there are *n* levels of susceptibility, denoted by R_{01} ... R_{0n}, and a different proportion of animals (ρ_1, ρ_2, ... ρ_n) corresponding to each of these groups. Thus, $R_0 = (R_{01}\rho_1 + R_{02}\rho_2 + \ldots + R_{0n}\rho_n)$.

Consider the two-genotype model. There are three possible scenarios. Firstly, $R_{01} < 1$ in which case there is no disease problem. Secondly, $R_{02}>1$, in which case the population will be susceptible to epidemics no matter what the genetic makeup of the population. Under most circumstances, to minimise the probability or severity of epidemics, $\rho 1$ should be set to zero, reducing the epidemic risk by having all animals homozygous for the resistance allele. Thirdly, $R_{01}>1$ and $R_{02}<1$. Here, the requirement is to reduce R_0 below 1.0, where $R_0 = R_{01}\rho_1 + R_{02}(1-\rho_1)$. Thus, the solution is:

$$\rho_1 = (R_0 - R_{02})/(R_{01} - R_{02})$$

where ρ_1 is the proportion of wild-type genotypes in the population, and R_0 is some desired value between 0 and 1. For the threshold case where R_0 is 1.0 and R_{02} is 0 (i.e. complete resistance) the proportion of resistant animals must be at least $1-1/R_{01}$, i.e. the previously described result.

For the *n* genotype model, there is no single explicit solution to the number of animals, required of each genotype. Rather, the aim is simply to combine the genotypic classes such that $R_0<1$. For example, for a 3 genotype model where $R_0 = (R_{01}\rho_1 + R_{02}\rho_2 + R_{03}(1-\rho_1-\rho_2))$, the requirement is simply to choose values of ρ_1 and ρ_2 that solve the equation (obtained by simple rearrangement):

$$\rho_1 + \rho_2(R_{02} - R_{03})/(R_{01} - R_{03}) = (R_0 - R_{03})/(R_{01} - R_{03})$$

Desired solutions to the n genotype model will usually combine as many of the genotypes as feasible, unless some are highly susceptible. There will usually be sound biological, production and epidemiological reasons why this should be the case, including minimisation of pathogen co-evolution risks (see section 7).

Epidemic probabilities

Consider a simple SIR model: in a population of N animals, made up of S susceptible, I infected and R recovered/removed animals, a transmission rate β, and a recovery rate γ, the rate of change of the numbers of animals in each category is described by the following equations: $dS/dt = -\beta SI$; $dI/dt = \beta SI - \gamma I$; $dR/dt = \gamma I$. As described in section 2, $\beta = R_0\gamma/N$. Genetic heterogeneity can be incorporated by replacing S, I and R with S(i,t), I(i,t) and R(i,t)

which represent the number of susceptible, infected or recovered animals of genotype i at time t such that $S(i,t)+I(i,t)+R(i,t) = N(i,t)$. The transmission rate is also genotype dependent and is denoted by $\beta(i) = R_0(i)\gamma/N(i)$ where $R_0(i)$ is the basic reproductive rate of the pathogen in a population with genotype i.

The stochastic implementation of this model enables epidemic probabilities to be estimated, as this allows us to determine the probability of each event type, i.e. the probability that a susceptible animal is infected and moves to the infected category or that an infected animal recovers. In a population where $R_0 < 1.0$, a minor epidemic is defined as one where there is more than a single infected animal. The probability of a minor epidemic upon the introduction of a single infected animal can be estimated from the probability that the first event, after the introduction of an infected index case, is an infection. First it is necessary to construct the 'rate' (K), of all possible events. In a situation where there is a single infected animal of genotype i, the possible events, in a population with 2 genotypes, i and j, are infection of an animal with genotype $i = \beta(i)S(i)c(i,i)$; infection of an animal with genotype $j = \beta(j)S(j)c(j,i)$; recovery of the infected animal = γ, where c(j,i) is the contact rate between animals with genotype i and animals with genotype j, which is simply the proportion of animals of type j animals in the population. Therefore, assuming for the purposes of illustration that the recovering rate (γ) is constant across genotypes:

$$K = \beta(i)S(i)c(i,i) + \beta(j)S(i)c(j,i) + \gamma$$

The probability of no epidemic is the probability that the first event is the recovery of the single infected animals, i.e. γ/K. The probability that an epidemic occurs is $1-(\gamma/K)$. For $R_0<1$ this is the probability of a minor epidemic, and for $R_0 > 1$ it is the combined probability of a major or a minor epidemic.

For generality, the probability of no epidemic may be solved using the two-genotype model. Let x_1 be the number of animals with the wild-genotype and x_2 be the number of resistant animals in the population, then $\beta_1 = R_{01}\gamma/x_1$ and $\beta_2 = R_{02}\gamma/x_2$ are the transmission rate in animals with the wild-genotype and resistance genotype respectively. With a single infected animal of genotype 1, the contact rate between this animal and the susceptible animals of each genotype is proportional to the number of animals of each genotype, which is defined above in terms of the R_0 values. Therefore, $c(1,1) = (R_0-R_{02})/(R_{01}-R_{02})$ and $c(2,1) = (R_{01}-R_0)/(R_{01}-R_{02})$, and the probability of no epidemic is:

$$\frac{\gamma}{\left[\left(\frac{R_0 - R_{02}}{R_{01} - R_{02}}\right)\left(\frac{R_{01}\gamma}{x_1}\right)x_1 + \left(\frac{R_{01} - R_0}{R_{01} - R_{02}}\right)\left(\frac{R_{02}\gamma}{x_2}\right)x_2 + \gamma\right]}$$

$$= \frac{1}{\left(\frac{(R_0 - R_{02})R_{01} + (R_{01} - R_0)R_{02}}{R_{01} - R_{02}} + 1\right)}$$

$= 1/(R_0+1))$

Therefore, when $R_0 \leq 1.0$, the probability of a minor epidemic is $R_0/(R_0+1)$. For $R_0 > 1.0$, this describes the probability of any epidemic. At $R_0 = 1.0$, the probability of either no epidemic

or a minor epidemic are both 0.5. When $R_0>1.0$, the probability of major and minor epidemics may be determined from the basic stochastic threshold theorem (Renshaw (1991), derived from extinction probabilities. In the case where an epidemic occurs and is started by a single infected animal, the probability of it being minor or major are $\gamma/N\beta$ and $(1-\gamma/N\beta)$. Using the relationship $R_0=N\beta/\gamma$, and multiplying $R_0/(R_0+1)$ by these two probabilities yields:

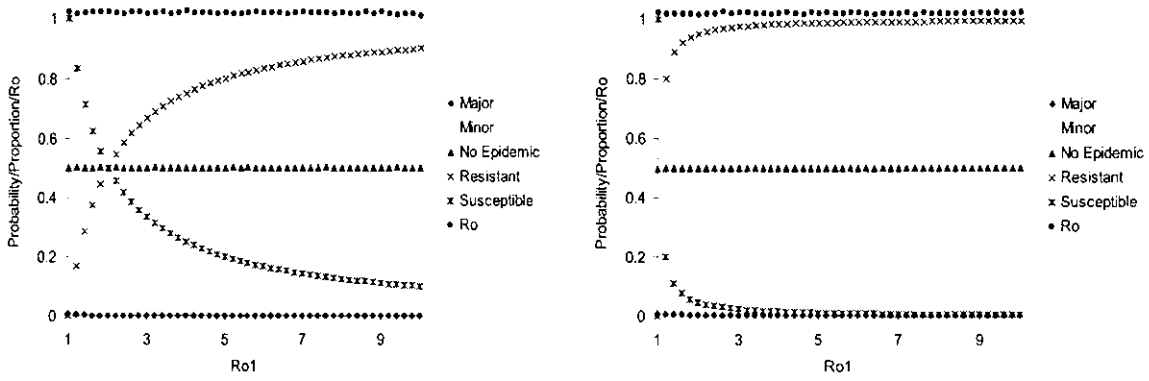
Probability (minor epidemic$|R_0>1$) = $1/(R_0+1)$
Probability (major epidemic$|R_0>1$) = $(R_0-1)/(R_0+1)$,

Robustness of results to parameter estimation
The models described above assume precise knowledge of the transmission rate. However, this is unlikely to be the case, especially for populations comprising the resistant genotype. As an example, assume that the transmission rate in a population entirely composed of the resistant genotype, is R_{02}, and is known accurately. However, assume that R_{01}, describing transmission in a population entirely composed of the wild-type genotype, has been estimated with error such that it has been underestimated by a proportion δ. Therefore, $R_{01}^*=R_{01}+\delta R_{01}$, is the actual value of the basic reproductive rate in the resistant population, and the true transmission in a mixed population is $R_0^*=(\rho R_{01}^*+(1-\rho)R_{02})$. However, ρ will have been determined under the expectation that the basic reproductive rate is $R_0=(\rho R_{01}+(1-\rho)R_{02})$, and the error in R_0^* and ρ is easily calculated.

The changes in epidemic probabilities are calculated similarly. Using the methodology above, we find that the probability of no epidemic is $1/(R_0+\delta R_{01}\rho+1)$, and the probability of a minor epidemic is (or major + minor epidemic if $R_0^*>1.0$) is $(R_0+\delta R_{01}\rho)/(R_0+\delta R_{01}\rho+1)$. When $R_0^*>1.0$, this breaks down into a probability of a minor epidemic of $1/(R_0+\delta R_{01}\rho+1)$ and a probability of a major epidemic of $(R_0+\delta R_{01}\rho-1)/(R_0+\delta R_{01}\rho+1)$.

Stochastic verification
These results may be verified using stochastic simulation, as described in section 4. For example, Figure 6.1 gives the probability of no epidemic, a minor or a major epidemic for two populations where the populations are constructed of two genotypes. The expected reproductive rate in a population composed entirely of resistant animals is (a) $R_{02} = 0.0$ and (b) $R_{02} = 0.95$. The population is composed to produce an overall $R_0 = 1.0$. It can be seen that the realised R_0 value is indeed close to 1.0 for all cases. Moreover, as predicted by theory, in both populations the probability that no epidemic occurs or that an epidemic is minor, dying out with a small number of infected animals, is 0.5 and the probability that an epidemic is major, infecting a large number of animals, is 0.0. However, the numbers of 'susceptible' and 'resistant' animals required to achieve the correct population structure differs considerably in the two scenarios. When $R_{02} = 0.95$, i.e. the resistance allele only confers partial resistance, then most animals in the population are required to carry this allele.

Figure 6.1 *Required proportions of resistant animals and resulting probability of no epidemic, minor or major epidemics for a two-genotype population. Expected basic reproductive rate in susceptible population, R_{01} is given on the x-axis. Expected basic reproductive rate for the resistant population is (a) R_{02} = 0.0, (b) R_{02} = 0.95. Observed basic reproductive rates for both populations is 1.0 and is based on 50 replicates per simulation. Each result is the average of 5000 stochastic simulations.*



Interpretation

Given knowledge of how each allele of the disease resistance gene affects the transmission of infection through a population of animals homozygous for that allele, then precise population structures can be set up to minimise the risk of epidemics. Whilst the risk and severity of epidemics will be absolutely minimised by making the population homozygous for the most resistant allele or genotype, this is not a specific requirement. Indeed, from a pathogen evolution risk perspective, as described below in section 7, it may not be desirable.

The strategy for defining optimal population structures does require knowledge of the transmission characteristics of the pathogen, for each genotype. This is likely to be difficult information to obtain with precision. However, quantifying the impact of imprecision of estimation is relatively straightforward.

Irrespective of these results, many breeding programs will probably err on the side of safety, and wish to treat the results obtained as lower limit predictors of the number of animals required. However, these results will inform breeding programs of how long it will take before a population can be assumed safe from an epidemic from a specific disease. Moreover, they will help planning in the efforts to minimise genotyping costs. Finally, in the case of mixtures of resistant and susceptible animals remaining on a farm, these results can also enable a producer to implement advanced genetic management. In this case, the aim is not only to protect the farm as a whole, but to also ensure that each subsection of the farm has a suitable mixture of resistant and susceptible animals.

