

INTERPRETATION AND PREDICTION IN DISEASE GENETICS

S.C. Bishop

Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, Midlothian EH25 9PS, UK

SUMMARY

This paper summarises general issues relating to the interpretation of field disease data and the prediction of responses to selection for disease resistance. Prioritisation of diseases to study is a non-trivial task as there are many criteria by which disease importance can be assessed; a structured approach is described herein. The interpretation of field data and the prediction of responses to selection require an understanding of disease epidemiology, as infection transmission dynamics will affect both the interpretation of genetic parameters and the likely observable outcomes. For many bacterial and viral diseases, field data will likely contain noise due to incomplete exposure to infection and imperfect diagnosis of infection. These factors will result in heritabilities or SNP associations being underestimated; hence a weak genetic signal from such data may mask stronger underlying genetic effects. Interpretation of field data for parasite resistance is more straightforward, provided animals have faced sufficient challenge. Mathematical models that predict responses to selection reveal non-linear relations between mean host genotype and outcomes such as observed infection levels, animal performance, and the likelihood and severity of epidemics. In many cases, total benefits are predicted to be larger than suggested by quantitative genetic theory alone, justifying inclusion of disease resistance in selection goals. These concepts are illustrated for nematode infections and footrot in sheep, and ruminant mastitis.

INTRODUCTION

Genetic variation in host resistance to infectious disease is ubiquitous, and disease resistance is now a major focus for animal geneticists. Further, the availability of dense single nucleotide polymorphism arrays (i.e. SNP chips) has given rise to hitherto unforeseen opportunities to dissect this between-host variation and identify genes contributing to this variation, by means of genome wide association studies. This, coupled with more traditional quantitative genetic variance-partitioning approaches, enables detailed descriptions of genetic aspects of disease resistance and the identification of individuals with extreme (high or low) risk of infection or disease.

It is usually necessary to use field data to get sufficient animals to reliably estimate genetic parameters for disease resistance or detect robust SNP associations. Data may be captured from a population undergoing an epidemic such as bovine tuberculosis, from an endemic disease such as mastitis, where herd-level disease occurrence is predictable, or from diseases such as nematode parasite infections where (depending on weather conditions) all animals will be challenged. However, such field data is very 'noisy': diagnosis of infection or disease may be imprecise; it can be difficult to determine when infection of an animal occurred; and it is often unclear whether or not apparently healthy animals have been exposed to the infection. This paper explores the nature of field disease data, assessing the opportunities and describing factors affecting its interpretation and analysis. Further, it explores predictions of the benefits of selecting animals for increased disease resistance, accounting for disease epidemiology and transmission of infection.

PRIORITISING DISEASES

There are many examples of host genetic variation in resistance to infection or in disease resistance, i.e. the side effects of infection, with even cursory literature reviews revealing more than 50 diseases for which genetic variation has been published (Bishop, 2005). It is a reasonable assumption that genetic variation will exist for resistance to most diseases and, therefore, most

Disease resistance

diseases are potentially targets for selective breeding of the host. In reality, only a small subset of diseases can ever be considered and prioritisation must be made. Davies et al. (2009) discuss and demonstrate criteria by which diseases can be ranked. These include economic impact, industrial concern, public concern, human health implications, animal welfare and international trade restrictions caused by the disease. Assigning scores to each category and summary across categories gives an empirical ranking of diseases. Although the ranking will differ between countries or production systems, it will tend to throw obvious suspects such as foot and mouth disease (FMD) to the top of the list. But from an animal genetics perspective, other factors such as feasibility of data collection, prior evidence for the extent of host genetic variation in resistance and the available genomic tools also influence the rankings. The combined rankings of Davies et al. (2009) suggest that for host genetic studies, the highest ranking diseases are salmonella infections, Marek's disease and coccidiosis in chickens, mastitis in cattle, *E. coli* infections and porcine reproductive and respiratory syndrome in pigs, and mastitis and gastrointestinal (GI) parasites in sheep. Whilst these are the most amenable diseases for genetic research, from a breeding perspective it is also important to consider the logistics of performing selection, the epidemiological benefits of selection and the compatibility with other control measures. If a disease meets these criteria, then is it a suitable candidate for breeding for host resistance.

