THE UNIVERSITY
of EDINBURGH

Coinfections in East African Shorthorn Zebu

Rebecca Callaby

University of Edinburgh

This dissertation is submitted for the degree of

Doctor of Philosophy

2015

Declaration

This dissertation is submitted in accordance with the requirements for a Doctorate of Philosophy by the School of Biological Sciences at the University of Edinburgh. The work included in this thesis has not been submitted for any other degree or professional qualification. The data described in this thesis was not collected by me, it was gathered by members of the Infectious Diseases in East African Livestock (IDEAL) project. The whole project is fully described in Bronsvoort et al. (2013) and papers written in connection with the IDEAL project are listed below. This thesis follows on from these manuscripts, so the list provides a reference to all members of the IDEAL project who have collected the data that I have used in this thesis. In addition, the selection analysis presented in Chapter 5 was completed by Hussain Bahbahani of the University of Nottingham. The rest of the data analysis presented in this thesis is my own work.

IDEAL publications

- Woolhouse, M. E. J., Thumbi, S. M., Jennings, A., Chase-Topping, M., Callaby, R., Kiara, H., Oosthuizen, M. C., Mbole-Kariuki, M. N., Conradie, I., Handel, I. G., Poole, E. J., Njiiri, E., Collins, N. E., Murray, G., Tapio, M., Auguet, O. T., Weir, W., Morrison, W. I., Kruuk, L. E. B., Bronsvoort, B. M. d. C., Hanotte, O., Coetzer, K., and Toye, P. G. (2015). Co-infections determine patterns of mortality in a population exposed to parasite infection. *Science Advances*, 1(2):e1400026

- Njiiri, N. E., Bronsvoort, B. M. d. C., Collins, N. E., Steyn, H. C., Troskie, M., Vorster, I., Thumbi, S. M., Sibeko, K. P., Jennings, A., van Wyk, I. C., Mbole-Kariuki, M., Kiara, H., Poole, E. J., Hanotte, O., Coetzer, K., Oosthuizen, M. C., Woolhouse, M., and Toye, P. (2015). The epidemiology of tick-borne haemoparasites as determined by the reverse line blot hybridization assay in an intensively studied cohort of calves in Western Kenya. *Veterinary Parasitology*, In Press, Corrected Proof

- Kiara, H., Jennings, A., Bronsvoort, B. M. d. C., Handel, I., Mwangi, S. T., Mbole-Kariuki Ndila, M., Conradie van Wyk, I., Poole, E. J., Hanotte, O.,

Coetzer, J. A. W., Woolhouse, M., and Toye, P. (2014). A longitudinal assessment of the serological response to *Theileria parva* and other tick-borne parasites from birth to one year in a cohort of indigenous calves in Western Kenya. *Parasitology*, 141:1289–1298

- Mbole-Kariuki, M. N., Sonstegard, T., Orth, A., Thumbi, S. M., Bronsvoort, B. M. d. C., Kiara, H., Toye, P., Conradie van Wyk, I., Jennings, A., Coetzer, K., Woolhouse, M. E. J., Hanotte, O., and Tapio, M. (2014). Genome-wide analysis reveals the ancient and recent admixture history of East African shorthorn zebu from Western Kenya. *Heredity*, 113:297–305

- Conradie van Wyk, I., Goddard, A., Bronsvoort, B. M. d. C., Coetzer, J., Handel, I., Hanotte, O., Jennings, A., Lesosky, M., Kiara, H., Thumbi, S., Toye, P., Woolhouse, M., and Penzhorn, B. (2014). The impact of co-infections on the haematological profile of East African short-horn zebu calves. *Parasitology*, 141(3):374–388

- Thumbi, S. M., Bronsvoort, B. M. d. C., Poole, E. J., Kiara, H., Toye, P. G., Mbole-Kariuki, M. N., Conradie, I., Jennings, A., Handel, I. G., Coetzer, J. A. W., Steyl, J. C. A., Hanotte, O., and Woolhouse, M. E. J. (2014). Parasite co-infections and their impact on survival of indigenous cattle. *PLoS One*, 9(2):e76324

- Murray, G. G., Woolhouse, M. E. J., Tapio, M., Mbole-Kariuki, M. N., Sonstegard, T. S., Thumbi, S. M., Jennings, A. E., van Wyk, I. C., Chase-Topping, M., and Kiara, H. (2013). Genetic susceptibility to infectious disease in East African shorthorn zebu: a genome-wide analysis of the effect of heterozygosity and exotic introgression. *BMC Evolutionary Biology*, 13(1):246

- Thumbi, S., Bronsvoort, B. M. d. C., Kiara, H., Toye, P., Poole, J., Ndila, M., Conradie, I., Jennings, A., Handel, I., Coetzer, J., Steyl, J., Hanotte, O., and Woolhouse, M. (2013a). Mortality in East African shorthorn zebu cattle under one year: predictors of infectious-disease mortality. *BMC Veterinary Research*, 9(1):175

- Thumbi, S. M., Bronsvoort, B. M. d. C., Poole, E. J., Kiara, H., Toye, P., Ndila, M., Conradie, I., Jennings, A., Handel, I. G., Coetzer, J. A. W., Hanotte, O., and Woolhouse, M. E. J. (2013b). Parasite co-infections show synergistic and antagonistic interactions on growth performance of East African zebu cattle under one year. *Parasitology*, 140:1789–1798

- Toye, P., Handel, I., Gray, J., Kiara, H., Thumbi, S., Jennings, A., van Wyk, I. C., Ndila, M., Hanotte, O., Coetzer, K., Woolhouse, M., and Bronsvoort, B. M. d. C. (2013a). Maternal antibody uptake, duration and influence on survival and growth rate in a cohort of indigenous calves in a smallholder farming system in Western Kenya. *Veterinary Immunology and Immunopathology*, 155(1-2):129–134

- Toye, P. G., Batten, C. A., Kiara, H., Henstock, M. R., Edwards, L., Thumbi, S., Poole, E. J., Handel, I. G., Bronsvoort, B. M. d. C., Hanotte, O., Coetzer, J. A. W., Woolhouse, M. E. J., and Oura, C. A. L. (2013b). Bluetongue and epizootic haemorrhagic disease virus in local breeds of cattle in Kenya. *Research in Veterinary Science*, 94(3):769–773

- Bronsvoort, B. M. d. C., Thumbi, S., Poole, E., Kiara, H., Auguet, O., Handel, I., Jennings, A., Conradie, I., Mbole-Kariuki, M., Toye, P., Hanotte, O., Coetzer, J., and Woolhouse, M. (2013). Design and descriptive epidemiology of the Infectious Diseases of East African Livestock (IDEAL) project, a longitudinal calf cohort study in Western Kenya. *BMC Veterinary Research*, 9(1):171

- Lesosky, M., Dumas, S., Conradie van Wyk, Ilana, I., Handel, I., Jennings, A., Thumbi, S., Toye, P., and Bronsvoort, B. M. d. C. (2013). A live weight-heart girth relationship for accurate dosing of East African shorthorn zebu cattle. *Tropical Animal Health and Production*, 45:311–316

- Conradie van Wyk, I., Goddard, A., Bronsvoort, B. M. d. C., Coetzer, J., Booth, C., Hanotte, O., Jennings, A., Kiara, H., Mashego, P., Muller, C., Pretorius, G., Poole, E., Thumbi, S., Toye, P., Woolhouse, M., and Penzhorn, B. (2012). Hematological profile of East African short-horn zebu calves from birth to 51 weeks of age. *Comparative Clinical Pathology*, 22:1029–1036

Publications as a result of this thesis

- Callaby, R., Hanotte, O., Conradie van Wyk, I., Kiara, H., Toye, P., Mbole-Kariuki, M. N., Jennings, A., Thumbi, S. M., Coetzer, J. A. W., Bronsvoort, B. M. d. C., Knott, S. A., Woolhouse, M. E. J., and Kruuk, L. E. B. (2015). Variation and covariation in strongyle infection in East African shorthorn zebu calves. *Parasitology*, 142(03):499–511

<div align="right">
Rebecca Callaby

2015
</div>

Acknowledgements

Abstract

The Infectious Diseases of East African Livestock (IDEAL) project followed 548 East African Shorthorn Zebu (EASZ) calves in Western Kenya for the first year of life and monitored the sequelae of infections by multiple parasites. More than 50 different parasites were identified during this time. The IDEAL project also gathered environmental information about the farm and collected phenotypic data on the calf and its dam. Calves were also genotyped for 55,777 single nucleotide polymorphisms (SNPs).

Recent research has looked at coinfection in rodents and humans but not in indigenous cattle. Here I investigate the evidence for coinfection in EASZ and study the associations occurring between coinfecting parasites. In addition, I examine the genetic and phenotypic factors which predispose an individual to infection with multiple parasites. Using information gathered by the IDEAL project, my thesis consists of the following chapters.

An investigation of the nature of concurrent associations and of lagged effects between different parasites. Using the parasites *Theileria* spp., *Coccidia* spp., *Strongyloides* spp., strongyles and *Calicophoron* spp. I show that the patterns of association between different parasites are complex: there is evidence for both positive and negative associations. For example, infection with *Strongyloides* spp. increased the risk of strongyle infection. Conversely, in other cases, being infected with one parasite decreased the calf's risk of infection with another parasite: for example, infection with *Strongyloides* spp. decreased the risk of infection with *Calicophoron* spp.

A study of the relationship between different respiratory viruses and their effect upon the host. I confirm that positive associations exist between Infectious Bovine Rhinotracheitis (IBR), Bovine Viral Diarrhoea Virus (BVDV) and Bovine Parainfluenza Virus Type 3 (PIV3) in a previously unstudied setting; being seropositive for any one of these three viruses means that an individual is more likely to be seropositive for the other two viruses than expected by chance. Being seropositive for IBR, BVDV or PIV3 did not affect the average daily weight gain of the calf, nor did PIV3 and BVDV serostatus have an effect on the calf ever experiencing a clinical

episode. However, IBR seropositive calves were less likely to experience a clinical episode of some form, suggestive of some protective aspect of IBR.

An examination of the sources of variation in faecal strongyle egg counts (EPG), and their association with body weight, host genetics and a suite of haematological measures. Using estimates of relatedness derived from the SNP data, I established that strongyle EPG has a genetic basis in EASZ, with a heritability of 23.9% (S.E. = 11.8%) and showed a consistently strong negative association between strongyle infection and the haematological parameters white blood cell count, red blood cell count, total serum protein and absolute eosinophil count. Furthermore, calf body weight at 1 week old was a significant predictor of strongyle EPG at 16-51 weeks, with smaller calves being predisposed to a higher strongyle EPG later in life.

A genome-wide association study (GWAS) to investigate if there is a genetic predisposition to East Coast Fever (ECF) death and a genetic basis to the packed cell volume at the time of seroconversion to *Theileria parva* (PCV_{TP}). I found no robust evidence for a relationship between genotyped single nucleotide polymorphisms (SNPs) and ECF death or PCV_{TP}. The effect of sample size upon GWAS and significance thresholds was investigated further through simulations. I conclude that the small number of cases influences the probability of association between a SNP and the phenotypic trait. Smaller case numbers produce more artifactual associations with SNPs, an effect not fully compensated for by the standard Bonferroni correction, suggesting that an empirical significance threshold should be used to directly account for sample size.

The results of this thesis provide an understanding of the associations occurring between different parasites, and of their causes and consequences. I discuss the results in the context of their implications for disease control strategies, suggesting the benefits of an integrated approach to control worm and *T. parva* alongside the possible genetic selection for parasite resistance and supplementary feeding of lightweight individuals to improve the health of EASZ.

Lay Summary

The Infectious Diseases of East African Livestock (IDEAL) project followed 548 newborn East African shorthorn zebu (EASZ) calves surrounding the town of Busia, on the edge of Lake Victoria in Western Kenya, for the first year of life. Farmers in this region rely heavily upon their livestock for milk, food and capital reserves etc. and there is very limited use of vaccination or other preventative measures against infectious diseases in this population. During the study period the calves became infected with multiple parasites and there was marked differences in individuals response to infection. Recent research has looked at coinfection in rodents and humans but not in indigenous cattle. Using data gathered by the IDEAL project this thesis aims to investigate the evidence for coinfection within EASZ and to generate a better understanding of the genetic and physical attributes of cattle which predisposes an individual to infection with multiple parasites.

I found evidence for both positive and negative associations amongst parasites infecting EASZ. Using different groups of worms found in the gastrointestinal tract of the cow as an example, in some instances, having one species of worm e.g. *Strongyloides* spp. increases a calf's risk of infection with another type of worm such as strongyles. Conversely, infection with *Strongyloides* spp., can also decrease a calf's risk of infection with other parasites such as *Calicophoron* spp. Infection with any one of the following viruses, Infectious Bovine Rhinotracheitis (IBR), Bovine Viral Diarrhoea Virus (BVDV) and Bovine Parainfluenza Virus Type 3 (PIV3), means that the calf is more likely to be infected with the other two viruses than you would expect by chance.

I established that there is a genetic basis to the number of worms a calf has (as measured by strongyle eggs per gramme of faeces (EPG)), which suggests that the number of worms a calf has is similar to the number its parents had and that which its own offspring will have. In addition, I found that it is possible to predict the number of worms a calf has at an older age by the birth weight of a calf, with smaller calves being predisposed to a higher strongyle EPG later in life. However, I found no robust evidence for a genetic basis to East Coast Fever (ECF) death or the calf's response to infection with *Theileria parva*. ECF is a tick-borne disease of cattle in Eastern Africa

caused by the parasite *T. parva* which results in death or severe illness.

I discuss the results of this thesis in the context of their implications for disease control strategies within East Africa, suggesting the benefits of an integrated approach to control worm and *T. parva* alongside the possible genetic selection for parasite resistance and supplementary feeding of lightweight individuals to improve the health of the calves.

Contents

List of Figures

List of Tables

Abbreviations

AEZ	Agro-ecological zone
BRD	Bovine respiratory disease complex
BVDV	Bovine viral diarrhoea virus
COV_P	Phenotypic covariance
COV_{PE}	Covariance in permanent environmental effects
EASZ	East African shorthorn zebu
ECF	East coast fever
ELISA	Enzyme-linked immunossorbent assays
EO	Absolute eosinophil count
EPG	Eggs per gramme of faeces
ET	European taurine
FEC	Faecal egg count
GLMM	Generalised linear mixed models
GWAS	Genome-wide association study
h^2	Heritability
HWE	Hardy-Weinberg equilibrium
IBR	Infectious bovine rhinotracheitis
IBS	Identity by state
IDEAL	Infectious diseases of East African livestock

LD	Linkage disequilibrium
LMM	Linear mixed models
MAF	Minor allele frequency
NDVI	Normalized difference vegetation index
PCV	Packed cell volume
PCV_{TP}	Packed cell volume at the time of seroconversion to *Theileria parva*
PIV3	Bovine parainfluenza virus type 3
PP	Percent positivity
QTL	Quantitative trait loci
r^2	Repeatability
RBC	Red blood cell count
SL	Sublocation
SNP	Single nucleotide polymorphism
TSP	Total serum protein
V_A	Additive genetic variance
V_P	Total phenotypic variance
V_{PE}	Permanent environmental variance
V_{SL}	Sublocation variance
V_{RES}	Residual variance
WBC	White blood cell count

Chapter 1

General Introduction

Coinfection is a descriptive term for being infected with multiple parasites. Parasites are organisms that live in or on the host and take their nourishment from or at the expense of the host. Infection means the invasion of the host with an organism that is not normally present in the body. Most hosts are commonly infected with multiple parasite species during their lifetime and many are often infected with more than one parasite species at once (Cox, 2001). I will extend the definition of coinfection to include individuals which are infected with multiple parasites through time, as the order of infection can be important for determining the consequence of infection and infection with other parasite species. For example, cattle infected with the protozoa *Theileria mutans* before *Theileria parva* experience a reduction in mortality associated with *T. parva* infection (Woolhouse et al., 2015). A similar pattern is observed with malaria infections in humans (Williams et al., 1996).

Coinfection is likely to occur by chance alone. Yet, is it possible for an individual to have more than just 'bad luck' that causes it to get multiple infections? Or is there any evidence for infection with multiple parasites being non-independent?

Coinfection is important as it can have many different consequences upon the host and the parasites concerned. Coinfection can affect the severity of disease (defined according to the signs or symptoms associated with a specific cause) and have unexpected impacts upon the host (Pedersen and Fenton, 2007). For example, Thumbi et al. (2013b) found both synergistic and antagonistic effects of coinfections on East Africa shorthorn zebu calves: calves had lower growth rates when they were coinfected with *Theileria parva* and *Anaplasma marginale* but a higher growth rate when they were coinfected with *T. parva* and *T. mutans*, compared to when the calf only experienced a single infection of *T. parva*. In this example, the effect of having the two parasites was greater than the sum of effects for each individual parasite (Thumbi et al., 2013b). Furthermore, coinfection can have great implications for

1

disease control and population health campaigns, if for example, the removal of one competing parasite inadvertently promotes another disease-causing parasite (Lafferty, 2010).

Accordingly, there are many important questions to ask about coinfection. What causes hosts to be infected with multiple parasites in the first place? What are the consequences of being infected with multiple parasites? In this thesis, I explore the evidence for coinfection in East African Shorthorn Zebu (EASZ) calves, in Western Kenya, using longitudinal cohort data gathered by the Infectious Diseases in East African Livestock (IDEAL) project (Bronsvoort et al., 2013). First, in this introductory chapter, I review the hypotheses concerning the causes and consequences of coinfection, as well as their importance, before discussing studies to investigate the occurrence of coinfection. Definitions for the infectious disease terms used in this thesis can be found in Table 1.1, at the end of this chapter, since the meanings of these terms vary in the veterinary medicine and ecological literature.

1.1   Causes of coinfection

There are three broad reasons used to explain why coinfection occurs. These are: 1) parasites share environmental requirements such as a similar spatial distribution, overlapping habitat or seasonal pattern in occurrence; 2) the characteristics of the host, for example their phenotype and/or genotype, may increase host susceptibility to infection with multiple kinds of parasites; and 3) coinfection could be a result of the host's history of infection, for example if prior infection with one parasite increases the chance of infection with another. The first reason is due to the characteristics of the parasite whilst the second is due to the characteristics of the host, whereas the third is due to characteristics of both the host and parasite. Below I consider each scenario in more detail.

1.1.1   Shared environmental requirements

The first explanation for the occurrence of coinfection is that it arises because individual parasites share common environmental risk factors. The occurrence of coinfection may depend upon the spatial and temporal distribution of the parasites within the environment (Batchelor et al., 2009; Davies and Pedersen, 2008; Wardrop et al., 2010). For example, temperature, elevation and distance to water are all predictors of *Schistosoma mansoni* and hookworm (*Ancylostoma duodenale / Necator americanus*) co-infection in humans since these are the habitat requirements of both

helminths (Brooker and Clements, 2009; Saathoff et al., 2005). Seasonal patterns in the presence of parasites may also occur. For instance, Cattadori et al. (2008) described the pattern of infection in rabbits with the gastrointestinal nematodes *Trichostrongylus retortaeformis* and *Graphidium strigosum* and showed that it was highly seasonal. Rabbits tend to have the highest prevalence of *T. retortaeformis* and *G. strigosum* between April and August and this seasonal variation contributed to the aggregation of parasites within the host (Cattadori et al., 2008). Therefore, the relevant spatial and seasonal scale at which coinfection occur depend upon the host and parasites in question (Viney and Graham, 2013).

Coinfection may also occur because parasites share the same transmission route. For example if two parasites are directly transmitted though soil then there is the potential for spatial and/or behaviour correlation in the hosts risk of exposure (Fenton et al., 2010). In addition, if vectors for different parasites share have similar environmental requirements then coinfection may also occur. For example, *S. mansoni* requires a aquatic-snail as part its lifecycle whilst malaria-transmitting mosquitoes require standing water thus people living in close proximity to bodies of water are at increase risk of infection with both parasites (Booth et al., 2004). Hence, it is important to account for the role of environmental effects in generating covariation between different parasites, to avoid making false inferences about effects of coinfecting parasites (Hellard et al., 2012). In addition, prior knowledge of the areas with high risk of multiple infections is important for the design of public health interventions since integrated/combined disease programs could be used to increase the cost effectiveness of such interventions (Schur et al., 2011).

1.1.2 Host characteristics/traits

Secondly, the characteristics of the individual may be important for determining the occurrence of coinfection. Firstly I will consider the effects of the host's phenotype on increasing its susceptibility to infection, before examining the different methods used to evaluate a host's genetic predisposition to infection.

Host phenotype

Individuals can be phenotypically classified by their nutritional or haematological status, which may influence their risk of infection. For example, malnutrition and anaemia can predispose hosts to infection. Humans lacking in micronutrients such as foliate are at increased risk of acquiring the human papillomavirus (Piyathilake et al., 2004). Likewise, Beldomenico et al. (2008) showed that rodents in poor condition (i.e.

individuals which are anaemic) had elevated haematological indicators of infection (neutrophilia and monocytosis) which were generally followed by a further decrease in indicators of host condition (red blood cell counts and lymphocytes), which in turn predisposes the individual to infection with other parasites such as cowpox, thus a "vicious cycle" is generated (Beldomenico and Begon, 2010; Beldomenico et al., 2008, 2009).

Even an individual's condition at birth may influence its infection status later in life. For example, low birth weight humans are more likely to be hospitalised for an infectious disease then heavier children and it has been proposed that lighter babies have depressed immune function (Chatrath et al., 1997; Hviid and Melbye, 2007). In ungulates, a positive phenotypic correlation has been found between parasite resistance and body size in Soay sheep (Coltman et al., 2001). These studies suggest that smaller or lighter individuals are at increased risk of infection and experience more severe consequences as a result. In addition, maternal effects can also influence parasite resistance in an individual. Soay sheep lambs born as twins or those born to the very young or old mothers show the highest nematode burdens (Hayward et al., 2010). This further supports the evidence that host phenotype and condition are important for determining the occurrence of infection.

Host genotype

a) Genetic variation in infection

There has been very limited work on how host genetics affects susceptibility to coinfection (Viney and Graham, 2013). One of the few studies which have looked for a common genetic control between the occurrence of co-infecting parasites was carried out by Pullan et al. (2010) on coinfection with hookworm (*Necator americanus*) and *Schistosoma mansoni* in a rural Brazilian community. The study found no evidence for a genetic correlation in the intensity of hookworm and schistosome infection, despite the infection intensities of both worms having high species-specific heritability. Infection intensity quantifies the number of worms a individual is infected with, this was measured indirectly by faecal egg counts. The authors attributed the lack of genetic correlation to exposure-related factors playing a greater role in determining coinfection risk than host genetic relatedness (Pullan et al., 2010).

a.i) Methods to identify genetic predisposition to a single infection

Quantitative genetics The genetic predisposition to a single parasite is more commonly investigated than the genetic predisposition to multiple infection. One

method to investigate the genetic disposition to a single infection is by using quantitative genetics. In quantitative genetics the variance of quantitative phenotypic traits within a population can be broken down into genetic and environmental components using quantitative genetics and therefore parameters such as heritability can be estimated. Heritability (h^2) is defined as the proportion of phenotypic variance (V_P) explained by the additive genetic variance (V_A), i.e. $h^2 = V_A/V_P$. Hence, heritability describes the extent to which the phenotypic variance between individuals is determined by additive genetic differences (Kruuk, 2004; Lynch and Walsh, 1998).

This technique can be useful in determining if an infection has a genetic component. For instance, it is known that variation in strongyle eggs can be due to variation in susceptibility (the ability of the host to become infected), resistance (the ability of the host to limit its parasite burden), tolerance (the ability of the host to limit harm from a parasite) or exposure (what the host experiences) to infection by strongyle worms. Evidence from domestic ungulates suggests that variation in strongyle faecal egg count (FEC) frequently has a heritable basis: for example, strongyle FEC has a heritability of approximately 30% in many cattle breeds (Leighton et al., 1989; Stear et al., 1990, 1988). However, with relevance to this thesis, the heritability of strongyle egg count is unknown in EASZ.

Genome Wide Association Studies An alternative method to identify a genetic predisposition to infection is to use genome wide association studies (GWAS). This method tries to identify associations between genotyped loci called single nucleotide polymorphisms (SNPs) within a population and a complex phenotypic trait. Since it is unlikely that the causal loci responsible for contributing to disease risk is genotyped, GWAS works by assuming that the genotyped SNPs are in linkage disequilibrium (LD) with ungenotyped causal variants that have a direct or indirect functional effect on disease risk (Visscher et al., 2012). LD is the non-random association between alleles at different loci caused by mutation, drift and it is broken down by recombination, meaning that loci closer to each other on the chromosome have higher LD (Wang et al., 2005).

Since the invention of GWAS, many significant discoveries have been made. In cattle, GWAS studies have mainly focused on production traits and resistance to endemic diseases (Bishop and Woolliams, 2014). One such example of GWAS in cattle is Bermingham et al.'s 2014 study of resistance to *Mycrobacterium bovis* (bovine tuberculosis, bTB). The study identified SNPs located in two novel bTB resistance loci that included the protein tyrosine phosphatase receptor T and myosin genes. SNPs surrounding these two genes could explain 21% of the phenotypic variance in bTB

resistance (Bermingham et al., 2014). However, I am unaware of any GWAS studies of coinfection in cattle, humans or other mammals.

a.ii) What does genetic variation in a single infection mean for coinfections?

Given, that the above examples show that it is possible to identify a genetic basis to the susceptibility or the response of a host to single infections; it is likely that coinfecting parasites may share a common genetic predisposition. One possible mechanism by which genetics may contribute to causing coinfection is through its control on the immune response (Viney and Graham, 2013). For example, the major histo-compatibility complex (MHC) is a diverse family of genes involved in the adaptive immune response (Janeway et al., 2005). It is hypothesised that, within a population, individuals have alleles on the MHC that confer susceptibility and resistance to an infection, however in heterozygotes the susceptible alleles are masked because resistance is generally on a dominant allele, therefore MHC heterozygotes may experience an advantage due to them being susceptible to fewer parasites than homozygotes (McClelland et al., 2003; Sin et al., 2014). For instance, MHC heterozygote mice had a lower standardised parasite load then homozygotes, suggesting that heterozygotes where superior to homozygotes when they were experimentally coinfected with *Salmonella enterica* and Theiler's murine encephalomyelitis virus (McClelland et al., 2003). However, natural coinfection status is not correlated with MHC heterozygosity coinfection in European badgers or water voles (Sin et al., 2014; Tollenaere et al., 2008). The difference between the two examples could be due to small sample sizes which lack the power to detect association between the MHC alleles and a parasite (Hedrick and Kim, 2000).

b) Breed effects on infection

In addition to specific causal variants and SNPs, it is well known that certain breeds of livestock tolerate infection better than other breeds. Examples of this include comparisons between the indigenous zebu (*Bos taurus indicus*) and European taurine (*Bos taurus taurus*) cattle populations in Africa. Ameni et al. (2007) found that indigenous zebu in Ethiopia have a lower prevalence of bovine tuberculosis (bTB) and a decreased severity of bTB disease compared to exotic Holstein cattle. In contrast, *Theileria parva*, the tick-borne pathogen which causes East Coast Fever (ECF) which occurs in eastern and central Africa, frequently causes fatal disease in exotic *Bos taurus*, but causes less severe (often subclinical) disease in indigenous *Bos indicus* cattle (Kariuki et al., 1995). In this thesis I defined pathogen as an infectious agent which can cause disease or damage to the host. Furthermore, *B. indicus x B. taurus*

crosses experience more severe clinical symptoms in response to *Babesia bigemina* infection then *B. indicus*, but this response is less severe in comparison to the *B. taurus* response (Bock et al., 1997). These findings suggests adaptation by each breed to different environments, driven by exposure to different parasites and selection pressures (Hanotte et al., 2010).

1.1.3 Infection history of the host

The occurrence of coinfection is context specific (Jackson et al., 2006; Telfer et al., 2008) and depends heavily upon the temporal pattern of infection or the host's history of exposure (Ezenwa and Jolles, 2011; Lello et al., 2008). Coinfecting parasites can have both positive and negative (antagonistic or synergistic) effects upon the parasites and host involved. These interactions can occur directly or indirectly via a variety of mechanisms as outlined in Pedersen and Fenton (2007).

Parasites sharing a similar physical location within the host have the potential to interact because they consume and compete for the same resources (Pedersen and Fenton, 2007). For example, Figure 1.1 shows the helminths and protozoa of the gastrointestinal tract (GI) of EASZ from the Infectious Diseases of East African Livestock (IDEAL) project, the longitudinal study on which this thesis is based (see Bronsvoort et al. (2013), section 1.4 or Appendix A for more details).