6.3 Impacts of Genetic Diversity on the Transmission of Infection

Genetic diversity is often argued as being of major importance when protecting animal populations against disease risks. The diversity of possible immune responses within an animal, and within populations as a whole, adds credence to this argument. However, from an epidemiological viewpoint the argument is harder to sustain. The impact of heterogeneity

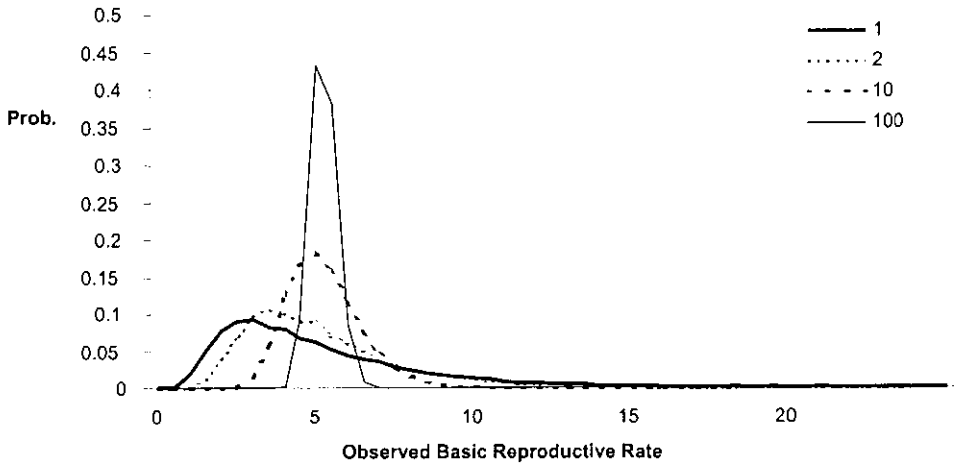
upon disease transmission in animal and human populations is well studied (Dushoff, 1999) and with unequal mixing between heterogeneous subgroups, the properties of disease invasion and transmission are dependent upon the heterogeneity and the mixing of subgroups (Dushoff and Levin, 1995). However, as described above, with equal mixing between subgroups expected disease transmission is simply the weighted average of the transmission properties of the subgroups. A typical outbred population of animals may simply be considered as a genetically heterogeneous population with equal mixing between subgroups, i.e. between individuals. Thus, from simple epidemiological principles, genetic heterogeneity *per se* should not be expected to influence expected disease invasion or transmission.

It is therefore necessary to address the apparent discrepancy between the intuitive consensus on the importance of genetic heterogeneity in animal populations and simple expectations derived from epidemiological principles. It is important to note that much of the elegant available epidemic theory is based upon a deterministic framework that yields expected or mean outcomes. Once again, a stochastic simulation approach is necessary to fully appraise the situation, although a simple discrete-time model can answer straightforward questions. The critical issue to address is the impact of genetic heterogeneity upon expected outcomes of future epidemics in domestic animal populations and, critically, the impact upon the variability of outcomes. For example, shown in Figure 6.2 are distributions of R_0 estimated from populations conforming to a very simple genetic model in which the population of 1000 animals is comprised of groups of genetically identical animals with regard to their resistance to a specific pathogen. In this case the population is comprised of 1, 2, 10 or 100 different genotypes (MacKenzie *et al.*, 2001). The genotypes are drawn at random from a log-normal distribution.

It can be seen in Figure 6.2 that genetic diversity, as indicated by the number of genotypes in the population, has no impact on the expected R_0. The mean values of each distribution are the same, despite the differences in the mode. Sampling from a Normal distribution results in equivalence of both the mean and the mode. This is in agreement with the results of Dushoff and Levin (1995), but apparently contradictory to the intuitive impression of many people. However, **variability** in outcomes is greatly affected by genetic diversity. As diversity decreases, a far greater variety of outcomes are possible. A genetically homogeneous population is more likely to be completely healthy **and** it is more likely to suffer a catastrophic epidemic. This result conforms to observations made of different farming systems. In general, intensive and uniform production systems (for either plants or animals) tend to be healthier on the whole than more extensive and diverse production systems. However, the intensive production systems are more prone to catastrophic breakdown when a health problem does arise. Therefore, in terms of genetic diversity, the questions to be addressed relate to the **risks** of specific outcomes, rather than mean outcomes.

The balance as to whether or not homogeneity is an advantage or disadvantage will depend upon the nature of expected disease challenge. If lowly contagious parasite challenges are expected, then homogeneity can turn a minor problem into a major one, and heterogeneity is more desirable. However, if extremely infectious challenges are expected, then with homogeneity there is a possibility that it may reduce the risk (as illustrated in Figure 6.2, with the occurrence of low R_0 values), whereas a risk of severe epidemics will always be present with heterogeneous populations.

Figure 6.2. *The distribution of the estimator of basic reproductive rate, R_0, for different numbers of genotypes with susceptibilities drawn from a log-normal distribution with mean 5.0 and standard deviation 3.75. The legend denotes the number of genotypes represented.*

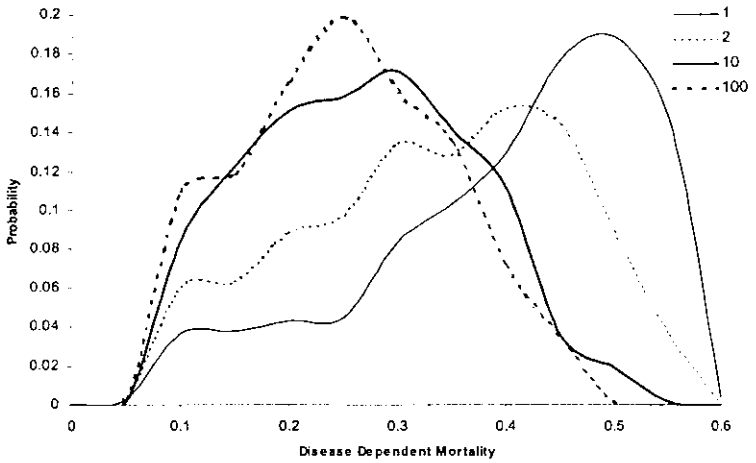


To predict future risk profiles, is it necessary to integrate results across variability in host genotype and variability in parasite type. Results from a stochastic implementation of this problem are given by Springbett *et al.* (2002). Pathogens were sampled from a meta-population of potential pathogens, and were characterized by their expected R_0. R_0 was drawn from one of two gamma distributions, a dispersed and a clumped distribution. Both distributions gave means of 1.5, but variances were 0.9 and 0.014, respectively. Genetic heterogeneity in the host population was conditioned upon the expected pathogen R_0. Individual susceptibility values were drawn from a log-normal distribution, centered on the expected R_0, assuming an infinitesimal model. Once again, the host population was modelled to consist of discrete groups of genetically identical individuals, ranging from n=1 (i.e. a population of clones) to n=100 (approximating a genetically heterogeneous population with large Ne).

As expected, mean realized R_0 was invariant to host population genetic diversity, however the variability in realized R_0 increased markedly as host population heterogeneity decreased. There was a two-fold increase in the variance of realized R_0 as n decreased from 100 to 1 for the dispersed distribution of pathogens, and an 8-fold increase in the clumped distribution (Springbett *et al.*, 2002). The numbers of infected animals, across the duration of the epidemic and at any given point in time, were also estimated. The equations matching the severity of the epidemic to R_0 (section 2) (*viz.*: $I=1-\exp(-R_0 I)$ and $y_{max}=1-(1+ln(R_0))/R_0$) match precisely for homogeneous populations, but describe upper bound relationships in the case of heterogeneity.

The nature of the changes in risk profiles with changing genetic heterogeneity implies that disease consequences, as opposed to mean infection levels, may change as heterogeneity changes. This is indeed the case, as verified by Springbett *et al.* (2002). Assuming a disease-dependent mortality of 0.08 deaths/infected animal/day, genetic heterogeneity does impact upon mortality when averaged across major epidemics, as shown in Figure 6.3.

Figure 6.3. *Disease-dependent mortality during major epidemics, with infection rates drawn from a gamma distribution (gamma distribution parameters: α = 2.5 and θ = 0.6) in populations with 1, 2, 10 and 100 genotypes.*



It can be seen that the probability density function for disease-dependent mortality depends critically upon genetic heterogeneity. For example, peak expected mortality doubled from 0.24 in a heterogeneous population to 0.49 in a genetically homogeneous population, given the same pathogen virulence and the same expected pathogen transmission rates. This general pattern of results also held for diseases with lower mortality rates. Therefore, decreased genetic diversity is predicted to substantially increase mortality, when major epidemics take hold.

In summary, whilst genetic diversity does not impact upon the expected transmission of infection, it does affect the likelihood, severity and impact of epidemics in animal populations. The result derives from second order effects, i.e. considering the variability of biological processes, rather than from the consideration of expected outcomes - the most common approach in epidemiological models. It should be noted that the impact of biodiversity upon epidemics has been previously experimentally demonstrated in plants (Zhu *et al.*, 2000). However, the plant context is somewhat different, as plants are stationary and heterogeneity is expressed both genetically and spatially; and spatial heterogeneity (in this case limited contacts between different genotypes) is a well-known factor affecting the transmission of infection (Dushoff, 1999; Dushoff and Levin, 1995).

6.4 Concluding Remarks

Some generalisations of basic genetic-epidemiological theory have been presented in this section, and it is interesting to note that both 1st order effects (i.e. means) and 2nd order effects (i.e. variability) are important, albeit in different contexts. Analogous to standard animal breeding theory, the 1st order effects determine basic genetic management strategies, and the 2nd order effects describe the accompanying risks and biodiversity/inbreeding/population size effects.

The aim should be for readers to take the general principles outline here and apply them to the necessary decision-making processes when considering disease genetics. In particular, a critical outcome is the desirability of combining different genotypes when addressing disease resistance, both to bring R_0 below 1.0 and also to minimise future disease epidemic problems. Whilst the lack of knowledge of R_0 may be a limitation when devising disease management strategies utilising major genes for disease resistance, some rules-of-thumb can be devised that will give general guidelines. For example, for highly infectious diseases (e.g. foot and mouth disease), nearly all animals would need to be resistant, for the genetic management strategy to be beneficial and used in preference to a non-genetic strategy. However, for less infectious diseases such as TGE, options where 50-75% of animals were resistant would probably suffice, as a starting point.

The potential for parasite coevolution is also important in the context of advanced genetic management strategies. Genetic diversity within the population makes effective parasite evolution less likely, and hence makes genetic control strategies more sustainable. These issues are covered in detail in the section 7.

7. Parasite Evolution and the Sustainability of Disease Control

7.1 Introduction

A common question addressed at geneticists aiming to breed disease-resistant animals is whether or not the parasite will evolve so as to overcome the genetic changes in the host. Note that the issue is correctly referred to as parasite evolution rather than parasite coevolution, as coevolution implies evolutionary (mutational) changes in both the host and the pathogen.

Potential parasite evolution is an issue of critical importance, yet it appears to have never been formally (theoretically) addressed by geneticists working in livestock disease resistance. Two experimental attempts to address parasite evolution will be described below.

Wider aspects of host-parasite coevolution have been extensively addressed by evolutionary geneticists and ecologists. This theory tends to be abstract, complex and not particularly relevant to animal breeding contexts, although there are exceptions as described in section 7.4. The theory is generally developed to account for naturally fluctuating host and parasite populations, and their interactions over evolutionary time periods, rather than parasite evolutionary dynamics in the context of a controlled host population. Conceptually, the controlled host population should make theoretical developments simpler rather than more complex. Additionally the vast bulk of the theory is concerned with the evolution of virulence, i.e. how damaging the parasite is to the host (there is a trade-off if the parasite is too damaging and kills the host), rather than the evolutionary dynamics of transmission per se or host resistance.

Absolute risks of parasite evolution are probably inestimable, and much effort could be wasted attempting to model this. However, a much wider and more relevant question is the **relative** risk of parasite evolution when comparing genetic control of disease with other competing control strategies. This relates to concept of the sustainability of disease control and is a major issue for all disease control measures. Therefore, the question of whether or not the parasite will evolve in response to genetic change in the host should perhaps be put to one side, and the question should be rephrased as follows: is parasite evolution more or less likely for genetic control strategies than it is for other competing (or complementary) strategies?

This section discusses general issues of host-parasite genetic interactions, co-evolutionary theory in relation to virulence, the transmission of infection and disease control. It then uses these arguments as a basis for determining relative risks of parasite evolution in different situations. Formal theory covering the application of these ideas to livestock disease genetics does not currently exist in a publishable form, so the arguments presented will generally be descriptive.

7.2 The Persistence of Genetic Variation in Host Resistance

The first key issue to address is why genetic variation in host resistance persists in populations when simple arguments suggest that natural selection should eliminate 'susceptibility' alleles and fix 'resistance' alleles, in the face of an infectious challenge.