FRAMEWORK FOR INTERPRETING FIELD DATA

Transmission of infection. The key to interpreting disease data and predicting responses to selection lies in understanding the infection transmission pathways. A schema for typical infections is shown in Fig. 1. Infection is transmitted from animals that are infected and infectious (i.e. capable of transmitting infection) to animals that are immunologically susceptible, i.e. capable of being infected. For some diseases infection may be transmitted via a reservoir, e.g. pasture or insect vectors. Animals advance through disease states according to the disease and the force of infection. For macroparasitic diseases, such as GI parasite infections in grazing ruminants, the infection pressure is essentially continuous and, depending on climate and treatment strategies, most animals are challenged almost continually. Thus, a population will mainly comprise animals in the *Infectious* state, and the *Recovered* state generally is not relevant. For microparasitic infections (bacteria or viruses), animals progress through some or most of the infection states in Fig. 1, and it may be difficult to quantify the extent to which animals are infected. A population will comprise animals in every state, with the proportion in each state depending on the time since the commencement of the epidemic and the resistance of the population to the disease in question.

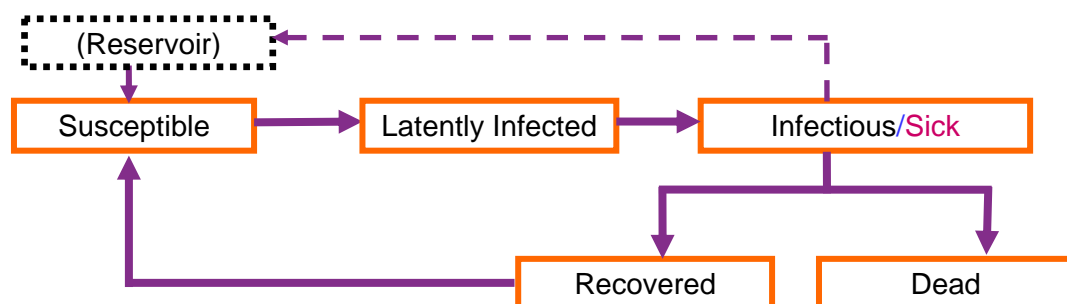


Figure 1. Potential Pathways of Transmission of Infection and Animal Infection States

Interpreting field data: GI nematode parasite infections. Continuous between-animal variation will be seen in indicator traits used to describe nematode resistance. These traits fall into three broad categories: (i) indicators of how heavily infected an animal is, such as faecal egg count (FEC), (ii) indicators of animals' immune responses, such as antibody levels or eosinophilia and (iii) indicators of the impact of infection on the animal, such as fructosamine or pepsinogen concentrations, packed cell volume (PCV) for *Haemonchus contortus* infections, or even growth rate. Provided it can be assumed that animals grazing the same pasture are equally challenged, then data analysis is straightforward: the data are transformed to render them approximately normally distributed, and analysed in the same way as any other trait. Animals with extreme EBVs are selected, i.e. for decreased FEC, or for increased PCV or growth rate.

Interpreting field data: bacterial or viral infections. Data collected will often be categorical, describing the state that an animal is in, e.g. alive vs. dead or diseased vs. apparently healthy. The severity of infection or disease is often not known or may be unobtainable. The apparently healthy category may comprise animals that have yet to be infected, either because they haven't had sufficient exposure or have a high degree of resistance to infection, are latently infected, or have been infected but have subsequently recovered.

The timing and extent of exposure to infection is generally unknown. Consider a population where a proportion e has faced sufficient exposure to pathogen challenge to have a chance of becoming infected. Let the virtual prevalence (p^*) be the prevalence if all animals were exposed to the pathogen. If exposure is random and independent of animal genotype, then observed prevalence is ep^* . Of the $1-ep^*$ proportion of animals that are healthy, $e(1-p^*)$ are exposed and apparently resistant, and $(1-e)$ have not yet been exposed. It can be shown (Bishop and Woolliams, in prep.) that incomplete exposure to infection has a near-linear impact on the heritability of liability to infection, as illustrated in Fig. 2. However, assuming exposure is random across families and genotypes, the impact on EBV estimation and hence genetic progress is somewhat less (Nieuwhof, 2009), because the incomplete exposure effects average out at the family level. Effects of non-random exposure, e.g. mother-offspring transmission, have yet to be explored.