Negative associations between parasites may occur through immune-mediated competition or cross-immunity between two antigenically-similar parasites, whereby previous infection with one parasite may cause an immune response to a second antigenically-similar parasite (Pedersen and Fenton, 2007). It has been proposed that cross-immunity may exist between yellow fever and dengue fever, and *Plasmodium vivax* and *P. falciparum* (Woolhouse et al., 2005). Furthermore, as the host immune system has a memory, temporally asynchronous species can be associated with each other. Lello et al. (2004) provides an illustration of this type of association in the free-ranging rabbit (*Oryctolagus cuniculus*). Rabbits have a 29% lower infection intensity of the stomach helminth *Graphidium strigosum* when they are infected with the small intestine helminth *Trichostrongylus retortaeformis*. Direct interaction between the two helminths is impossible, as they reside in different regions of the body, therefore an immune-mediated response is the most likely explanation for this finding (Lello et al., 2004).

Indirect immunological responses can also however have positive synergistic results on coinfecting parasites (yet negative results on the host), via both immuno-suppression and Th1-Th2 trade-offs (Graham et al., 2007; Pedersen and Fenton, 2007). For instance, infectious bovine rhinotracheitis virus (IBR) and the
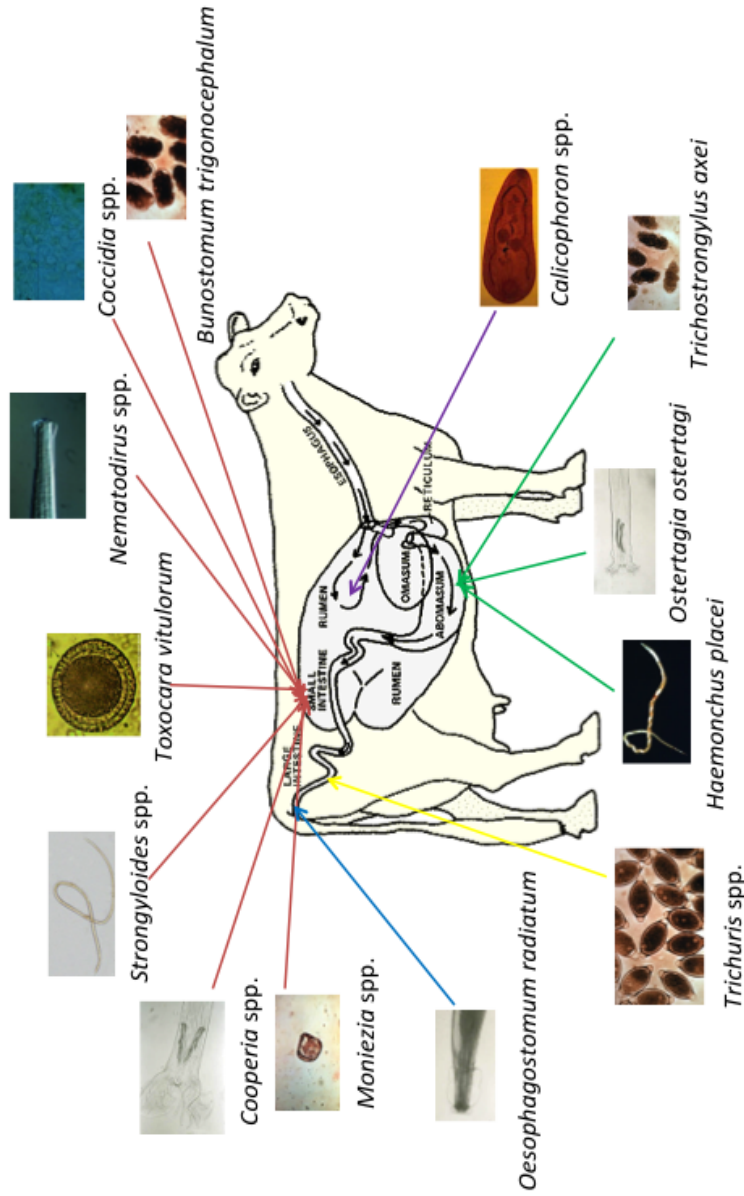
Figure 1.1 Helminths and protozoa of the gastrointestinal (GI) tract of East African shorthorn zebu calves from the Infectious Diseases of East African Livestock (IDEAL) project. The colour of the arrow indicates that the parasites are found within the same site of the GI tract. Samples with >2000 coccidia oocysts present were cultured to the type of coccidia present (Bronsvoort et al., 2013). The location of different *Coccidia* spp. within the host depends upon the species. *Eimeria bovis*, *Eimeria zuernii* and *Eimeria alabamensis* are found within the small and large intestine, *Eimeria auburnensis* is found in the small intestine only. Within these locations, endogenous development of *E. bovis* occurs in the endothelial cells of the central lacteals in the small intestine whereas for *E. zuernii* and *E. auburnensis* endogenous development occurs in the connective tissue cells and mesodermal cells of the lamina propria, respectively (Hammond et al., 1946; Stockdale, 1976). The endogenous phase of *Eimeria alabamensis* is in the nucleus of the intestinal cells (Davis et al., 1957). The location of *Eimeria subspherica* and *Eimeria cylindrica* within the host is unknown (Austrialian Society for Parasitology Inc., 2015).

protozoa *Trypanosoma* spp. are both thought to cause immuno-suppression in cattle, which results in a higher susceptibility to other parasites (Chiejina et al., 2003; Hutchings et al., 1990; Wellenberg et al., 2002). For example, immuno-supression from *Trypanosoma* spp. results in enhanced pathogenicity of gastrointestinal worms such as *Haemonchus contortus* were as immuno-suppression from IBR leads to secondary bacterial pneumonia (Hutchings et al., 1990; Kaufmann et al., 1992).

A hosts susceptibility to coinfection may also be increased by chronic helminth infection skewing the immune response of the host towards a Th2 response (van Riet et al., 2007). It has been inferred that the association between bacterial infections, such as bovine tuberculosis and anthrax, and gastrointestinal worms in African buffalos and zebra occur as a result of Th1-Th2 trade-offs (Cizauskas et al., 2014; Jolles et al., 2008). Typically, T helper 1 (Th1) cells produce cytokines (e.g. interferon (IFN)-γ and Interleukin (IL)-12) which defend against intracellular parasites such as microparasites (viruses and protozoa) by stimulating cell-mediated inflammatory responses. In contrast, T helper 2 (Th2) cells defend the host against extracellular parasites (macroparasites such as helminths) by stimulating the production of antibodies, mucus secretions and the production of cytokines (e.g. IL-4, IL-5, IL-10 and IL-13, (Abbas et al., 1996; van Riet et al., 2007)). Cytokines from Th1 and Th2 cells are cross-regulatory, IFN-γ selectively inhibits proliferation of Th2 cells and IL-10 inhibits cytokine synthesis by Th1 cells (Abbas et al., 1996; Graham et al., 2007; Jolles et al., 2008; Mosmann and Sad, 1996). Consequently, when an individual is coinfected with intra and extracellular parasites, the host's ability to mount an effective immune response to both parasites is impaired. Thus, the impact of coinfection upon the host will be affected by the type of parasite-parasite association occurring (Cox, 2001).

1.2 Consequences of coinfection

Coinfection can increase the severity of disease. For example, humans coinfected with *Babesia microti* and *Borrelia burgdorferi* experience more symptoms than those infected with either parasite alone (Krause et al., 1996). Similarly, mice coinfected with the helminth *Litomosoides sigmodontis* and malaria *Plasmodium chabaudi* had more severe outcomes as indicated by increased anaemia and loss of body mass than mice with malaria or *L. sigmodontis* alone (Graham et al., 2005). The length and intensity of infection can also be increased when the parasite of interest is present in a coinfected host, compared to when it is present on its own (Fenton and Perkins, 2010). For instance, field voles (*Microtus agrestis*) coinfected with *Babesia microti* experienced shorter *Bartonella taylorii* infections however when the voles

were coinfected with cowpox they experienced longer *B. taylorii* infections (Telfer et al., 2008). In addition, a meta-analysis of helminth-malaria coinfections in mice has shown that coinfection resulted in an increased peak parasitaemia (indicating that the malaria parasites are replicating within the host) for helminth-malaria coinfections which involve the normally resolving malaria parasites, *Plasmodium chabaudi* and *Plasmodium yoelii* (Knowles, 2011).

In the previous section, I showed that the condition of the host was an important consideration as to whether or not it was susceptible to infection with a parasite; the example I provided looked specifically at anaemia and a vicious circle of coinfection. Anaemia is also one of the outcomes of coinfection. In the Infectious Diseases of East African Livestock (IDEAL) project (described in section 1.4), calves coinfected with *Trypanosomes* spp. and strongyles have a decreased packed cell volume (the percentage of red blood cells in the blood), whilst platelet counts were decreased by infection with the tick-borne pathogens such as *Anaplasma marginale*, *Babesia bigemina*, *Theileria mutans* and *Theileria parva* (Conradie van Wyk et al., 2014). Therefore two simultaneous infections may result in pathological effects greater than the additive effect of each pathogen in a single infection (Tompkins et al., 2011).

Other research on the IDEAL population has shown that the growth rates of calves are altered when they were coinfected with *T. parva*, *A. marginale* and *T. mutans* as previously discussed (Thumbi et al., 2013b). In addition, calves were at increased risk of death from East Coast Fever (which is caused by *T. parva* infection) if they were coinfected with *Trypanosoma* spp. or had a high worm burden (the number of worms the calf had, measured indirectly via faecal egg counts and so is equivalent to the egg burden (Thumbi et al., 2014)). In Appendix A, I provide a more detailed discussion of the outcomes of the IDEAL project. To summarise the current findings of the IDEAL project with regards to coinfection, the IDEAL project has show that coinfection can have both positive and negative consequences for the host. However my thesis concentrates on what causes a host to be infected with multiple parasites and to identify parasites which are associated with each other.

1.3 Implication of coinfection in population health campaigns

It is important to understand the associations between any given parasite species as they may influence the host's wellbeing and the epidemiology and transmissibility of coinfecting parasites (Pedersen and Fenton, 2007). The removal of one coinfecting parasite species may yield additional beneficial results if its elimination decreases

the severity or prevalence of a second parasite. For example, Brooker et al. (2012) identified a positive association within individuals for coinfection with the malaria parasite (*Plasmodium falciparum*) and hookworm (*Necator americanus*), implying that the introduction of hookworm control into an area could result in a decrease in the prevalence of *P. falciparum* infection, although this suggestion needs further experimental investigation. However treatment of individuals for one parasite may yield unexpected negative consequences for the population in the presence of coinfection. For instance, Ezenwa and Jolles (2015) study of free ranging African buffalos showed that although individuals treated with antihelminthics were 9 times more likely to survive bovine tuberculosis (bTB) infection than untreated individuals, the treatment also resulted in a 8-fold increase in the reproductive number of bTB and enhanced the spread of bTB (Ezenwa and Jolles, 2015). Thus, it is important to understand the consequences of coinfections.

1.4 Coinfection studies and the IDEAL Project

The majority of coinfection studies are observational and cross-sectional (Tompkins et al., 2011), making it difficult to establish the directionality of any associations or the timescale on which associations operate. Although some longitudinal and experimental studies of wild populations have been carried out (e.g. Knowles et al. (2013); Pedersen and Antonovics (2013); Telfer et al. (2010)), the number of animal recaptures in these studies are relatively low, making it difficult to track the time course of multiple infections in detail. Other studies have used laboratory experiments to study coinfection e.g.. Cox (2001); Graham (2008); Rodriguez et al. (1999), see also Hartgers and Yazdanbakhsh (2006) for a review of coinfection experiments. It is relatively easy to add or remove a parasite species during a laboratory experiment, therefore the nature of interspecific association between parasites can be characterised. However these studies lack the ability to investigate the impact of natural environmental conditions and the wider parasite community upon the coinfection and host-parasite interactions. Comprehensive data on a host's full parasite burden and over the time scale on which parasite-parasite associations operate are therefore rare.

The Infectious Diseases of East African Livestock (IDEAL) project provides a unique opportunity to study the role of coinfection in a cattle population. The IDEAL project is a longitudinal study of 548 indigenous calves in Western Kenya. Details on the study design and data collected are given in Bronsvoort et al. (2013) and Appendix A. Only free-grazing indigenous East African shorthorn zebu cattle were

eligible for inclusion in this study. These cattle are minimally managed and there is very limited use of vaccination or other preventative measures against infectious diseases. Thus the risk of exposure to many different parasites is high, so the full diversity of parasites can be seen and coinfection is very likely to occur. The IDEAL study site covers a wide range of agro-ecological zones running from the shores of Lake Victoria to the base of Mt. Elgon and calves are grouped into sublocations (defined by different administrative areas), hence the environmental effects on parasite occurrence can also be investigated (Figure 1.2).

Calves enrolled into the IDEAL project had their infectious disease burden (the range of parasites, pathogens and exposures that the calf is infected with, a calf is said to be infected in this thesis if it tested positive for the organism of interest) followed for the first year of life, during which time they would have acquired most of the parasites which occur in cattle in East Africa (Wymann et al., 2007). The timing of infections can be investigated, as calves are unexposed at birth and infection was followed every 5 weeks as they aged. Calves were also monitored for episodes of clinical disease and their sequelae investigated. More than 50 different parasites were identified during this time (Bronsvoort et al., 2013). Information has also been gathered on the herd demography, husbandry and management history. Genetic information, in the form of 50K Illumina® BovineSNP50 beadchip v1, is also available. Consequently, the IDEAL project has provided me with a wealth of information to which I can examine the role of coinfection in indigenous East African cattle.
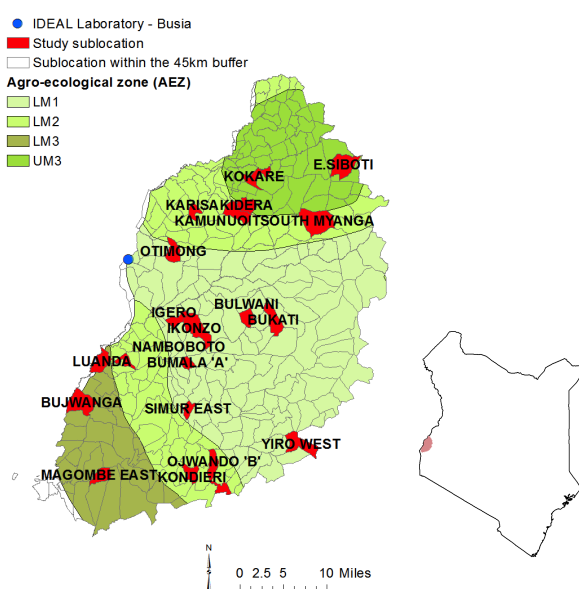
Figure 1.2 Map showing the IDEAL project study area in western Kenya, highlighting the selected sublocations in red, the five agro-ecological zones in green and study laboratory in Busia (the blue point). The small insert map shows the study area in relation to the whole of Kenya. Figure is taken from Bronsvoort et al. (2013).

1.5   Aims and objectives

The objectives of this thesis is to use data gathered by the Infectious Diseases of East African Livestock (IDEAL) project to investigate the evidence for coinfection and to generate a better understanding of the genetic and phenotypic factors affecting coinfection and their consequences within the study population. This thesis consists of the following chapters:

Chapter 2: Evidence for parasite-parasite associations in East African shorthorn zebu calves

> This chapter aims to investigate the nature of concurrent associations and of lagged effects between different pathogens, both over short and relatively longer term, and to examine the extent to which such associations are driven by other characteristics of either the animal or its environment.

Chapter 3: Seroprevalence of respiratory viral pathogens of indigenous calves in Western Kenya

> The aim of this chapter is to estimate the seroprevalence of antibodies against respiratory viruses and to investigate the associations occurring between them and their effect upon the host, by examining the relationship between serostatus and experiencing a clinical episode and its effect on growth rate. In addition this chapter investigates the environmental risk factors for being seropositive and the association between the respiratory viruses and other parasites identified during the IDEAL study.

Chapter 4: Variation and covariation in strongyle infection in East African shorthorn zebu calves

> This chapter aims to dissect the potential genetic and non-genetic sources of between- and within-individual level variation in strongyle faecal egg count. Its objectives are therefore to identify the causes of variation in strongyle worm burden within East African Shorthorn Zebu (EASZ), and in particular to investigate: 1) the components of variation in strongyle EPG; 2) the components of variation in traits which might be affected by strongyle infection, specifically weight and haematological parameters; and 3) the association between strongyle infection and these other traits.

Chapter 5: A genome-wide association study of East Coast Fever death and packed cell volume at the time of seroconversion to *Theileria parva*

The aim of this chapter is to firstly identify candidate SNPs associated with East Coast Fever (ECF) death and packed cell volume at the time of seroconversion to *Theileria parva* using GWAS and to identify genes and QTL associated with the candidate SNPs. Secondly, this chapter compares candidate SNPs to regions already known to be under selection in EASZ. Lastly, this chapter compares the minor allele frequencies of the SNPs associated with ECF death in shorthorn zebu to other cattle breeds with varying degrees of exposure and tolerance to ECF.

Chapter 6: Discussion

This chapter discusses the main findings of the thesis and relates the results to disease control strategies.

Table 1.1 Definitions of infectious disease terms used in my thesis. The definitions for these terms vary within the veterinary medicine and ecological literature therefore this table provides a definition for each term within each field as well as the definition I choose to use in this thesis

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Coinfection	The simultaneous infection of a host by more than one parasite strain or species (Méthot and Alizon, 2014).	A descriptive term for being infected with multiple parasites (Cox, 2001).	A descriptive term for being infected with multiple parasites. I extend this definition to include individuals that are infected with multiple parasites through time.
Commensal	Microbes found on the skin or within the body that do not usually cause disease (Thrusfield, 2005). Other definitions historically used in microbial pathology include 1) a harmless parasite and 2) organisms of the normal flora (Casadevall and Pirofski, 2000).	A partnership between species which coexist without detriment or obvious benefit (Hooper and Gordon, 2001).	Refers to bacteria, which are found on the skin or within the body of the calf that do not normally cause disease, therefore they are part of the normal flora.
Cross-immunity	A form of immunity where immunity to one parasite or strain is effective in protecting the host against another antigenically similar organism (Wearing and Rohani, 2006).	A form of immunity where one parasite or strain induces an immune response to another parasite or strain (Poletto et al., 2015).	A form of immunity where one parasite or strain induces an immune response to another parasite or strain.

15

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Disease	Historically defined according to the signs or symptoms associated with a specific cause. However there are no universally accepted criteria for establishing 'disease' (Temple et al., 2001; Tikkinen et al., 2012).	As in the veterinary medicine definition, there are no established criteria for defining disease apart from using the signs and symptoms associated with a specific cause (Temple et al., 2001). However, Campbell et al. (1979) suggests that disease is the sum of the abnormal phenomena displayed by a group of living organisms in association with a specified common characteristic or set of characteristics by which they differ from the norm for their species in such a way as to place them at a biological disadvantage.	Is defined according to the signs or symptoms associated with a specific cause.
Exposure	Is a risk factor for a particular disease, can include infectious agents and non-infectious determinants of disease such as the environment (White et al., 2008).	Something that the host has been or is subjected to (Hampson et al., 2011)	Something that the host is subjected to. In the case of the IDEAL project this can include infectious agents and non-infectious determinants such as the environment (Bronsvoort et al., 2013).

16

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Immunity	The phenomenon whereby exposure to a parasite induces changes in the host which protects it from disease upon subsequent challenge with the same or a related organism (Murphy, 1996).	The ecological definition is the same as the veterinary medicine definition whereby exposure to a parasite induces changes in the host which protects it from disease upon subsequent challenge with the same or related organism (Hawley and Altizer, 2011; Murphy, 1996).	A change in the host which protects it from disease upon subsequent challenge with the same or related organism.

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Infection	Invasion of the host with an organism that is not normally present in the body. Individuals can respond in a variety of ways to infection: inapparent infection is infection without any clinical signs, subclinical infection is infection without overt clinical signs and clinical infection produces clinical signs (Thrusfield, 2005). Other veterinary medicine definitions put more emphasis on the outcome of infection and proliferation within the host. For example, infection has been defined as 1) when parasites have passed through the normal barriers such as the skin or mucous membranes and have invaded and proliferated in the host and 2) the invasion of the host tissues by a microorganism resulting in disease (Casadevall and Pirofski, 2000).	The ecological definition is similar to the veterinary medicine definition whereby an organism which is not normally present in the host is identified as being within the host (Hawley and Altizer, 2011; Schmid-Hempel, 2011).	Invasion of the host with an organism that is not normally present in the body. If a calf tested positive for the parasite of interest in the IDEAL project then I classified that individual as being infected with the organism.

18

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Infection or infectious disease burden	Refers to a quantitative estimate of the impact of disease on an individual, population or geographical region, using a multitude of indicators such as Quality-Adjusted Life Years (QALY) and Disability-Adjusted Life Years (DALY) (Mangen et al., 2013).	Quantifies the number of individuals being affected by a disease (Brooker, 2010).	The range of parasites, pathogens and exposures that the calf is infected with. In the IDEAL project infection or infectious disease burden is used to mean the range of parasites, pathogens and exposures that the cohort is observed to be affected by (Bronsvoort et al., 2013).
Infection intensity	Used to quantify the infection levels of a specific organism within an individual or population (Suchdev et al., 2014).	The number of individuals (determined directly or indirectly) of a particular parasite species in each infected host (Margolis et al., 1982). Infection intensity is frequently expressed as a numerical range, for example, a high infection intensity of *Theileria* spp. in the IDEAL project is defined as ≥ 2 infected cells per microscope field, in multiple fields (Thumbi et al., 2013a).	The number of individuals (determined directly or indirectly) of a particular parasite species in each infected host.

19

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Parasite	An organism that lives on or in another (the host) and takes its nourishment from or at the expense of its host. Some parasites can cause disease but not all do (CDC website, 2015).	Any organism that lives in or in close association with a host organism, showing some a degree of adaptation that carries a cost to the host. In the ecological literature parasites may include both micro-parasites and macro-parasites as well as behavioural parasites such as birds which exploit their hosts by stealing food (Méthot and Alizon, 2014; Poulin, 2011).	An organism that lives on or in another (the host) and takes its nourishment from or at the expense of its host. I include micro-parasites (e.g. viruses, protozoa and bacteria) and macroparasites (e.g. worms) in this definition.
Pathogen	An agent (e.g. bacteria, protozoa, virus etc.), which causes or can cause disease or damage to the host (Pirofski and Casadevall, 2012). This definition is currently being debated in the post-genomic era as some pathogens do not cause disease in all hosts. Therefore the word pathogen has been redefined by one manuscript to mean an organism that causes virulence to the host upon infection (Méthot and Alizon, 2014).	A parasite which is capable of causing damage to the host; this definition can include classical pathogens as well as opportunistic pathogens; damage to the host can be a result of direct action or via the host immune response (Casadevall and Pirofski, 1999). In the ecological context an infection can be defined as a pathogen even if it is not pathogenic (Méthot and Alizon, 2014).	An infectious agent (e.g. bacteria, protozoa, virus etc.), which causes or has the potential to cause disease or damage to the host. In accordance with the ecological definition, the pathogen can be non-pathogenic.

20

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Pathogenicity	There are many different variations on the definition of pathogenicity. These include 1) the ability of an infectious agent to cause disease Thrusfield2013; 2) the capacity of an organism to cause damage in a host (Casadevall and Pirofski, 1999); 3) the amount of damage caused by the organism to the host (Shaner et al., 1992) or 4) the physiological process underlying the development of virulence (Schmid-Hempel, 2011).	In plant and insect pathology, pathogenicity has been defined in terms of the ability of an organism to infect a host and cause disease (Shapiro-Ilan et al., 2005; Thomas and Elkinton, 2004).	The ability of an infectious agent to cause disease.
Resistance	Is a function of the immune system, which works to protect the infected host by reducing its parasite burden through the detection, neutralization, destruction or expulsion of the parasite (Medzhitov et al., 2012).	The ability to limit parasite burden, therefore resistance means that the host is protected at the expense of the parasite (Råberg et al., 2009). In plant ecology resistance is defined as the inverse of parasite burden; as resistance increases parasite burden decreases (Schneider and Ayres, 2008).	The ability of the host to limit its parasite burden.

21

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Secondary Infection	Opportunistic infection with a second parasite due to infection with a previous parasite causing damage to the host or impairing the clearance of the second parasite (Coetzer and Tustin, 2004). In epidemiology, secondary infection can also be used to mean the number of infections arising from a single infected individual in a completely susceptible population (Antia et al., 2003).	Often applied to microbial communities. If the microbial community is viewed as a network of interacting species then a perturbation in the community may open a niche in which opportunistic infection with a secondary parasite can occur. For example, the application of antibiotics can remove mutualistic bacteria from the gut and so secondary infection with other bacteria such as *C. difficile* can occur (Lemon et al., 2012).	Opportunistic infection with a second parasite due to infection with a previous parasite causing damage to the host or impairing the clearance of the second parasite.
Seroconversion	When antibodies to the antigen of interest become detectable in the blood, therefore an individual moves from a seronegative to a seropositive state (Toye et al., 2013b).	Is an immunological term, therefore it has the same definition in the ecological literature as in the veterinary medicine literature.	The movement from a seronegative to a seropositive state which occurs when antibodies to the antigen of interest become detectable in the blood.

22

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Seropositive / Seronegative	Antibodies to the antigen of interest are / are not detectable in the blood (Toye et al., 2013b).	Is an immunological term, therefore it has the same definition in the ecological literature as in the veterinary medicine literature.	Antibodies to the antigen of interest are / are not detectable in the blood. In the IDEAL project ELISA tests are used to detect antibodies to the parasite of interest, individuals are said to be seropositive for the parasite if their test result are above the manufacturer's cut-off and seronegative if their test result are below the manufacturer's cut-off.
Susceptibility	The ability of an individual to become infected. A susceptible individual is an individual who is at risk of being infected with a parasite (Thrusfield, 2005).	The state of being at risk for an event occurring. Therefore, in the case of infection, a susceptible individual is an individual who is at risk of being infected with a parasite (Barlow, 1996).	The ability of the host to become infected. A susceptible host is at risk of being infected with the parasite of interest.

23

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Tolerance	The ability to limit harm; therefore tolerance saves the host from harm without having any direct negative effect on the parasite (Råberg et al., 2009). However tolerance has many other definitions in veterinary and medical communities, these are discussed in detail by Råberg et al. (2009). For example, indigenous cattle breeds are tolerant for infection with *Theileria parva* whereas European breeds are non-tolerant, therefore tolerance can refer to the overall effect of disease severity irrespective of parasite burden (Kariuki et al., 1995). Whereas in immunology, tolerance is taken to mean not mounting an immune response to an antigen (Råberg et al., 2009).	The ability to limit harm from a parasite. In plant ecology, tolerant plants have a smaller decrease in their overall health as parasite burden increases compared to less tolerant plants; therefore tolerance is described according to the slope of a reaction norm when fitness is plotted against parasite burden (Schneider and Ayres, 2008).	The ability of the host to limit harm from a parasite.

24

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Transmission	The spread of an infectious organism through contact between susceptible individuals. The word transmitted can be used to describe the mechanism by which an infectious organism is passed between individuals (Thrusfield, 2005).	The ecological definition is the same as the medical definition. The spread of an infectious organism through contact between susceptible individuals (Levin et al., 2009).	The spread of an infectious organism through contact between susceptible individuals. The word transmitted can be used to describe the mechanism by which an infectious organism is passed between individuals.
Virulence	Like pathogenicity, there are many variations on the definition of virulence. It has been described as 1) the degree of pathogenicity; 2) the relative capacity of a microbe to cause damage to the host; 3) the relative power and degree of pathogenicity possessed by an organism to produce disease and 4) the disease producing capacity of any infectious agent (Casadevall and Pirofski, 1999; Shaner et al., 1992).	In evolutionary ecology, virulence is a quantitative trait that measures the decrease in fitness due to infection (Méthot and Alizon, 2014).	The relative capacity of a parasite to cause damage in a host.

Table 1.1 – continued from previous page

Term	Veterinary Medicine Definition	Ecological Definition	Definition used in my thesis
Worm burden	The infection intensity of worms and so worm burden quantifies the number of worms present in the host. Worm burden is normally measured indirectly by counts of eggs expelled in faeces (eggs per gram of faeces, (Anderson et al., 2013)).	The ecological definition is the same as the veterinary medicine definition. Worm burden quantifies the number of worms within the host and is normally measured indirectly via the number of eggs per gram of faeces (Stear et al., 1995).	The number of worms present within the host. In this thesis worm burden is measured indirectly via faecal egg counts and so is equivalent to the egg burden.

Chapter 2

Evidence for parasite-parasite associations in East African shorthorn zebu calves

Summary

Associations between parasites may be important determinants of infection dynamics; however comprehensive data on parasite burden are rare. Here, I investigate if infection with one parasite species is associated with the subsequent infection of another parasite species in a longitudinal study of infection burdens in East African Shorthorn Zebu calves in Kenya, called the Infectious Diseases of East African Livestock (IDEAL) project. Calves were screened every 5 weeks, from birth until 51 weeks of age, for a wide range of parasites. Generalised linear mixed models, with and without time lags were used to test for associations between 5 groups of parasites: *Theileria* spp., *Coccidia* spp., *Strongyloides* spp., *Calicophoron* spp. and strongyles. I found evidence of substantial age-related variation in parasite prevalence. Evidence exists for both positive and negative associations between parasites. For example, infection with *Strongyloides* spp. increased the risk of strongyle infection. Conversely, being infected with *Strongyloides* spp. decreased the calf's risk of infection with *Calicophoron* spp. The majority of associations between parasites persisted after accounting for variation in seasonal, environmental and calf level risk factors.