Complex host-parasite interactions guide the evolution of both hosts and parasites, and are considered to be one of the major reasons for the maintenance of genetic variation in natural

populations. Combining genetic theory and epidemiology can give insight into why genetic variation in host resistance to infection persists. In simple terms these arguments include the following. Firstly, selection pressures, especially those for disease resistance, will act in different ways across time and environments. This maintains genetic variation for all traits, in both wild and domestic populations. Secondly, natural selection will not make populations of animals completely resistant to infection. This is because as natural selection moves a host population towards resistance, the selection pressure for resistance decreases. There exists, for each disease, a certain proportion of susceptible animals that can be carried in the population without exposing the population as a whole to risks of epidemics. Once the proportion of genetically susceptible animals falls below this level (e.g. $1\text{-}1/R_0$), selection pressure for resistance ceases. This is a form of density-dependent selection. Hence, genetic variation for resistance to infection remains in the population. Thirdly, modern domestic livestock populations have been selected for other characteristics. Together, these factors explain why genetic variation in resistance has been observed for diseases that have been comprehensively studied.

Many of these concepts are more formally encapsulated in the Red Queen hypothesis (van Valen, 1973) that predicts an evolutionary arms race between host and parasite. This hypothesis is based on the observation to Alice by the Red Queen in Lewis Carroll's 'Through the Looking Glass' that '*in this place it takes all the running you can do, to keep in the same place*'.

The Red Queen hypothesis states that interactions among species (such as hosts and parasites) lead to constant natural selection for adaptation and counter-adaptation, resulting in coevolving populations having fluctuating non-equilibrium dynamics. If evolution is slower than ecological changes that affect the distribution of hosts or parasites (which often it will be), then these Red Queen dynamics will arise (Khibnik and Kondrashov, 1997). The consequence of this will be genetic variability observed in both the host and the parasite. May and Anderson (1983) show the chaos dynamics of simple polymorphisms in both host and parasite, demonstrating how interactions between hosts and parasites tend to promote polymorphisms in both populations and how the changes in gene frequency may be steady, cyclic or chaotic.

In contrast to resistance, when there is genetic variation between animals within a population for tolerance, natural selection will push populations towards enhanced tolerance. This explains the presence of breeds with disease tolerance in areas where there is consistent and predictable exposure to infection. This phenomenon is particularly common in tropical animal production systems. It has been well described, for example, for tolerance to trypanosomosis in African cattle.

7.3 General Features of Co-evolutionary Theory

Traditionally co-evolution models have been built up using simple gene-for-gene models, in which a gene (or allele) in the host is matched by a gene in the parasite. However, greater insight into coevolution is obtained when coevolution is considered in an epidemiological context. May and Anderson (1990) consider the impact of virulence on the overall survival of the parasite, and demonstrate that whilst high levels of virulence are often counter-productive (the host dies and the parasite can not transmit to the next host), the conventional wisdom of parasite evolution towards lowered virulence is not necessarily the case.

The May and Anderson (1990) argument is essentially as follows. Natural selection will tend to maximise the R_0 of the parasite, and (all other things being equal) R_0 will be maximised by

having the host survive as long as possible. However, generally the harm done to the host will be related in some way to the production of the transmission stages of the parasite. Therefore, maximising R_0 involves trade-offs between the production of the transmission stages (therefore ultimately β) and the damage to the host (i.e. virulence, which should be small). If high fecundity can be obtained with minimal harm to the host, then conventional wisdom will apply, and coevolution will be towards reduced virulence. The outcome depends upon this relationship. Adapting the notation of May and Anderson (1990) to be consistent that that used previously in these notes, the following equation is obtained:

$$R_0 = \beta N/(\varepsilon + b + \gamma)$$

Where βN describes transmission in a population of size and density described by N, ε is the virulence (formally the disease-induced death rate of the host), b is the host death rate from other causes, and γ is the recovery rate (as previously defined). If β and γ are independent of the virulence ε, then clearly R_0 is maximised by minimising parasite virulence. However, if β, γ and ε are interrelated, then intermediate virulence may be optimal and hence selected for.

It may be noticed that the arguments of Anderson and May (1990) are focussed upon the parasite and are independent of the host. No assumption is made about the host, other than it is susceptible to infection. Gandon and Michalakis (2000) address the impacts of whether the host has qualitative (i.e. 'all or none' resistance) or quantitative resistance to the parasite. Not surprisingly, the qualitative 'all or none' resistance leads to an evolutionary reduction in virulence that is dependent upon the fraction of hosts in the population that are resistant. In other words, the fewer susceptible hosts there are, the less the damage that the parasite can afford to do. In contrast, with quantitative variability in resistance, the hosts that are more resistant are harmed less but allocate more of their resources to the immune section (in agreement with the costs *vs.* consequences arguments in section 5). Because the within-host growth rate of the parasite or parasite population is often correlated with the deleterious effects of parasites, such resistance can directly affect parasite virulence. The theoretical result obtained by Gandon and Michalakis (2000) is that the parasite will tend to evolve towards the same level of deleterious effect upon its local host, therefore increased resistance (i.e. decreased β) will lead to increased virulence.

The arguments put forward by Gandon and Michalakis (2000) at first sight appear to be unfortunate for animal breeders aiming to increase disease resistance. However, they are still incomplete results, ignoring a number of factors of importance, including the nature of the parasite, the mode of resistance (as described in section 7.4) and the nature of the genetic resistance. Moreover, they are asymptotic results, describing long-term expectations rather than the dynamics of the pathway towards the final outcome.

An important extension to coevolutionary theory is that of multiple infections (or super-infections) of the host, i.e. when the host is simultaneously infected by more than one species or strain of parasite. This has important applications in epidemic theory for animal breeders, when considering diseases such as mastitis. It is also important when formalising parasite evolution risks is relation to host genotype because in the case of incomplete resistance infections by the evolved parasite will essentially be super-infections, over and above the existing infection. It is interesting to note that the major impetus for super-infection coevolutionary theory came from HIV infections in humans, as HIV is a virus with a fast mutation rate and hosts are often infected by more than one strain.

Super-infection dynamics and their impacts upon parasite evolution are considered by many authors, including Gandon *et al.* (2001a). In the case of super-infections, within-host competition between the different parasite types arises, and as a first generalisation, this favours enhanced virulence. However, the ultimate outcome of multiple infections will be a complex interrelationship between (i) the virulence of the different strains, (ii) the relationships between virulence and dominance of one strain over another, and (iii) the impact of virulence on host mortality - hence the density of susceptible animals and the potential for transmission of infection to other susceptible animals. It is probably sufficient to say that the evolutionary dynamics of super-infections can differ qualitatively from those of single infections. However, the exact natures of the differences are complex and depend upon the input assumptions.

7.4 Application of Joint Evolutionary and Epidemic Theory to Disease Control

The theory briefly outlined so far is put into better perspective when applied to a specific disease control problem. An excellent example is provided by Gandon *et al.* (2001b). These authors use an epidemic model to predict parasite evolutionary potential against different types of vaccines, especially 'imperfect' vaccines that have less than 100% efficacy. The models are then applied to anti-malarial vaccines, however the arguments would be equally valid in animal production contexts. An example would be for vaccines against Marek's disease in chickens, where each new generation of vaccine has apparently resulted in a new, more virulent (pathogenic) strain of the virus (Witter *et al.*, 1998).

The model of Gandon *et al.* (2001b) is a super-infection epidemic model, accounting for the dynamics of both the original strain of virus and a mutant variety. Trade-offs are incorporated into the model such that an arbitrary and empirical relationship between transmission and virulence is defined (e.g. $\beta = a\varepsilon^c$), and similarly an arbitrary and empirical relationship between the recovery rate and virulence is defined (e.g. $\gamma = d\varepsilon^e$). Critically, however, vaccines with different modes of action are considered. The mode of action of the vaccine was assumed to be either (i) anti-infection, corresponding to the host genetic susceptibility parameter **g** in section 4.2.2, (ii) anti-growth (population growth in a microparasitic context) (a parameter not used in the genetic context so far), (iii) anti-transmission, corresponding to the host genetic infectivity parameter **f** in section 4.2.2, (iv) anti-toxin, corresponding to genetic tolerance, or (v) a combination of these modes of action.

Under the assumption that each of the vaccines had 80% efficacy, the mode of action of the vaccine had marked impacts upon the long-term likelihood of parasite evolution. Vaccines that decreased susceptibility resulted in slightly reduced parasite virulence; vaccines that decreased infectivity had no impact; vaccines that increased tolerance lead to slightly greater virulence; vaccines that reduced parasite population growth lead to markedly greater virulence. Under some circumstances, the last example actually led to the vaccine increasing rather than decreasing malarial deaths (in the long term).

Timescales over which these effects are likely to occur are more difficult to predict than the actual long-term expectations. In the simulations of Gandon *et al.* (2001b), at 90% vaccine coverage with an anti-growth vaccine of 80% efficacy (a worst-case scenario), it took 38 years for a mutant more than twice as virulent to increase from 1% to 50%, after which time it spread rapidly towards fixation. However, the speed of invasion of mutants is very susceptible to the shape of the trade-off functions described above. These are generally assumed, although

there is laboratory animal data on these relationships (e.g. Mackinnon and Read, 1999). Different relationships and trade-offs for different host-parasite situations will potentially lead to very different outcomes.

The methodology of Gandon *et al.* (2001b) provides a potential starting point for considering parasite evolution in an animal breeding context, especially when relative risks compared to other disease control strategies are considered. The vaccine context is analogous to a single gene effect in animal breeding, where the 'resistant' allele fails to reduce β to zero. Therefore, additional complexities in the genetic model would almost certainly be required. Additional factors that would have to be considered are outlined in section 7.6.

7.5 Sustainable Disease Control

It is true to say that all disease control measures that aim to reduce parasite numbers can lead to genetic changes in the parasite to evade the control strategy. This is best documented in cases of control by chemicals and antibiotics. Examples previously given in section 1 include:

- The evolution of resistance to anthelmintics by nematodes in all major sheep producing countries (e.g. Jackson, 1993; Waller, 1997). This threatens sustainable sheep production throughout the world. Resistance has developed to all classes of anthelmintic, and no new classes are currently being produced by pharmaceutical companies.
- The evolution of resistance to antibiotics by bacteria (e.g. Usera *et al.*, 2002). This threatens the efficacy of antibiotics in many production systems. This is especially problematic in intensive production systems where antibiotics are used to control unknown and sometimes sub-clinical disease problems.
- The evolution of resistance to anti-protozoals. For example, there is widespread resistance to drugs used to control animal trypanosomosis (e.g. Eisler *et al.*, 2001).
- The evolution of acaricide resistance by populations of ticks (e.g. Spickett, 1994), leading to greater problems of tick infestation and tick-borne diseases.
- The evolution of vaccine resistance. For example, each new generation of vaccines against Marek's disease has apparently led to strains of ever more virulent viruses (e.g. Witter, 1998).

Sustainable disease control is characterised by avoiding reliance upon a single control strategy and attempting to combine complementary approaches for controlling the transmission of infection. A strategy that is reliant upon a single mechanism, e.g. using a chemical that kills the parasite in a specific way, essentially gives the parasite a single target against which to mutate. The parasite that has a mutation that allows it to avoid that control strategy will have a huge selective advantage over the rest of the parasite population, hence that genotype will propagate and become dominant, leading to the often observed breakdown in the control strategy.

Disease control strategies that combine different approaches will generally be more sustainable, as parasites with the mutation allowing them to escape one strategy will still be susceptible to other strategies. Genetic and non-genetic strategies may be combined complementarily, and this would be advocated under most circumstances. Moreover, genetic strategies depending upon more than one gene may also be considered to be a form of combined control strategy.

The point to note is that essentially all disease control strategies (indeed, all technologies)

have a finite shelf life. All strategies are prone to breakdown eventually, and a realistic aim should be to mix and match strategies to maximise their longevity as well as their current utility. It must be assumed that in the long-term, parasite populations will evolve. Therefore the question to pose is what is the likely longevity of genetic control measures in relation to other control measures, and what are the factors affecting this relative longevity?

7.6 Considerations for Animal Genetics

It should be noted at the outset that although there are examples where genetic resistance to diseases in plants has failed, there are currently no recorded examples of this occurring in domestic animal populations. Factors which affect relative risk are now discussed.

Resistance vs. Tolerance
The primary nature of the disease resistance will have a large impact upon parasite evolution risks. If the disease resistance is by means of resistance to infection, then selection pressure will be placed upon the parasite such that parasites with mutations enabling them to overcome the host resistance will have a selective advantage.

Risk of genetic change in the parasite is minimal when it is tolerance of infection that is being selected for. In this case the parasite can complete its lifecycle unimpeded. Thus, there is no selection pressure placed upon the parasite to change its transmission characteristics. Many apparent differences seen between populations in disease occurrence are due to tolerance of infection, and these differences have often been stable for very long time periods, e.g. thousands of years. This is especially true when observing the tolerance characteristics of locally adapted tropical breeds of livestock in regions where certain diseases are endemic, e.g. for tick worry, tick-borne diseases and trypanosomosis.