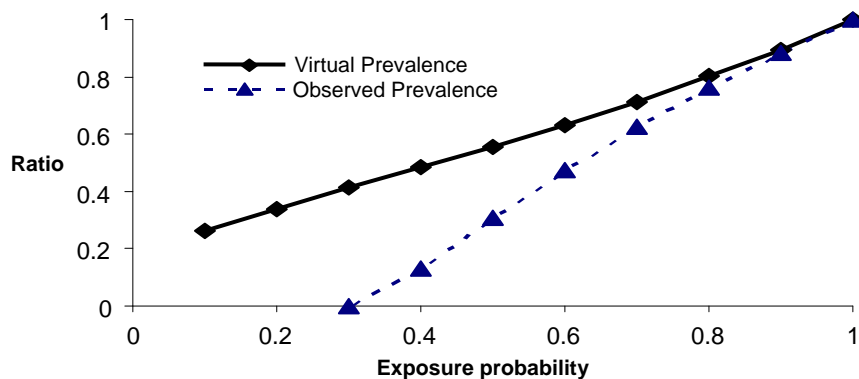


Figure 2. Ratio of estimated heritability to true heritability on the liability scale for varying levels of incomplete exposure, assuming virtual or observed prevalences of 0.3.

Imperfect diagnosis also impacts upon inferences from field data and heritability estimation. Fundamental to all diagnostic tests are the concepts of specificity and sensitivity. Specificity (F) is the probability that truly healthy animals (i.e. not infected by the pathogen of interest) are classified as such, and sensitivity (T) is the probability that truly infected or diseased animals are

correctly classified. True prevalence (p) may be estimated from estimated prevalence (p') as: $p = (p' + F - 1) / (F + T - 1)$. Imperfect specificity and sensitivity impact on the heritability of liability to infection (Bishop and Woolliams, in prep.). In general, for true disease prevalence less than 0.5, imperfect specificity will markedly reduce the estimated heritability and imperfect sensitivity will result in a minor reduction in the estimated heritability. For example, for a true prevalence of 0.3, sensitivity or specificity of 0.8 will result in heritability of liability being underestimated by 20% and 45%, respectively. The effects of imperfect sensitivity and specificity are reversed for prevalences greater than 0.5. The implication of this underestimation is that even low heritabilities under field conditions may be indicative of substantial true heritable variation in resistance.

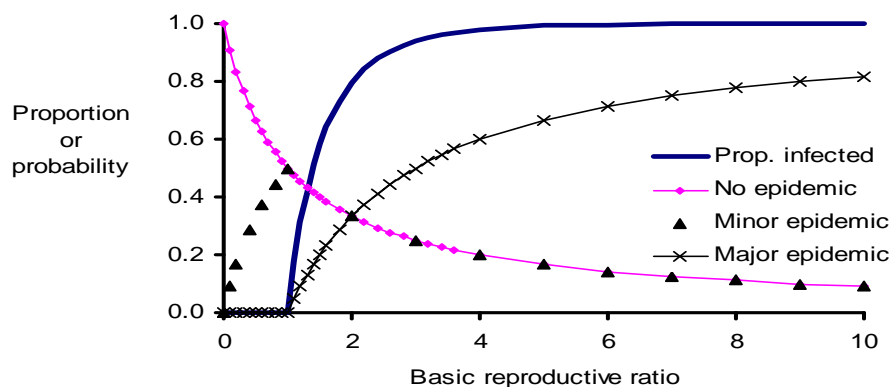


Figure 3. Epidemic probabilities, given the presence of an infected animal, and severity of major epidemics, as a function of the basic reproductive ratio, for hypothetical viral or bacterial diseases.