2.1   Introduction

Most hosts are commonly infected with multiple parasite species during their lifetime and many are often infected with more than one parasite species at the same time

(Cox, 2001). Associations amongst parasites can occur directly, or indirectly via a variety of mechanisms such as competition between parasites for similar resources or for location within the host (Pedersen and Fenton, 2007). Associations may also occur via the immune system, through immuno-suppression, cross-immunity and Th1-Th2 trade-offs, where the immune response to a previous infection of one parasite alters the immune environment prior to infection with a successive parasite (Graham et al., 2007; Lello et al., 2004; Pedersen and Fenton, 2007).

Many studies have investigated coinfection, although the majority of these are observational and cross-sectional (Tompkins et al., 2011), making it difficult to establish the directionality of any associations (Fenton et al., 2014). In particular, the relative timing of infection by different parasites can have important implications for understanding the dynamics of multi-parasite infections: for instance, a positive association between the occurrence of parasite A at a prior time point and parasite B at the current time point would suggest that prior infection by A is a risk factor, whereas a negative association would imply that prior infection by A is a protective factor. For example, in wild field voles, individuals newly infected with *Babesia microti* are approximately 5 times more likely to become subsequently infected with *Anaplasma phagocytophilum* than an uninfected individual, whilst field voles with *B. microti* infections have reduced susceptibility to infection with *Bartonella* spp. (Telfer et al., 2010).

Although some experimental and longitudinal studies of wild populations have been carried out (e.g. Ezenwa and Jolles (2015); Knowles et al. (2013); Pedersen and Antonovics (2013); Telfer et al. (2010)), the number of recaptures per individual animal in these studies are relatively low, making it difficult to track the time course of multiple infections in detail. Other studies have used laboratory experiments to study coinfection (e.g. Cox (2001); Graham (2008); Hartgers and Yazdanbakhsh (2006); Rodriguez et al. (1999) and many more). It is relatively easy to add or remove a parasite species during a laboratory experiment, therefore the nature of interspecific association between parasites can easily be characterised. However laboratory studies lack the ability to investigate the impact of environmental conditions and the wider parasite community upon coinfection and host-parasite interactions. Consequently, comprehensive field data on a host's full parasite burden and the time scale over which parasite-parasite associations operate is rare.

The condition of the host or the environment in which it resides can also influence the occurrence of coinfection. For example, Beldomenico et al. (2008) showed that rodents in poor condition (i.e. individuals which were anaemic) had elevated haematological indicators of infection (neutrophilia and monocytosis) which were

generally followed by a further decrease in condition, which in turn predisposed the individual to infection with other parasites such as cowpox, thus a vicious cycle was generated (Beldomenico and Begon, 2010; Beldomenico et al., 2008, 2009). Similarly, environmental factors such as temperature, elevation, distance to large water bodies and the amount of vegetation cover present, have been associated with increased risk of helminth coinfections in humans (Brooker and Clements, 2009; Saathoff et al., 2005). Therefore it is important to account for the role of environmental effects in generating covariances, and thus to avoid making false inferences about associations between coinfecting parasites (Hellard et al., 2012).

The Infectious Diseases of East African Livestock (IDEAL) project (Bronsvoort et al., 2013) provides a unique opportunity to evaluate parasite-parasite associations. This longitudinal study tracked the infectious disease burden through the first year of life (the period of highest natural mortality (Phiri et al., 2010)) of East African shorthorn zebu calves in Western Kenya, including infection status through the year with regard to a total of more than 50 different parasites. Calves were recruited within a week of birth and were followed routinely every 5 weeks until they were 51 weeks old, therefore the timing and sequelae of multiple infections could be investigated. Furthermore, cattle in the study region are minimally managed and there is very limited use of vaccination or other preventative measures against infectious diseases. Thus the risk of exposure to many different parasites is high, meaning that coinfection is very likely to occur. Additionally, the study area covers a range of agro-ecological zones from Mount Elgon in the north to Lake Victoria in the south (Bronsvoort et al., 2013), allowing investigation into the effect of environmental variation upon coinfection. See Bronsvoort et al. (2013) and Appendix A for more detail on the IDEAL project.

Within the IDEAL project it has been shown that coinfecting parasites can have both positive and negative effects upon the host. For example, Thumbi et al. (2013b) found that calves had lower growth rates when they were coinfected with *Theileria parva* and *Anaplasma marginale* and they had a higher growth rate when they were coinfected with *T. parva* and *T. mutans*, compared to when the calf only experienced a single infection of *T. parva*. In this example, the effect of the two parasites was greater than the sum of the effects for each individual parasite (Thumbi et al., 2013b). However, the association occurring between parasites was not investigated.

This study aims to use the uniquely-detailed data of the IDEAL project to track the fine-scale dynamics of coinfection by multiple parasite species in East African shorthorn zebu calves, and in particular to characterise associations between parasites, concurrently or sequential, over time scales of 0-10 weeks, and of the extent to which such associations are driven by other characteristics of either the calf or its

environment.

2.2 Material and Methods

2.2.1 Study population

The Infectious Diseases of East African Livestock (IDEAL) project recruited 548 indigenous East African Shorthorn Zebu (*Bos indicus*) calves in Western Kenya when they were 3-7 days old and followed them for the first year of life (Bronsvoort et al., 2013). Calves were selected using a stratified two-stage random cluster study design. In the first stage, 20 'sublocations' (the smallest administrative unit in Kenya) were selected from 5 agro-ecological zones, within a radius of 45km from the town of Busia. During the second stage, 28 calves were recruited at less than one week old from each sublocation, at staggered points across the entire study period. Data collection ran from October 2007 to September 2010 (see Bronsvoort et al. (2013) and Appendix A for a detailed description of the study design, including a map of the study area). At the recruitment visit, a questionnaire was completed by the calf's owner. This questionnaire collected information about the farm and farmer, other livestock, water sources and animal husbandry practices. Calf locations were geo-referenced using hand-held GPS devices (Garmin 12, Garmin Kansas, USA). Calves received routine visits every 5 weeks until they were 51 weeks old. During these visits, individuals were given a clinical examination in which biological samples were taken for laboratory analysis, hematological profiling and parasite identification. This study presents the parasite-parasite associations from the 455 calves which survived until the end of the study period at 51 weeks of age; however, repeating the analysis using data from all calves enrolled into the study period produced very similar results.

2.2.2 Data collection

A full description of the methods used and a list of parasites identified during the IDEAL project is provided by Bronsvoort et al. (2013) and Appendix A, so here I provide only a brief summary of the methods used to identify the parasites on which this study concentrates (detailed below). All tests described here were carried out at the IDEAL project field laboratory in Busia, Kenya. At each 5 weekly visit, thin and thick blood smears were taken from marginal ear veins and screened for haemoparasites by microscopic examination of at least 100 fields. The species of parasite was identified and recorded. Two faecal samples were also collected from the rectum. The McMasters counting technique (Hansen and Perry, 1994) was used to

identify the presence of *Coccidia* oocytes and nematode and cestode eggs. As different nematode species have morphologically similar eggs, which are hard to differentiate using microscopy, the second faecal sample collected was used for faecal culture, as described by Hansen and Perry (1994). During faecal culture, nematode eggs were permitted to hatch and develop into the infective larval stage 3 (L3), which was then used to speciate any infecting nematodes.

Indicators of calf condition were gathered at each five weekly visit. Blood samples were collected and blood cell analysis was automatically performed using the pocH-100iV Diff (Sysmex® Europe GMBG); see Conradie van Wyk et al. (2012) for more details. Profiles were produced for the total white cell count, red blood cell count and platelet count. Packed cell volume (PCV) was measured manually with a Hawksley micro-haematocrit reader. Total serum protein (TSP) was determined using a refractometer and absolute eosinophil count (EO) was quantified by differential counts from thin EDTA blood smears stained with Diff Quick. Calves were weighed (in kilogrammes) at recruitment, then again every five weeks until 31 weeks of age, and once again at a last visit at 51 weeks. Furthermore, all calves were genotyped using a 50K Illumina® BovineSNP50 beadchip v.1 and these genotypes were then used to determine their level of introgression from European taurine cattle and heterozygosity (Mbole-Kariuki et al., 2014; Murray et al., 2013). Additional information on the condition of the calf's dam was gathered at the recruitment visit in the form of dam girth (cm) and body condition score. Body condition score is a method of assessing the body condition of the dam using a standard 10 point score (Bronsvoort et al., 2013).

Environmental data at the farm level and environmental factors were gathered at the recruitment visit using a questionnaire. Additional information on the amount of vegetation cover present at each location, as measured by the normalized difference vegetation index (NDVI) within a 250m by 250m pixel size of the satellite imagery surrounding each calf's homestead, was also obtained for each visit (MODIS satellite image processed by the Geographical Information System (GIS) group at the International Livestock Research Institute).

2.2.3 Statistical analysis

Parasite selection and non-metric multidimensional scaling

As over 50 parasites were identified during the IDEAL project (Bronsvoort et al., 2013), only those parasites tested at every 5 weekly visit with a prevalence of greater than 10% across the whole study period were included in this analysis. This avoided spurious results from those parasites with a very low prevalence and reduced the

number of comparisons made. Seven parasites met my inclusion criteria: these were the helminths *Strongyloides* spp. (treated as a single group), *Oesophagostomum radiatum*, *Calicophoron* spp. (treated as a single group), *Trichostrongylus axei* and *Haemonchus placei* and the protozoans *Coccidia* spp., and *Theileria* spp. (both treated as single groups as identified by microscopy). A descriptive summary of each of these seven parasites or parasite groups is provided in Table C.1.

To increase the statistical power of the study, groups of parasites were identified using non-metric multidimensional scaling (NMS) in PC-ORD software version 6.03 (MjM Software Design, Gleneden Beach, OR, USA). The matrix consisted of the response variables, which were the presence or absence of each parasite at every 5-weekly visit. I ran NMS with a Euclidian distance measure after relativizing by the standard deviates of each column (the presence/absence of each parasite at each visit). The dimensionality of the dataset was determined by plotting an inverse measure of fit ('stress') to the number of dimensions; stress is a measure of the departure from monotonicity in the plot of original space *vs.* that of the ordination space (McCune and Mefford, 2011). Optimal dimensionality was based on the number of dimensions with the lowest stress. 500 iterations of NMS were run with random starting points to ensure that the solution was stable.

The optimal dimensionality was a 3-dimensional solution (stress=12.57) and explained 84% of the variance in the dataset (Axis 1 r^2=37.5, Axis 2 r^2=17.1, Axis 3 r^2=29.3), more than expected by chance (Monte Carlo test P value= 0.004). Joint plots were generated and variables used in the NMS analysis were shown as vectors, the direction indicating a positive or negative correlation along a given axis (Figure 2.1). Based on this graph, five distinct groups were identified: *O. radiatum*, *T. axei* and *H. placei* formed one group (strongyles) whereas *Calicophoron* spp., *Strongyloides* spp., *Theileria* spp. and *Coccidia* spp. each formed separate groups, Figure 2.1. It is biologically plausible that *O. radiatum*, *T. axei* and *H. placei* cluster together as they are all strongyle type worms and have morphologically similar eggs (Urquhart et al., 1996).

Therefore, in the following analyses, I considered infection by each of these five groups of the different families (with the strongyles group defined as being infected with one or more of the three species). Within the strongyles group, positive associations were identified between all three species (detected using the multivariable model structure described below).
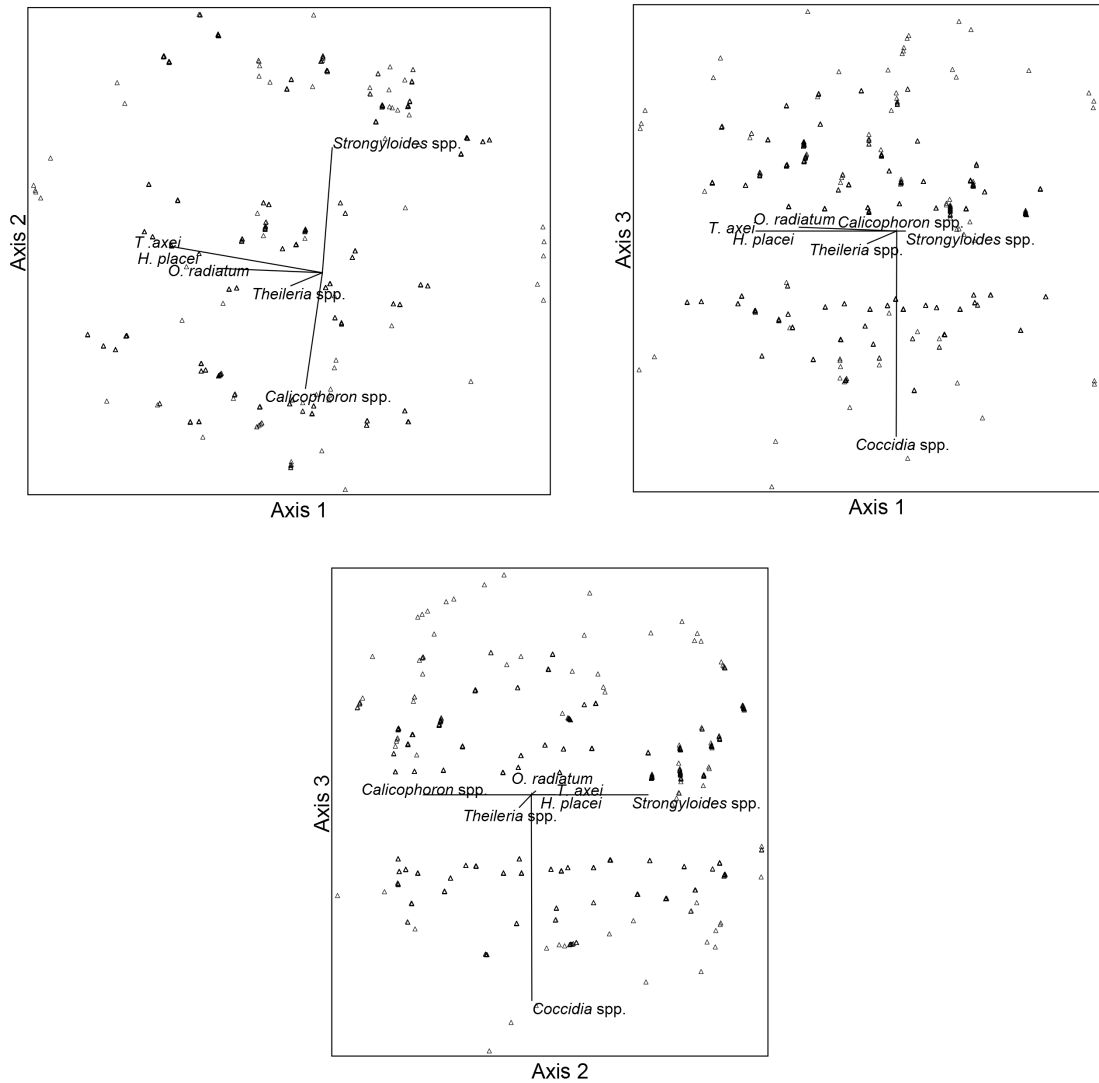
Figure 2.1 Joint graph illustrating the groups of parasites from the NMS scaling. The optimal 3-dimensional solution had a stress of 12.57 and explained 84% of the variance in the dataset. Five distinct groups were identified: the strongyles species; *Oesophagostomum radiatum*, *Trichostrongylus axei* and *Haemonchus placei* formed one group (strongyles) whereas *Calicophoron* spp., *Strongyloides* spp., *Theileria* spp. and *Coccidia* spp. each formed separate groups.

1. Analysis of temporal associations between multiple parasites

I investigated temporal associations between parasites by creating lagged time variables (T_{-5} and T_{-10}) which describe the presence or absence of the parasite at previous visits, either 5 weeks prior (T_{-5}) or 10 weeks (T_{-10}) prior to the current visit (Table C.2).

Generalised linear mixed models (GLMMs) were used to investigate whether infection of a susceptible calf with parasite A at the present time point (T_0) was associated with that calf being infected with parasite B at the present time point (T_0) or at a previous time point (T_{-5} or T_{-10}) or other parasites (parasite C etc.) at the different time points. Infection at T_0 in calf i at time j was modelled as a binary (presence/absence) response variable with a logit link function with binomial errors, the adaptive Gauss-Hermite Quadrature approximation to the maximum likelihood estimation and optimised using the 'Bound Optimization By Quadratic Approximation' algorithm in R version 3.1.2 using the *lme4* package (Bates et al., 2014). The models had the following structure:

$$logit(ParasiteA.T_{0_{ij}}) =$$
$$\alpha + \beta_1 Age_{ij} + \beta_2 ParasiteA.T_{-5_{ij}} + beta_3 ParasiteA.T_{-10_{ij}} +$$
$$\beta_4 ParasiteB.T_{0_{ij}} + \beta_5 ParasiteB.T_{-5_{ij}} + \beta_6 ParasiteB.T_{-10_{ij}} +$$
$$\beta_7 ParasiteC.T_{0_{ij}} + \beta_8 ParasiteC.T_{-5_{ij}} + \beta_9 ParasiteC.T_{-10_{ij}} +$$
$$+ \ldots + b_{1_i} Calf_i + b_{2_i} Sublocation_i + \varepsilon_{ij} \quad (2.1)$$

where α is a intercept, *ParasiteY.T_0* is the presence/absence of parasite Y at the current time point and *ParasiteY.T_{-X}* is the presence/absence of parasite Y at time point X weeks earlier and ε is the residual variation. This model structure allows the testing of the effects of parasite B at time T_{-X} on infection with parasite A at time T_0 whilst ensuring that the relationship is not confounded by the effects of infection with parasite A at T_{-X}. No evidence for temporal autocorrelation was observed, beyond that accounted for using this model structure, as tested by the autocorrelation function in R (Box et al., 1994; Pinheiro and Bates, 2000).

Calf age (*Age*) was included as a categorical, multi-level, fixed effect to account for non-linear changes in exposure risk as the calf matures: for example, immature calves aged 6-12 months the highest strongyle egg counts, followed by young calves (<6 months) and adults (>12 months) (Waruiru et al., 2000). Calf identity (*Calf*) and sublocation (*Sublocation*, 20 levels) were included in the model as random effects (b_{1i}, b_{2i}) to account for the repeated measures upon individuals and the clustering

introduced by the study design. A separate model was built for each of the five parasite groups.

I constrained the analysis to time lags T_0, T_{-5} and T_{-10} (current visit, 5 weeks and 10 weeks prior to the current visit, respectively) because for each additional time lag after this, the number of missing values increased, resulting in a decrease in the total number of observations per visit with full information and giving rise to a much reduced dataset for analysis: for example, by including a time lag of 15 weeks, the dataset was reduced to 50% of its original size. Results from preliminary pairwise models also indicated that lag effects over longer than this time were exceptionally rare.

Backwards stepwise selection was used to sequentially remove variables from the maximum model, starting with variables which had the least significant P value. Given the large number of tests involved, I chose the most parsimonious model (stage 1 model) to be the model in which all explanatory variables were associated with the response variables with a P value less than 0.01. By using a significance threshold of P<0.01 I observe more significant associations then you would expect by chance, thus I reduce the risk of making a type 1 error given the large number of comparisons made.

2. The role of calf or seasonal or environmental variation in determining associations

Associations between the occurrence of two different parasites may be driven by their joint correlation with some other characteristic, either of the calf or of the environment: for example, if calves of a given size or in a given location are more likely to be infected by multiple parasites, this will generate positive associations. To explore this possibility, each potential environmental covariate (see Table C.3 for a full list) was added to the most parsimonious multiple parasite models. Furthermore, seasonal or temporal variation in parasite exposure may drive the parasite-parasite associations, so month and year of sampling (2007-2010) were also included as factors in the environmental analysis. The fraction of calves seropositive for the parasite during the dry and wet season (March and May) is shown in Figure C.1. Environmental variables associated with the presence of the parasite of interest with a P value <0.2 were added to the parasite-time lag model from stage 1. Backwards stepwise selection was used to sequentially remove variable to achieve the most parsimonious model. I then checked whether the inclusion of the significant calf, environmental or seasonal variables changed conclusions about the associations between parasites, with the expectation that if associations between parasite A and B were driven by their joint correlation with a different variable, inclusion of the latter would reduce the significance of the association between A and B. For ease of comparison, in the

Appendix C, I present three different types of models showing the individual effects of indicators of calf condition, environmental factors and seasonal variables upon the parasite-parasite associations.

2.3 Results

2.3.1 Age-related changes in prevalence

455 individual calves were included in this study giving a total of 4847 observations (with 158 missing observations; see Bronsvoort et al. (2013) and Appendix A for details). In addition, it was not possible to obtain faecal samples at all visits, and so 722 out of the 4847 visits had missing faecal samples, these were included in the analysis as not applicable values.

All calves were infected with *Theileria* spp. at some point during the study, and all but one calf was infected with *H. placei* (Table 2.1). Out of the parasites included in this study, strongyles were the most prevalent parasite group with a prevalence of 64.4% across all tested visits, whereas *Strongyloides* spp. were the least common with a prevalence of 11.6% (Table 2.1). The fraction of calves positive for each parasite at each age and the cumulative fraction of ever having been infected with each parasite are presented in Figure 2.2. There were marked differences in the patterns of prevalence with age. *Strongyloides* spp. appeared early on in a calf's life and reached peak prevalence at 6 weeks old, after which the prevalence decreased. In contrast, the prevalence of strongyles showed a sharp increase at 11 weeks old and plateaued out to around 75% of calves infected from age 21 weeks. Similarly, *Theileria* spp. also infected a large proportion of calves across all ages, with all calves being infected with *Theileria* spp. by 51 weeks of age. *Calicophoron* spp. and *Coccidia* spp. had intermediate prevalences, but the fraction of calves infected also increased with age. The median number of parasite groups (from the NMS scaling) with which a calf was infected with during any one visit was 2 (range 0-5), and this also increased with age, Figure 2.3.

Table 2.1 Prevalence of each parasite. Prevalence is followed by the 95% confidence intervals in brackets. The total number of calves is 455.

Parasite	Number of calves which were ever infected by the parasite	Percentage of calves ever infected by the parasite (95% Confidence Interval)	Number of tested visits	Number of visits where the parasite was present	Prevalence at Tested Visits (95% Confidence Interval)
Strongyloides spp.	263	57.8 (53.1-62.4)	4125	477	11. 6 (10.6-12.6)
Oesophagostomum radiatum	310	68.1 (63.6-72.4)	4125	604	14.6 (13.6-15.8)
Calicophoron spp.	336	73.8 (69.5-77.8)	4125	751	18.2 (17.0-19.4)
Trichostrongylus axei	420	92.3 (89.5-94.6)	4125	1334	32.3 (30.9-33.8)
Coccidia spp.	434	95.4 (93.0-97.1)	4125	1399	33.9 (32.5-35.4)
Haemonchus placei	453	99.6 (98.4-99.9)	4125	2555	61.9 (60.4-63.4)
Strongyles	454	99.8 (98.8-100.0)	4125	2656	64.4 (62.9-65.9)
Theileria spp.	455	100.0 (99.2-100.0)	4820	2861	59.4 (58.0-60.8)
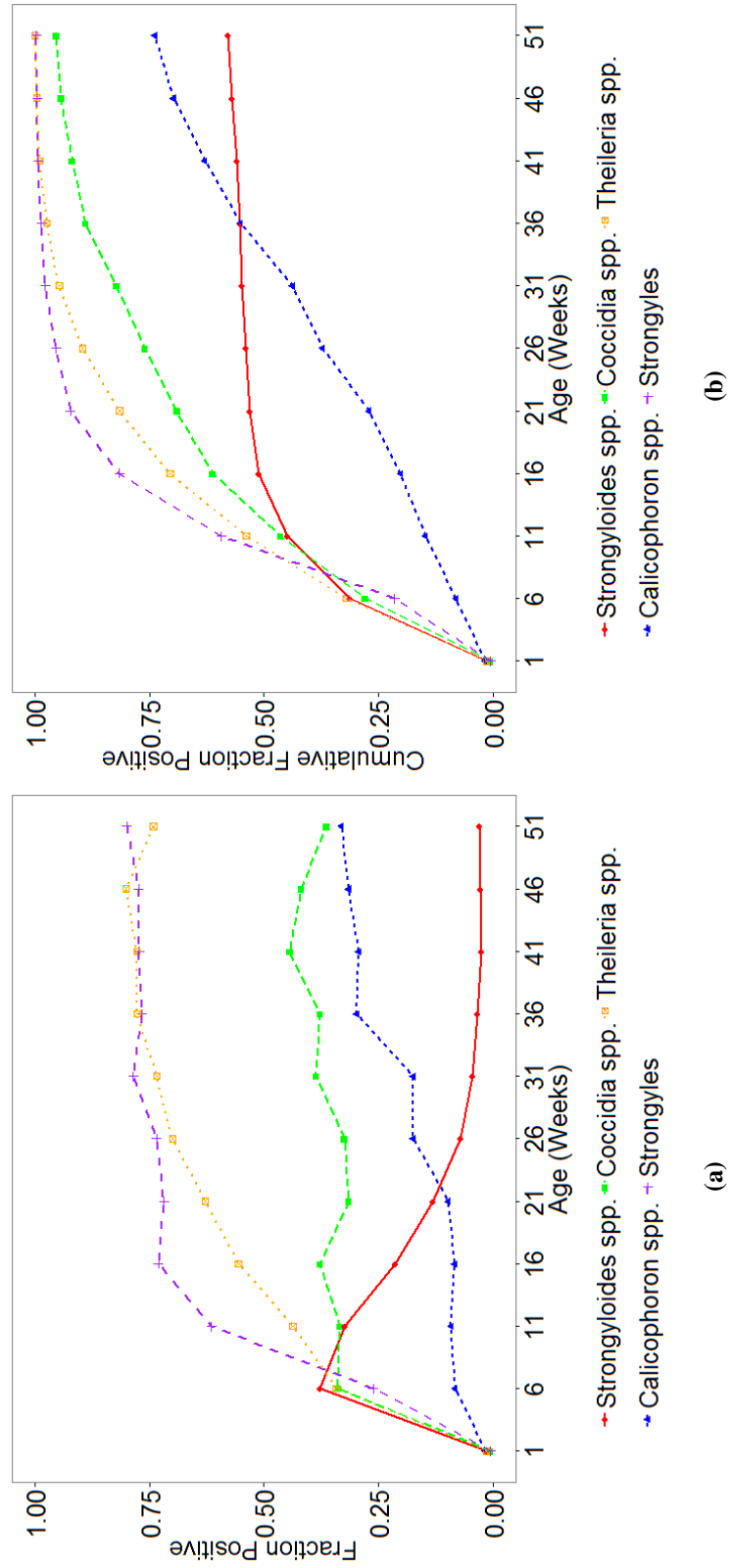
Figure 2.2 a) Fraction of calves infected with the parasite group at each age and b) cumulative fraction of calves ever infected by the parasite group up until the age stated.