Selection pressure
Rates of genetic change in the parasite population will be proportional to the selection pressure placed upon them. This is extreme in the situation of interventions such as vaccines or chemicals that kill essentially all the parasites except those that are resistant.

For genetic change of host populations, the selection pressure dynamics differ considerably from those for chemical intervention for several reasons: (i) genetic resistance is seldom complete, (ii) genetic resistance varies between hosts, and (iii) genetic change is slow (even when introgressing resistance alleles). Therefore, selection pressures on the parasite are usually somewhat weaker than when chemical intervention is used.

The impact of genetic variability amongst hosts upon selection pressure dynamics is best illustrated for macroparasitic infections that show an aggregated distribution amongst hosts. In this case, the majority of the next generation of parasites are shed from the most susceptible hosts, i.e. the dynamics of the parasite population genetics are determined not by the most resistant animals but by the least resistant animals. Mutations will generally be deleterious for parasite fitness, therefore mutant parasites from the most resistant hosts will be at a selective disadvantage compared to wildtype parasites shed by susceptible hosts. The change in parasite population dynamics will be proportional to the weighted average of the number of parasites shed by each genotype of host multiplied by the relative fitness of each the parasite sub-populations. This will result in only slow change in the parasite population.

Genetic complexity of resistance

The evolutionary challenge the parasite faces will be a function of the number mechanisms of resistance that the host employs, i.e. the number of genes governing resistance. This is directly analogous to the sustainable disease control arguments given above. The greater the number of genes governing resistance, the greater the difficulty the parasite will have in overcoming all barriers. From a sustainability viewpoint, this mitigates against single genes for resistance.

Genetic diversity

The genetic diversity of host animals poses challenges to the parasite at the between-host rather than the within-host level. Mutations that are beneficial in a host of one type will not necessarily be beneficial in hosts of different genotypes. Thus at the population level, enhanced genetic diversity should help to protect the host population from mutant parasites gaining advantage.

Host and parasite generation intervals

Rates of successful parasite evolution will depend upon the time-dependent opportunity for parasite to propagate mutant genotypes, in relation to genetic change in the host. This is a function, amongst other things, of the relative generation intervals of the host and the parasite. These will invariably favour the parasite, but the relative generation intervals will vary dramatically for different types of parasite. Viruses and bacteria replicate very quickly, hence will have many generations per host generation. However, for some macroparasites such as nematodes, the ratio of parasite to host generation time can be surprising low. For example, many nematodes may typically have two generations per year whereas their host may have generation intervals as low as two years, giving a ratio of only 4 to1.

Mechanism of resistance

In addition to the distinction between resistance and tolerance described above, the actual mechanism of resistance is important, in a manner analogous to the vaccine-induced evolution arguments of Gandon *et al.* (2001b). If resistance is mediated by genes controlling population growth within the host or the transmission of infection from the host, then the resident population of parasites within the host has an excellent opportunity to propagate beneficial mutations. However, if resistance truly is the ability to resist infection, then opportunities for the parasite population to evolve are much less.

Parasite meta-population dynamics

The meta-population dynamics of the parasite, i.e. the interactions between the different sub-populations, will also determine the likelihood of a successful mutant genotype becoming dominant in the population. The basic concept is that the parasite population in which there is evolutionary pressure is only a subset of the total parasite population, and the interaction of the sub-population with the whole population will determine the success of the mutant genotype. This concept holds for both micro- and macroparasites and will be illustrated separately.

Consider nematode infections in ruminants. At any point in time, a proportion of the total parasite population will be within the host and a proportion will be on the pasture. Only the parasites within the host will be subject to selective pressure. The overall rate of parasite evolution will be proportional to the relative sizes of the two sub-populations. This is well illustrated by the relative rates of evolution of anthelmintic resistance in goats compared to sheep. Goats are considerably more susceptible to nematode parasites than sheep when forced

to be obligate grazers, and as such a greater proportion of the meta-population is found within the host than for sheep. A simple prediction from this is that the rate of evolution of anthelmintic resistance would be greater in goats than in sheep, and this has indeed been shown to be the case in a number of studies (e.g. Jackson, 1993; Chartier, 1998). Therefore, in terms of parasite evolution against host genotype, husbandry or disease management strategies that minimise the proportion of the meta-population that is within the host should also reduce the risks of successful parasite evolution.

Meta-population dynamics for microparasitic infections depend upon the interactions of the selected host population and neighbouring unselected populations. Parasites that have previously been hosted by unselected animals are extremely unlikely to have evolved resistance necessary to enable them to compete in the selected host population. The risk comes from within the population itself. In principle, complete resistance to infection should result in coevolution risks being minimal, with disease eradication resulting. This is because there is no actual circumstance in which the parasite can effectively evolve. However, partial resistance does pose a risk as the parasite can potentially gain a foothold and evolve accordingly. In this case, minor epidemics pose the greatest coevolution risk, as parasites with the desirable mutation can then spread unimpeded through the sub-population of uninfected hosts.

Genetic markers vs. indicator traits
As discussed in section 8, selection for disease resistance may either by means of genetic markers linked to genes conferring resistance, phenotypic indicator traits describing the response of an animal to infective challenge, or a combination of both. Arguably, in the case of indicator traits, parasite evolution is not a major concern, as selection will always be against the current parasite genotype. With genetic markers, there is a danger that parasite evolution may go unnoticed and marker-based selection may no longer be effective.

7.7 Experimental Investigations of Parasite Evolution

Several studies have experimentally investigated parasite evolution in response to host resistance status, for nematode (*Haemonchus contortus*) infections in sheep. In an Australian study using unselected sheep, passaging *H. contortus* through immunologically susceptible and resistant animals for six and nine parasite generations, respectively, failed to elicit an evolutionary response in the parasite (Albers and Burgess, 1988). A similar study involving five serial passages of the same parasite in (immunologically) resistant and susceptible hosts also resulted in no host-acquired responses (Adams, 1988). Likewise, a study using immunologically resistant and susceptible Barbados Black belly sheep in which *H. contortus* was passaged for 10 generations through the respective hosts also failed to show signs of parasite evolution (Saulai *et al.*, 2001). Lastly, Woolaston et al. (1992) failed to find any evidence that *H. contortus* is adapting to sheep bred for resistance to this parasite.

Whilst these studies are inevitably small scale and short term from an evolutionary perspective, and do not constitute proof that parasite evolution will not occur in these situations, they are nevertheless encouraging for the geneticist. The number of parasite generations in these studies is equivalent to approximately three years, although somewhat longer if the meta-population dynamics are taken into account. Additionally, the difference between immunologically susceptible and resistant hosts would take many years to achieve through selection. Thus, together these results provide reassurance that parasite evolution in

the context of nematode infections should not be a primary concern in the short to medium term.

7.8 Overview

The theory relating parasite evolution to host genotype in an animal breeding is not well developed; however some broad qualitative conclusions can be drawn.

Firstly, parasite evolution is a risk, however the risk is very dependent upon the situation being investigated. It should be remembered that all technologies have a limited shelf life, including other non-genetic control strategies. Long-term possibilities of failure should not stop the short and medium term benefits being exploited.

Secondly, under most circumstances genetic approaches to controlling disease should be at least as sustainable as other control measures. In particular, they are likely to be more sustainable than control by chemicals or vaccines.

Thirdly, risks of parasite evolution vary between different genetic strategies and disease scenarios. As a broad approximation, macroparasite resistance is likely to be less prone to being overcome by parasite evolution than microparasite resistance. Also, genetic control based on the combined effects of several genes is likely to be sustainable for longer than control based on a single gene. Selection based on measured response to the infectious challenge is always likely to be appropriate, as it will select animals against the current parasite strain or genotype.

In summary, it appears that risks of parasite evolution following selection for disease resistance are insufficient to cast doubt on the use of genetic selection as a disease control strategy. Indeed, when compared with competing control strategies, selection for disease resistance becomes an attractive option in terms of parasite evolution risks. However, caution should be exercised, and sustainable disease control strategies will include as diverse a combination of control strategies as is feasible.

8. From Theory to Practice

8.1 Introduction

The aim of the preceding chapters has been to develop the theory and conceptual framework within which rational decisions can be made regarding disease genetics. Application of the theory should enable the breeder to decide (i) whether or not a disease is likely to be amenable to genetic strategies and whether these will truly result in sufficient benefits within a reasonable time scale, (ii) plan and implement the necessary research, (iii) analyse the data in the most meaningful way, and (iv) design and implement effective breeding programs. This section aims to consider each of these topics in turn.

Almost certainly, the nature of the application will vary for different diseases and production systems; thus it is not possible to describe strategies that will be generally applicable to all situations. Therefore these notes are not prescriptive, i.e. they do not give recipes. Rather, they aim to outline basic concepts that will be extended in the group sessions. It is assumed that the principles of animal genetics and genetic improvement are known. Therefore, the emphasis is on the additional knowledge and concepts required for the implementation of disease genetics.

The key to the correct application of genetic-epidemiological theory and concepts to actual diseases situations is a good understanding of the disease biology. In particular, this is critical to ensure that concepts are not misapplied to inappropriate disease situations. For example, the results from macroparasitic infections are generally not directly applicable to microparasitic infections.

8.2 Current Status of Disease Resistance Genetic Improvement Programs

The starting point for considering the transition from theory to practice is an appraisal of the current situation regarding the implementation of breeding programs for disease resistance. Despite the large number of well-documented examples of genetic variation in resistance (Table 1.3), there are surprisingly few cases of structured commercial breeding programs for disease resistance. The best-known examples are for nematode resistance in sheep, tick resistance, mastitis and Marek's disease. For nematode resistance there are breeding programs underway in New Zealand, Australia and the UK; tick resistance is an integral part of cattle breeding programs in subtropical Australia; mastitis resistance, either in terms of clinical cases of mastitis or reduced somatic counts, is incorporated into many dairy cow and sheep breeding programs; and Marek's disease resistance is a feature of modern chicken breeding. Breeding for resistance to post-weaning *E. coli* diarrhoea in pigs, where the causal mutation or closely linked markers are known, has begun in some selection programs. Additionally, scrapie resistance selection is now underway in several countries in Western Europe, although arguably this is imposed rather than a choice by breeders. Although this list is not exhaustive, it must be concluded that despite many opportunities there are limited convincing examples of breeding schemes for disease resistance.

The reasons for the relative lack of convincing examples of breeding schemes for disease resistance are many and varied. They probably include:

- A belief by many practitioners: veterinarians, scientists and animal producers in the adequacy of existing disease control strategies, even when these are failing.

- The lack of awareness of the opportunities. This is partly a function of the interdisciplinary nature of the skills required to exploit these opportunities. For example, veterinarians traditionally have had a poor appreciation of genetics and genetic opportunities, whereas geneticists have often had a poor understanding of diseases and are not able to prioritise diseases or predict the consequences of selection for resistance. Additionally, communication between these groups has often been poor. Geneticists are now generally becoming more aware of the opportunities for breeding for disease resistance, however this awareness is relatively recent and there has been limited opportunity to translate awareness into actual outputs.
- The lack of relevant and effective infrastructure to exploit the genetic opportunities. Exploitation often requires considerable information, infrastructure (technology), cooperation and will. Most obviously, geneticists have often had trouble obtaining animal health data. Such data is often collected in large amounts by veterinary scientists, but not in a way that is amenable to genetic analyses.
- The lack, until now, of a framework for decision-making, i.e. the prioritising of diseases or health scenarios in terms of genetic opportunities.

These issues are all relevant to livestock production in developed countries. They are exacerbated in developing countries where often a lot of the basic technology and infrastructure may be absent. Ironically, the appropriate technology to implement rational breeding strategies for disease resistance may be much simpler in developing countries than in developed countries. For example, the most appropriate breeding decision may simply be to reject exotic genotypes that are almost certainly susceptible to local endemic diseases, and utilise locally adapted breeds, instead.

8.3 Decision-Making: Choosing the Right Disease

There are many potential target diseases for the geneticist. Indeed, there are many more diseases than can ever feasibly be addressed. Therefore, the first use of genetic-epidemiological modelling techniques, either formally or simply by using the logic that underpins these techniques, will be in the choice of suitable target diseases. Two simple examples will illustrate the informal application of this logic.