Predicting responses to selection. The flip side of the interpretation of field data is the prediction of responses to selection, using field data or genetic markers. The key point is that genetic theory has to be extended to include the epidemiological impact of infection, i.e. to account for the altered transmission of infection along the pathways shown in Fig. 1. This will result in non-linear relationships between average host genotype for resistance and disease outcomes at the population level. For macroparasitic infections, such as ruminant nematode infections, the outcome traits are simply indicators of the severity of infection or disease, measured at the individual animal level, and these impacts are described below. In general the feedback loops will lead to additional genetic progress (Bishop and Stear, 1997). The situation for microparasitic infections is more complex, requiring the development of genetic-epidemiological models to quantify the outcomes (MacKenzie and Bishop, 1999). Here, the outcomes may be assessed in terms of the probability and severity of epidemics, and the benefits depend on the infectiousness of the disease which is described by R_0 , the basic reproductive ratio (i.e. the expected number of secondary cases after the introduction of a single infected animal). Increasing host resistance to infection has a largely linear impact in terms of reducing R_0 , however the relationship between R_0 and epidemic outcomes is non-linear, as illustrated by Mackenzie and Bishop (2001). For highly infectious diseases (e.g. FMD) a large decrease in R_0 would be required before any noticeable improvement in epidemic outcomes would be observed, whereas for diseases whose R_0 is only slightly above 1.0 (e.g. scrapie) modest improvements in resistance may lead to large reductions in epidemic likelihood or severity. These impacts are illustrated in Fig. 3 for the probabilities of no, minor or major epidemics, given the introduction of a single infected animal into the population, and the

proportion of animals infected during the course of a major epidemic (Bishop and MacKenzie, 2003). A minor epidemic is one that dies out through stochastic events without intervention.

APPLICATIONS

Nematode infections in sheep. Genetic selection has often been used to help to control nematode infections in sheep, usually based on FEC (see summary by Bishop and Morris, 2007). In almost all cases FEC, once appropriately transformed, is a moderately heritable trait and one which responds to selection. Genome scans to detect QTL are now well advanced in many countries. With the exception of a QTL near the interferon gamma locus on chromosome 3, a feature of these studies is the difficulty in detecting QTL that are consistent between studies. Selection based on either phenotypic data or whole genome results obtained using a dense SNP chip would appear to be the most promising ways of achieving genetic progress.

An important feature of selection for nematode resistance is the interaction between host genotype and disease epidemiology, as altering host genotype can also change the force of infection faced by the population as a whole. In this case, by creating a population of animals that has lower mean FEC, the return of eggs to the pasture will decrease and the larval contamination on pasture will tend to decrease. This, in turn, will lead to reduced parasite challenge to all animals, furthering the benefits of selection. Therefore, the total benefits from selection are larger than those arising directly from genetic change in the host. This phenomenon was quantified *in silico* by Bishop and Stear (1997 and 1999). The benefits are manifested to some extent by decreased FEC, but more strongly by improved performance due to decreased larval challenge. Experimental verification of this phenomenon has been provided by Gruner et al. (2002) and Leathwick et al. (2002). This experimental demonstration is important, as the extra benefits are largely invisible if all animals are grazing the same pasture.

The model of Bishop and Stear (1997) has recently been extended to include a more mechanistic description of the development of immunity and the interactions between host and parasite (Vagenas et al., 2007a and b). In particular, this model incorporates host nutrition and, thus, when parameterised at the population level (Vagenas et al., 2007c) it can be used to explore interactions between host genotype and nutrition, and their impact on genetic parameters. This is important as it may go some way to resolving apparent contradictions seen across datasets and countries for genetic correlations between FEC and performance; for example, is the genetic correlation between FEC and growth rate favourable (negative) or unfavourable (positive). Assuming no linkage or pleiotropy between the genes underlying resistance to nematodes, Vagenas et al. (2007c) showed a marked predicted impact of level of nutrition; for poor nutrition moderate favourable correlations were predicted between FEC and lamb growth rate, however these correlations became essentially zero on high protein diets. These results were extended by Doeschl-Wilson et al. (2008) who demonstrated likely impacts of genetic correlations between underlying growth and immunological traits on predicted genetic parameters for production and resistance traits. Extreme genetic correlations observed from field studies could only be reproduced by assuming genetic relationships between the underlying input resistance traits. Altering preferences in the resource allocation between growth and immune response functions had less pronounced effects on the genetic parameters for the same traits. Effects were stronger when the allocation priority shifted towards growth, in which case worm burden and faecal egg counts increased and genetic correlations between these resistance traits and body weight became stronger. The results suggest that moderate pleiotropy and linkage may have large impacts on observed genetic parameters, and hence on outcomes of selection for nematode resistance.