38
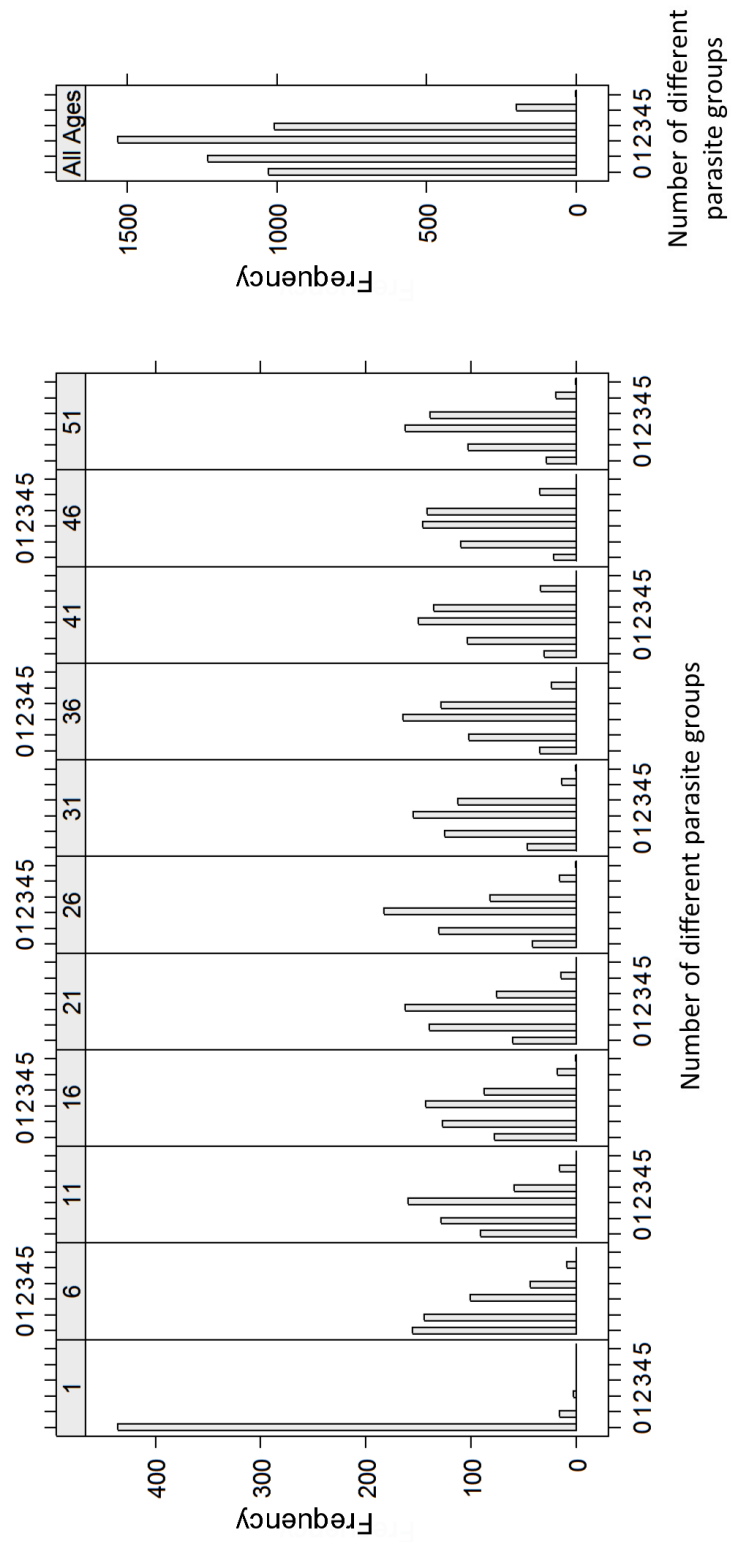
Figure 2.3 Histogram showing the number of different parasite groups from the NMS that a calf had at a) each age and b) across the whole of the study period.
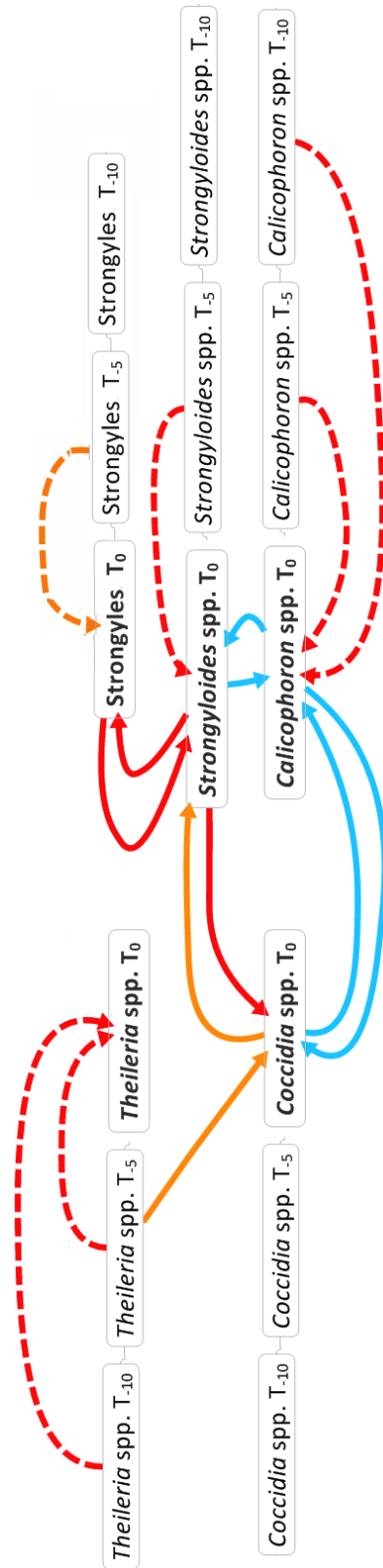
Figure 2.4 A schematic diagram showing the significant associations between parasites identified from the analysis of parasite presence at the current visit (T_0) or at a 5 and 10 week lag (T_{-5} and T_{-10}, respectively) before accounting for indicators of calf condition, environmental and seasonal variation. Positive associations (risk factors) are shown in orange/red and are significant a $P < 0.01$ and $P < 0.001$, respectively. Negative associations (protective factors) are shown in blue and are significant at $P < 0.01$. Interspecific associations are shown by solid lines; intraspecific associations are shown by dashed lines.

40
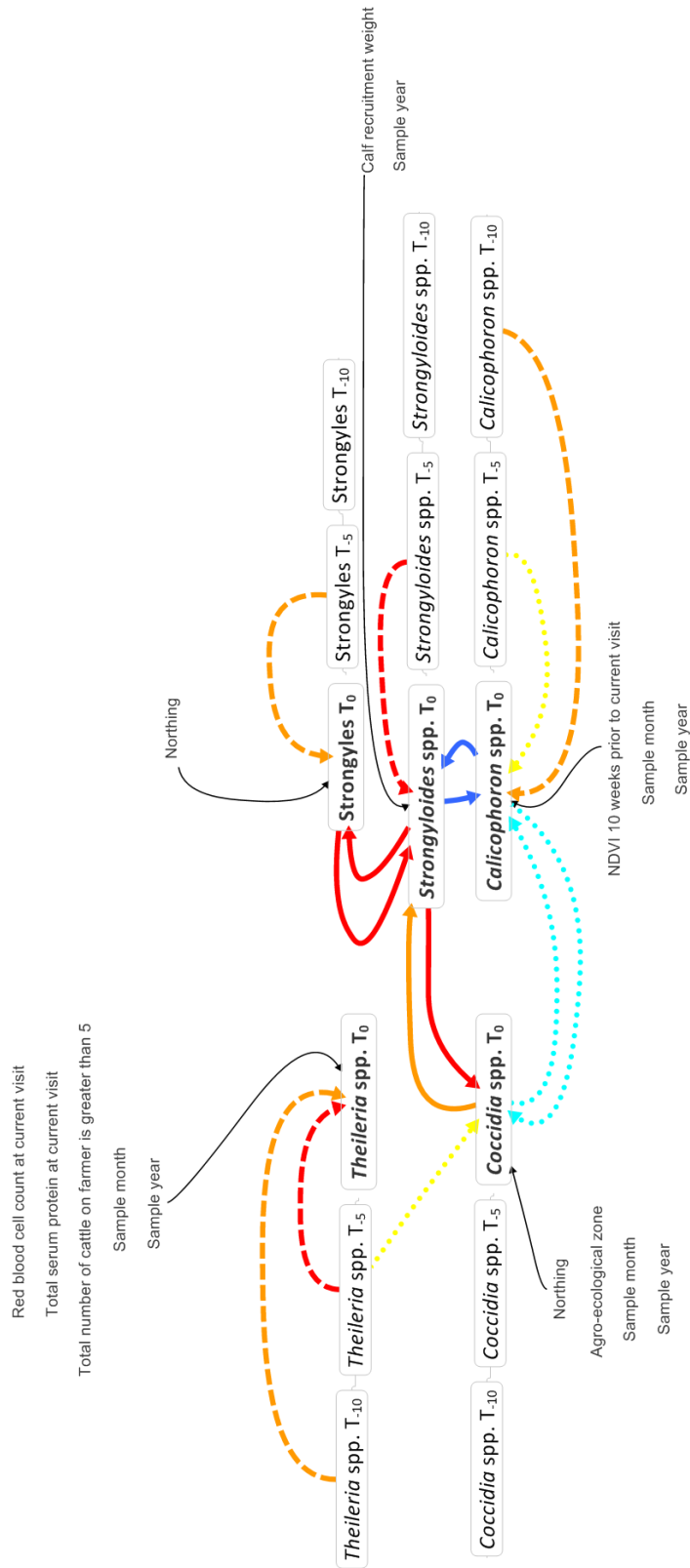
Figure 2.5 A schematic diagram showing the significant associations between parasites identified from the analysis of parasite presence at the current visit (T_0) or at a 5 and 10 week lag (T_{-5} and T_{-10}, respectively) after accounting for indicators of calf condition, environmental and seasonal variation. Positive associations (risk factors) are shown in orange/red and are significant a $P < 0.01$ and $P < 0.001$, respectively. Negative associations (protective factors) are shown in blue and are significant at $P < 0.01$ and $P < 0.001$, respectively. Interspecific associations are shown by solid lines; intraspecific associations are shown by dashed lines. Following the addition of indicators of calf condition, environmental and seasonal variation associations between parasites with P>0.01 are shown by dotted lines, with yellow dotted lines indicating positive associations and light blue dotted lines showing negative associations. Solid black arrows indicate other significant environmental or temporal variables.

41

Table 2.2 Associations between multiple time lags and parasites identified from the analyses of the parasites present at the current visit (T_0) or at a 5 and 10 week lag (T_{-5} and T_{-10}, respectively) in (a) a model with only parasites and (b) after adding indicators of calf condition, environmental and seasonal variation to the parasite-only model in (a). *Summary P value from the ANOVA F ratio of the GLMM.

Response Parasite (T_0)	Explanatory Variable	(a) Parasite Only Model						(b) Calf, Seasonal and Environmental Variation Model					
		OR	95% CI	P value	AIC	N_O	N_C	OR	95% CI	P value	AIC	N_O	N_C
Strongyloides spp.	Strongyles T_0	2.56	1.85 - 3.56	<0.001	1687	3114	454	2.55	1.83 - 3.54	<0.001	1626	3105	453
	Calicophoron spp. T_0	0.47	0.29 - 0.74	0.001				0.53	0.33 - 0.85	0.008			
	Coccidia spp. T_0	1.44	1.11 - 1.87	0.006				1.53	1.17 - 2.00	0.002			
	Strongyloides spp. T_{-5}	3.31	2.4 - 4.56	<0.001				2.55	1.84 - 3.55	<0.001			
	Recruitment Weight							0.95	0.91 - 0.98	0.003			
	Sample year*							NA	NA	<0.001			
	Calf age*	NA	NA	<0.001				NA	NA	<0.001			
	Random effects:	Calf variance = 0.10; Sublocation variance = 0.12						Calf variance = 0.00; Sublocation variance= 0.14					
Calicophoron spp.	*Strongyloides* spp. T_0	0.41	0.23 - 0.73	0.002	2233	2339	445	0.38	0.21 - 0.69	0.001	2100	2339	445
	Coccidia spp. T_0	0.71	0.57 - 0.89	0.003				0.79	0.62 - 1.00	0.052			
	Calicophoron spp. T_{-5}	1.88	1.47 - 2.41	<0.001				1.36	1.04 - 1.78	0.025			
	Calicophoron spp. T_{-10}	1.86	1.43 - 2.43	<0.001				1.55	1.16 - 2.08	0.003			
	NDVI 10 weeks prior to current visit							18.66	3.00 - 116.03	0.002			
	Sample month*							NA	NA	<0.001			
	Sample year*							NA	NA	<0.001			
	Calf age*	NA	NA	<0.001				NA	NA	<0.001			
	Random effects:	Calf variance = 0.14; Sublocation variance = 0.32						Calf variance = 0.20; Sublocation variance = 0.10					

42

Table 2.2 – continued from previous page

Response Parasite (T_0)	Explanatory Variable	(a) Parasite Only Model						(b) Calf, Seasonal and Environmental Variation Model					
		OR	95% CI	P value	AIC	N_O	N_C	OR	95% CI	P value	AIC	N_O	N_C
Coccidia spp.	*Strongyloides* spp. T_0	1.52	1.2 - 1.92	0.001	4730	3623	455	1.61	1.26 - 2.05	<0.001	4668	3623	455
	Calicophoron spp. T_0	0.78	0.65 - 0.93	0.006				0.85	0.70 - 1.02	0.088			
	Theileria spp. T_{-5}	1.25	1.06 - 1.47	0.008				1.13	0.96 - 1.34	0.142			
	Northing							0.15	0.05 - 0.47	0.001			
	AEZ							NA	NA	<0.001			
	Sample month*							NA	NA	<0.001			
	Sample year*							NA	NA	<0.001			
	Calf age*	NA	NA	0.001				NA	NA	0.032			
	Random effects:	Calf variance = 0.10; Sublocation variance = 0.07						Calf variance = 0.09; Sublocation variance = 0.01					
Strongyles	*Strongyloides* spp. T_0	2.59	1.89 - 3.57	<0.001	3420	3114	454	2.67	1.94 - 3.68	<0.001	3414	3114	454
	Strongyles T_{-5}	1.37	1.13 - 1.66	0.002				1.36	1.12 - 1.65	0.002			
	Northing							2.59	1.37 - 4.91	0.003			
	Calf age*	NA	NA	<0.001				NA	NA	<0.001			
	Random effects:	Calf variance = 0.06; Sublocation variance = 0.05						Calf variance = 0.06; Sublocation variance = 0.02					
Theileria spp.	*Theileria* spp. T_{-5}	1.84	1.56 - 2.16	<0.001	4054	3608	455	1.63	1.37 - 1.93	<0.001	3838	3576	451
	Theileria spp. T_{-10}	1.49	1.26 - 1.78	<0.001				1.29	1.07 - 1.54	0.007			
	Red blood cell count at current visit							0.91	0.87 - 0.95	<0.001			
	Total serum protein at current visit							1.23	1.07 - 1.42	0.005			
	Total number of cattle on farm is >5							1.43	1.19 - 1.72	<0.001			
	Sample month*							NA	NA	<0.001			
	Sample year*							NA	NA	<0.001			
	Calf age*	NA	NA	<0.001				NA	NA	<0.001			
	Random effects:	Calf variance = 0.05; Sublocation variance = 0.08						Calf variance = 0.07; Sublocation variance = 0.07					

2.3.2 Analysis of temporal associations between multiple parasites

a) Intraspecific interactions

Positive intraspecific associations between each parasite group and its occurrence at an earlier time point (e.g. between T_0 and either T_{-5} or T_{-10}) were common (occurring 5 out of 10 possible times, Figure 2.5). This reflects the fact that calves infected with a parasite at time point T_{-5} or T_{-10} are more likely to still be positive for the parasite at T_0 due to the chronic nature of these infections.

b) Interspecific interactions

The associations occurring between parasites are illustrated in Figure 2.5. In total, there were 3 statistically significant interspecific associations occurring between parasites at one or more time points (T_0, T_{-5} and T_{-10}) out of the possible 60 interspecific pairwise combinations. These associations included 3/10 of all possible pairwise associations which could occur between the parasite at the current time point (T_0). I observed both positive and negative interspecific associations as illustrated in Figure 2.4 and Figure 2.5 (odds ratios for their associations are presented in Table 2.2), of which 2/3 associations were positive (between *Coccidia* spp. and *Strongyloides* spp. and strongyles) and 1/3 associations were negative (between Strongyloides spp. and *Calicophoron* spp.). However, after accounting for indicators of calf condition, environmental and seasonal variation no interspecific associations between parasites involving a time lag with a P<0.01 were observed (Figure2.5).

2.3.3 The role of calf, environmental or seasonal variation in determining associations

The parasites shared similar seasonal patterns of occurrence, 4/5 of the parasites were associated with sample year and 3/5 parasites were associated with sample month (Figure 2.5 and Table 2.2). However, a different set of calf and environmental variables were associated with each parasite. In addition to sample year, the occurrence of *Strongyloides* spp. was positively associated with calf recruitment weight, heavier calves were less likely to have *Strongyloides* spp. *Calicophoron* spp. was associated with NDVI 10 weeks prior to the current visit in addition to the sampling month and year. Strongyles were associated with northing and *Coccidia* spp. was associated with agro-ecological zone, northing, sample month and year. Factors associated with *Theileria* spp. also included sample month and year as well as the calf factors red blood cell count and total serum protein count at the current visit. Individuals from

farms with more than 5 cows in the herd were more likely to be infected with *Theileria* spp. than individuals from smaller farms (Table 2.2).

Inclusion of the indicators of calf condition and environmental and seasonal variation changed the coefficients for each of the associated parasites very little, despite the AIC indicating that the inclusion of these additional variables improved model fit (Table C.4 - Table C.9).

The majority of significant associations between parasites remained significant after accounting for variation in seasonal, environmental and calf level risk factors. 5 out of the 6 lagged, intraspecific associations and 3/10 of all possible pairwise associations which could occur between the parasite groups at the current time point (T_0) remained, at significance threshold P<0.01. The only interspecific association which involved a time lag between *Theileria* spp. (T_{-5}), and *Coccidia* spp. in the parasite only (stage 1 model) was removed after accounting for environmental variation (Figure 2.5 and Table 2.2).

2.4 Discussion

2.4.1 Associations between parasites

Akin to many other studies of coinfection (e.g. Lello et al. (2004) and Telfer et al. (2010)) and in particular to studies of coinfection with gastrointestinal parasites and helminths (Behnke et al., 2005; Christensen et al., 1987; Moreno et al., 2013), I have identified both positive and negative associations between coinfecting parasites in East African shorthorn zebu calves. Most associations remained after accounting for measured environmental variation. This makes it less likely that hosts were coinfected simply because the parasites share the same environmental requirements. Furthermore, inclusion of indicators of calf condition such as calf weight and hematological parameters did not affect the association between parasites, suggesting that the relationship observed is not solely due to, for example, the calf being in ill health, thereby increasing the risk of infection and generating a vicious circle as seen by Beldomenico et al. (2008). In addition, no influence of sex on coinfections was identified, even though male calves have higher burdens of strongyles than female calves in this population (Callaby et al., 2015). However, inclusion of seasonal variation reduced the estimated effect size for the association between *Calicophoron* spp. and *Coccidia* spp., consistent with this relationship being confounded by the increased prevalence of *Calicophoron* spp. during the wet season (Figure C.1). This likely reflects the *Calicophoron* spp. lifecycle, which involves snails as its intermediate

host (Sanabria and Romero, 2008). Seasonal fluctuations in helminth-*Coccidia* associations have also been identified in African buffalo and other ungulates (Gorsich et al., 2014; Turner and Getz, 2010). In addition, the association between *Theileria* spp. (T$_{-5}$), and *Coccidia* spp. (T$_0$) was removed after accounting for environmental variation in the form of northing, agro-ecological zone, sample month and sample year. My results suggest that it is important to account for environmental and seasonal variation when analyzing parasite-parasite association data to avoid drawing false conclusions about the relationships amongst coinfecting parasites.

Modelling the parasite associations before accounting for the environmental variation could lead to a overestimation of the number of associations occuring. However, since I have used P<0.001 to account for multiple comparisons, this is unlikely. In addition, for a subset of data, were sufficient information was available I ran the full models with all parasites, time lags and environmental variables included in the model, significant factors remained in the minimal model consistent with the two stage model. This suggests that my findings are not very sensitive to the two stage analysis.

The positive and negative parasite-parasite associations identified in this study occurred between the gut parasites (strongyles, *Strongyloides* spp., *Calicophoron* spp. and *Coccidia* spp.), which share the same host resources (Table C.1 and Figure 2.5). In a network analysis of human parasites, the majority of associations identified occurred amongst parasites which shared the same resources (Griffiths et al., 2014). This is also likely to be the case in the IDEAL calves (Table C.1).

As in Cizauskas's 2013 study of coinfection in other ungulate species, zebra and springbok, I identified positive associations between *Coccidia* spp. and *Strongyloides* spp. Surprisingly, neither study identified any association between strongyles and *Coccidia* spp. It has previously been hypothesized that these two parasites should interact as the immune response to the protozoa *Coccidia* spp. is controlled by a Th1 type immune response whereas the immune response to helminths, including strongyles is controlled by a Th2 type immune response, which is thought to down-regulate the response to *Coccidia* spp. (Cizauskas, 2013; Craig et al., 2008; Hong et al., 2006). In addition, associations between *Coccidia* spp. and helminths has been observed in wild mice (Fenton et al., 2014; Knowles et al., 2013; Pedersen and Antonovics, 2013). Experimental treatment of wild mice with Ivermectin (a broad-spectrum anthelmintic drug) lead to a shift in the parasite community of the host which resulted in a reduced nematode infection and an increased *Coccidia* spp. infection (Pedersen and Antonovics, 2013), however the mechanism behind this association remains unclear.

Theileria spp. was separated from the other parasites in this study (Figure 2.5). This is fitting as *Theileria* spp. is a blood-borne parasite, whilst the other parasites in this study are found within the gut. However, it is still possible to expect Th1-Th2 type associations (Ezenwa and Jolles, 2015; Graham, 2008). The mouse and human equivalent of *Theileria* spp. is *Plasmodium* spp. which causes malaria. A meta-analysis of helminth-malaria coinfections in mice has shown that coinfection resulted in an increased peak parasitaemia (indicating that the malaria parasites are replicating within the host) for ordinarily resolving malaria infections (*P. chabaudi* and *P. yoelii*), whilst coinfection had no effect on parasitaemia with lethal malaria infections *P. berghei* (Knowles, 2011). Since my analysis used the presence or absence of *Theileria* spp., I could be masking the effect of coinfections between helminths and individual *Theileria* species such as *T. parva* and *T. mutans*.

2.4.2 Temporal associations between multiple parasites

Concurrent infections were identified as the main drivers of the associations found in this chapter, rather than there being long-term carry over effects. One possible explanation for the predominance effect of concurrent infection is that time lags of 5 week intervals are too long and a finer time scale is needed. For example, Knowles et al.'s 2013 experiment on wood mice (*Apodemus sylvaticus*) only detected associations between nematodes and *Eimeria* within a 1-3 week period following treatment with Ivermectin by the fourth week, infection by both parasites had returned to pre-treatment levels. Also many other studies have shown that the timing and order of previous infections is important (Hoverman et al., 2013; Jackson et al., 2006; Telfer et al., 2010; Thomas et al., 2003; Tompkins et al., 2011). For instance, an experiment in mice showed that prior infection by influenza virus increased the burden of *Listeria monocytogenes* (Jamieson et al., 2010), whilst protective immunity occurring between two temporally separated parasites has been exploited in the development of vaccines such as the BCG and smallpox (Fenner et al., 1988; Fine, 1995).

Since the identified parasite-parasite associations occurred between infections at the same points it means that direction of effect is unknown. It is not possible to tell if parasite A is causing a decreased/increase risk of parasite B or if in fact, it is parasite B changing the risk of having parasite A. Therefore a finer timescale and further information in the form of experiments would be needed to identify the mechanisms behind the associations.

To conclude, I have found evidence for both positive and negative associations between coinfecting parasites in zebu calves. My results indicate that most associations involve concurrent infections and that the majority of these associations between

parasites remained (3/10 pairwise associations) after accounting for variation in seasonal, environmental and calf level risk factors. The findings could have important implications for population health campaigns as the removal of *Coccidia* spp. or strongyles could have knock-on benefits, as the interaction between these parasites was positive, so the elimination of one species may result in a decrease in the severity or the prevalence of a second parasite. In contrast, the removal of other species such as *Calicophoron* spp. could have detrimental effects on infection by other parasites, as their elimination could result in an increase in the prevalence of the second parasite, whereas the removal of *Strongyloides* spp. could have both positive and negative effects on the burden of *Coccidia* spp. and *Calicophoron* spp., respectively. Furthermore, the prevalence of the parasite also affects the power of the analysis to detect the relationship. Therefore it is important to understand the prevalence and the directionality of effects of the multiple parasites involved in a population before planning population health campaigns.

Chapter 3

Seroprevalence of respiratory viral pathogens of indigenous calves in Western Kenya

Summary

This chapter investigates the seroprevalence, distribution and relationship between the viruses involved in the bovine respiratory disease complex (BRD), infectious bovine rhinotracheitis (IBR), bovine parainfluenza virus Type 3 (PIV3) and bovine viral diarrhoea virus (BVDV) in East African Shorthorn Zebu calves at 51 weeks of age. In addition, I evaluate their effect upon the host. Calves experience similar risks of infection for IBR, PIV3, and BVDV with a seroprevalence of 20.9%, 20.1% and 19.8% respectively. I confirm that positive associations exist between IBR, PIV3 and BVDV in a previously unstudied setting; being seropositive for any one of these three viruses means that an individual is more likely to be seropositive for the other two viruses than expected by chance. This relationship remained after accounting for environmental variation. Infection with other viruses, protozoa or helminths did not increase the risk of being seropositive for IBR or PIV3; however individuals seropositive for the protozoa *Babesia bigemina* were more likely to seropositive for BVDV than those *B. bigemina* seronegative individuals. Being seropositive for any one of the three respiratory viruses does not affect the average daily weight gain of the calf. Nor does being PIV3 or BVDV serostatus have an effect on the calf ever experiencing a clinical episode. However, calves which were IBR seropositive were less likely to experience a clinical episode of some form, suggestive of some protective aspect of IBR.

3.1 Introduction

Most studies of infectious diseases in East African cattle have concentrated on gastro-intestinal parasites and vector-borne diseases. As a result, to date relatively little is known about viral diseases, except for those that are clinically symptomatic or which affect international trade such as foot and mouth disease, bluetongue and epizootic haemorrhagic disease (Bronsvoort et al., 2003; Toye et al., 2013b). Yet, in the rest of the world, other viral diseases are known to have a large impact upon the livestock industry. For example, bovine respiratory disease complex (BRD) is responsible for major economic losses in cattle from intensive farming systems as a result of pneumonia (Bowland and Shewen, 2000). Furthermore, studies of intensively farmed feedlot and beef cattle in North America have shown that individuals infected with BRD exhibit a decreased growth rate and some individuals show signs of clinical illness (Gardner et al., 1999; Schneider et al., 2009; White and Renter, 2009).

There are many factors which contribute to the BRD, including stress, management practices and biological agents (Snowder et al., 2006; Taylor et al., 2010). Viruses have been implicated in contributing to BRD by causing lesions in the bovine respiratory tract and/or impairing the clearance of bacteria from the lower respiratory tract (Coetzer and Tustin, 2004). Some of the viruses contributing to BRD include infectious bovine rhinotracheitis virus (IBR), bovine parainfluenza virus type 3 (PIV3) and bovine viral diarrhoea virus (BVDV). Each of these viruses has specific clinical signs, consequences and economic importance (see Appendix D.1). In addition, both IBR and BVDV are immunosuppressive (Hutchings et al., 1990; Koppers-Lalic et al., 2001; Roth et al., 1986; Wellenberg et al., 2002). Numerous epidemiological studies have reported positive associations between these viruses (Fulton et al., 2000; Martin and Bohac, 1986) and the relationship has been reviewed since the 1980s (Yates, 1982). Although these viruses have been documented in many parts of the world (Coetzer and Tustin, 2004), little is known about their distribution within sub-Saharan Africa, particularly in East African extensive farming systems.

Furthermore, the Infectious Disease of East African Livestock (IDEAL) project has shown that individuals within East African cattle populations are coinfected with many different parasites (Bronsvoort et al., 2013). Both positive and negative associations exist between the coinfecting parasites in the IDEAL project (see Chapter 2). Previous research into the IDEAL study population has also shown that coinfections with protozoans such as *Theileria parva*, *Theileria mutans* and *Trypanosoma* spp., and gut nematodes such as strongyles have impacts upon calf survival, growth rate and hematological parameters (Conradie van Wyk et al., 2014; Thumbi et al., 2013b, 2014; Woolhouse et al., 2015).

The purpose of this study is to improve current knowledge on the seroprevalence and distribution of IBR, PIV3, and BVDV in cattle in sub-Saharan Africa by estimating the seroprevalence of antibodies to these viruses in East African shorthorn zebu calves from Western Kenya using data gathered by the IDEAL project. This chapter also aims to quantify the associations between these viruses within this study population and to investigate their association with infection with other parasites prevalent in the study region. In addition, the effect of the viruses on calf growth rate and the risk of developing clinical illness will be evaluated.

3.2 Material and Methods

3.2.1 Study population

The study reported here was part of the Infectious Diseases of East Africa Livestock (IDEAL) project, the design of which has previously been reported by Bronsvoort et al. (2013), see also Appendix A. Briefly, the IDEAL project was an intensive cohort study of 548 indigenous shorthorn zebu calves from 3-7 days old which were followed during their first year of life between October 2007 and September 2010.

Calves were selected using a stratified two-stage random cluster study design. In the first stage, 20 sublocations (the smallest administrative unit in Kenya) were selected from 5 agro-ecological zones, in a radius of 45km from the town of Busia for logistical reasons. This area is dominated by smallholder mixed crop-livestock production systems, with an average farm size of 2 hectares and about 5 cattle per farm (Bronsvoort et al., 2013). During the second sampling stage, 28 calves from each sublocation were recruited. To be recruited into the study, calves were aged between 3-7 days old, their dam had to have been on the farm for at least one year, the calf was not a result of artificial insemination and the herd should have been under open grazing management. Only one dam and calf from a farm could be in the study at any one time.