- It is unlikely that breeding for resistance to foot and mouth disease (FMD) would be a sensible strategy for the UK livestock industries, even if it were possible. FMD is a highly infectious disease with a high R_0. Simple considerations show that for diseases of this nature it is necessary to have a high proportion of animals completely resistant to the disease before the population as a whole is protected. This would take many decades to achieve. In the meantime, any epidemic (from which the population would NOT be protected) would result in large-scale slaughter of animals, if the disease control strategy was the same as in 2001. Thus all breeding efforts would be undone. In this example, current disease control strategies override genetic approaches.
- For a zoonotic disease it would be unwise to breed animals for apparent resistance, if this apparent resistance were in fact tolerance of infection. Such breeding would ignore the cause of the problem and merely hide the symptoms of disease, thereby potentially exacerbating the human health problem.

From a disease genetics viewpoint, decision-making involves correctly matching (a) tools and technologies with (b) disease and environmental scenarios with (c) animal genetic resources. Initially, a series of primary questions must be answered:

- Are there diseases of major (economic) importance needing to be addressed?
- Are current control strategies adequate, sustainable and cost-effective?
- Do current animal genetic resources cope with these disease challenges?

If the answer is "yes" to the first question and "no" to either of the other questions, then genetics (e.g. selection) may have an important role to play in dealing with the infection and the subsequent disease.

At this stage it is necessary to do an economic appraisal of the disease (formal or informal) to ensure that the disease issue is indeed sufficiently important to warrant either a research program or (and) a breeding program. Further questions may then be asked of the animal genetic resources. For example, is there evidence of differences between breeds or between animals within breeds in their resistance or tolerance? The answers to these questions may come from hard data, from anecdotes, or from inferences made from other populations or closely related diseases.

Investigation of the disease using a genetic epidemiological model is now strongly recommended. At this stage there will almost certainly be a gap in the biological data necessary to parameterise the model. However, attempting to construct the model will particularly focus attention of the biology of the disease, the transmission pathways and the areas where knowledge is lacking. This exercise will have many benefits, including:

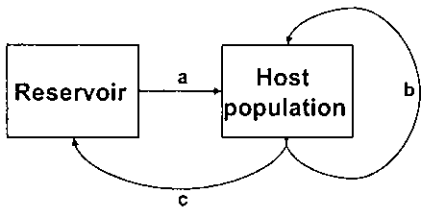
1. Indicating the nature of the benefits that can be achieved through using genetics to help control the disease, e.g. will the benefits be reduced animal mortality, increased performance or decreased risks of epidemics.
2. Enabling economic appraisals of the benefits of selection to be achieved
3. Helping to derive economic weights for resistance, if resistance is to be included in a multi-trait breeding objective. This discussed further in section 8.7.
4. Focusing attention on areas where knowledge is lacking in terms of the disease biology and disease genetics
5. Helping to plan effective experimental studies.

8.4 Research Strategies

Nature of challenge

Research strategies to investigate disease genetics will depend critically upon the nature of the disease. Consider the pathways of transmission of infection, shown again in Figure 8.1.

Figure 8.1. *Summary of pathways of infection for diseases in domestic livestock.*



The critical first question is whether or transmission of infection from the reservoir to the host population along pathway a is sporadic or predictable. This will often determine whether or not field data describing individual animal response to the infectious challenge is adequate, or whether deliberate challenge experiments need to be set up.

If there is some degree of continuous or predictable challenge, i.e. animals in the field are challenged by the parasite at specific predictable times, then phenotypic measurement of animals' responses to natural challenge is a possibility. In that case the next questions to be addressed are whether or not there are measurements that show (i) if the individual animal has been challenged, (ii) the degree to which it has been challenged, (iii) the response of the animal to the challenge and (iv) the degree of disease suffered by the animal? Such measurements should also be informative about the state of the infection, e.g. is the infection latent or is the individual animal fully infectious? Additionally, carrier state animals should also be identifiable.

Field data will mainly be appropriate for endemic infections, such as tick infestations, mastitis, nematode infections, or microparasitic infections that have reached an endemic equilibrium within the host population, such as is often seen with PRRS infected pig herds. Field data from epidemics may sometimes be appropriate for use, but it will require the breeder to be opportunistic, i.e. to be able to collect the disease data in herds or flocks where the recording infrastructure for pedigrees and performance is already in place.

In the case of infectious challenges not being predictably present in the field, then deliberate challenge experiments may need to be set up. This will almost certainly be considerably more expensive than field data obtained from natural challenges, performed on far fewer animals, and disease containment facilities may be required. It may also raise additional issues, such as animal wastage because infected animals may be not suitable for further performance or breeding (e.g. the challenge protocol might require subsequent slaughter of infected animals). Ethical concerns are also obvious with deliberate challenges, and ethical 'cost-benefit' analyses may be required before experimentation is permitted. In terms of breeding scheme design, deliberate challenges can lead to inefficiency if the challenged animals are subsequently not available for breeding.

Under some circumstances, an alternative to challenges with the live parasite may be to measure animal response to a vaccine. This may work because the vaccine challenge should resemble parasite challenge, albeit with a considerably reduced severity. The advantage of deliberate challenges over natural challenges lies in the control of the challenge; it is known exactly how much parasite (or vaccine) that the animal has been challenged with, and it is known that all animals have faced the same challenge as far as it is possible. Therefore, assumptions about the degree and timing of challenge are not required, and the data should be more accurate.

Interpretation of data

No matter whether it is a natural or artificial challenge that the animal faces, considerable care must be taken over the integrity and interpretation of the data. Several issues are raised here, some of which are not obvious to geneticists or to animal health researchers.

Firstly, wherever possible, continuous data (e.g. bacterial or nematode count) should be collected in preference to binary data such as 'infected or not'. In general, continuous data is more informative, easier to analyse and usually more heritable than binary data.

Secondly, the statistical properties of the measurements should be understood, particularly the accuracy and repeatability of the measurement. Often, the imprecision of 'disease' measurements

may not be fully appreciated by veterinary or immunological researchers. For example, in a detailed assessment of the precision of faecal egg count measurements in lambs infected with nematodes, it was demonstrated that 1/3 of the phenotypic variation in egg counts was due solely to imprecision in the laboratory measurements – this is variation between replicated measurements performed on the same samples (Bishop *et al.*, 1996). This highlights the importance of replication of measurements. It also highlights differences in results (e.g. heritabilities) that can occur if some laboratories replicate their measurements, reporting the average measurement, and others do not to this.

Another example of the scrutiny of the statistical properties of measurements comes from the consideration of various immunological measurements that were considered as candidates in a pig generalised immunity study (Clapperton *et al.*, 2003). Here, many potential measurements were assessed according to three criteria: (i) the precision of measurement (i.e. the repeatability of replicated measurements made on the same blood sample), (ii) the across-time repeatability and (iii) the stability of measurements across time for groups of animals. The last statistic is a measure of the constancy of the invisible 'continuous challenge' across time. According to these three criteria, several measurements were rejected as being measured too imprecisely (e.g. $r < 0.9$), insufficiently repeatable (e.g. $r < 0.5$) (potentially reflecting a low heritability), or too variable across time for the pigs as a whole. This procedure will tend to reject measurements that are too complex and favour simpler measurements – this is advantageous as simpler measurements are more easily scaled up for larger studies. Scaling up measurements to whole populations may be a conceptual challenge to scientists used to working with small-scale experiments, e.g. 8 animals per treatment.

A third issue is the interpretation of phenotypic indicator traits: do they give an appropriate (rather than accurate) assessment of the infection or disease status of the host animal? The appropriateness of indicator traits is often assumed, but it does need to be checked. The disease biology can sometimes lead to unexpected results. For example, investigating nematode egg counts in lambs naturally infected with *T. circumcincta*, Bishop and Stear (2000) found that egg counts rose linearly with low worm burdens, reached a plateau, then actually declined at high worm burdens due to severe density dependent constraints. This result is specific for this nematode species and should NOT be extrapolated to other nematode species; however such relationships should be investigated for other parasite species. From a geneticists viewpoint, relationships between disease traits should be investigated at the genetic rather than phenotypic levels, as the imprecision inherent in many disease indicator measurements can lead to very low phenotypic relationships. For example, returning to the nematode example, egg counts measured across time on the same animal are surprisingly lowly correlated phenotypically, with this correlation little higher than the heritability, yet equivalent genetic correlations are close to unity indicating that it is the same underlying trait being measured. Therefore: make sure the indicator trait makes biological sense, don't get too gloomy if phenotypic relationships are disappointing, and investigate relationships at the genetic level.

Finally, it is necessary to understand the actual interpretation of the measurements collected. For example, resistance (as strictly defined in section 1) may be indicated by measurements such as parasite burden or the number of parasites (or eggs) shed by the host, provided the degree of infectious challenge is known. Performance of infected animals is not a measure of resistance; rather it **may** be an indicator of tolerance (or resilience). Strictly speaking, tolerance will be defined as performance conditional upon infection status. Likewise, mortality-related traits in infected animals are indicators of parasite virulence, when comparing alternative parasites. When comparing animals infected with the same parasite they are crude indicators of tolerance (or

extreme **lack** of tolerance). Time until infection in naturally infected populations is an indicator of resistance. However, it must be realised that there are several epidemiological assumptions implicit in this measurement. These relate to the transmission pathways and the force of infection (the degree of infectious challenge), and these factors may be confounded with the expression of the trait. Therefore, time until infection might be expected to be an underestimate of the heritability of resistance, hence somewhat lowly heritable – as indeed it is (e.g. Henryon *et al.*, 2001).

Phenotypes vs. genetic markers

The aim of many researchers is to find disease resistance genes of QTL, thus enabling (i) selection in the absence of infection and (ii) potentially much faster responses to selection. Whilst responses to selection for disease indicator traits may be relatively fast, e.g. up to 5% per annum (because these traits often have large standard deviations), resistance genes may allow substantially faster progress. This arises simply from the nature of resistance, where there is an upper bound to the trait. For example, a gene may confer complete resistance to infection. This is quantitatively a much larger step than that normally achievable though directional selection. This favourable comparison of the efficiency of marker or gene assisted selection with phenotypic selection contrasts with performance traits. For performance traits, usually there is no upper bound on the trait value and gene-based selection is (at best) only marginally better than phenotypic selection, in situations where the trait phenotype can be measured.

The disadvantage of selection based on genetic markers is the enhanced risk of parasite coevolution (N.B. this is only a risk, not a certainty), accompanied by the danger that selection may no longer be based on animal response to the current relevant parasite genotype. These issues were discussed in section 7.

Experimental population designs to detect genes (or QTL) differ from those used to investigate the quantitative genetic properties (e.g. heritabilities) of traits. Gene/QTL detection using linkage analyses generally requires the creation of large families within which genes are likely to be segregating. Genes/QTL are then detected by the co-segregation of alleles with phenotype. For this reason, diverse crosses between populations thought to differ for disease resistance are often used as the starting point for creating segregating families. Conversely, studies aiming to estimate heritabilities or other genetic parameters require a comparison of within- and between-family variation, **within** a population. Such genetic parameters are population-specific, and using crosses of divergent populations will result in biased parameters that are difficult to interpret and seldom applicable. Within-population studies that allow simultaneous estimation of genetic parameters and genes/QTL can be designed, however to meet the requirements of both objectives they will need to be very large scale and hence expensive.

Whilst genes/QTL can allow selection for resistance or tolerance in the absence of infection, hence avoiding phenotypic measurement, it must not be forgotten that phenotypes of challenged animals are required to detect genes/QTL in the first instance (or to verify the effects of genes/QTL previously detected in different populations). Hence, genetic marker technology will not negate the need for phenotypic measurement of infected animals at some stage.

8.5 Data Collection

The collection of disease and health data presents one of the biggest challenges to animal geneticists. This is generally not such a problem with planned experiments, where the challenge lies in simply scaling up measurements normally performed on a few animals to several hundred animals (minimum!). However, the collection of animal health data can pose a considerable challenge for field studies and breeding programs. Also, data collection must always be planned with the data interpretation in mind.

At the field scale, the key to successful data recording lies in successful combining of animal health and animal performance recording protocols. Often, data may be collected for epidemiological studies with similar information recorded as geneticists would typically record. However, factors not recorded may include actual individual animal performance and parentage. Bridging this gap, or refocusing the studies onto herds where this information is collected remains the challenge.

This problem has been solved in different ways for different diseases. For example, mastitis represents one of the simplest diseases to record under field conditions, as an easily assessed indicator traits (somatic cell count) is readily and cheaply measured on small milk samples. Additionally, clinical mastitis is relatively easily diagnosed, leading to comprehensive datasets being created in a fairly straightforward manner. Likewise, for degree of infection by nematode parasites, faecal samples are easily collected by farmers and sent to laboratories with expertise in enumerating egg counts. In this case, the cost of measurement becomes a greater issue, and it may be difficult for this to be borne by individual farmers. However, within structured group breeding schemes the costs may be shared by several participants, making it feasible. For the main tropical nematode parasite problem, *Haemonchus contortus*, eyelid colour may be easily assessed as an indicator of anaemia, i.e. tolerance of infection.