Disease resistance

Footrot. Footrot is a common cause of lameness in both lambs and mature sheep and it is a major welfare problem in sheep. Footrot is a highly contagious bacterial disease caused by *Dichelobacter (Bacteroides) nodosus*. In addition to the welfare concerns, it is also a major cause of economic loss. In the UK context it is estimated to have economic costs to the UK sheep industry of £24 million per annum (Nieuwhof and Bishop, 2005).

Substantial genetic variation in resistance to footrot has been demonstrated by Raadsma et al. (1994) using deliberate challenge data, and by Nieuwhof et al. (2008) using field data. Nieuwhof et al. (2008) found that data describing ‘affected or not’ was at least as heritable as data giving more detailed descriptions of the severity of infection, possibly because these data only describe the ~10% of animals that have clinical signs of disease and not the ~90% of animals that do not. Essentially this heritability describes the probability of animals being in the ‘infectious/sick’ category in Fig. 1. The heritabilities would rise if the non-affected grouping of animals could be distinguished into the other non-affected categories in Fig. 1, and if those that have not been exposed to infection could be identified. A further finding from Nieuwhof et al. (2008) was that heritability of liability to footrot appeared to increase with flock disease prevalence, even when corrections were made for prevalence effects. This suggests that the greater force of infection in high prevalence flocks has allowed genetic variation in resistance to be more strongly expressed.

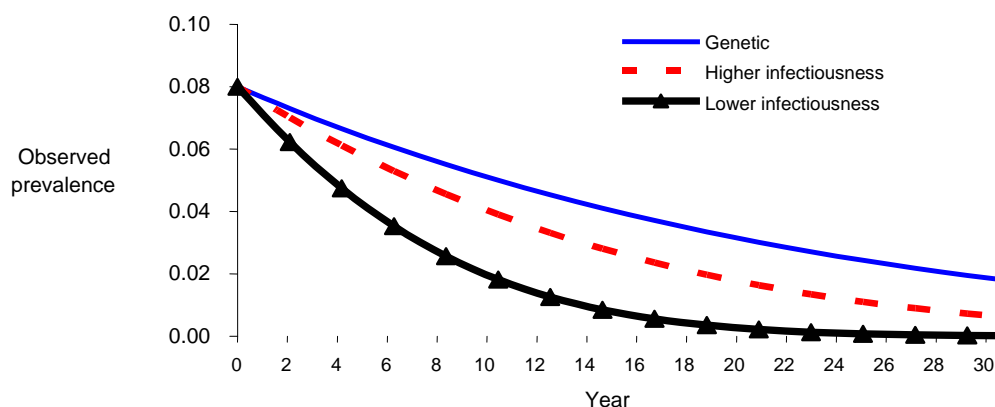


Figure 4. Responses to selection for footrot resistance, as assessed by observed prevalence of disease, predicted either by standard genetic theory or by a genetic epidemiological model, for situations where the bacteria have high or low infectivity.

Potential responses to selection for footrot have been quantified using a genetic-epidemiological model which accounts for transmission of infection via pasture as well as host genotype (Nieuwhof et al., 2009). Because footrot affected sheep can recover, show short-lived immunity to infection and ultimately become susceptible to reinfection (as shown in Fig. 1), epidemic properties are complex and an endemic equilibrium state can be reached in which the proportion of animals affected at any time point is largely constant, given constant weather conditions, stocking densities, management strategies, etc. Nieuwhof et al. (2009) derived mathematical terms describing the expected endemic equilibrium infection levels, and explored the properties of these equilibria as genetic progress was made for footrot resistance, assuming that resistance is expressed through the acquired immune response (hence the speed with which animals recover from infection). The results are shown in Fig. 4 for cases where genetic progress is estimated ignoring epidemiological impacts, and cases where the disease is assumed to have relatively high or low levels of infectiousness. For both assumptions about infectiousness, genetic

progress was somewhat faster than predicted by quantitative genetic theory alone, making selection a more attractive proposition. The more infectious the pathogen, hence the more quickly the disease spreads through the flock, the more closely the selection dynamics resemble those for normal production traits.