3.2.2 Data collection

At the recruitment visit, the calf's owner completed a questionnaire. This questionnaire collected environmental information about the farm, other livestock, water sources and animal husbandry practices. Calf locations were geo-referenced using hand-held GPS devices (Garmin 12, Garmin Kansas, USA). Calves then received routine visits every 5 weeks, until they were 51 weeks old, at which a clinical examination took place and biological samples were taken for laboratory analysis. If the calf

presented with a clinical illness during the study period then the calf received an extra clinical examination and further biological samples were taken for analysis. Jennings (2013) provides more detail on the clinical episodes within this population. In addition, calves were weighed (measured to the nearest 0.5kg) at recruitment, then again every five weeks until 31 weeks of age, and once again at the last visit at 51 weeks. The dam's girth was also measured at the recruitment visit (measured in cm). Each calf was genotyped using a 50K Illumina® BovineSNP50 beadchip and the percentage of European taurine introgression and level of heterozygosity was calculated. Mbole-Kariuki et al. (2014) and Murray et al. (2013) describe the calculation of the percentage of European taurine introgression and heterozygosity within this population.

3.2.3 Laboratory analysis

Viruses

Jugular vein blood samples were collected in plain Vacutainer™ (Becton Dickinson, England) tubes. Blood samples were allowed to clot, the serum recovered and aliquots were stored at $-20\,^{\circ}$C, until serological analysis could take place.

The serological samples from 455 calves at 51 weeks old were screened using SVANOVIR kits obtained from Svanova Biotech AB (Uppsala, Sweden) to identify antibodies specific to the following viruses: a) infectious bovine rhinotracheitis (IBR); b) bovine viral diarrhoea virus (BVDV) antibody and c) bovine parainfluenza virus type 3 (PIV3). In addition, BVDV antigens were screened using kits obtained from IDEXX (Montpellier SAS, France). All the kits (apart from the BVDV antigen kit) were designed to detect virus specific IgG antibodies in serum using a procedure based upon a solid phase indirect Enzyme Linked Immunosorbent Assay (Indirect ELISA). All assays were performed and analysed according to the manufacturer's instructions using single wells. The laboratory analysis was performed at the International Livestock Research Institute, Nairobi, Kenya.

The SVANOVIR ELISA tests are based on the measurement of corrected optical density (OD_{corr}). To calculate the OD_{corr}, values in wells coated with a particular antigen are corrected by subtracting the OD values of corresponding wells containing the control antigen ($OD_{test\ antigen}$ - $OD_{control}$ = OD_{corr}). The Percent Positivity (PP) values are calculated as follows:

$$Percent\ Positivity = \frac{Test\ Sample\ (OD_{corr}) \times 100}{Positive\ Control\ (OD_{corr})}$$

IDEXX ELISA tests results are interpreted using a similar index called the "sample to positive control percentage" (S/P%), which was calculated according to the manufacturer's instructions for the test.

To convert the continuous PP (or S/P%) values into a serostatus which is a binary seropositive/seronegative outcome, the PP (or S/P%) values were interpreted according to the manufacturer's cut-offs specified in Appendix D.2. Justification for grouping the inconclusive IBR results with the IBR seropositive individuals (as defined in Appendix D.2) is presented in the Appendix D.3. Therefore, seropositive/seronegative means antibodies to the antigen of interest are detectable / not detectable in the blood of the host, whereas seroconversion is the movement from a seronegative to a seropositive state.

Other parasites

In addition to the screening of viruses, diagnostic tests were performed to screen for more than 100 parasites on samples collected during each calf's involvement in the IDEAL project. A description of all diagnostic tests used, the timing of testing, and parasites identified are provided in Bronsvoort et al. (2013) and Appendix A. To make parasite infection status at each visit comparable with the serostatus of the calf to the viruses at 51 weeks old, for this chapter I created a new set of variables: if the calf tested positive for a parasite at any visit then the calf is classified as ever being infected with the parasite.

3.2.4   Statistical analysis

Seroprevalence of the viruses

The crude seroprevalence for each virus was calculated using the *epicalc* package (Chongsuvivatwong, 2012) in R v.2.15.2. The weighted adjusted seroprevalence for each virus was calculated using the R *survey* package (Lumley, 2012). The weighting adjusted for the number of breeding dams in each sublocation. Sublocation-specific seroprevalence was mapped using ArcGIS.

Associations between viruses

The association between the three viruses was assessed using generalised linear mixed models (GLMM) fitted with a logit link function and binomial errors and a Laplace approximation to the maximum likelihood estimation in R v.2.15.2 using the *lme4* package (Bates et al., 2014). I investigate the association between the serostatus of

virus A at 51 weeks old with serostatus of multiple viruses at 51 weeks with the following model structure:

$$logit(\textit{Virus A serostatus}_i) = \alpha + \beta_1 \textit{Virus B serostatus}_i + \beta_2 \textit{Virus C serostatus}_i$$
$$+ b_{1_i} \textit{Sublocation}_i + \varepsilon_i \qquad (3.1)$$

where α is the intercept and ε is the residual variation. *Virus X serostatus$_i$* is the serostatus of the calf (*i*) at 51 weeks of age for each virus, as defined in section 3.2.3, and is included in the models as a fixed effect (β). Sublocation (*Sublocation*, 20 levels) is included in the model as a random effect (*b*) to account for the study design and environmental similarity between calves clustered into each sublocation.

A separate model was constructed for each virus. Viruses were said to be co-distributed if they occurred in an individual more often than expected by chance. The same patterns of association were examined using continuous measures of PP values instead of serostatus (Appendix D.4).

Following construction of the virus-only model, environmental confounders (Appendix D.5), which may affect respiratory parasites or contact between calves, were added into each of the models. By using a two stage statistical approach it allowed me to examine whether the associations between being seropositive for the different viruses was due to coinfection or if the associations were being driven by a joint underlying correlation such as a shared environmental factor or calf characteristic.

Associations between virus serostatus and ever being infected with other parasites

To identify associations between virus serostatus at 51 weeks old (the response variable) and being infected with other parasites, Equation 3.1 was extended to include a binary fixed effect variable denoting whether the calf ever had the parasite. A calf was classified as positive for the parasite of interest if it tested positive for the parasite at any time point during its 51 weeks of observation in the IDEAL study (see Appendix D.6 for a list of parasites identified in the IDEAL study). The confounding effects of seroconversion to the other viruses were accounted for by including the serostatus of the other viruses as binary fixed effects in the model, for example:

$$logit(\textit{BVDV serostatus}_i) = \alpha + \beta_1 \textit{PIV3 serostatus}_i + \beta_2 \textit{IBR serostatus}_i$$
$$+ \beta_3 \textit{Parasite A}_i + b_{1_i} \textit{Sublocation}_i + \varepsilon_i \qquad (3.2)$$

where α is the intercept and ε is the residual variation. Fixed effects are symbolized by β. *[VIRUS NAME] serostatus$_i$* is the serostatus of the calf (*i*) at 51 weeks of age and

Parasite A describes whether or not the calf has ever been infected with the parasite of interest at any time point during its 51 weeks of involvement in the IDEAL project. Sublocation (*Sublocation*, 20 levels) is included in the model as a random effect (*b*) to account for repeated measures at each location in the study design and environmental similarity between calves clustered into each sublocation. A separate model was built for each of the viruses.

Using the approach of Hosmer et al. (2000), parasites which were associated with the serostatus of the virus with a value of P<0.2 in the stage 1 analysis were included in the stage 2 model. Backwards-stepwise selection was used to sequentially remove variables from the maximum model starting with variables which had the least significant P value. Given the large number of combinations involved, I chose the most parsimonious model to be the model in which all explanatory parasites were associated with the serostatus of the virus of interest with a P value <0.01. If a virus-parasite association was detected then potential confounding environmental information was included in the virus-parasite association model using the same method as that of the virus-only models.

To examine the association between virus serostatus and egg burden, as measured by the number of strongyle eggs per gramme of faeces (EPG), the same model format as Equation 3.2 was used. However, the transformed median strongyle EPG ($\log_{10}$(median EPG + 1)) across all observations for each calf was fitted as the explanatory variable.

Association between virus serostatus, growth rate and clinical episodes

To assess the association between IBR, PIV3 or BVDV and clinical illness, GLMMs with a logit link function and binomial errors were built. Firstly, the effect of being seropositive for the three viruses on ever experiencing any kind of clinical episode was modelled, where the response variable was if the calf presented with a clinical episode at least once during the study period:

$$logit(Clinical\ episode\ ever_i) = \alpha + \beta_1 IBR\ serostatus_i + \beta_2 PIV3\ serostatus_i$$
$$+ \beta_3 BVDV\ serostatus_i + b_{1_i} Sublocation + \varepsilon_i \quad (3.3)$$

where α is the intercept and ε is the residual variation. Fixed effects are symbolized by β. *[VIRUS NAME] serostatus$_i$* is the serostatus of the calf (i) at 51 weeks of age. Sublocation (*Sublocation*, 20 levels) is included in the model as a random effect (*b*).

Secondly, the same model structure (Equation 3.3) was used to investigate the effect of virus serostatus on ever experiencing a specific type of clinical episode or

clinical sign (see Appendix D.7 for a list of the gross types of clinical episode and the main clinical signs observed). I was particularly interested in whether being seropositive for any of the three viruses was associated with a calf ever experiencing a respiratory clinical episode or cough and fever since IBR, PIV3 and BVDV contribute to the BRD causal web by causing lesions in the respiratory tract and/or impairing the clearance of bacteria from the lower respiratory tract and these are the symptoms which usually occur when the syndrome is acute (Coetzer and Tustin, 2004). A respiratory clinical episode is one where the calf shows signs of respiratory problems such as nasal discharge, costo-abdominal respiration or shallow/rapid breathing or deep laboured breathing and was classified as a clinical episode by the visiting animal health assistant or veterinarian. Thirdly, the effect of virus serostatus on the *number* of clinical episodes a calf experienced (see Appendix D.7 for a description of this response variable) was investigated by GLMM with negative binomial errors using the lme4 package in R v.2.15.2 (Bates et al., 2014). Again, the serostatus of the virus was included in the model as a binary fixed effect, whilst sublocation was included in the model as a random effect:

$$
\begin{aligned}
log(\textit{Number of clinical episodes}_i) = \\
\alpha + \beta_1 \textit{IBR serostatus}_i + \beta_2 \textit{PIV3 serostatus}_i + \beta_3 \textit{BVDV serostatus}_i + \\
b_{1_i} \textit{Sublocation}_i + \varepsilon_i \quad (3.4)
\end{aligned}
$$

The effect of virus serostatus on the age at which the calf experienced a clinical episode was also investigated. If a constant exposure to the virus is assumed, then I predict that there would be a constant risk of experiencing a clinical episode. However if there was a age related effect of exposure on experiencing a clinical episode then I hypothesize that the risk of observing a clinical episode will be different in young, medium and old age classes. Therefore three different binary variables of no/yes for experiencing a clinical episode at <16 weeks (young), 16-36 weeks (middle age) and >36 weeks (old) were created and I fitted a GLMM with binomial errors to model the effect of virus serostatus on each of the age groups, e.g.:

$$
\begin{aligned}
logit(\textit{Clinical episode within age class}_i) = \\
\alpha + \beta_1 \textit{IBR serostatus}_i + \beta_2 \textit{PIV3 serostatus}_i + \beta_3 \textit{BVDV serostatus}_i + \\
b_{1_i} \textit{Sublocation}_i + \varepsilon_i \quad (3.5)
\end{aligned}
$$

Therefore the model in Equation 3.5 was fitted three times for each age group.

As exposure to other parasites may influence an individual's risk of experiencing a clinical episode, parasites which were associated with ever experiencing a clinical episode with P<0.2 were included as a fixed effect in Equation 3.3 and Equation 3.4. Backwards-stepwise selection was used to sequentially remove variables from the maximum model starting with variables which had the least significant P value. This allows me to examine if the association between clinical episode and serostatus was due to the exposure from the virus of interest, rather than being driven by infection with a different parasite species.

Lastly, the effect of virus serostatus on a calf's growth rate was examined using a linear mixed model. Growth rate was measured by the average daily weight gain (ADWG) of a calf, which is defined as:

$$\text{Average daily weight gain (kg)} = \frac{\text{Weight at end of study (kg) - Weight at start of study (kg)}}{\text{Total number of days in study period (days)}}$$

ADWG was $\log_{10}$ transformed so that it was normally distributed. Sublocation was included in the model as a random effect and the fixed effects included in the model were the binary serostatus of each virus and the environmental variables from Appendix D.5.

Genotype and phenotype at birth and infection risk

The relationship between virus serostatus and clinical episode is potentially confounded by the effect of recruitment weight: previous research has shown that individuals which have a lower birth weight have an increased infection risk later in life (Lu et al., 2013; Raqib et al., 2007). Within the IDEAL study population, it has been observed that lighter calves at birth have a higher strongyle EPG (Callaby et al., 2015). Therefore, I investigate the effect of an individual's phenotype at birth on infection risk in later life by examining the association between dam girth at recruitment and the recruitment weight of the calf on being seropositive for IBR and ever experiencing a clinical episode and the number of clinical episodes. In addition, an individual's genotype may affect recruitment weight, infection risk and clinical outcome. For example, inbred red deer are lighter at birth and are less likely to survive the first year of life (Walling et al., 2011). Likewise, inbred Soay sheep have increased susceptibility to parasitism by gastrointestinal helminths (Coltman et al., 1999). Furthermore, Murray et al. (2013) identified a negative association between heterozygosity (a measure of inbreeding) and death and a positive association between outbreeding in the form of European taurine introgression and clinical illness in East

African shorthorn zebu. Therefore, I investigate the effect of an individual's genotype on infection risk in later life by examining the association between genotypic traits in the form of percentage European taurine introgression and heterozygosity (see Chapter 4 and 5 or Mbole-Kariuki et al. (2014) and Murray et al. (2013) for more details) on ever experiencing a clinical episode and number of clinical episodes. This was achieved by including the genotypic and phenotypic traits as fixed effects in the clinical episode models (Equation 3.3 and Equation 3.4). The confounding effects of IBR, PIV3 and BVDV serostatus were included in the model as binary fixed effects. Since IBR was the only virus associated with clinical episodes, I also investigated the effect of the same genotypic and phenotypic traits on IBR serostatus by fitting them as fixed effects in the IBR-virus-only association model from Equation 3.1.

Lastly, the effects of dam girth at recruitment, European taurine introgression and heterozygosity on calf recruitment weight were investigated using linear mixed models. Initially, the univariable association between recruitment weight and the genotype or phenotype was modelled; sublocation was included in the model as a random effect. Then a multivariable model was built which included all genotypes and phenotypes from the recruitment weight univariable analysis associated with recruitment weight with P<0.2. Backwards-stepwise selection was used to obtain the most parsimonious model.

3.3 Results

3.3.1 Seroprevalence of the viruses

Estimation of the seroprevalence of each virus, indicates that calves experience a similar risk of infection for IBR, PIV3, and BVDV antibodies, with adjusted seroprevalences of 20.9%, 20.1%, 19.8%, respectively (Table 3.1). Twenty three (5.1%) of calves were seropositive for all three virus antibodies, whilst 285 (62.6%) of calves were seronegative for all three virus antibodies. No calves were seropositive for the BVDV antigen. The seroprevalence of the viruses in each sublocation is plotted in Figure 3.1.

Table 3.1 Crude and adjusted seroprevalence of each virus. Seroprevalence is followed by the 95% confidence interval in brackets. IBR = infectious bovine rhinotracheitis; PIV3 = bovine parainfluenza virus type 3; BVDV = bovine viral diarrhoea virus

Virus	N_{calves}	N_{tests}	Crude Seroprevalence	Adjusted Seroprevalence
IBR	91*	455	20.00 (16.42-23.98)	20.90 (16.41-25.39)
PIV3	80	455	17.62 (14.23-21.44)	20.08 (15.52-24.63)
BVDV	79	454	17.36 (13.99-21.16)	19.76 (15.17-24.36)

*66 of the IBR inconclusive calves (according to the manufacturers cut-offs) were classified as seropositive following the case-case analysis in Appendix D.3
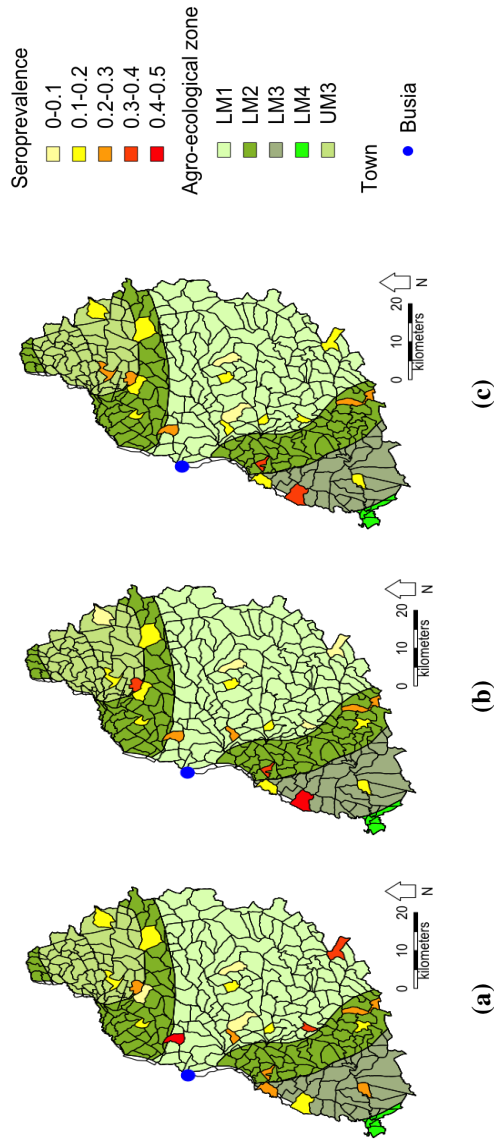
Figure 3.1 Map of western Kenya showing the distribution of calves seropositive for a) IBR; b) PIV3 and c) BVDV at 51 weeks of age. The map also shows the five agro-ecological zones coloured different shades of green and the 20 sublocations in the study area are shaded according to the observed seroprevalence for each virus within them. The blue circle indicates the location of the project laboratory in Busia.

59

Table 3.2 Odds ratio and 95% confidence interval from the virus-only analyses of the association between IBR, PIV3 and BVDV serostatus.

| Response Virus | Explanatory Virus | | | | | | | | | Random effect: Sublocation | |
| | IBR | | | PIV3 | | | BVDV | | | | |
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value	Variance	SD
IBR	-	-	-	2.77	1.53-5.03	0.001	2.77	1.54-5.00	0.001	0.15	0.38
PIV3	2.76	1.53-4.98	0.001	-	-	-	5.87	3.30-10.46	<0.001	0.14	0.38
BVDV	2.77	1.56-4.92	0.001	5.98	3.40-10.52	<0.001	-	-	-	0.00	0.00
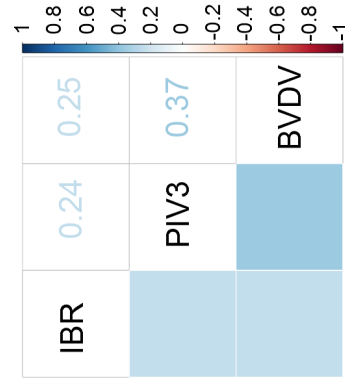
Figure 3.2 Correlation matrix showing the correlation IBR, PIV3 and BVDV serostatus without correcting for any other variables.

3.3.2 Associations between viruses

The virus-only GLMMs indicated that IBR, PIV3 and BVDV were co-distributed (Table 3.2). There was an increased risk of an individual being seropositive for one of these viruses, if the calf was also seropositive for one or both of the other viruses (Table 3.2). In addition, a positive correlation was observed between the three viruses (Figure 3.2). The same pattern of results were observed using PP value as a continuous variable (Appendix D.4). Inclusion of environmental confounders in the virus-only models did not affect the relationship observed between the serostatus of the three viruses. The association between viruses after accounting for environmental variation is presented in Appendix D.5.

3.3.3 Associations between virus serostatus and ever being infected with other parasites

Ever being infected with any of the other parasites in the IDEAL project did not increase the risk of being IBR or PIV3 seropositive (see Appendix D.6 for results of the univariable analysis). In addition, no association was found between IBR, PIV3 or BVDV serostatus and strongyle EPG (Appendix D.6). However, *Babesia bigemina* seropositive individuals had 2.88 times the odds of being BVDV seropositve than those which were *B. bigemina* seronegative (OR=2.88, 95% CI=1.67-4.97, P<0.001; Appendix D.6). The association between BVDV and *B. bigemina* remained when confounding environmental variables were included in the model.

3.3.4 Association between virus serostatus, growth rate and clinical episodes

No association was identified between IBR, PIV3 and BVDV and average daily weight gain (ADWG), i.e. being seropositive for IBR, PIV3 or BVDV did not alter a calf's ADWG (Table 3.3 and Figure 3.3).

In total, 45% (206/455) of the calves in this study experienced a clinical episode. No association was found between PIV3 and BVDV serostatus and ever experiencing a clinical episode, respiratory clinical episode, or any other gross category or clinical disorder (Figure 3.4, Appendix D.7). Unexpectedly, IBR seropositive calves were less likely to experience any type of clinical episode than IBR seronegative calves (OR$_{\text{IBR seropositive}}$=0.43, 95% CI=0.25-0.73, P=0.002, Table 3.4). Furthermore, IBR seropositive calves experienced fewer clinical episodes than IBR seronegative calves (IBR seronegative: Mean=0.85, SE=0.07, Range=0-11; IBR seropositive: Mean=0.47,

Table 3.3 The effect of IBR, PIV3 and BVDV serostatus on average daily weight gain ($\log_{10}$(ADWG)). a) without accounting for environmental, calf and dam variation and b) after accounting for environmental, calf and dam variation.

	Virus Serostatus	Estimate	SE	OR	95% CI	P value
a)	IBR Seropositive	0.02	0.02	1.02	0.98-1.06	0.242
	PIV3 Seropositive	0.02	0.02	1.02	0.98-1.06	0.384
	BVDV Seropositive	0.01	0.02	1.01	0.97 -1.05	0.572
b)	IBR Seropositive	0.01	0.02	1.01	0.98-1.05	0.456
	PIV3 Seropositive	0.02	0.02	1.02	0.98-1.06	0.343
	BVDV Seropositive	0.02	0.02	1.02	0.98-1.06	0.431
	Water provision (animals go to water)	-0.03	0.01	0.97	0.94-1.00	0.042
	Dam girth at recruitment (cm)	1.10	0.35	3.00	1.52-5.94	0.002
	Dam condition score at recruitment	0.03	0.01	1.03	1.02-1.05	<0.001
	Calf sex (effect of being male)	0.03	0.01	1.03	1.01-1.06	0.017
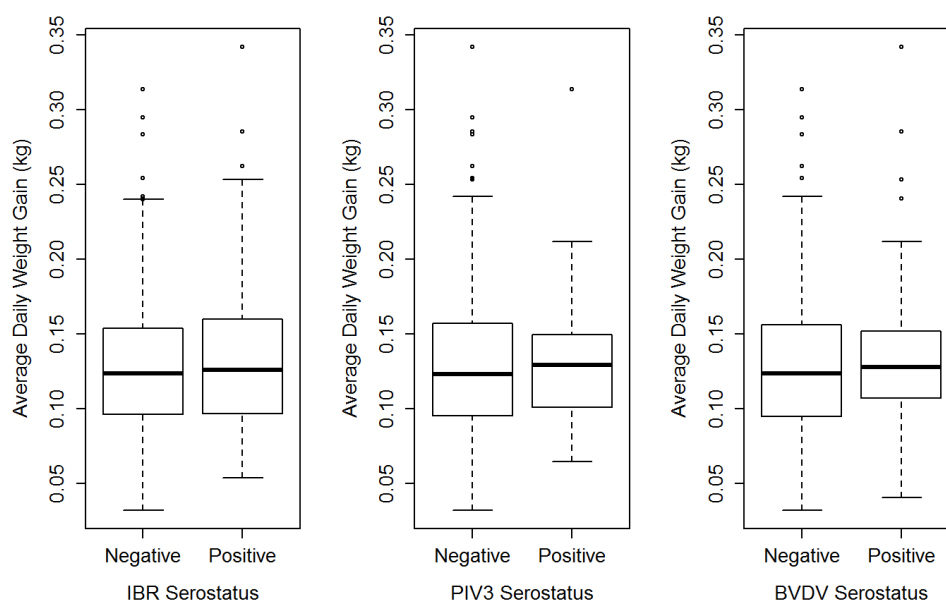
Figure 3.3 Box and whisker plots showing the average daily weight gain (ADWG) for calves which were seropositive or seronegative for IBR, PIV3 and BVDV. The heavy solid black line represents the median ADWG for each group, whilst the bottom and top of the box represents the 25th and 75th percentiles, respectively. The whiskers represent 1.5 times the interquartile range. Points beyond the whiskers are outliers.
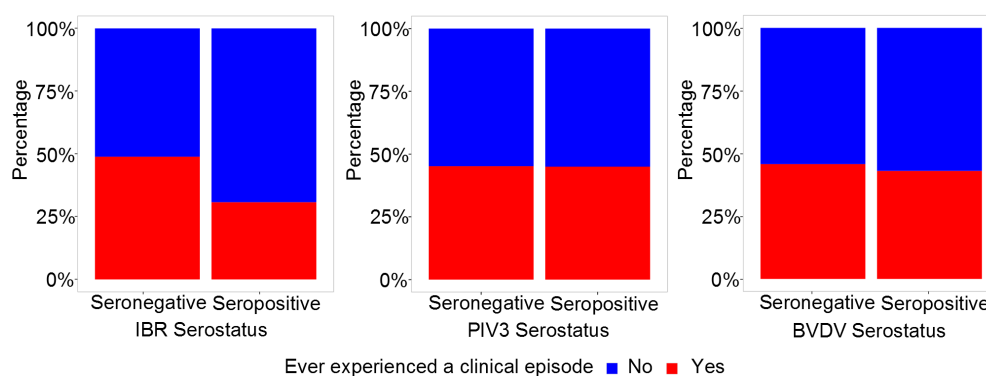
Figure 3.4 Percentage of individuals which ever experienced any type of clinical episode for each serostatus of IBR, PIV3 and BVDV.

SE=0.09, Range=0-4). This effect was independent to the effect of the serostatus for the other two viruses and the relationship was not confounded by any of the environmental, husbandry, dam or calf level factors described in Appendix D.5. No association was observed between IBR and respiratory clinical episodes, cough and fever however IBR seropositive calves where less likely to experience mild fever, gastrointestinal problems and clinical disorders of unknown aetiology than IBR seronegative calves (Table 3.5 and Table 3.6).

In addition, the association between IBR and clinical episode remained after accounting for exposure to different parasites (Table 3.7). Calves which were seropositive for *T. parva* were at increased risk of experiencing a clinical episode; however *B. bigemina* serostatus was not associated with ever experiencing a clinical episode (Table 3.7).

Being IBR seropositve at 51 weeks of age is correlated with clinical episodes in young and old calves: IBR seropositive calves are less likely to experience a clinical episode at old and young ages (>36 weeks and <16 weeks, respectively) but there was no significant effect of IBR serostatus on the probability of experiencing a clinical episode for middle aged calves (16-36 weeks old, Table 3.8). Clinical episodes of any kind most commonly occur in middle aged calves and respiratory clinical episodes tend to occur in younger individuals (Appendix D.7).

Table 3.4 The association between serostatus and experiencing any type of clinical episode and the number of clinical episodes a calf experienced (a) before and (b) after accounting for confounding from recruitment weight. c) The effect of recruitment weight on experiencing any type of clinical episode and the number of clinical episodes a calf experienced without accounting for serostatus. Each line represents a separate model. A GLMM with binomial errors was used to fit the ever experienced any type of clinical episode model whilst a GLMM with negative binomial errors was used to fit the number of clinical episodes model.

Explanatory Variable	Ever experienced any type of clinical episode					Number of clinical episodes a calf experienced				
	Estimate	SE	OR	95% CI	P value	Estimate	SE	OR	95% CI	P value
a) IBR Seropositive	-0.85	0.27	0.43	0.25-0.73	0.002	-0.63	0.20	0.53	0.36-0.79	0.002
PIV3 Seropositive	0.08	0.29	1.08	0.61-1.89	0.793	0.07	0.19	1.07	0.73-1.56	0.710
BVDV Seropositive	-0.02	0.28	0.98	0.56-1.71	0.938	-0.07	0.19	0.93	0.64-1.37	0.730
b) Recruitment weight (kg)	-0.03	0.03	0.97	0.92-1.02	0.260	-0.04	0.02	0.96	0.93-1.00	0.038
IBR Seropositive	-0.79	0.27	0.45	0.26-0.77	0.004	-0.59	0.20	0.56	0.38-0.83	0.004
PIV3 Seropositive	0.09	0.29	1.09	0.62-1.92	0.759	0.04	0.19	1.04	0.71-1.52	0.836
BVDV Seropositive	-0.05	0.29	0.95	0.54-1.68	0.870	-0.04	0.20	0.96	0.65-1.41	0.840
c) Recruitment weight (kg)	-0.04	0.03	0.96	0.91-1.01	0.096	-0.05	0.02	0.95	0.92-0.99	0.013

Table 3.5 Number of calves and the odds ratio for seropositive calves which ever experienced each type of gross category of clinical disorder during their enrolment in the IDEAL project (for specific clinical signs see Table 3.6). Each gross category of clinical disorder was used as the response variable in GLMMs with the serostatus of all three viruses was included in the model as explanatory variables. GLMMs were only carried out on disorders which affected more than 20 individuals. N= Number of calves with the gross category of clinical disorder; N_{SN} = Number of seronegative calves; N_{SP} = Number of seropositive calves.