Alterative assessments of disease problems that are sometimes used are individual animal treatments. This may be effective when animals are diagnosed and treated on an individual animal basis (see Henryon *et al.*, 2001), however in some situations it can cause interpretational problems, especially if animals are group treated – e.g. in the cased of attempting to use treatment requirements as an indicator of nematode resistance. This method of data collection obviously also requires accurate diagnoses of animals.

In general, the more clearly defined the disease problem, the more meaningful will be the collected data. Data describing general aspects of health or onset of a specific disease condition, when this is one disease out of many, will generally be harder to interpret and less heritable than data describing a specific infection. This arises both from the non-specific nature of the data and also from the epidemiological assumptions implicit in the data (the severity and timing of infectious challenge will be unknown). Censoring is also implicit in this data – if an animal gets one disease then it hasn't had a full opportunity to get competing diseases! Therefore, we often find that non-specific studies of animal health conclude that disease resistance is lowly heritably, whereas focussed studies often conclude that resistance or immune response to a particular infection is moderately to highly heritable.

The challenge to the geneticist is to focus, wherever possible, on well-defined problems using verified measurements that are meaningful from a disease biology and epidemiological viewpoint. Larger less focussed studies may be successful; indeed they may often be necessary first steps in the identification of important health problems and in the identification

of diseases that are likely to be amenable to genetic approaches. However, in order to be successful, they will require considerable cooperation between geneticists, veterinarians and epidemiologists.

8.6 Data Analysis

As with all animal breeding and genetics data, data analysis and data interpretation can present challenges. In general, the methodological challenges are similar to those encountered generally in genetic data analysis, however the interpretation challenges may differ somewhat. This section discusses several general issues relating to the methodological approaches, specific techniques, interpretation of data in relation to the underlying epidemiological models and specific problems encountered in some analyses.

To be a Bayesian or a frequentist?
Two contrasting approaches implemented by animal geneticists for analysing their data are 'frequentist' or 'Bayesian' techniques. Sometimes there is tension between the advocates of the two approaches. Frequentists utilise the data at hand to obtain point estimates of the parameter of interest, and usually construct standard errors to describe the distributional properties of the estimator. The interpretation of the standard error is that it is the standard deviation of the distribution of estimated parameters, if the experiment were to be repeated many times. Properties of the estimator are usually based on asymptotic assumptions, i.e. that the data set is extremely large. Typically, frequentists use techniques based on Maximum Likelihood (ML), e.g. Restricted or Residual ML (REML). Bayesian techniques provide marginal posterior distributions of parameters of interest, given some a priori assumption about the parameter distribution and the data itself. Note that the outcome is a distribution of parameters, rather than a point estimate. In the case of asymmetric posterior distributions a further choice may have to be made whether to use the mode (analogous to the ML solution), the median or the mean if a single description of a parameter is required. The latter two imply 'loss functions' describing the impact of (or risk associated with) incorrect choices. The median is chosen if the loss function is linear, the mean if the loss function is quadratic. Bayesian techniques are generally implemented through the use of computer-intensive Monte-Carlo Markov Chain (MCMC) techniques such as Gibbs Sampling or the Metropolis-Hastings algorithm.

An excellent summary of the issues associated with Bayesian and frequentist techniques is given by Blasco (2001). To paraphrase from the conclusions of this paper, if the geneticist is not concerned with abstract philosophical issues but is more concerned with tools to solve problems, then either approach can be used satisfactorily and justifiably. For both approaches there is software available to solve a wide variety of problems. The important issue for the geneticist to address is whether or not there is software or amenable techniques for the specific problem that she is interested in.

A subjective comparison of frequentist *vs.* Bayesian techniques is that frequentist techniques such as REML will give answers to straightforward problems more quickly and easily than Bayesian techniques, and with less danger of going horribly wrong. However, there are many techniques for which explicit frequentist techniques or solutions are not available. A well-known example is segregation analyses in complex pedigrees; here MCMC techniques will provide solutions whereas frequentist solutions may not be available. Data augmentation is a further situation where application of MCMC techniques will provide solutions. For example,

in the search for disease resistance genes or QTL there may be incomplete genotyping of a population: here MCMC techniques may be used to infer genotypes for animals without genotypes using pedigree information, adding to the precision of the parameter estimates (PongWong and Woolliams, 1996). In general terms, Bayesian techniques may be applicable to any complex analytical situation, provided that posterior distributions of parameters given the data can be derived. This may be particularly beneficial when complex epidemiological parameters are being derived. Further, from an epidemiological viewpoint, the posterior probability distribution supplied by Bayesian techniques may be useful for indicating credible upper and lower bounds for certain parameters.

The nature of disease data

The most fortunate situation for the geneticist is where the data is continuous and shows a Normal distribution (or a Normal distribution after simple transformation). In this case, standard linear model techniques can usually be applied to the data. Often, however, disease data is not of this nature. For example, it may be binary (e.g. infected or not; dead or alive), a mixture of binary and continuous, or survival type data (time until an event). In the last case, the data will almost certainly also be censored.

Typically disease data is either static or dynamic. Static data will describe, for example, whether or not a cow had mastitis at a particular time point. Statistical techniques capable of handling binary data will be appropriate in this case. Dynamic data will ask the question: when did the cow get mastitis? Survival analyses may be appropriate for this type of data, and cows not getting mastitis during the measurement period may be regarded as censored observations. Mixture distributions will arise from static data, when measurements of disease severity are continuous. In this case, the data will comprise (apparently) uninfected animals, along with animals infected to a varying degree. Repeated static observations, e.g. nematode egg count measurements across time of the same lambs, will lend themselves to longitudinal analysis techniques such as random regression.

Binary data analysis

Genetic parameter estimation from binary data using a threshold liability model has been the subject of much research and debate. Estimation of genetic parameters from binary data is now well established in a Bayesian framework (e.g. Kadarmideen *et al.*, 2001), and also through the application of frequentist generalised linear mixed models using, for example, a probit link function (e.g. Kadarmideen *et al.*, 2000, on the same data).

However, it is instructive to consider the simple application of logistic regression for analysing binary data (in a non-variance component context). In the disease context, logistic regression is often used to analyse the risk of developing disease as a function of the predictive variables. Studies that report risks, e.g. of developing a disease if you consume a certain product, invariably use logistic regression.

Let the trait be coded: healthy=0, diseased=1, and with a single predictor variable, then the probability of disease is:

$$P(Y=1|X) = \pi(X) = \exp(\beta_0 + \beta_1 X)/(1+\exp(\beta_0 + \beta_1 X)) = 1/(1+\exp(-\beta_0 - \beta_1 X))$$

Based on this, a 2x2 contingency table can be constructed to calculate the probability of observing diseased or healthy status depending upon whether or not an animal has been exposed to infection:

	Predictor variable	
	X=1 (Exposed)	X=0 (Not exposed)
Outcome variable		
Y=1 (diseased)	$\pi(1)= \exp(\beta_0 + \beta_1 X)/(1+\exp(\beta_0 + \beta_1))$	$\pi(0)= \exp\beta_0/(1+\exp(\beta_0))$
Y=0 (healthy)	$1-\pi(1)= 1/(1+\exp(\beta_0 + \beta_1))$	$1-\pi(0)= 1/(1+\exp(\beta_0)$

The odds of being diseased amongst exposed animals is $\pi(1)/(1-\pi(1))$
The odds of being diseased amongst the non-exposed animals is $\pi(0)/(1-\pi(0))$
The odds ratio (OR) is defined as the ratio of the odds for exposed animals to the odds for non-exposed animals. Taking formulae from above and cancelling out:
$OR = \exp(\beta_1)$.

Thus, the log of the odds ratio is: $ln(OR)=ln(\exp(\beta_1)) = \beta_1$.

The logit transformation is:

$g(X) = ln(\pi(X)/(1-\pi(X)) = ln(\exp(\beta_0+\beta_1 X) = \beta_0+\beta_1 X$

As above, the outcome β_1 values give the log OR values, which can then be transformed into OR values, thus predicting the relative change in risk as the predictor variable is increased by one unit, when the rarer outcome event is coded as 1.

Other transformations can also be used to transform the binomial data, for example the probit transformation based on the normal cumulative density function used by Kadarmideen *et al.* (2000). However, the attraction of the logistic regression approach (at least for straightforward analyses) is that the outcome is a meaningful parameter, i.e. the odds ratio.

Survival analyses
Survival analyses are now becoming widely used and applied in animal genetics contexts, following the availability of packages enabling genetic parameter estimation for longevity or survival time (Ducrocq and Sölkner, 1998). Survival analyses need not be based upon the death of the animal, they can equally well be applied to data that describes the time until any specific event, e.g. the time until disease is observed.

Survival analyses use complex proportional hazards models, such as the Weibull or Cox models. Extending these models to include random genetic effects results in mixed proportional hazards models, the so-called frailty models. An example of a Weibull frailty model is given by Henryon *et al.* (2001), analysing time until diagnosis of certain diseases or health conditions. From this methodology, sire (and/or dam) variances are estimated, and approximate terms describing the heritability can be constructed. For example, the heritability of log-frailty will be of the form (extending from Ducrocq and Casella, 1996):

$h^2(\log) = \sigma^2_a/(k\sigma^2_a + V^* + \pi^2/6)$

where σ^2_a is the additive genetic ('frailty') variance, k will be ¼ for a sire model and ½ for a sire-dam model, $\pi^2/6$ is the error variance of an extreme value distribution (which implicitly includes the unaccounted for genetic variance) and V* indicates any other variance component estimated in the analysis and required within the phenotypic variance.

An alternative to the parametric Weibull model is the semi-parametric Cox model that utilises an arbitrary baseline hazard function. In cases where the Weibull model is adequate, analysis using a Weibull hazard function will be preferred as it will produce more precise parameter estimates and it will be less computer-intensive. The Cox model may be more applicable when circumstances (e.g. environments or infectious challenge) change across time, as these fluctuations may be better accounted for. However, for Cox model the heritability cannot be simply expressed as it is for the Weibull model.

Mixture distributions

Mixture distributions often arise from parasite burden data, when the data comprises a group of animals with burdens of zero and a group of infected animals, where the actual burden is counted. Correct methods of analysing such data are unclear, depending upon the actual interpretation of the data. Much relies upon the accuracy of the ascertainment of the trait. For example, do the zero parasite burdens truly indicate uninfected animals or are they only apparently uninfected? If the trait ascertainment was perfect, would these zero burden animals remain at zero or would they have positive parasite burdens. These questions are critical for determining whether or not the zero counts are part a continuum of resistance, or whether zero counts are qualitatively different from non-zero counts.

Approaches to analysing such mixture distributions may include the following. Firstly, the data may be analysed as if it were a continuously distributed trait. However, if the number of zero counts is large, even after transformation the data will not be normally distributed (or more importantly, the residuals will not be normally distributed), violating one of the basic assumptions usually made in such analyses. Additionally, this makes the assumption that the same interpretation may be given to zero and non-zero counts. Secondly, the data may be condensed down to a binary trait, 'infected or not'. This will result in a valid analysis, providing zero counts are true zeroes, however it does ignore the information contained in the variability of non-zero values. Thirdly, the zero data may be disregarded, and analyses performed only on infected animals; again throwing information away. Bishop and Stear (2001) give an example of these approaches for analysing data of this nature. To protect against trait ascertainment problems these authors used a somewhat stringent definition of whether the animal was infected or not: whether or not the animal was **ever** infected during the observation period, rather than whether or not it was infected at any particular time point.

Parameter estimation in relation to epidemiological models

As has been repeatedly emphasised, the key to effective disease data analysis lies in deriving and analysing traits of epidemiological and biological relevance to the disease scenario. An excellent, albeit extremely complex example of this, is the analysis of data describing the BSE epidemic in the UK during the 1980s and early 1990s (Anderson *et al.*, 1996 and subsequent papers). Based on observational data and epidemic theory, an epidemic model describing the transmission of BSE throughout the UK dairy industry was derived. This model comprised several critical parameters describing the infection dynamics. Numerical values for these parameters were then derived from a thorough analysis of the data. Exploration of the parameterised model then gave insight into the epidemic dynamics and allowed evaluation of various control strategies. This is directly analogous to the creation of genetic-epidemiological models, and the use of these models to investigate the consequences of selection.