The footrot example illustrates several key lessons. Firstly, simple scoring criteria, such as presence or absence of hoof lesions, are heritable traits which should respond to selection. Secondly, responses to selection for disease resistance may be greater in higher prevalence herds. Lastly, total gains may exceed those predicted by quantitative genetic theory alone.

Mastitis. Mastitis, inflammation of the mammary gland, is usually caused by bacterial organisms such as *Staphylococcus spp.*, *Streptococcus spp.*, *Pseudomonas spp.*, *Mycoplasma spp.* and various coliforms such as *E. coli*. Mastitis incidence in the European dairy industries has been estimated at 30% of cows per year, and each case has been estimated to cost between 150 to 300 euros per diseased cow. Selection for increased milk yield will generally worsen the incidence of mastitis, due to the unfavourable genetic correlation between milk yield and mastitis susceptibility. Therefore, efforts to reduce mastitis, or prevent its incidence from rising, are a part of most dairy cattle evaluations. Currently, selection to reduce the incidence of mastitis is based on udder conformation, somatic cell count (SCC) and mastitis infection history. SCC and clinical mastitis generally have low heritabilities, usually in the range 0.05 to 0.15 (Rupp and Boichard, 2003). Mastitis resistance is probably due to structural attributes of the udder or teat, as well as immune responses. QTL associated with mastitis resistance traits have been reported on almost all of the 29 bovine chromosomes, in a variety of populations and breeds (see Khatkar et al., 2004). This large number of QTL suggests that gene-assisted selection, using causative mutations underlying these QTL, may be inefficient if each of these mutations explains only a small proportion of the observed variation in SCC or clinical mastitis. A genome-wide selection approach using a dense SNP chip would be advantageous.

However, there is also likely to be considerable benefit in redefining traits describing mastitis resistance, and this may also address concerns as to whether continued selection for reduced SCC is a long-term solution to mastitis. SCC is used as an indicator of mastitis, with high SCC values indicating that an animal is likely to be infected. However, SCC measurements on a group of animals comprise a mixture distribution trait describing baseline SCC in unaffected animals as well as elevated SCC in infected animals. The concern is that reducing SCC too far may reduce baseline SCC levels, hence animals' ability to respond to infection (Rupp and Boichard, 2003). In actual fact, the trait that is of interest to the breeder is liability to mastitis. Therefore, a rational approach when considering SCC data is to decompose it into baseline SCC values for uninfected animals, response SCC values for animals that are infected, along with the probability that a particular animal falls into one distribution and not the other. This concept was introduced by Odegard et al. (2005), and the quantitative genetic properties of such mixture distribution traits were formalised by Gianola et al. (2006). The primary selection criterion arising from this data decomposition is the liability of an animal to be affected by mastitis. The secondary question is whether to increase or decrease SCC; however, this question must be asked separately for baseline and response SCC. To answer this, it is necessary to calculate genetic correlations between SCC and mastitis liability, separately for baseline and response SCC. It is suggested that selection for resistance should not be for SCC *per se*, but for liability to infection, either conditional upon SCC or assessed independently through diagnoses of infection or clinical signs of mastitis.

The fact that mastitis is caused by different species of bacteria raises further issues of potential importance. Can these separate categories of infection be teased apart and is it beneficial to do so? Further, for the infectious (as opposed to 'environmental') sources of mastitis, further insight may be gained by assessing cow liability to mastitis in relation to the force of infection. For example,

Disease resistance

are there genetic influences on the order in which animals become infected, and do genetic effects alter as disease prevalence changes? If these effects exist, they may point to additional epidemiological benefits from selection for increased resistance. However, it may require considerable quantities of detailed data to assess these effects.