Gross category of disorder observed	N	IBR					PIV3					BVDV				
		N_{SN}	N_{SP}	OR	95% CI	P value	N_{SN}	N_{SP}	OR	95% CI	P value	N_{SN}	N_{SP}	OR	95% CI	P value
Feeding/drinking problems	31	29	2	0.27	0.06-1.20	0.086	28	3	0.56	0.16-2.00	0.371	26	5	1.35	0.48-3.85	0.569
Fever	164	142	22	0.42	0.23-0.75	0.003	136	28	0.88	0.49-1.60	0.673	134	30	1.20	0.67-2.15	0.545
Gastrointestinal problems	32	31	1	0.12	0.02-0.89	0.039	27	5	1.23	0.43-3.48	0.698	28	4	0.87	0.28-2.70	0.806
Ill Thrift	141	119	22	0.64	0.36-1.13	0.122	115	25	1.10	0.61-1.98	0.755	119	22	0.82	0.45-1.51	0.528
Nervous/behaviour problems	58	51	7	0.50	0.21-1.18	0.111	48	10	1.15	0.52-2.54	0.728	49	9	0.96	0.42-2.17	0.917
Problems involving swelling	46	40	6	0.66	0.25-1.74	0.407	41	5	0.41	0.13-1.23	0.112	39	7	0.90	0.34-2.37	0.828
Respiratory problems	42	36	6	0.66	0.25-1.75	0.409	34	8	0.95	0.38-2.43	0.922	35	7	0.86	0.33-2.27	0.760
Respiratory problems or cough and fever	39	34	5	0.54	0.19-1.54	0.250	31	8	1.08	0.42-2.77	0.873	32	7	0.97	0.37-2.58	0.957
Skin/coat problems	25	21	4	0.50	0.15-1.68	0.265	18	7	1.73	0.61-4.94	0.305	18	7	1.82	0.64-5.15	0.258
Unknown problems	166	141	25	0.57	0.33-0.98	0.044	135	30	1.17	0.67-2.06	0.586	139	27	0.90	0.51-1.59	0.720
Any type of clinical episode	206	178	28	0.43	0.25-0.73	0.002	169	36	1.08	0.61-1.89	0.793	172	34	0.98	0.56-1.71	0.938

Table 3.6 Number of calves and odd ratios for seropositive calves which experienced the following clinical signs during their enrolment in the IDEAL project. Each clinical sign was used as the response variable in GLMMs with the serostatus of all three viruses was included in the model as explanatory variables. GLMMs were only carried out on clinical signs which affected more than 20 individuals. N= Number of calves with the gross category of clinical disorder; N_{SN} = Number of seronegative calves; N_{SP} = Number of seropositive calves.

Main clinical signs	N	IBR					PIV3					BVDV				
		N_{SN}	N_{SP}	OR	95% CI	P value	N_{SN}	N_{SP}	OR	95% CI	P value	N_{SN}	N_{SP}	OR	95% CI	P value
Decreased appetite	27	26	1	0.16	0.02-1.20	0.074	25	2	0.42	0.09-1.94	0.266	23	4	1.34	0.43-4.23	0.615
Fever: mild (39.5°C-40.5°C)	105	92	13	0.42	0.21-0.84	0.014	87	18	0.92	0.48-1.78	0.812	85	20	1.28	0.68-2.42	0.444
Fever: high (>40.5°C)	81	68	13	0.65	0.33-1.29	0.223	66	15	1.01	0.50-2.02	0.987	65	16	1.23	0.62-2.43	0.552
Lethargy (behaviour)	55	49	6	0.45	0.18-1.12	0.087	46	9	1.11	0.49-2.51	0.802	47	8	0.91	0.39-2.13	0.824
Lymph nodes	44	38	6	0.69	0.26-1.83	0.459	39	5	0.42	0.14-1.27	0.123	37	7	0.93	0.35-2.47	0.885
Rough/staring coat	130	108	22	0.72	0.41-1.27	0.251	104	25	1.23	0.68-2.22	0.487	108	22	0.92	0.50-1.68	0.781
Shallow or rapid breathing	30	28	2	0.29	0.06-1.29	0.104	26	4	0.68	0.21-2.26	0.530	26	4	0.86	0.26-2.81	0.801
Weight loss/loss of condition	66	61	5	0.33	0.12-0.86	0.024	61	5	0.41	0.15-1.12	0.083	60	6	0.58	0.23-1.48	0.257
Unknown	169	144	25	0.56	0.32-0.96	0.035	138	30	1.13	0.64-1.99	0.677	142	27	0.87	0.49-1.54	0.637

Table 3.7 Most parsimonious model showing the association between ever experiencing a clinical episode and exposure to different parasites.

Parasite	Prevalence (%)	OR	95% CI	P value
Seropositive for IBR	20.00	0.42	0.25-0.69	0.001
Seropositive for *Theileria parva*	85.27	2.77	1.50-5.11	0.001

Table 3.8 a) The effect of virus serostatus on whether or not the calf experienced any type of clinical episode in each of the three age classes and (b) the effect of exposure to other parasites on age of experiencing a clinical episode.

Explanatory Parasite	Experienced a clinical episode at less than 16 weeks old			Experienced a clinical episode between 16-36 weeks old			Experienced a clinical episode at greater than 36 weeks old		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
a)									
IBR	0.34	0.12-0.92	0.034	1.29	0.71-2.35	0.401	0.25	0.12-0.54	<0.001
PIV3	1.01	0.44-2.30	0.981	1.13	0.58-2.20	0.714	0.96	0.51-1.81	0.899
BVDV	1.38	0.63-3.02	0.427	1.20	0.62-2.33	0.584	0.87	0.46-1.67	0.682
b)									
IBR	0.31	0.11-0.87	0.026	1.19	0.63-2.26	0.592	0.18	0.08-0.41	<0.001
PIV3	0.95	0.40-2.22	0.897	1.09	0.54-2.18	0.815	0.97	0.49-1.93	0.932
BVDV	1.62	0.70-3.73	0.262	1.14	0.57-2.27	0.710	0.94	0.47-1.87	0.863
Cooperia spp.	3.76	1.37-10.34	0.010	0.57	0.18-1.77	0.332	1.81	0.76-4.30	0.178
Median strongyle EPG	0.87	0.52-1.47	0.614	1.33	0.85-2.08	0.217	1.84	1.16-2.91	0.010
Moniezia spp.	1.64	0.77-3.52	0.203	0.44	0.20-0.98	0.044	0.64	0.33-1.23	0.178
Microfilaria spp.	1.99	0.24-16.34	0.524	8.45	1.30-54.97	0.026	0.51	0.06-4.73	0.556
Strongyloides spp.	0.94	0.48-1.85	0.856	1.31	0.77-2.24	0.315	2.07	1.27-3.37	0.004
Seropositive for *Theileria parva*	1.59	0.52-4.84	0.415	2.80	1.05-7.45	0.039	2.11	1.00-4.46	0.051
Theileria taurotragi (RLB)	3.18	1.56-6.47	0.001	0.90	0.44-1.86	0.781	0.66	0.33-1.33	0.246

3.3.5 Genotype, phenotype at birth and infection risk

The relationship between IBR and ever experiencing a clinical episode is potentially confounded by the effect of recruitment weight. Heavier calves are marginally less likely to experience a clinical episode then lighter calves: for each 1kg increase in recruitment weight calves have 0.96 times the odds of experiencing a clinical episode (OR=0.96, 95% CI=0.91-1.01, P=0.093, Table 3.4). The same pattern of association was observed with the number of clinical episodes experienced by a calf (Table 3.4; see also Jennings (2013) for more details on clinical episodes within this population). However, the effect of IBR on experiencing any type of clinical episode was not removed when recruitment weight was fitted in the model (Table 3.4).

Calves which were heavier at recruitment were at increased risk of being IBR seropositive at 51 weeks of age (OR=1.11, 95% CI=1.04-1.19, P=0.002, Table 3.9). Yet, recruitment weight has no effect on the serostatus of the other two viruses or on the association observed between the three viruses (Table 3.9). Furthermore, the effect of recruitment weight on IBR serostatus was not confounded by calf sex, level of European taurine introgression or heterozygosity or the girth of the calf's dam at recruitment (Table 3.10). This is despite the observations that male calves are heavier than female calves at recruitment (male mean recruitment weight=19.72, SE=0.24; female mean recruitment weight=18.66, SE=0.24) and that larger dams produce heavier calves (per 1 cm increase in dam girth at recruitment it is predicted that a calf experiences a 1.2 fold increase in its recruitment weight; OR=1.20, 95% CI=1.15-1.25, P<0.001, Table 3.11). The relationship between IBR, recruitment weight and ever experiencing a clinical episode is illustrated in Figure 3.5.

Table 3.9 Effect of recruitment weight on the association between virus serostatus. Recruitment weight was included as a continuous fixed effect in the virus-only association model.

Response Virus	Explanatory Variable											
	IBR			PIV3			BVDV			Recruitment Weight		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
IBR	-	-	-	2.88	1.57-5.28	0.001	2.79	1.53-5.09	0.001	1.11	1.04-1.19	0.002
PIV3	2.82	1.54-5.15	0.001	-	-	-	6.08	3.40-10.90	<0.001	0.97	0.90-1.05	0.446
BVDV	2.78	1.55-4.99	0.001	6.16	3.49-10.87	<0.001	-	-	-	1.01	0.94-1.09	0.720

Table 3.10 Effect of recruitment weight, sex and genotype on IBR serostatus. BVDV and PIV3 serostatus is included in all these models as a factor. Those variables in each section (separated by a horizontal line) are fixed effects in a single multivariable model.

Explanatory Variable	OR	95% CI	P value
Recruitment weight	1.11	1.04-1.19	0.002
PIV3 seropositive	2.88	1.57-5.28	0.001
BVDV seropositive	2.79	1.53-5.09	0.001
Recruitment weight	1.12	1.05-1.20	0.001
Calf sex (effect of being male)	0.95	0.57-1.58	0.853
Moderate European taurine introgression	0.61	0.29-1.28	0.190
Substantial European taurine introgression	0.59	0.18-1.98	0.393
PIV3 seropositive	2.89	1.57-5.34	0.001
BVDV seropositive	2.96	1.61-5.43	<0.001

Table 3.11 The effect of being male and the dam's girth at recruitment on the weight of the calf at recruitment (kg), when the calf is less than 1 week old.

Explanatory Variable	Estimate	SE	OR	95% CI	P value
Calf sex (effect of being male)	1.37	0.32	3.92	2.10-7.31	<0.001
Dam girth at recruitment (cm)	0.18	0.02	1.2	1.15-1.25	<0.001

Dam's girth at recruitment

Recruitment weight: Estimate=0.18, SE=0.02, OR=1.20, 95% CI=1.15-1.25, P<0.001

Calf sex (effect of being male)

Recruitment weight: Estimate=1.37, SE=0.32, OR=3.92, 95% CI=2.10-7.31, P<0.001

Recruitment weight

IBR: OR= 1.11, 95% CI=1.04-1.19, P=0.002

Ever experienced any type of clinical episode: OR=0.96, 95% CI=0.91-1.01, P=0.096

Number of clinical episodes experienced: Estimate=-0.05, SE=0.02, OR=0.95, 95% CI=0.92-0.99, P=0.013

Ever experienced any type of clinical episode

Other infections effect on ever experiencing a clinical episodes

Theileria parva (serology): OR=2.77, 95% CI=1.50-5.11, P=0.001

Infectious Bovine Rhinotracheitis

Clinical episode ever: OR=0.43, 95% CI=0.25-0.73, P=0.002

Number of clinical episodes: Estimate=-0.63, SE=0.20, OR=0.53, 95% CI=0.36-0.79, P=0.002

PIV3: OR=2.76, 95% CI=1.53-4.98, P=0.001

BVDV: OR=2.77, 95% CI=1.56-4.92, P=0.001

Bovine Viral Diarrhoea Virus

IBR: OR=2.77, 95% CI=1.54-5.00, P=0.001

PIV3: OR=5.87, 95% CI=3.40-10.46, P<0.001

Babesia bigemina: OR=2.95, 95% CI=1.79-4.85, P<0.001

Babesia bigemina

BVDV: OR=2.88, 95% CI=1.67-4.97, P<0.001

Bovine Parainfluenza Virus Type 3

IBR: OR=2.77, 95% CI=1.53-5.03, P=0.001

BVDV: OR=5.98, 95% CI=3.40-10.42, P<0.01

Figure 3.5 Schematic diagram showing the associations between IBR, PIV3, BVDV serostatus, recruitment weight and ever experiencing a clinical episode. Male calves or calves from larger dams have a higher recruitment weight. Heavier calves at recruitment are more likely to be IBR seropositive but they are less likely to experience a clinical episode. IBR is associated with PIV3 and BVDV, testing seropositive for any one of these viruses increases the risk of being seropositive for the other viruses. IBR seropositive calves are less likely to experience a clinical episode of any type; however, infection with *Theileria parva* increases the individual's risk of experiencing a clinical episode. Furthermore, *Babesia bigemina* seropositive calves are more likely to be seropositive for BVDV but *B. bigemina* serostatus is not associated with ever experiencing a clinical episode.

71

3.4   Discussion

This study aimed to describe the prevalence and association of viruses involved in the bovine respiratory disease complex (BRD) in Western Kenya. I have shown that IBR, PIV3 and BVDV all have an estimated seroprevalence of around 20%. The observed seroprevalence of IBR is within the range (16%-54%) that McDermott et al. (1997) estimated for three districts in Kenya (not including the Busia district) in 1991-1992. However the seroprevalence of the three viruses is lower than that observed in traditionally managed herds in Zambia, which range from 42%-76% (Ghirotti et al., 1991). All three viruses are transmitted via sections or aerosols, in addition, vertical and horizontal transmission of BVDV can also occur. Therefore one explanation for the difference observed is that the herd sizes in the Zambian study range from 20-100 cattle, compared to a median of 5 (range 1 to 131) here (Bronsvoort et al., 2013; Ghirotti et al., 1991). IBR, PIV3 and BVDV are observed at higher prevalences in larger herd sizes (Snowder et al., 2006) and they are more common in intensively farmed animals, where there is a high level of contact between individuals. Cattle in the IDEAL project are extensively farmed, with a median herd size of 5 and so the risk of contact between a susceptible individual with an infected or persistently infected individual is lower. Furthermore, the cattle in Ghirotti et al. (1991) included individuals aged 3 months to adults, whereas the IDEAL calves were aged 51 weeks old, so there may also be differences in age-related seroprevalence. In addition, no BVDV antigen positive calves were identified in this study, suggesting that there are no persistently infected individuals (Brock, 2003).

Cross-reactivity may occur between a virus and its related viruses. For example, BVDV is cross reactive to other pestiviruses such as Classical Swine Fever and Border Disease Virus of sheep; IBR is cross-reactive to four herpesviruses from other animals including goats and buffalo; and PIV3 cross-reacts with human strains of the virus (Coelingh et al., 1986; Handel et al., 2011; Lyaku et al., 1992). However, since none of the above list were reported from the study population, the majority of seropositivity was likely due to exposure to the virus of interest.

In accordance with numerous other epidemiological studies, but in a previously unstudied setting, this analysis has found that IBR, PIV3 and BVDV are associated (Durham and Hassard, 1990; Fulton et al., 2000; Martin and Bohac, 1986). Therefore, this chapter provides proof of principle that the use of generalised linear mixed model method to identify parasite-parasite associations works. In addition, the GLMM method has been suggested for use and utilised in other coinfection analysis e.g. Hellard et al. (2012); Telfer et al. (2010) and Fenton et al. (2014).

Inclusion of environmental confounders into the models quantifying the relationship between the serostatus of the three viruses had little effect on the association observed between them. Other studies have suggested that at the herd level the main risk factors for BRD are the production type, herd size, housing and management practices such as animal movement and hygiene (Gay and Barnouin, 2009). Risk factors for IBR include increased movement into the herd and distance to neighbouring farms, which increases the risk of infection through contacts with infected individuals (van Schaik et al., 1998). Risk factors for PIV3 and BVDV included age (Figueroa-Chávez et al., 2012). In addition, the presence of a persistently infected individual can increase BVDV risk (Mainar-Jaime et al., 2001). Since all the calves in this study were the same age when testing took place and sublocation was fitted as a random effect, this variation was removed from the study.

I did not find any associations between ever being infected with any of the other viruses, protozoas or helminths screened for in the IDEAL project (with the exception of *B. bigemina* - see below) and an increased risk of being seropositve for IBR, PIV3 or BVDV. Unlike intensive systems where associations between BRD and secondary bacterial infection are common, I did not detect an association of IBR, PIV3 or BVDV with any bacterial infections (Coetzer and Tustin, 2004; Griffin et al., 2010; Hotchkiss et al., 2010). There are several possible explanations for this: firstly, this study population is an extensive farming system, with fewer animals in close proximity, and thus the risk of infection is lower. In addition, the IDEAL cattle are not housed so they do not experience the respiratory distress associated with poorly ventilated housing where really acute severe respiratory outbreaks occur (Coetzer and Tustin, 2004). Furthermore, bacteriology using swabs was only carried out during clinical episodes in the IDEAL project, presumably some bacterial infections may have been missed if individuals had a sub-clinical episode (Bronsvoort et al., 2013). Moreover, it is likely that the bacteria species associated with BRD are ones which the IDEAL project did not test for as it required broncholaveolar lavage followed by culture in clinical cases to establish this connection. Lastly, the bacterial species in question e.g. *Pasteurella* spp. are ubiquitous and commensals (Hotchkiss et al., 2010).

B. bigemina is an immunosuppressive protozoa which is transmitted by *Boophilus* spp. ticks and causes the disease babesiosis. Babesiosis clinical signs include severe anaemia and haemoglobinuria (Coetzer and Tustin, 2004). *B. bigemina* seropositive individuals were at increased risk of being BVDV seropositive relative to *B. bigemina* seronegative calves and vice versa. Calves which were BVDV seropositive were more likely to be *B. bigemina* seropositive than those calves which were seronegative for BVDV. The relationship between *B. bigemina* and BVDV was not removed when

potential environmental confounders were included in the model, suggesting that it is not driven by common environmental factors. It could be possible that I am missing some underlying variable which is driving this association.

I am unaware of any studies investigating the relationship between *B. bigemina* and BVDV in the literature. However, previous research on a related *Babesia* species, *B. divergens*, found that BVDV persistently infected cattle which were inoculated with *B. divergens* did not experience any more severe clinical symptoms of babesiosis than individuals free of BVDV (Brun-Hansen et al., 1998). This research was carried out to investigate the effect of the differences in haematological parameters between cattle with persistent BVDV and non-infected cattle following infection with *B. divergens* as both parasites affect blood parameters which gives them the potential to interact (Brun-Hansen et al., 1998).

Since the order of *B. bigemina* and BVDV infection is unknown, it is not possible to determine the causality. BVDV could be associated with *B. bigemina* because BVDV is immunosuppressive, and so it may increase risk of infection (Roth et al., 1986). Yet it is unclear as to why only seroconversion to *B. bigemina* is affected by BVDV, if the action was through immunosuppression then I would expect that other parasites would also be affected by BVDV. More studies are needed to explain the mechanism of association between *B. bigemina* and BVDV.

Another difference in my findings and studies of intensive systems is that coinfection with IBR, PIV3 and BVDV does not affect the growth (ADWG) of a calf. Studies of feedlot and beef cattle have shown that individuals with bovine respiratory disease complex exhibit a decreased growth rate and have lower weight gains (Gardner et al., 1999; Martin et al., 1999; Schneider et al., 2009). This difference may exist because there are other factors that are having a more limiting effect on growth in the study system, such as the lack of high energy food in comparison to that available in intensive systems in addition to parasitism, and so being seropositive for these viruses has little effect on the ADWG.

I found no association between PIV3 and BVDV with clinical episodes. Surprisingly, I observed the opposite result to what might be expected between IBR serostatus and clinical episodes: IBR seropositive calves were less likely to experience any type of clinical episode and overall had fewer clinical episodes then IBR seronegative calves. In addition, this chapter shows that heavier calves are more likely to be IBR seropositive and have fewer clinical episodes; this is one possible explanation as to why IBR seropositive calves were less likely to experience clinical episodes then IBR seronegative calves. If it is assumed that recruitment weight is an indicator of condition, then this finding suggest the hypothesis that heavier calves

are in a better condition to fight infection and so seroconvert to IBR. Meanwhile, as heavier individuals are in better condition it also means that they experience fewer clinical episodes as they can mount a immune response to the infection before it results in clinical illness. Yet the association between IBR and clinical episodes remained after accounting for recruitment weight. Furthermore, the association between IBR and clinical episode remained after accounting for environmental variation although, like the association between BVDV and *B. bigemina*, it is possible that IBR could be sharing some unknown factor with clinical episodes which was not accounted for.

The association between IBR and clinical episode was robust to the inclusion of exposure to other parasites which cause severe clinical signs such as *Anaplasma marginale*, *B. bigemina* and *T. parva*. In addition, IBR was the only parasite to show a protective effect, however the reason for this difference is unclear.

Since IBR is known to be an immunosuppressive virus (Ghram et al., 1989) which causes clinical illness with clear clinical signs and similar results are not reported in the literature, it is possible that selection bias is occurring here as only calves which survived to 51 weeks of age were included in the study and dead calves could be systematically different from alive ones. Therefore, similar to Berkson's (1946) paradox of selecting cases and controls from a hospital population, where the results are only attributable to hospital populations rather than the community at risk, the relationship maybe conditional on using data from calves which survived until 51 weeks of age; if the dead calves had been analysed then the outcome may be different. By including dead calves in the analysis it may be possible to establish that IBR is a risk factor for clinical episodes, as death is the most extreme outcome of a clinical episode. However, it should be noted that none of the calves in the IDEAL project died as a result of IBR infection as far as could be determined from *post mortem*. Nonetheless, IBR may have been a co-factor in some of the deaths, although this was not tested for. More information is needed on the serostatus of all calves to evaluate the effects of IBR on clinical episodes in the whole of the IDEAL study population.

IBR seropositive calves were less likely to experience a clinical episode when they were less than 16 weeks old or greater than 36 weeks old, but there was no evidence for an effect in middle-aged calves (16-36 weeks old). This suggests that there may be different explanations between the association between IBR and CE in young, middle aged and old calves. In young calves, maternally derived antibodies can confer immunity to IBR which wanes by seven months old in Friesian cattle breeds (Bradshaw and Edwards, 1996). Therefore, young calves are protected by maternal immunity and so they are less likely to experience clinical illness as a result. In contrast, the IDEAL study shows that clinical episodes most commonly occur in

middle aged calves (Appendix D.7), this maybe a result of waning maternal immunity and increased exposure to parasites in general (not specifically IBR) from weaning (Bronsvoort et al., 2013). In the oldest age group, calves are less likely to experience clinical episodes at an older age if it is assumed that it is the first exposure to IBR which makes an individual ill. Therefore, on the second exposure to IBR they are less likely to experience a clinical episode. This idea assumes that the force of infection is high so that calves become infected at an early age and that seroconversion is short-lived, so that in the old age group I am observing the results of the second exposure to IBR. However contradictory evidence outlined at the end of this chapter suggest that this is not the case.

Maternal antibodies may also affect the IBR serostatus of calves at 51 weeks of age. Previous studies have shown that calves with early exposure to IBR, whilst maternal antibodies are present, are at risk of becoming seronegative latent carriers once the maternal antibodies have declined because the virus can replicate in the presence of passive immunity (Lemaire et al., 2000; Levings and Roth, 2013). This could cause me to underestimate the seroprevalence of IBR and its effect. Therefore, more testing is needed to separate out the effect of maternal immunity and to investigate the age-specific changes in IBR infection risk on clinical episodes.

The association between experiencing any type of clinical episode and IBR may also be due to the large number of clinical episodes of unknown aetiology. If it were easier to define the type of clinical episode a calf experienced then it may be possible to identify the relationship between IBR and specific kinds of clinical episode. This relationship may also be dependent on the strain of IBR virus circulating within the population (Kaashoek et al., 1996). It is known that some strains of IBR virus are less virulent than others and that low virulent strains have a poor ability to induce clinical signs, therefore further work is also needed to identify the strain of virus circulating within this region (Kaashoek et al., 1996; Muylkens et al., 2007).

Lastly it is possible that the relationship between IBR and clinical episode could be a type 1 error. However, given that (a) the P values is low (P=0.002) and (b) that the results is replicated in the two tests, splitting by age, there is little chance of it being a type 1 error. Therefore some other as yet unidentified age-related factor such as changes in the immune response or exposure to other parasites as the calf ages may be causing the association between IBR and clinical episodes.

The use of serology to infer infection history from antibodies in the blood needs to be treated with some caution, as test sensitivity and specificity, cross-reactivity and cut-off thresholds, may lead to false classification of seronegative and seropositive individuals (Gilbert et al., 2013). Since individuals in this study were only tested at 51

weeks of age some misclassification may occur which could cause spurious results to occur if they were bias. For example misclassification may occur if an individual was infected with a virus early on in the study period but have lost the antibody response by the end of the study period, in which case, the isotype that the ELISA test identifies is important, as isotypes decay at different rates. The ELISA tests used to produce the results for this chapter identify IgG, which is secreted later in the immune response and persist longer in the circulation then other isotypes such as IgM which is the first antibody secreted in response to infection (Gilbert et al., 2013). A loss of antibody response would cause me to underestimate the seroprevalence and the effect of the virus. In contrast, calves may be exposed to the viruses multiple times during the study period and this could affect the antibody levels observed, as the more often a calf is infected with a parasite then the more likely it is for a calf to be classified as seropositive. However, both of these points are likely to have little effect on the results observed in this chapter as the duration of immunity to IBR, PIV3 and BVDV is greater than 6 months following vaccination with attenuated strains of IBR, PIV3 and BVDV (Peters et al., 2004). In addition, a study by Van der Poel et al. (1995) has found that antibodies to live IBR vaccine persist for up to 30 months after vaccination.

To conclude, this study shows that the viruses IBR, PIV3 and BVDV are co-circulating in East African shorthorn zebu calves in Western Kenya. I identified positive associations occurring between IBR, PIV3 and BVDV in an unstudied setting. Further studies are needed to identify the long-term impact of these viruses on cattle productivity and their interactions with other parasites, given that I have identified previously undiscovered and unexpected associations between BVDV and *B. bigemina* and between IBR and clinical episodes.

Chapter 4

Variation and covariation in strongyle infection in East African shorthorn zebu calves

This chapter has been published in Callaby et al. (2015) and has been reproduced with the publishers permission. See Appendix B for a copy of the published manuscript.

Summary

Parasite burden varies widely between individuals within a population, and can covary with multiple aspects of individual phenotype. Here I investigate the sources of variation in faecal strongyle eggs counts, and its association with body weight and a suite of haematological measures, in a cohort of indigenous zebu calves in Western Kenya, using relatedness matrices reconstructed from SNP genotypes. Strongyle egg count was heritable (h^2=23.9%, SE=11.8%) and I also found heritability of white blood cell counts (h^2=27.6%, SE=10.6%). All the traits investigated showed negative phenotypic covariances with strongyle egg count throughout the first year: high worm counts were associated with low values of white blood cell count, red blood cell count, total serum protein, and absolute eosinophil count. Furthermore, calf body weight at one week old was a significant predictor of strongyle faecal egg count at 16-51 weeks, with smaller calves having a higher strongyle egg count later in life. My results indicate a genetic basis to strongyle faecal egg count in this population, and also reveal consistently strong negative associations between strongyle infection and other important aspects of the multivariate phenotype.

4.1   Introduction

Gastrointestinal parasite infections of livestock are responsible for large economic losses in pastoral systems (Keyyu et al., 2003). They reduce weight gain and fertility, and may even cause direct losses through mortality (Wymann et al., 2008). Reduction of gastrointestinal parasite infections would therefore improve animal health and remove some of the constraints on livestock enterprises in developing countries, thereby reducing poverty (Perry and Sones, 2007). However, management of parasite infection requires an understanding of the causes of variation in parasite burdens, variation which can be substantial even between individuals within a population. For example, in indigenous East African Shorthorn Zebu (*Bos indicus*, EASZ) calves in Western Kenya, most individuals experience an apparently low level of strongyle worm infection, whilst others experience a high worm burden and suffer severe consequences such as death (Thumbi et al., 2013a). In this chapter, I explore possible causes of this variation, and quantify its covariation with other variables.