An example of epidemic parameter estimation for a disease of interest to animal breeders is given by White *et al.* (2001), in which parameters are estimated describing the dynamics of

multi-species mastitis infections in dairy cattle. A mathematical model of the transmission and interaction of so-called major and minor bacterial species was created, with the parameters being the respective transmission rates of each species and the cross-protection afforded by the immune response to each pathogen respectively. These were then estimated from available experimental data describing the levels of bacterial infection per teat per cow in a dairy herd.

Parameter estimation in a genetic-epidemiological context has an additional level of complexity, as the geneticist is often interested not only in first order statistics, i.e. parameter means (as described above), but also in second order statistics, i.e. parameter variances when the parameter is expressed at the individual animal level. Interpretation of individual measurements in an epidemiological context is described above. There are fewer studies of parameter estimation for a series of sequential host-parasite interactions, such as is often required in epidemic models. However, an example is given by Stear *et al.* (1997) for nematode parasite infections in lambs, as described in section 3. In this study, detailed parasite measurements enabled genetic parameter estimation for the probability of parasite establishment (assuming all lambs faced equal parasitic challenge), for parasite growth rates and for parasite fecundity, all expressed as traits of the host. Only for parasite mortality were variance components not estimable.

8.7 Breeding Program Implementation

By this stage all the information and knowledge required to implement breeding programs should be available. Appropriate diseases should have been chosen according to sound economic and epidemiological criteria, research will have demonstrated that utilisable genetic effects exist, and statistical analyses will have been applied appropriately to analyse and interpret the experimental data – e.g. to calculate variance components, verify marker linkages or demonstrate breed differences. Additionally, all stakeholders (i.e. geneticists, breeders, animal producers, animal health practitioners) will have been convinced and willing to be involved in the process. It is the assumed that a genetic epidemiological model of the genetic strategy (e.g. selection, introgression or breed choice/substitution) has demonstrated that the genetic strategy will indeed help to provide a sustainable solution to the disease problem and won't be compromised by other disease control strategies.

Key to the success of the breeding program will be effective data recording protocols, as described in section 8.5. This is a critical step in the process, and may be a place where the breeding program can break down, especially if the data recording protocols are difficult, time-consuming or expensive. Goodwill from those involved in the measurement process may diminish if costs are excessive or if no apparent progress has been made. Selection for disease resistance using verified genetic markers may require less recurrent cost and input.

Effective and appropriate breeding strategies should ideally be as simple as is feasible to meet the breeding objective, and low-technology strategies may often be more appropriate than technologies using more complex technology. The most appropriate strategy will differ for different situations, however some examples are:

- Breed choice or substitution. This will be particularly appropriate in many situations where debilitating endemic diseases are present, and the choice is between locally adapted breeds or apparently more productive exotic breeds. Under most circumstances, the locally adapted breed will lead to more sustainable livestock production. Examples

are the rejection of Holstein cattle by small-scale farmers in tropical regions and the use of indigenous dairy breeds, the use of sheep breeds that are tolerant to high levels of nematode parasite challenge, or the use of trypanotolerant cattle or sheep breeds.

- Selection based on phenotypic measurement. Again this will be most easily implemented when there is a predictable infectious challenge. It will be particularly appropriate for selection for resistance to macroparasitic diseases such as nematode infections, tick infestations or fly infestation. An obvious microparasitic disease where this can also be implemented is mastitis.
- Selection or introgression utilising genetic markers alone. This may be the only solution for some diseases where providing infectious challenges is impracticable. This is the case for scrapie resistance in sheep and Marek's disease resistance. It will also be the exploitation route for resistance to various E. coli infections in pigs, not because infection is impractical, but because there are genetic makers for complete resistance.
- Marker-assisted selection, using a combination of genetic markers and phenotypic measurement. Potentially, this is feasible in all situations where phenotypic measurement is possible, and in principle it should allow the breeder greater flexibility regarding breeding procedures. The current worldwide activity in attempting to detect QTL for a variety of diseases suggests that many geneticists share this vision. However, practical implementation of marker-assisted selection is proving somewhat rare and elusive.

Under most circumstances, a disease resistance (or tolerance) trait will only be part of a larger selection goal. Hence, the issue to be addressed is the relative weighting to be applied to the disease resistance trait *vs.* the performance traits. Here again, there is a need for genetic-epidemiological models to provide the input variables into the economic calculations. These models will predict the changes in disease incidence or severity as the goal trait changes; hence enable relative economic weights to be calculated.

The use of genetic epidemiological model to estimate relative economic weights has been applied to nematode infections in sheep (Bishop and Nagel, unpublished), where the indicator trait is nematode egg counts. There are a number of benefits from reducing egg counts, including enhanced animal health and potentially reduced anthelmintic costs. However, reduced egg counts lead to enhanced growth rates through the epidemiological impact of reduced parasite challenge (see section 3) and this provides a conservative method for estimating economic weights. To do the calculations, the common currency of the profit equation was changed from monetary units (e.g. Dollars) to kg live weight. Using the model of Bishop and Stear (1999), the predicted gains in live weight attributable solely to the epidemiological effect of reducing egg counts were estimated. This enabled the calculation of the change in profit (kg live weight) per unit change in *ln*(egg count). Therefore, this gives an economic weight for egg counts relative to that for live weight. It was found that the impact of one unit change in *ln*(egg count) was similar to that for one kg change in live weight (ignoring reduced animal health costs). Subsequent selection index calculations demonstrated that an index containing live weight and egg counts with these relative economic weights substantially outperformed selection on live weight alone, with the gains being robust to changes in the relative economic weight given to egg counts. In this case, improvements in performance came from three sources: (i) the direct improvements through having performance traits in the index, (ii) the correlated improvements in performance arising from the fact that egg counts were favourable genetically correlated with performance ($r_g \sim -0.2$), i.e. egg counts were being used as a predictor trait and (iii) the epidemiological effects, as described above.

In principle, genetic-epidemiological models could contribute to the calculation of economic weights for disease resistance traits in a variety of circumstances. If the models include animal performance, they potentially allow a unit change in the resistance trait to be translated into a change in the overall output of the system as a whole. However, little appears to have been done in this subject area using models that account for the disease epidemiology. It is a topic that is ripe for investigation.

Lastly, having implemented a genetic improvement strategy, an often-forgotten requirement is to monitor and appraise the success of the breeding program. A breeding program that is implemented and then left alone may eventually fail, unless participants have a strong motivation in ensuring its success, as in programs run by commercial breeding companies. It is critical that a breeding program for disease resistance is regularly assessed in terms of whether or not it is meeting its objectives of helping to improve animal health or control the targeted disease problem, and whether or not it is contributing to sustainable livestock production.

Bibliography

Adams, D.B. (1988). Infection with *Haemonchus contortus* in sheep and the role of adaptive immunity in selection of the parasite. *International Journal of Parasitology*, 18: 1071-1075.

Adams, L.G. and Templeton, J.W. (1998). Genetic resistance to bacterial diseases of animals. *Revue Scientifique et Technique*, 17: 200-219.

Albers, G.A.A. and Burgess, S.K. (1988). Serial passage of *Haemonchus contortus* in resistant and susceptible sheep. *Veterinary Parasitology*, 28: 303-306.

Anderson, R.M., Donelly, C.A., Ferguson, N.M., Woolhouse, M.E.J., Watt, C.J., Udy, H.J., MaWhinney, S., Dunstan, S.P., Southwood, T.R.E., Wilesmith, J.W., Ryan, J.B.M., Hoinville, L.J., Hillerton, J.E., Austin, A.R. and Wells, G.A.H. (1996). Transmission dynamics and epidemiology of BSE in British cattle. *Nature*, 382: 779-788.

Anderson, R.M. and May, R.M. (1992) Infectious Diseases of Humans. Dynamics and Control. Oxford University Press, Oxford.

Baker R. L., Mugambi, J.M., Audho, J. O., Carles, A. B. and Thorpe, W. (2002). Comparison of Red Maasai and Dorper sheep for resistance to gastro-intestinal nematode parasites, productivity and efficiency in a sub-humid and a semi-arid environment in Kenya. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production,* Communication 13-10.

Barger, I.A. (1989) Genetic resistance of hosts and its influence on epidemiology. *Veterinary Parasitology* 32, 21-35.

Barnes, E.H. and Dobson, R.J. (1990) Population dynamics of *Trichostrongylous colubriformis* in sheep: computer model to simulate grazing systems and the evolution of anthelmintic resistance. *International Journal for Parasitology*, 20, 823-831.

Bishop, S.C. (1999). Modelling relationships between host genotype and disease epidemiology in domestic animals: nematode infections in ruminant livestock. *Animal Breeding Abstracts*, 67: 549-553.

Bishop, S.C., Bairden, K., McKellar, Q.A. and Stear, M.J. (1996). Genetic parameters for faecal egg count following mixed, natural predominantly *Ostertagia circumcincta* infection and relationships with live weight in young lambs. *Animal Science*, 63: 423-428.

Bishop, S.C. and MacKenzie, K. (2003). Genetic management strategies for controlling infectious disease in livestock populations. *Genetics, Selection, Evolution*, 35: S1-S15.

Bishop, S.C and Stear, M.J. (1997). Modelling responses to selection for resistance to gastrointestinal parasites in sheep. *Animal Science* 64, 469-478.

Bishop, S.C. and Stear, M.J. (1999). Genetic and epidemiological relationships between productivity and disease resistance: gastrointestinal parasite infection in growing lambs. *Animal Science*, 69: 515-525.

Bishop, S.C. and Stear, M.J. (2000) The use of a gamma-type function to assess the relationships of the between the number of adult *Teladorsagia circumcincta* and total egg output. *Parasitology*, 121: 435-440.

Bishop, S.C. and Stear, M.J. (2001). Inheritance of, and factors affecting, egg counts during early lactation in Scottish Blackface ewes facing mixed, natural nematode infections. *Animal Science*, 73: 389-395.

Blasco, A. (2001). The Bayesian controversy in animal breeding. *Journal of Animal Science*, 79: 2023-2046.

Bliss and Fisher, R.A. (1953). Fitting the negative binomial to biological data and a note on the efficient fitting of the negative binomial. *Biometrics*, 9: 176-200.

Chartier, C., Pors, I., Hubert, J., Rocheteau, D., Benoit, C. and Bernard, N. (1998). Prevalence

of anthelmintic resistant nematodes in sheep and goats in Western France. *Small Ruminant Research*, 29: 33-41.

Clapperton, M., Bishop, S.C. and Glass, E.J. (2001). Leucocyte sub-sets and acute phase proteins are associated with productivity in Large White pigs. *Proceedings of the BSAS Winter Meeting 2003*, 32.

De Jong, M.C.M., Diekmann, O. and Heesterbeek, J.A.P. (1994). The computation of R_0 for discrete-time epidemic models with dynamic heterogeneity. *Mathematical Biosciences* 119: 97-114.

Diekmann, O., Heesterbeek, J.A.P. and Metz, J.A.J. (1990). On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28: 365-382.

Ducrocq, V. and Casella, G. (1996). A Bayesian analysis of mixed survival models. *Genetics, Selection, Evolution*, 28: 505-529.

Ducrocq, V. and Sölkner, J. (1998). 'The Survival Kit', a package for large analyses of survival data. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*, 27: 447-448.

Dushoff, J. (1999). Host heterogeneity and disease endemicity: a moment-based approach. *Theoretical Population Biology*, 56: 325-335.

Dushoff, J. and Levin, S. (1995). The effects of population heterogeneity on disease invasion. *Mathematical Biosciences*, 128: 25-40.

Edfors-Lilja, I., Wattrang, E., Andersson, L. and Fossum, C. (2000). Mapping quantitative trait loci for stress induced alterations in porcine leukocyte numbers and functions. *Animal Genetics*, 31: 186-193.

Edfors-Lilja, I., Wattrang, E., Marklund, L., Moller, M., Andersson-Eklund, L., Andersson, L., and Fossum, C. (1998). Mapping quantitative trait loci for immune capacity in the pig *Journal of Immunology*, 161: 829-835.

Eisler, M.C., Brandt, J., Bauer, B., Clausen, P.H., Delespaux, V., Holmes, P.H., Ilemobade, A., Machila, N., Mbwambo, H., McDermott, J., Mehlitz, D., Murilla, G., Ndung'u, J.M., Peregrine, A.S., Sidibe, I., Sinyangwe, L. and Geerts, S. (2001). Standardised tests in mice and cattle for the detection of drug resistance in tsetse-transmitted trypanosomes of African domestic cattle. *Veterinary Parasitology*, 97: 171-182.

Fisher, R.A. (1930) The Genetical Theory of Natural Selection. Clarendon, Oxford.