CONCLUSIONS

This key take-home message from this paper is that the interpretation of data and estimation of benefits of selection differ for disease data compared to standard performance data, and these interpretations also differ between categories of disease. It is important to understand the disease in question, rather than simply treat data as if it were performance trait data. From a breeding perspective, the impacts of accounting for transmission of infection and disease epidemiology are generally positive. For example, weak genetic signals in field data may often be masking stronger underlying genetic effects. Further, true benefits of selection may well be greater than those predicted using naïve quantitative genetic theory that ignores disease epidemiology.

In summary, considerable opportunities exist to breed animals for enhanced disease resistance, often simply using noisy field data either directly for selection or to calibrate SNP chip data.

ACKNOWLEDGEMENTS

This paper has summarised work funded by the BBSRC, Defra, MLC, REERAD, GENUS and the EU. I wish to thank many colleagues for inputs, most of whom have been cited as coauthors.

REFERENCES

- Bishop, S. C. (2005) In: "Encyclopedia of Animal Science", p. 288. editors W. G. Pond and A.W. Bell, Marcel Dekker Inc., New York.
- Bishop, S.C. and MacKenzie, K.M. (2003) *Genet. Sel. Evol.* **35**:S3.
- Bishop, S.C. and Morris, C.A. (2007) *Small Rum. Res.* **70**:48.
- Bishop, S.C. and Stear, M.J. (1997) *Anim. Sci.* **64**:469.
- Bishop, S.C. and Stear, M.J. (1999) *Anim. Sci.* **69**:515.
- Davies, G., Genini, S., Bishop S.C. and Giuffra, E. (2009) *Anim.* **3**:415.
- Doeschl-Wilson, A.B., Vagenas, D., Kyriazakis, I., Bishop, S.C. (2008) *Genet. Sel. Evol.* **40**:241.
- Gianola, D., Heringstad, B. and Odegaard, J. (2006) *Genet.* **173**:2247.
- Gruner, L., Cortet, J., Sauve, C., Limouzin, C. and Brunel, J.C. (2002) *Vet. Parasit.* **109**:277.
- Khatkar, M.S., Thomson, P.C., Tammen, I. and Raadsma, H.W. (2004) *Gen. Sel. Evol.* **36**:163.
- Leathwick, D.M., Atkinson, D.S., Miller, C.M., Brown, A.E. and Sutherland, I.A. (2002). Proc. of Novel Approaches III, a workshop on helminth control in livestock in the new millennium. Moredun Research Institute, Scotland. July 2-5, 2002.
- MacKenzie, K. and Bishop, S.C. (1999) *Anim. Sci.* **69**:543.
- MacKenzie, K. and Bishop, S.C. (2001) *J. Anim. Sci.* **79**:2057.
- Nieuwhof, G.J. (2009) PhD Thesis, University of Edinburgh.
- Nieuwhof, G.J. and Bishop, S.C. (2005) *Anim. Sci.* **81**:23.
- Nieuwhof, G.J., Conington J. and Bishop, S.C. (2009) *Genet., Select., Evol.* **41**:19.
- Nieuwhof, G.J., Conington, J., Bünger, L. Haresign, W. and Bishop, S.C. (2008) *Anim.* **2**:1289.
- Odegard, J., Madsen, P., Gianola, D., Klemetsdal, G., Jensen, J., Heringstad, B. and Korsgaard, I.R. (2005) *J. Dairy Sci.* **88**: 652.
- Raadsma, H.W., Egerton, J.R., Wood, D., Kristo, C. and Nicholas, F.W. (1994) *J. Anim. Breed. Genet.* **111**:367.
- Rupp, R. and Boichard, D. (2003) *Vet. Res.* **34**:671.
- Vagenas, D., Bishop, S.C. and Kyriazakis, I. (2007a) *Parasit.* **134**:1263.
- Vagenas, D., Bishop, S.C. and Kyriazakis, I. (2007b) *Parasit.* **134**:1279.
- Vagenas, D., Doeschl-Wilson, A., Bishop, S.C. and Kyriazakis, I. (2007c) *Int. J. Parasit.* **37**:1617.