Strongyles are a group of nematode gut worms which produce morphologically similar eggs and include, the species *Haemonchus placei*, *Trichostrongylus axei* and *Oesophagostomum radiatum* (Urquhart et al., 1996). The most common method used to estimate worm burden is a count of the number of strongyle eggs per gramme of faeces (EPG), a non-invasive, relatively easily-measured variable. It has been shown that faecal egg counts can be used index of parasite burden in Australian cattle, although the relationship between the two may not be exactly linear (Bryan and Kerr, 1989). Variation in strongyle faecal egg count can be due to a variation in susceptibility, resistance, tolerance or exposure to infection by strongyle worms. Evidence from other domestic ungulates suggests that variation in strongyle EPG frequently has a heritable genetic basis: for example, strongyle faecal egg count (FEC) has a heritability of 18% (95% CI =10-25%) in West African N'Dama cattle (Zinsstag et al., 2000), and the heritability is approximately 30% in many other cattle breeds (Leighton et al., 1989; Stear et al., 1990, 1988). Similarly, strongyle EPG in Scottish Blackface sheep lambs has a heritability of 32% (Riggio et al., 2013); see also Beraldi et al. (2007) and Crawford et al. (2006).

In addition to additive genetic effects, there may also be consistent environmental-based causes of variation in parasite burden between individuals. These 'permanent environmental effects' comprise all variance of non-(additive) genetic origin that persist throughout an individual's life-time, and so for example may include long-running effects of maternal environment or of how an individual was raised and housed: for instance, in a feral Soay sheep population, lambs born as twins or born to very young or old mothers have higher parasite burdens than those born

as singletons or to prime-age mothers (Hayward et al., 2010). Stear et al. (1996) also found higher parasite burdens in Scottish Blackface sheep twins in comparison to singletons. The physical environment that an individual resides in will also be important for determining its exposure to a particular parasite, which in turn can affect the burden of infection observed (e.g. Batchelor et al. (2009)). Finally, there may be variation between measures made on an individual at different time points, due to, for example, effects of ageing, immediate climatic effects, or simply stochastic variation and measurement error.

Variation in parasite burden may also have implications for the expression of other important traits, especially if parasite resistance is costly and may therefore be traded off against investment in other traits (Norris and Evans, 2000). Such associations can be quantified within individuals by looking at the covariation of parasite burden and other traits, for example morphological variables such as growth rates or weight, or physiological variables such as haematological parameters, to test for any costs associated with high parasite burdens (e.g. Coltman et al. (2001)). In particular, one of the strongyle species, *Haemonchus placei*, is an important cause of anaemia in ruminants (Kaufmann et al., 1992). Conradie van Wyk et al. (2014) and Vanimisetti et al. (2004) have shown negative correlations between parasite burden and various haematological parameters in EASZ and sheep. Finally, it is possible that an individual's phenotype at birth may influence their infection risk later in life. For example, in humans, babies that have a lower birth weight are more likely to develop lower respiratory tract infections when they are coinfected with hand, foot and mouth disease (Lu et al., 2013). Likewise, Read et al. (1994) showed there is a higher risk of childhood infectious disease mortality in lower birth weight babies than heavier ones.

Traditionally, pedigree information has been used to estimate quantitative genetic parameters such as the heritability of a trait (Falconer and Mackay, 1996). More recently, the development of high density SNP beadchips means that novel alternative approaches can be used without reference to pedigree records (Visscher et al., 2014; Yang et al., 2010). This has reduced previous constraints faced during estimation of heritability in wild populations due to the lack of accuracy and completeness of the pedigree (Pemberton, 2008). Bèrènos et al. (2014) compared heritability estimates produced from using both pedigrees and SNPs from related Soay sheep and demonstrated that heritability estimates obtained from dense SNP data are in correspondence with pedigree estimates.

The Infectious Diseases of East African Livestock (IDEAL) project (Bronsvoort et al., 2013) provides a unique opportunity to study natural variation and covariation in strongyle EPG in indigenous EASZ from Western Kenya. Cattle in this region are

80

minimally managed and there is very limited use of vaccination or other preventative measures against infectious diseases. Therefore the study population is similar to a wild population in that, unlike other estimates of genetic variation in FEC in domestic populations (e.g. Bishop et al. (1996)), animals have not been treated for anthelminthics (those individuals which were treated with anthelminthics were retrospectively removed from the cohort as part of the IDEAL study design); variation therefore reflects natural diversity in parasite burden. Calves were enrolled in the study at birth and their infectious disease burden, haematological profiles and growth were tracked for the first year of life (Bronsvoort et al., 2013; Conradie van Wyk et al., 2012). Strongyle worm burdens (assessed via EPG) have a major impact upon the calves in the study population: for example, an increase in strongyle EPG by a count of 1000 eggs is associated with a 3.3% reduction in weight gain over the first year (Thumbi et al., 2013b), and an increase in the hazard of death by 1.5 (95% CI =1.4-1.7, P<0.001; Thumbi et al. (2013a)). Moreover, genome-wide genetic information is available in the form of SNPs as each calf enrolled in the IDEAL project was genotyped with a 50K Illumina$^{®}$ BovineSNP50 beadchip (Mbole-Kariuki et al., 2014; Murray et al., 2013), providing the opportunity to exploit this information to estimate a relatedness matrix and thereby derive estimates of variance components, including additive genetic variance of different traits.

The aim of this study is to dissect the potential genetic and non-genetic sources of between- and within-individual level variation in strongyle faecal egg count. I present a multivariate analysis of associations between strongyle faecal egg count (EPG), body size, and a suite of haematological measures. I quantified the variance components of five physiological traits and their covariation with strongyle EPG. Finally I investigated whether the characteristics of newborn calves could be used to predict subsequent EPG levels, by looking at the association between weight at birth and strongyle EPG later in life.

4.2 Material and Methods

4.2.1 Study population

Five hundred and forty-eight free-grazing indigenous East African Shorthorn Zebu calves in Western Kenya were selected using a stratified two-stage random cluster study design. In the first stage, 20 sublocations (the smallest administrative unit in Kenya) were selected from 5 agro-ecological zones, across an area of roughly 45 x 90km. Around twenty-eight 3 – 7 day old calves were recruited from each

sublocation, all from different mothers and different farms; see Bronsvoort et al. (2013) and Appendix A for a detailed description of the study design. Recruited calves were followed for their first year of life. They were visited every 5 weeks for a clinical examination at which they were weighed and at which blood and faecal samples were taken for parasite identification and haematological profiling. A total of 446 calves that survived to 51 weeks of age (and had passed the SNP quality control checks, see SNP quality control section below) were included in this analysis, giving a total of 4727 observations and an average of 10.6 visits per calf.

4.2.2 Data collection

The McMaster counting technique (Hansen and Perry, 1994) was performed on the faecal samples from each visit to each calf to quantify the number of strongyle eggs per gramme of faeces (EPG) present. The other traits considered in this study were: white blood cell count (WBC), red blood cell count (RBC), total serum protein (TSP), absolute eosinophil count (EO), and body weight. Blood cell analysis was automatically performed using the pocH-100iV Diff (Sysmex$^®$ Europe GMBG); see Conradie van Wyk et al. (2012) for more details. Haematological profiles were produced for the total white cell count (WBC) and red blood cell count (RBC). Total serum protein (TSP) was determined using a refractometer and absolute eosinophil count (EO) was quantified by differential counts from thin EDTA blood smears stained with Diff Quick. Previous studies have shown that higher RBC and heavier body weights are associated with lower faecal egg counts (Conradie van Wyk et al., 2014; Thumbi et al., 2013b; Vanimisetti et al., 2004). All laboratory analysis was completed at IDEAL project field laboratory in Busia, Kenya.

Calves were weighed (in kilogrammes, measured to the nearest 500g) at recruitment, then again every five weeks until 31 weeks of age, and once again at a last visit at 51 weeks. The number of observations for each trait is presented in Table 4.1.

4.2.3 SNP quality control and construction of the relationship matrix

All calves were genotyped using a 50K Illumina$^®$ BovineSNP50 beadchip v1. The beadchip contained 55,777 SNPs before quality control, spread evenly throughout the genome with an average of 1895 SNPs on each autosome and 1362 SNPs on the X chromosome (Murray et al., 2013). Quality control was applied to all SNP data prior to analysis using GenABEL (Aulchenko et al., 2007), with the following criteria: SNP call rate cut-off of 0.9; individual call rate of 0.9 and an identity by state

(IBS) threshold cut-off of 0.9. The IBS threshold means that if a pair of individuals is estimated to be exceptionally highly related (e.g. identical twins) then one of the individuals would be removed. The minimum minor allele frequency for SNPs was set to 0.005, to include all SNPs with a minor allele count of 5 or more. Any X chromosome genotypes that were inconsistent with the phenotype were removed. This quality control resulted in 42,119 autosomal and X markers (41419 autosomal markers plus 700 X markers) and 446 calves for analysis. I explored the effect of varying the quality control parameters and the number of SNPs included in the IBS matrix on the resulting estimates of heritability; details are given in Table E.1 and Table E.2; in general, estimates of heritability for strongyle EPG increased with increasing marker density. Plots of the distribution of the minor allele frequencies at SNP markers and the association between linkage disequilibrium and the distance between pairs of SNPs are presented in Figure E.1.

All SNPs and calves which passed the quality control checks were then used to construct an identity-by-state matrix in GenABEL (Aulchenko et al., 2007) using the allele frequency weighted option, giving the kinship coefficients for use in the variance component and heritability analyses described below. The average genomic estimate of kinship between calves as given by the IBS matrix ranged from -0.02 to 0.24. Three pairs of calves had a genomic estimate of relatedness greater than 0.2 and 6 pairs of calves had a genomic estimate of relatedness between 0.15 and 0.2.

Approximately 20% of the calves in the IDEAL study cohort were shown to have some level of introgression from European taurine cattle, although calves that were first generation offspring from European taurine were explicitly excluded from the study (Bronsvoort et al., 2013; Mbole-Kariuki et al., 2014). These calves with lower levels of European taurine introgression were included in my study since the aim of the study was to describe the components of variation in strongyle faecal egg count in the population. The effect of excluding the introgressed calves on the heritability estimates is presented in Table E.3.

4.2.4 Statistical analysis

Trait distributions

In order to account for the distribution of the strongyle EPG counts, I used generalised linear mixed models with a negative binomial distribution and log link function; as observations of Strongyle EPG were in multiples of 50, they were first divided by 50 so that the data resemble typical count data. Note that estimates of variance components for EPG are therefore on a latent scale rather than on the original data

scale (Nakagawa and Schielzeth, 2010). All other variables were analysed assuming Gaussian distributions. Body weight was first transformed to $\log_{10}$ (weight) and EO to $\log_{10}$ (EO + 1) to account for their slightly-skewed distribution.

A significant increase in RBC was found between the calves aged 1 versus aged 6 weeks old, followed by a general decreasing trend in calves aged 6-51 weeks (Figure E.2 and Conradie van Wyk et al. (2012)). I therefore focused my analysis of RBC on calves aged 6-51 weeks old. Removal of the records from one week old calves did not affect the direction of associations observed and only resulted in small changes to the variance and heritability estimates.

Random effects and variance components estimation

I used an animal model to estimate the variance components of each trait (Kruuk, 2004; Lynch and Walsh, 1998). Animal models are a form of mixed model, with fixed and random effects, that can break phenotypic variation down into the different components via a model of the form:

$$y = Xb + Za + Pc + Sd + e \qquad (4.1)$$

where y is the phenotype of interest and b is a vector of fixed effects that are unknown constants that affect the mean of the distribution. The random effects, which determine the variance of the trait, were additive genetic (a), permanent environment (c), sublocation (d) and residual effects (e). In particular, a is a vector associated with the identity-by-state matrix (see Powell et al. (2010) and Visscher et al. (2008) for more details on calculating heritabilities using identity-by-state matrices rather than pedigrees) and is derived from the principle that if a trait has a high degree of genetic variance relative to its other components of variance, pairs of relatives will have high phenotypic similarity. X, Z, P and S are all design matrixes corresponding to the appropriate fixed or random effects. Permanent environmental effects are measurable because of the repeated observations on the same individual; this between-individual variation is likely to result from long-term environmental or non-additive genetic effects, and in this case will probably incorporate most of any maternal effects (Kruuk and Hadfield, 2007). The total phenotypic variance (V_P) for a trait was therefore broken down into the additive genetic variance (V_A), permanent environmental variance (V_{PE}), sublocation variance (V_{SL}) and residual variance (V_R):

$$V_P = V_A + V_{PE} + V_{SL} + V_R \qquad (4.2)$$

The narrow-sense heritability of a trait (h^2) is defined as the proportion of phenotypic variance (V_P) explained by the additive genetic variance (V_A), $h^2 = V_A/V_P$. It describes the extent to which differences between individuals are determined by additive genetic effects (Falconer and Mackay, 1996). I also report the repeatability (r^2) of each trait, defined as the proportion of the phenotypic variance due to consistent differences between individuals and is given by the ratio of the between individual variance to the total variance, $r^2 = (V_A + V_{PE} + V_{SL})/V_P$.

The covariances between traits can be investigated using multivariate models. By extending the above approach of variance partitioning to multiple traits, and linking them through a covariance term in the random effects, I can ask how much of the phenotypic covariance (COV_P) between traits is due to covariance of the different random effects described above, e.g. covariance in the permanent environment effects (COV_{PE}).

All statistical analyses were carried out in ASReml version 3.0.5 (Gilmore et al., 2006).

Components of variation in strongyle EPG

Estimation of the components of variance of strongyle EPG at each visit indicated that there was insufficient statistical power to analyze measures at every visit separately. In order to overcome this, I used a univariate animal model fitted with a negative binomial distribution to estimate the heritability of strongyle EPG across all ages. Age (as a multi-level factor) was fitted as a fixed effect to account for changes across visits in mean EPG with age. Sex was also included in this model as a fixed effect and V_A, V_{PE} and V_{SL} were fitted as random effects. Unlike other studies which have estimated genetic variation in FEC in domestic animals (e.g. Bishop et al. (1996)), individuals in this study population have not been treated with anthelminthics, and so represent natural levels of variation. Repeated observations on individuals are therefore not necessarily independent assessments of resistance, because nematodes might persist between sample dates. However my mixed models account for the repeated measures structure of the data by fitting a permanent environment effect, defining the number of individuals as the appropriate number of independent observations (Kruuk and Hadfield, 2007).

The significance of V_A was evaluated by comparing the component estimate to the standard error, as it is not advisable to carry out likelihood ratio tests in generalized linear mixed models with negative binomial errors in ASReml (Gilmore et al., 2006). Finally, for comparison with previous studies which have analysed strongyle faecal egg counts assuming Gaussian errors (Beraldi et al., 2007; Bishop et al., 1996; Coltman

et al., 2001; Stear et al., 1990), I also present analyses of linear mixed models assuming a normal distribution of $\log_{10}$ (Strongyle EPG + 50) in the Appendix E.

Components of variation in physiological traits

The components of variance in the physiological traits were examined by constructing a univariate Gaussian repeated measures animal model for each trait. As above, age and sex were included as fixed effects, and V_A, V_{PE} and V_{SL} were fitted as random effects in all models. The significance of V_A for each trait was assessed with a likelihood ratio test comparing the full animal model to one in which the additive genetic variance was set to zero.

Associations between strongyle infection and physiological traits

I assessed associations between strongyle infection and the physiological traits (and body size) in three different ways, by: (i) testing whether infection affected mean levels of the physiological traits; (ii) testing whether size at birth predicted levels of strongyle infection later in life; and (iii) assessing components of covariance between all traits.

The effects of strongyle infection on the physiological traits were therefore first quantified by univariate animal models with the trait as the response variable and explanatory variables of age at visit, calf sex and strongyle EPG classified into two categories of 'high' and 'low' EPG. A 'high' strongyle EPG was defined as a value above the median strongyle EPG across all visits (200 EPG), and a 'low' strongyle EPG one below the median. This categorization was chosen to reflect the non-linearity in effect of strongyle EPG estimate of effect (Conradie van Wyk et al., 2014). All of the explanatory variables were coded as factors and V_A, V_{PE} and V_{SL} were fitted as random effects.

Secondly, I tested whether a calf's phenotype very early in life was an informative predictor of the index of infection burden, EPG, later in life, and specifically whether the calf's recruitment weight predicted strongyle EPG later in the first year of life. This was achieved by constructing a univariate animal model with a negative binomial distribution to evaluate the effect of calf weight at recruitment (when the calf is less than 1 week old) on strongyle EPG in older calves (aged 16-51 weeks, following a plateau in median strongyle EPG after 16 weeks). This model includes calf age and sex as fixed effects and V_A, V_{PE} and V_{SL} as random effects. The magnitude and directionality of association between the trait and strongyle EPG is given by the parameter estimate, whilst its significance was assessed using Wald F statistics.

Thirdly, the covariances and correlations between strongyle EPG and the physiological traits were assessed by constructing a multivariate model of all six traits (strongyle EPG, WBC, RBC, TSP, EO and Weight), using measures across the whole year for all traits. Calf age and sex were included as fixed effects and Strongyle EPG was fitted with a negative binomial error distribution, whilst the other traits were fitted with a Gaussian error distribution. The resulting six-trait multivariate model was computationally much more demanding than the univariate models described above, due to the much greater number of parameters (an extra 80 parameters) being estimated. Therefore, I had to take several steps to facilitate reliable convergence. Firstly, I was unable to separate between-individual differences into genetic versus permanent environment effects, so I restricted the analysis to separating between- versus within-individual-level variances and covariances, omitting the genetic relationship matrix from the model. By only including calf identity as a random effect, I obtained estimates of the individual- (phenotypic-) level variance, which reflects consistent differences between individuals; similarly, the model partitions the total phenotypic covariance between two traits into that due to between-individual versus within-individual (residual) components. Secondly, I was unable to fit sublocation as a random effect in the models, so it was omitted from the multivariate analysis. Note however that sublocation was never significant in any of the univariate models Table 4.1, and its effects will be included in the permanent environment effect (co)variance. Since likelihood ratio tests are not advisable with generalized linear mixed models in ASReml (Gilmore et al., 2006), significance of estimates was assessed based on their magnitude relative to the standard error.

4.3 Results

4.3.1 Summary statistics

Out of the 4032 visits with faecal samples taken from the 446 live calves that passed the genetic quality control checks, strongyle eggs were detected in 3071 (76.2%) visits using the McMasters technique. The overall median number of strongyle eggs per gramme of faeces was 200 EPG (range: 0-12250 EPG). All calves were infected with strongyle eggs at some point during their 51 weeks of inclusion in the study. Infection rates increased up to 16 weeks of age, and then levelled off afterwards, with an average of 89.8% of visits showing non-zero EPG between the ages of 16-51 weeks, and a median strongyle EPG of 300 EPG (range: 0-12250 EPG). The median strongyle EPG and the fraction of calves positive at each age are shown in Figure 4.1.
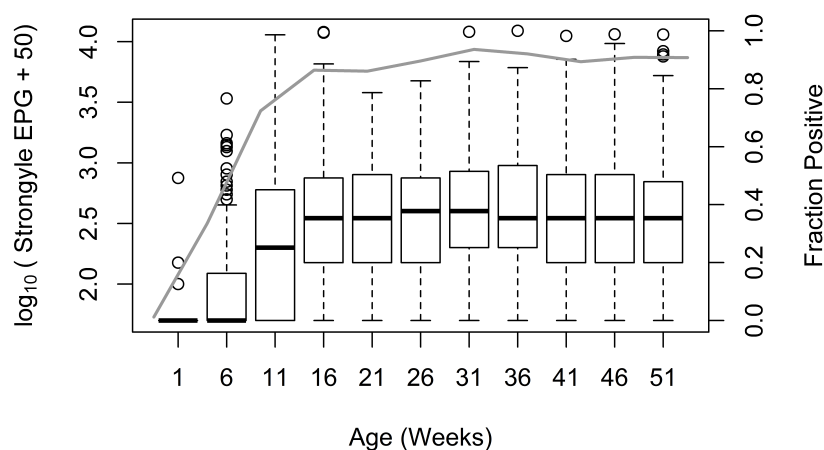
Figure 4.1 Distribution of strongyle EPG (box plots, left-hand axis) and the fraction of calves which tested positive at each age (right-hand axis). The black heavy solid lines in each box are the median EPG at each age group, the bottom and top of the box represent the 25th and 75th percentiles, respectively, and the whiskers represent 1.5 times the interquartile range. Points beyond the whiskers are outliers. Strongyle EPG is transformed as $\log_{10}$(strongyle EPG + 50). The solid grey line represents the fraction of tested calves positive for strongyle eggs at each age.

4.3.2 Components of variation in strongyle EPG

Additive genetic variance contributed the most (after residual variance) to the overall variance in strongyle EPG, resulting in heritable variation in Strongyle EPG in East African shorthorn zebu calves (h^2 = 23.9%, SE = 11.8%, Table 4.1). In contrast, the contribution of permanent environmental effects to the overall variance was relatively low (4.3%, SE = 11.5%, Table 4.1 and Figure 4.2). Strongyle EPG had a repeatability of 31.4% (SE = 2.2%). In addition, male calves had a higher strongyle EPG than female calves (effect estimate = 0.23, SE = 0.08, P value = 0.01).

Complete removal of the 'introgressed' calves from the study resulted in a lower heritability estimate and larger standard errors, whilst inclusion of the European taurine introgression as a fixed effect did not alter the heritability estimate (with ET introgressed calves included h^2 = 23.9%, SE = 11.8%, N calves= 446; with ET introgression included as a fixed effect, h^2 = 25.7%, SE = 11.9%, N calves = 446; with ET introgressed calves excluded h^2 = 13.3%, SE = 13.4%, N calves= 353; see Table E.2 and Table E.3).

For comparison of the negative binomial errors model with models assuming Gaussian errors, I present analyses of linear mixed models assuming a normal distribution of $\log_{10}$ (Strongyle EPG + 50) in Table E.4. Both methods produce similar estimates of heritability of strongyle EPG although, notably, the standard errors are much larger with the GLMM.
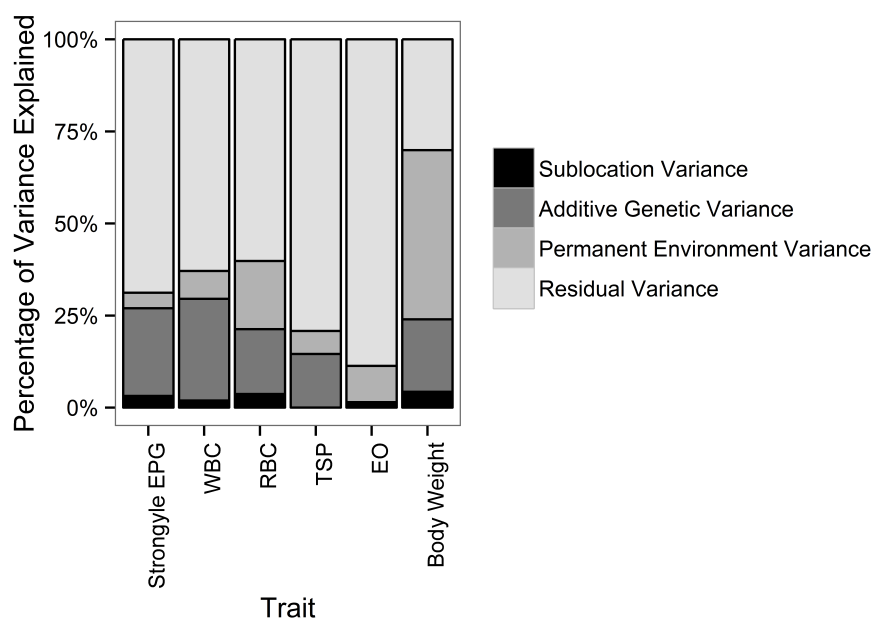
Figure 4.2 Percentage of variance explained by each component for all traits investigated. Details of the full model (including SEs on variance component estimates) are given in Table 4.1. Strongyle EPG uses an animal model fitted with a negative binomial distribution. WBC = White Blood Cell Count; RBC = Red Blood Cell Count; TSP = Total Serum Protein; EO = Transformed Absolute Eosinophil Count ($\log_{10}(EO + 1)$); Weight = Transformed Body Weight ($\log_{10}(Weight)$).

Table 4.1 Variance components ($\pm$ standard error) for all traits considered in univariate repeated measures models which include calf age and sex as fixed effects and V_A, V_{PE}, V_{SL} as a random effect. The proportion of total variance (V_P) explained by the permanent environment variance (V_{PE}) is also presented. The total number of calves for each trait is 446. WBC = White Blood Cell Count; RBC = Red Blood Cell Count; TSP = Total Serum Protein; EO = Transformed Absolute Eosinophil Count ($log_{10}(EO + 1)$); Weight = Transformed Body Weight ($log_{10}(Weight)$); V_{SL} = sublocation variance; V_A = additive genetic variance; V_{PE} = permanent environment variance; V_{RES} = residual variance; h^2 = heritability; V_{PE}/V_P (%) = Proportion of the total phenotypic variance explained by the permanent environment variance expressed as a percentage; r^2 = repeatability.

Trait	N	Mean ($\pm$ SD)	V_{SL}	V_A	V_{PE}	V_{RES}
Strongyle EPG	4032	542.72 ± 15.79	0.06 ± 0.03	0.45 ± 0.23	0.08 ± 0.22	1.30 ± 0.03
WBC	4693	11.08 ± 3.44	0.22 ± 0.15	3.15 ± 1.22	0.86 ± 1.17	7.17 ± 0.16
RBC	4281	8.79 ± 2.03	0.12 ± 0.06	0.57 ± 0.41	0.60 ± 0.04	1.95 ± 0.04
TSP	4721	7.98 ± 0.90	$2.56e^{-3} \pm 3.15e^{-3}$	0.07 ± 0.04	0.03 ± 0.04	0.38 ± 0.01
EO	3690	0.14 ± 0.14	$1.71e^{-4} \pm 1.12e^{-4}$	$9.44e^{-5} \pm 8.78e^{-4}$	$1.74e^{-3} \pm 9.14e^{-4}$	$1.57e^{-2} \pm 3.89e^{-4}$
Weight	3338	1.55 ± 0.18	$4.19e^{-4} \pm 2.41e^{-4}$	$1.91e^{-3} \pm 1.87e^{-3}$	$4.46e^{-3} \pm 1.87e^{-3}$	$2.92e^{-3} \pm 7.69e^{-5}$

Trait	h^2 (%)	V_{PE}/V_P (%)	r^2 (%)	Sex effect estimate
Strongyle EPG	23.92 ± 11.83	4.32 ± 11.53	31.44 ± 2.16	0.23 ± 0.08
WBC	27.63 ± 10.56	7.54 ± 10.28	37.17 ± 1.97	-0.15 ± 0.21
RBC	17.60 ± 12.46	18.50 ± 12.36	39.68 ± 2.10	-0.26 ± 0.11
TSP	14.50 ± 8.20	7.05 ± 8.09	22.08 ± 1.63	-0.04 ± 0.04
EO	0.53 ± 4.97	9.84 ± 5.15	11.34 ± 1.44	-0.01 ± 0.01
Weight	19.64 ± 19.22	45.93 ± 19.06	69.89 ± 1.67	0.02 ± 0.01

4.3.3 Components of variation in physiological traits

The age-related profiles for the physiological traits are shown in Figure E.2 (split according to whether the calf had high or low EPG at the time). WBC, EO and weight all increased with age, as expected. However RBC increased rapidly until 6 weeks old and then declined sharply. A decline from birth in TSP was observed until 21 weeks of age when TSP started to increase again. These distributions and the effect of coinfections on WBC are discussed in Conradie van Wyk et al. (2012) and Conradie van Wyk et al. (2014) respectively.

Estimates of the variance components and the heritability of each trait are shown in Table 4.1. WBC was the only physiological trait to show evidence for significant V_A (LRT: $\chi^2 = 8.8$, d.f. = 1, P = 0.003, Table 4.1; $h^2 = 27.6\%$, SE = 10.6%). There were large differences between traits in the proportion of the total variance V_P explained by each variance component: for example, permanent environment effects explained most (45.9%, SE = 19.1%) of the total variance in body weight, but only a relatively small proportion of the variance in the other parameters (7.1%-18.5%; Table 4.1). Weight had the highest repeatability of 69.9% (SE = 1.7%); repeatability otherwise ranged from 11.3% (SE = 1.4%) for EO to 37.2% (SE = 2.0%) for WBC.