Frank, S.A. (1994). Coevolutionary genetics of hosts and parasites with quantitative inheritance. *Evolutionary Ecology*, 8: 74-94.

Gandon, S., Jansen, V.A.A. and van Baalen, M. (2001). Host life history and the evolution of parasite virulence. *Evolution*, 55: 1056-1062.

Gandon, S., MacKinnon, M.J., Nee, S., and Read, A.F. (2001). Imperfect vaccines and the evolution of pathogen virulence. *Nature*, 414: 751-756.

Gandon, S. and Michalakis, Y. (2000). Evolution of parasite virulence against qualitative or quantitative host resistance. *Proceedings of the Royal Society of London, Series B*, 267: 985-990.

Grenfell, B.T. (1988). Gastrointestinal nematode parasites and the stability and productivity of intensive ruminant grazing systems. *Philosophical Transactions of the Royal Society of London B*, 321: 541-563.

Grenfell, B.T. and Dobson, A.P. (1995). Ecology of Infectious Diseases in Natural Populations. Cambridge University Press.

Grenfell, B.T., Smith, G. and Anderson, R.M. (1987) A mathematical model of the population biology of *Ostertagia ostertagi* in calves and yearlings. *Parasitology*, 95: 389-406.

Grenfell, B.T., Wilson, K., Isham, V.S., Boyd, H.E.G. and Dietz, K. (1995). Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode –

ruminant interactions as a case study. *Parasitology*, 111: S135-S151.
Hamer, W.H. (1906). Epidemic disease in England. *The Lancet*, i: 733-739.
Henryon, M., Berg, P., Jensen, J. and Sorensen, S. 2001. Genetic variation for resistance to clinical and subclinical diseases exists in growing pigs. *Animal Science*, 73: 375-387.
Hone J. (1994). A mathematical model of detection and dynamics of porcine trans-missible gastroenteritis. *Epidemiology and Infection* 113: 187-197.
Jackson, F. (1993). Anthelmintic resistance - the state of play. *British Veterinary Journal*, 149: 123-138.
Kadarmideen, H.N., Rekaya, R. and Gianola, D. 2001. Genetic parameters for clinical mastitis in Holstein-Friesians in the United Kingdom: a Bayesian analysis. *Animal Science*, 73: 229-240.
Kadarmideen, H.N., Thompson, R. and Simm, G. (2000). Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle. *Animal Science*, 71: 411-419.
Kermack, W.O. and McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society*, A115: 700-721.
Khibnik, A.I. and Kondrashov, A.S. (1997). Three mechanisms of Red Queen dynamics. *Proceedings of the Royal Society Of London Series B-Biological Sciences,* 264: 1049-1056.
Klasing, K.C., Johnstone, B.J. and Benson, B.N. (1991). Implications of an immune response on growth and nutrient requirements of chicks. In *Recent advances in animal nutrition* (Eds. W. Haresign and D.J.A. Cole). Butterworth, London. Pp. 135-146.
Knap, P.W. and Bishop, S.C. (2000). Relationships between genetic change and infectious disease in domestic livestock. BSAS Occasional Meeting: The Challenge of Genetic Change in Animal Production. Edinburgh, 26-27 October, 1999. In: The challenge of genetic change in animal production. BSAS Occasional Publication, number 27: 65-80. Ed. W.G. Hill, S.C. Bishop, B. M^{c}Guirk, J.C. M^{c}Kay, G. Simm and A.J. Webb. © 2000 British Society of Animal Science. At www.BSAS.org.uk/publs/genchng/contents.pdf
Kyriazakis, I., Houdijk, J. and Coop, R.L. (2002). Immunonutrition: the nutritional control of acquired immunity to parasites. *Proceedings of the British Society of Animal Science* 2002: 232.
Leathwick, D.M., Barlow, N.D. and Vlassof, A. (1992). A model for nematodiasis in New Zealand lambs. *International Journal for Parasitology*, 22: 789-799.
Leslie, P. (1945). On the use of matrices in certain population mathematics. *Biometrika*, 35: 183-212.
MacKenzie, Katrin (2000). Quantifying selection for resistance to infectious diseases in pigs using genetic epidemiological models. Ph.D. Thesis. University of Edinburgh.
MacKenzie, K.M. and Bishop, S.C. (1999). A discrete-time epidemiological model to quantify selection for disease resistance. *Animal Science*, 69: 543-551.
MacKenzie, Katrin and Bishop, S.C. (2001). Developing stochastic epidemiological models to quantify the dynamics of infectious diseases in domestic livestock. *Journal of Animal Science*, 79: 2047-2056.
MacKenzie, Katrin and Bishop, S.C. (2001). Utilising stochastic genetic epidemiological models to quantify the impact of selection for resistance to infectious diseases in domestic livestock. *Journal of Animal Science*, 79: 2057-2065.
MacKenzie, K., Woolliams, J.A. and Bishop, S.C. (2001). A discrete-time model to evaluate the spread of infectious disease in livestock populations of differing genetic diversity. *Proceedings of the 52nd Meeting of the European Association for Animal Production*, 165
Mackinnon, M.J. and Read, A.F. (1999). Genetic relationships between parasite virulence and transmission in the rodent malaria *Plasmodium chabaudi. Evolution*, 56: 689-703.

Mallard, B.A., Wilkie, B.N., Kennedy, B.W., Gibson, J. and Quinton, M. 1992. Immune responsiveness in swine: eight generations of selection for high and low immune response in Yorkshire pigs. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*, **27:** 257-264.

May, R.M. and Anderson, R.M. (1983). Epidemiology and genetics in the coevolution of parasites and hosts. *Proceedings of the Royal Society, Series B*, 219: 281-313.

May, R.M. and Anderson, R.M. (1990). Parasite-host coevolution. *Parasitology*, 100: S89-S101.

Pareto, V. (1906). Manuale d'Economica Politica. Milan.

Perry, B.D., McDermott, J.J., Randolph, T.F., Sones, K.R. and Thornton, P.K. (2002). Investing in animal health research to alleviate poverty. International Livestock Research Institute (ILRI), Nairobi, Kenya.

Pinard, M.H., Van Arendonk, J.A.M., Nieuwland, M.G.B. and Van der Zijpp, A.J. (1992). Divergent selection for immune responsiveness in chickens: estimation of realized heritability with an animal model. *Journal of Animal Science,* 70: 2986-2993.

Pinard, M.H., Janss, L.L.G., Maatman, R., Noordhuizen, J.P.T.M. and Van der Zijpp, A.J. (1993). Effect of divergent selection for immune responsiveness and of major histocompatibility complex on resistance to Marek's disease in chickens. *Poultry Science,* 72: 391-402.

PongWong, R. and Woolliams, J.A. (1996). Estimating major gene effects with partial information using Gibbs sampling *Theoretical and Applied Genetics*, 93: 1090-1097.

Renshaw, E. (1991). Modelling Biological Populations in Space and Time, Cambridge University Press, Cambridge.

Ross, R. (1908). Report on the Prevention of Malaria in Mauritious. London.

Sarker, N., Tsudzuki, M., Nishibori, M., Yamamoto, Y. (1999). Direct and correlated response to divergent selection for serum immunoglobulin M and G levels in chickens. *Poultry Science*, 78: 1-7.

Saulai, M., Cabaret, J., Hostache, G., Mandonnet, N. and Aumont, G. (2001). Life-trait evolution of a parasite strongyle nematode in response to host resistance: an experimental approach using *Haemonchus contortus* in black belly lambs. *Genetics Selection Evolution*, 33: S25-S44.

Schat, K.A. and Davies, C.J. (2000). Viral Diseases. In: Breeding for Disease Resistance in Farm Animals. 2nd Edition. Pp 271-300. Ed. R.F.E. Axford, S.C. Bishop, F.W. Nicholas and J.B. Owen. CABI Publishing.

Smith, G. and Guerrero, J. (1993). Mathematical models for the population biology *of Ostertagia ostertagi* and the significance of aggregated parasite distributions. *Veterinary Parasitology*, 46: 243-257.

Spickett A.M. (1994). Tick ecology. *International Journal for Parasitology*, 24: 845-849.

Springbett, A.J., MacKenzie, K., Woolliams, J.A. and Bishop, S.C. (2002). The contribution of genetic diversity to the spread of infectious diseases. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Communication 13-31.

Spurlock, M.E. (1997). Regulation of metabolism and growth during immune challenge: An overview of cytokine function. *Journal of Animal Science*, 75: 1773-1783.

Stahly, T.S. (1996). Impact of immune system activation on growth and optimal dietary regimens of pigs. In *Recent advances in animal nutrition* (eds. P.C. Garnsworthy, J. Wiseman and W. Haresign). Nottingham University Press, Nottingham. Pp. 197-206.

Stear, M.J., Bishop, S.C., Bairden, K., Duncan, J.L., Gettinby, G., Holmes, P.H., McKellar, Q.A., Park, M., Strain, S. and Murray, M. (1997). How hosts control worms. *Nature*, 389: 27

Strain, S., Bishop, S.C., Henderson, N., Holmes, P.H., McKellar, Q.A., Mitchell, S. and Stear, M.J. (2002). The genetic control of IgA activity and its association with parasite resistance

in naturally infected sheep. *Parasitology*, 124: 545-552.
Stringer, S.M., Hunter, N. and Woolhouse, M.E.J. (1998) A mathematical model of the dynamics of scrapie in a sheep flock. *Mathematical Biosciences*, 153: 79-98.
Usera, M.A., Aladuena, A., Gonzalez, R., De la Fuente, M., Garcia-Pena, J., Frias, N. and Echeita, M.A. (2002). Antibiotic resistance of Salmonella spp. from animal sources in Spain in 1996 and 2000. *Journal of Food Protection*, 65: 768-773.
Vagenas, D. (2002). Genetic and quantitative aspects of resistance to gastrointestinal nematode parasites in small ruminants. PhD Thesis, University of Edinburgh.
Vagenas, D. and Bishop, S.C. (2002) Modelling the joint effects of genotype and protein intake on resistance to gastrointestinal parasite infections in sheep. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Communication 13-08
van der Waaij, E.H., Bijma, P., Bishop, S.C. and van Arendonk, J.A.M. (2000). Modeling selection for production traits under constant infection pressure. *Journal of Animal Science*, 78: 2809-2820.
van der Waaij, E.H., Bijma, P., Bishop, S.C. and van Arendonk, J.A.M. (2002). Utilising genetic markers for disease resistance to improve production under constant infection pressure. *Journal of Animal Science*, 80: 322-329.
Van Valen, L. (1973). A New Evolutionary Law. *Evolutionary Theory*, 1: 1-30.
Waller, P.J. (1997). Anthelmintic resistance. *Veterinary Parasitology*, 72: 391-405
Wells, S.J., Ott, S.L. and Hillbeerg Sitzinger, A. (1998). Key health issues for dairy cattle – new and old. *Journal of Dairy Science*, 81: 3029-3035.
White, L.J., Evans, N.D., Lam, T.J.G.M., Schukken,mY.H., Medley, G.F., Godfrey, K.R. and Chappell, M.J. (2001). The structural identifiability and parameter estimation of a multispecies model for the transmission of mastitis in dairy cows. *Mathematical Biosciences*, 174: 77-90.
Williams, N.H., Stahly, T.S. and Zimmerman, D.R. (1997). Effect of chronic immune system activation on nitrogen retention, partial efficiency of lysine utilization and lysine needs of pigs. *Journal of Animal Science,* 75:2472-2480.
Williams, N.H., Stahly, T.S. and Zimmerman, D.R. (1997). Effect of level of chronic immune system activation on the growth and dietary lysine needs of pigs fed from 6 to 112 kg. *Journal of Animal Science,* 75: 2481-2496.
Witter, R.L. (1998). The changing landscape of Marek's disease. *Avian Pathology*, 27: S46-S53.
Woolaston, R.R., Elwin, R.L., Barger, I.A., 1992. No adaptation of *Haemonchus contortus* to genetically resistance sheep. *International Journal for Parasitology*, 22, 377-380.
Woolhouse, M.E.J., Stringer, S.M., Matthews, L., Hunter, N. and Anderson, R.M. (1998). Epidemiology and control of scrapie within a sheep flock. *Proceedings of the Royal Society of London, Series B*, 265:1205-1210.
Zhu, Y.Y., Chen, H.R., Fan, J.H., Wang, Y.Y., Li, Y., Chen, J.B., Fan, J.X., Yang, S.S., Hu, L.P., Leung, H., Mew, T.W., Teng, P.S., Wang, Z.H. and Mundt, C.C. (2000). Genetic diversity and disease control in rice. *Nature*, 406: 718-722.