4.3.4 Associations between strongyle infection and physiological traits

Effect of strongyle infection on physiological traits

I found significant effects of strongyle infection on all the physiological traits considered. The impact of strongyle EPG on every trait at each age is illustrated in Figure E.2 and quantified in Table 4.2. Table 4.2 shows that calves with a higher strongyle EPG at a given age tended to have a lower RBC, TSP, and EO than those with a lower strongyle EPG. Furthermore, calves with a high strongyle EPG were also lighter than those with a lower EPG (by -0.03 $\log_{10}$ (kg); SE = 0.01 on average). Similar results were observed when a continuous measure of EPG ($\log_{10}$ (StrongyleEPG + 50)) rather than a binary measure was used as an explanatory variable.

Table 4.2 The effect of a high or low strongyle EPG at a given age on the trait of interest, using univariate animal models. A high EPG is defined as being above the median Strongyle EPG whilst a low EPG is defined as being below the median strongyle EPG. The median is the overall median taken across all visits. The significance is given by the Wald F statistic. NA = Not applicable, as multiple factor level estimates are not reported.

Trait	White Blood Cell Count				Red Blood Cell Count			
	Estimate	SE	F	P value	Estimate	SE	F	P value
Intercept	9.676	0.217	8049.5	<0.001	10.56	0.124	10999.24	<0.001
High Strongyle EPG	-0.471	0.102	0.35	0.55	-0.362	0.054	236.11	<0.001
Sex (effect of being male)	-0.046	0.21	0.05	0.819	-0.224	0.111	4.04	0.047
Calf age (11 factor levels)	NA	NA	25.94	<0.001	NA	NA	184	<0.001

Trait	Total Serum Protein				Absolute Eosinophil Count			
	Estimate	SE	F	P value	Estimate	SE	F	P value
Intercept	9.69	0.041	200000	<0.001	0.055	0.009	1166.54	<0.001
High Strongyle EPG	-0.09	0.023	646.29	<0.001	-0.007	0.005	10.96	0.001
Sex (effect of being male)	-0.034	0.036	0.88	0.349	-0.009	0.006	2.22	0.140
Calf age (11 factor levels)	NA	NA	269.4	<0.001	NA	NA	37.15	<0.001

Trait	Weight			
	Estimate	SE	F	P value
Intercept	1.267	0.008	72274.08	<0.001
High Strongyle EPG	-0.015	0.003	3579.83	<0.001
Sex (effect of being male)	0.018	0.008	5.44	0.021
Calf age (11 factor levels)	NA	NA	2682.21	<0.001

Does weight at first visit predict strongyle infection in older calves?

Weight at the recruitment visit (when the calf was less than a week old) was significantly associated with later strongyle EPG: calves that were lighter at the first visit had a higher strongyle EPG when aged 16-51 weeks old than calves that were heavier (Table 4.3). As above, males also had higher levels of EPG.

Table 4.3 Association between strongyle EPG in older calves (aged 16-51 weeks old) and the calf's weight at the recruitment visit (calf aged <1 week), using a univariate animal model.

Variable	Estimate	SE	P value
Calf weight at recruitment (Kg)	-0.03	0.012	0.011
Calf sex (effect of being male)	0.276	0.087	0.006
Calf age (7 factor levels)	NA	NA	0.185

Covariances between strongyle EPG and physiological traits

The individual-level and residual covariances between strongyle EPG and the physiological traits of interest are shown in Table 4.4. All traits had a negative individual-level covariance with strongyle EPG whilst positive covariances were found amongst all the blood parameters and weight. This indicates that an increase in strongyle EPG was associated with a decrease in blood parameters and weight, whilst an increase in weight, etc. was associated with an increase in blood parameters and *vice versa*. Comparison of the between-individual *vs.* residual (within-individual) variance showed that both follow the same pattern, but that there were higher levels of between-individual than residual (within-individual) level correlations.

Table 4.4 Covariance/variance/correlation matrix for the between-individual and residual (within-individual) level variance between strongyle EPG and trait. Covariances are shown below the diagonal (in *italics*), the associated correlations above the diagonal and variances on the diagonal. Standard errors are in brackets. WBC = White Blood Cell Count; RBC = Red Blood Cell Count; TSP = Total Serum Protein; EO = Transformed Absolute Eosinophil Count ($\log_{10}(EO + 1)$); Weight = Transformed Body Weight ($\log_{10}(Weight)$).

Individual Level Variance	Strongyle EPG	WBC	RBC	TSP	EO	Body Weight
Strongyle EPG	*0.620 (0.065)*	-0.120 (0.068)	-0.386 (0.062)	-0.192 (0.075)	-0.140 (0.098)	-0.182 (0.060)
WBC	*-0.204 (0.118)*	*4.678 (0.426)*	0.133 (0.064)	0.284 (0.068)	0.468 (0.079)	0.280 (0.054)
RBC	*-0.355 (0.067)*	*0.336 (0.168)*	*1.365 (0.128)*	0.251 (0.069)	0.138 (0.092)	0.399 (0.050)
TSP	*-0.046 (0.019)*	*0.188 (0.049)*	*0.090 (0.027)*	*0.094 (0.011)*	0.030 (0.104)	0.273 (0.061)
EO	*-0.005 (0.004)*	*0.048 (0.010)*	*0.008 (0.005)*	*0.0004 (0.001)*	*0.002 (0.0004)*	0.368 (0.077)
Body Weight	*-0.013 (0.004)*	*0.053 (0.011)*	*0.041 (0.006)*	*0.007 (0.002)*	*0.002 (0.0003)*	*0.008 (0.001)*

Residual Level Variance	Strongyle EPG	WBC	RBC	TSP	EO	Body Weight
Strongyle EPG	*1.124 (0.041)*	-0.095 (0.025)	-0.140 (0.025)	-0.124 (0.025)	-0.022 (0.026)	-0.137 (0.025)
WBC	*-0.255 (0.069)*	*6.472 (0.239)*	0.345 (0.023)	0.152 (0.025)	0.246 (0.024)	0.126 (0.026)
RBC	*-0.214 (0.040)*	*1.262 (0.101)*	*2.065 (0.046)*	0.265 (0.024)	0.082 (0.026)	0.209 (0.025)
TSP	*-0.067 (0.014)*	*0.199 (0.034)*	*0.196 (0.020)*	*0.265 (0.010)*	0.032 (0.026)	0.182 (0.025)
EO	*-0.003 (0.003)*	*0.078 (0.009)*	*0.015 (0.005)*	*0.002 (0.002)*	*0.016 (0.001)*	0.018 (0.026)
Body Weight	*-0.007 (0.001)*	*0.016 (0.003)*	*0.015 (0.002)*	*0.005 (0.001)*	*0.0001 (0.0002)*	*0.002 (0.0001)*

4.4 Discussion

My analyses of data from zebu calves in Western Kenya quantified several sources of variation: firstly, in strongyle worm burdens, and secondly in body size and a suite of haematological parameters that I anticipated might be affected by strongyle infection. Measures of associations between strongyle EPG and the physiological traits were consistently negative, suggesting a possible cost of increased parasite burdens. Below, I discuss each of these aspects of my results in turn.

4.4.1 Components of variation in strongyle EPG

My results show, firstly, substantial changes with age in median levels of strongyle EPG in EASZ calves. The difference in median strongyle EPG in young (age 1-11 weeks) and old (age 16-51 week) calves is possibly due to weaning, with calves moving more once they are weaned and so older calves being at higher risk of becoming infected due to sampling more areas. I observe lower median faecal egg counts then might normally be expected for *Haemonchus* infections (e.g. compare to Hansen and Perry (1994)). However, Kanyari et al. (2010) study of cattle from a peri-urban area in a neighbouring area of Western Kenya (which included exotic breeds) observed a similar prevalence and mean strongyle EPG (mean = 296, range = 0-8300 EPG (Kanyari et al., 2010) and Figure 4.1 for comparison). Secondly, as in other studies e.g. Hayward et al. (2009) and Moore and Wilson (2002)); male calves have a higher strongyle EPG then female calves.

Thirdly, my analyses indicated that strongyle EPG was heritable (h^2 = 23.9%, SE = 11.8%). Similar heritabilities have been observed in feral Soay sheep lambs on St Kilda (h^2 = 26%, SE = 12%, Beraldi et al. (2007)) and in Scottish Blackface sheep ewes (h^2 = 23%, SE = 9%, Bishop and Stear (2001)). These estimates are from models which included the same fixed effects of age and sex as used in my analysis, but they also included additional fixed effects such as weight and twin status, so direct comparisons of heritability need to be treated cautiously (Wilson, 2008). As I found evidence for the presence of heritable variation in strongyle EPG, it may therefore be possible for selection for parasite resistance to occur. Quantitative trait loci and SNPs associated with strongyle faecal egg count have been identified in Soay sheep (Beraldi et al., 2007) and Blackface lambs (Riggio et al., 2013), but so far have not yet been tested for in indigenous African cattle.

Lastly, complete removal of the 'introgressed' calves from the study resulted in a lower heritability estimate and larger standard errors. The decrease in heritability is possibly due to European introgressed calves having a higher genetic variance whilst

the larger standards errors are likely to be due to a decrease in sample size. Inclusion of the European taurine introgression as a fixed effect did not alter the heritability estimate. However, as the focus of the aim of this study is to describe the components of variation in Strongyle EPG in the study cohort I wish to include as much variance in the population as possible in the dataset. Furthermore, the level of European taurine introgression is on a continuous scale and the cut-off to determine what level of introgression should be excluded is somewhat arbitrary.

4.4.2 Components of variation in physiological traits

I found evidence for significant additive genetic variation in WBC in the IDEAL study population (WBC, $h^2 = 27.6\%$, SE = 10.6%; $V_A = 3.1$, SE = 1.2). This is in accordance with other analyses of WBC count, which have found it to be heritable in both humans and pigs ($h^2 = 35\%$, SE = 9%, Pankow et al. (2001); $h^2 = 29\%$, SE = 10%, Clapperton et al. (2009), respectively). None of the other traits investigated showed evidence for significant additive genetic variation. However Rowlands et al. (1995) showed that packed red-cell volume was heritable in zebu ($h^2 = 32\%$, SE= 7%, sample size = 936) and body weight is known to be highly heritable in many other species, including in a much larger study of beef cattle ($h^2 = 41\%$; Marshall (1994)). More generally, haematological parameters are highly heritable in humans, for example haemoglobin levels, RBC, WBC, and platelet numbers have heritability estimates of 37%, 42%, 62% and 57%, respectively (Garner et al., 2000). The difference with my results may reflect limited statistical power. In addition, age may be playing an important role in determining the overall (co)variance seen, as heritability (of for example weight and hindleg length in Soay sheep) changes with age (Wilson et al., 2007). Furthermore, all of these traits are likely to be polygenic, and so are influenced by many loci of small effect (Goddard and Hayes, 2009), and so it is unlikely that all of the causal loci were detected given the low linkage disequilibrium in EASZ (see below).

4.4.3 Possible biases in heritability estimation

Using the SNP data, I have demonstrated here that it is possible to estimate the heritability of select traits without the need for pedigree information or even the presence of close relatives. I found evidence of heritable variation in strongyle EPG and in WBC. However it is worth noting that my estimates may be slightly lower than the true heritability because of the ascertainment bias of the SNP chip (Matukumalli et al., 2009). Additionally, in the absence of close relatives (such as in the IDEAL study sample, as all the calves had different mothers and the average genomic relatedness

from the IBS matrix ranged from -0.02 to 0.24, and only 9 pairs of calves out of the 446 individuals had a genomic estimate of relatedness greater than 0.15), the heritability estimated is determined by the variance explained by causal variants that are in linkage disequilibrium with the genotyped SNPs (Yang et al., 2010). Mbole-Kariuki et al. (2014) showed that EASZ have lower levels of average linkage disequilibrium between adjacent SNP pairs on the SNP chip than other cattle breeds (Nelore and N'dama cattle). Therefore the residual relatedness (i.e. between two 'unrelated' individuals) is low; consequently unrelated individuals (by known pedigree) will only share very short proportions of the genome. Furthermore, as marker density increases, my estimate of heritability also increased (Table E.1). These factors suggest that my estimates of heritability may be lower than those which would be estimated using more closely related individuals and more dense markers (Bèrènos et al., 2014; Yang et al., 2011). Similarly, Robinson et al. (2013) found marker-based estimates to be as low as 60% of the value of pedigree-based estimates of heritability of wing length in a wild bird population.

As such, the estimates presented here should be taken as lower limits on the true estimates of heritability of the different traits in this population, which may also explain why I did not find significant heritability (h^2 = 19.6% SE = 19.2%) for body weight, a trait which is commonly found to have significant additive variance. However, conversely, use of known relatives can result in an overestimation of the true heritability as relatives may share non-additive effects such as dominance, epistasis and shared environmental conditions, which may then confound estimates of similarity due to genetic effects if not adequately accounted for (Kruuk and Hadfield, 2007). Since the study does not include close relatives, my estimates will not be affected by this issue.

Care needs to be taken in distinguishing additive genetic effects from other sources of variance in this analysis as maternal or shared environment effects may be important. The IDEAL dataset has information on only one calf per mother; therefore I cannot estimate maternal effects explicitly. However this data structure also means that maternal effects are less likely to confound estimates of additive genetic variance, as the most usual scenario is that covariance between full-sibs or maternal half-sibs due to maternal effects is mistaken for additive genetic effects (Kruuk and Hadfield, 2007). Any maternal effects are most likely to be contained in the permanent environment effect variance; however there is also the possibility that if the maternal effects themselves are to any extent genetically based and if related mothers are in the same sublocation, they may also contribute to the sublocation variance. Note however that all calves were from different farms, so very immediate local effects will not be generating

any covariance between individuals.

It is also worth pointing out that my estimates had relatively large standard errors, especially for the parameters associated with additive genetic effects. This may be a result of the relatively small sample size (446 individuals) and a lack of relatedness structure between calves, though the sample sizes are relatively standard for similar analyses on wild animal populations (e.g. sample sizes are between 306-576 for Soay sheep (Coltman et al., 2001) and 333-634 for red deer (Clements et al., 2011) for some heritability estimates on wild mammal populations).

4.4.4 Associations between strongyle infection and physiological traits

Previous work on this study population has also found associations between EPG and other key components of individual phenotypes, specifically survival rates and body size (Thumbi et al., 2013a,b). Thus strongyle EPG has a major impact on life history in this population. I have added to this information the contribution of the different components of variance in each of these traits, and the observation that birth weight predicts subsequent worm infection.

Calves with a higher strongyle EPG tended to have lower mean EO, WBC, RBC, and TSP than those with fewer eggs: these associations applied both to average values across all observations on a calf (the 'individual-level' covariances in Table 4.4), and within each visit ('residual' covariances in Table 4.4). Some strongyle species, such as *Haemonchus placei*, are important causes of anaemia in cattle (Kaufmann et al., 1992). Since anaemia is defined as an erythrocyte count, haemoglobin concentration or packed cell volume below the reference value for that species (Jain, 1993), it is expected that RBC will decrease in association with strongyle infection, as I observed in this study. Furthermore, as some strongyles such as *Haemonchus placei* are blood sucking parasites, a reduction in all blood parameters at the same time is likely to be due to total blood loss in calves with high burdens. The loss in total serum protein will probably also contribute to the reduction in weight. Meanwhile, the negative association between EO and strongyle EPG could be explained by EO having been implicated in the resistance to infection in ruminants. For example, Bricarello et al. (2007) found a negative association between nematode faecal egg count and blood eosinophil counts in Nelore-breed cattle.

Calves that were lighter weight at less than 1 week old had a higher strongyle EPG than heavier calves when they are aged 16-51 weeks old. In a study of humans, Raqib et al. (2007) observed altered immune function in low birth weight babies which may

increase vulnerability to infection later in life. Alternatively, the association could be generated by correlations of both early weight and subsequent strongyle infection with some other unmeasured aspect of individual condition, without requiring any causal component. It is also possible that lighter calves may be eating less and therefore might be expected to have lower intensities of infection, due to sampling fewer areas, but I observe the opposite direction of effect, with lighter calves having a higher strongyle EPG. However since the calves' consumption of food was not monitored during the IDEAL project, I cannot investigate this further.

4.4.5 Concluding remarks

To conclude, in this study I have used relationship matrices reconstructed from SNP genotypes to demonstrate evidence for heritable variation in strongyle EPG in EASZ. I also found significant additive genetic variation in WBC. All additional traits investigated showed negative phenotypic covariances with strongyle EPG throughout the first year of life: high strongyle EPG was associated with low WBC, RBC, TSP, and EO. Weight at one week old was significantly associated with strongyle faecal egg count at 16-51 weeks: smaller calves had a higher strongyle faecal egg count later in life. My results indicate that additive genetic variation in strongyle faecal egg count is present in this population, and that strongyle FEC is associated with variation in other important variables. Further investigation is needed to understand the physiological mechanisms of the interactions between strongyle EPG and haematological parameters that allow EASZ calves to tolerate a high strongyle faecal egg count.

Chapter 5

Genome-wide association study of East Coast Fever death and packed cell volume at the time of seroconversion to *Theileria parva*

Summary

The tick-borne disease East Coast Fever (ECF), caused by the protozoa *Theileria parva*, is responsible for high mortality and morbidity rates of cattle in East Africa. In the IDEAL calves ECF is responsible for 40% of the reported infectious disease mortality (Thumbi et al., 2013a). It has been observed that European cattle breeds exposed to the protozoa *T. parva*, frequently show classical signs of ECF and are severely affected by the pathogen. In contrast, indigenous cattle may experience insignificant or subclinical effects of *T. parva* infection. Furthermore, tolerant cattle are better at maintaining their packed cell volume following infection. Here I present a case-control genome-wide association study (GWAS) to investigate if there is a genetic predisposition to ECF death in East African shorthorn zebu (EASZ) in Western Kenya. In addition, I present a continuous GWAS to investigate the genetic basis to packed cell volume at the time of seroconversion to *T. parva* (PCV_{TP}).

My GWAS suggests that there is no robust evidence for a genetic predisposition to ECF death in the IDEAL population; the GWAS results follow the same distribution of that of a simulated null model. In addition, I found no evidence for a genetic association between PCV_{TP} and genotyped loci. However, the minor allele frequency (MAF) for candidate SNPs associated with ECF death and PCV_{TP} differs between breeds. These candidate SNPs are in areas of the genome which have been selected for

and these regions show a reduction in the level of heterozygosity of alleles in EASZ compared to Holstein, Jersey, N'Dama and Nellore breeds. Moreover, the candidate SNPs are close to genes and QTL with relevance to *T. parva* tolerance. I suggest that the difference between the GWAS and selection analysis and MAF results is likely to be due to sample size.

Therefore, this chapter also investigates the effect of sample size upon GWAS and significance thresholds through simulations. I conclude that the number of cases is important in case-control GWAS as the probability of a SNP being associated with being a case is reflected by the number of cases in the sample, with a lower proportion of cases falsely identifying more associated SNPs. This suggest that an empirically-defined significance threshold level is more suitable than the typical Bonferroni correction when analysing data from small sample sizes, since it reflects the distribution of the data, the number of SNPs, the sample size and the proportion of cases and controls.

5.1 Introduction

East Coast Fever (ECF) is a fatal illness of cattle caused by the protozoa *Theileria parva*, which is transmitted by the tick *Rhipicephalus appendiculatus*. ECF is responsible for major economic losses through mortality, morbidity, loss of production and cost of control in East Africa (Homewood et al., 2006; Kiara et al., 2014). The main clinical signs of ECF are pyrexia, enlarged superficial lymph nodes, wasting and severe pulmonary oedema (Coetzer and Tustin, 2004).

European cattle breeds exposed to infected ticks frequently show classical signs of ECF and are severely affected by the pathogen (Kariuki et al., 1995). In contrast, cattle of African origin show variable responses to infection, for example zebu calves born to immune dams or raised in endemic areas may experience insignificant or subclinical effects of *T. parva* infection (Coetzer and Tustin, 2004; Moll et al., 1984; Mukhebi et al., 1992). The difference in different breed's response to *T. parva* infection could be due to differences in their origin. It is thought that humpless (*Bos taurus taurus* or taurine) breeds such as the European Holstein and Jersey originated from cattle domesticated from North East Africa or the Near East, whereas humped breeds, otherwise known as zebu breeds (*Bos taurus indicus*) originated from Asia and gained entry into Africa via the East Coast and the Horn of Africa (Decker et al., 2014; Hanotte et al., 2002). *T. parva* is endemic to East Africa (Lessard et al., 1990; Norval et al., 1991).

As different cattle breeds show differing levels of tolerance (the ability of the to

limit harm from a parasite Råberg et al. (2009)), it is likely that there is a genetic component to tolerance. Previous research by Murray et al. (2013) has shown that indigenous East African cattle with European taurine (ET) introgression experience more clinical illnesses generally (not specifically ECF) than non-introgressed animals. A similar pattern of tolerance has been described in connection to infection with a related *Theileria* species, called *Theileria annulata* which is found in North Africa and Asia (Coetzer and Tustin, 2004). Cattle originating from *T. annulata* endemic areas produce fewer clinical signs and are able to prevent the over-stimulation of cytokine pathways in comparison to exotic Holstein cattle (Glass et al., 2005). In addition, studies have identified quantitative trait loci (QTL) and single nucleotide polymorphisms (SNPs) associated with health-related traits in cattle such as trypanotolerance and tick resistance (Machado et al., 2010; Noyes et al., 2011). Where trypanotolerance is defined as the ability of an individual to remain apparently healthy whilst being infected with a potentially lethal trypanosome infection, thus individuals tolerate the trypanosome infection (Murray et al., 1984).

It is known that trypanotolerant individuals are better at maintaining their packed cell volume (PCV, the percentage of red blood cells in the blood), mounting an immune response and controlling parasite proliferation than susceptible individuals (Hanotte et al., 2003; Paling et al., 1991; Trail et al., 1989). Furthermore, QTLs associated with decreased PCV in cattle after trypanosome infection challenge have been identified, suggesting that there is a genetic bases to an individual's response following trypanosome challenge (Hanotte et al., 2003). In the case of ECF, PCV is negatively associated with *T. parva* infection (Conradie van Wyk et al., 2014). Furthermore, a lower PCV is associated with calf mortality, although in this case it was thought that PCV was acting as a measure of health related to mortality rather then a risk factor (Thumbi, 2012). However, it is unknown if there is a genetic basis to ECF or the change in PCV following *T. parva* in East African Shorthorn Zebu (EASZ).

One method, to look for a genetic predisposition to a trait is using genome-wide association studies (GWAS). GWAS tries to identify associations between genotyped loci called single nucleotide polymorphisms (SNPs) within a population and a phenotypic trait such as death from ECF or PCV. GWAS works by assuming that the genotyped SNPs are in linkage disequilibrium (LD) with ungenotyped causal variants that have a direct or indirect functional effect on disease risk (Visscher et al., 2012). LD is the non-random association between alleles at different loci caused by mutation or drift and it is broken down by recombination, meaning that loci closer to each other on the chromosome have higher LD (Wang et al., 2005). Once SNPs associated with a specified trait have been identified then the region surrounding the candidate SNP is

evaluated to identify genes and QTL in order to postulate a mechanism which allows a genetic predisposition to the trait. I am unaware of any GWAS for ECF or related traits.

Population-level differences in allele frequencies associated with susceptibility, resistance or tolerance to other pathogens have also been identified. For example, human populations exposed to the malaria-causing protozoa, *Plasmodium* spp., have higher frequencies of red cell and haemoglobin-related alleles than non-exposed populations (Durand and Coetzer, 2008). Since the minor allele frequency (MAF) determines the power to detect the genotypes associated with a particular phenotype, the extent to which an allele explains the risk of a disease in a population depends, in part, on how common the allele is in that population (Guthery et al., 2007). Consequently, the MAF should be considered when calculating allele frequencies in a given population to estimate the burden of disease associated with a particular polymorphism (Cross et al., 2010; Periasamy et al., 2014).

The Infectious Diseases of East African Livestock (IDEAL) project provides a unique opportunity to identify the genetic determinates of ECF death and low PVC. 548 EASZ calves were recruited into the study at birth and their infection status monitored for the first year of life. If the calf died during the study period then the cause of death was determined *post mortem*. All individuals enrolled into the study were genotyped using a 50K Illumina® BovineSNP50 beadchip, providing us with the opportunity to identify SNPs associated with death due to ECF. Thumbi et al. (2013a) has shown that ECF is responsible for 40% of the reported infectious disease mortality in the IDEAL calves, whilst Callaby et al. (2015) reported evidence for additive genetic variation in infection traits in the form of strongyle egg count within the same population. Furthermore, Bahbahani (2015) has identified areas in the IDEAL calves genome which are thought to be under selection: his study has shown that there are substantial differences in the genomes of EASZ in comparison to European (Holstein-Friesian and Jersey), African (N'Dama) and Asian (Nellore) cattle populations.

This chapter is made of five distinct sections. In the first section, I use genome-wide association studies (GWAS) to identify candidate SNPs associated with ECF death and PCV at the time of seroconversion to *T. parva* (PCV_{TP}) in the IDEAL calves. In the second section, empirical simulations are used to create null models to explore whether the Bonferroni correction gives appropriate levels of significance for the ECF death GWAS. The third section compares the candidate SNPs identified in the GWAS to regions identified by Bahbahani (2015) which are thought to be under selection in EASZ compared to other breeds: if the candidate SNPs were truly associated with

ECF death or PCV_{TP}, then I hypothesize that regions containing these SNPs are under selection. The fourth section identifies genes and QTL associated with the candidate SNPs. Lastly, the fifth section compares the minor allele frequency of the SNPs associated with both traits in EASZ to other cattle breeds with varying degrees of tolerance and exposure to *T. parva*.

5.2 Materials and Methods

5.2.1 Study population and data collection

The Infectious Diseases of East African Livestock (IDEAL) project was a longitudinal study of 548 indigenous calves in Western Kenya. Calves were selected using a stratified two-stage random cluster study design. In the first stage, 20 sublocations (the smallest administrative unit in Kenya) were selected from 5 agro-ecological zones, across a rough area of 45 x 90km. Around 28 3-7 day old calves were enrolled from each sublocation; see Bronsvoort et al. (2013) and Appendix A for a detailed description of the study design. Recruited calves were followed for their first year of life and received routine clinical examinations every five weeks (routine visit), at which biological samples were taken. If the calf experienced a clinical episode during the study period, this was reported and the calf received an additional clinical examination (clinical visit) in which additional biological samples were obtained. If the calf died during the study period, then a full post-mortem examination took place, and the cause of death was determined.

5.2.2 Phenotypes of interest

East Coast Fever death

Blood samples obtained from the calves at each 5 weekly routine visit were screened for antibodies against *T. parva* using indirect enzyme-linked immunossorbent assays (ELISA) at the International Livestock Research Institute, Nairobi, Kenya. The percentage positivity (PP) was calculated for each sample. The PP is a ratio of the test serum optimal density (OD) compared to the OD of a positive control serum. The age of seroconversion to *T. parva* was the minimum age at which the PP value from the ELISA test for *T. parva* was greater than 20, and the titre rose between two consecutive visits (Kiara et al., 2014). This rule means that the minimum age of seroconversion is 6 weeks old, and calves can only seroconvert once during the study period. In addition, this rule prevents seroconversion from being confounded by the presence of maternal

antibodies in the calf and gave a seroprevalence of 77% (423/548) in the IDEAL project cattle (Kiara et al., 2014).

ECF-like clinical episodes (ECF CE) were identified as visits at which the calf had a rectal temperature greater than or equal to 40 °C and had its precrural, suprascapular or parotid lymph nodes bilaterally enlarged plus the calf had evidence of *T. parva* infection (seroconverted to *T. parva* or reverse line blot or microscopy positive for *T. parva* (Woolhouse et al., 2015). An acute ECF death was defined as the scenario when an individual had its definitive aetiological cause or contributing cause of death identified as ECF and died prior to seroconverting to *T. parva*. These observations were used to define the cases and controls used in the ECF death GWAS presented here:

> **Cases:** Acute ECF deaths

> **Controls:** Calves which were still alive at 51 weeks old and had seroconverted to *T. parva* and did not have an ECF CE within 2 routine visits (equivalent to approximately 10 weeks) prior to seroconversion to *T. parva*. Therefore controls had subclinical *T. parva* infection.

Other case-control definitions were considered in the preliminary analyses. As I discuss in further detail below, surprisingly, these produced different candidate SNPs, indicating several issues with regard to GWAS methodology which I explore in more detail below (see section 5.2.4 for definition of candidate SNP and section 5.2.5 for comparison of different case-control definitions).

Packed cell volume (PCV) at the time of seroconversion to *Theileria parva*

As described in the introduction, a decreased PCV has been associated with both *T. parva* infection and calf mortality whilst a higher or maintained PCV has been found in cattle which tolerate infection with other protozoas such as *Trypanosome* spp. Therefore I aim to investigate if there is a genetic association between genotyped SNPs and PCV following *T. parva* infection.

Jugular blood was taken for hematology in 5-ml ethylenediaminetetraacetic acid (EDTA) plastic tubes at each 5 weekly routine visit. From this the packed cell volume (PCV) was measured using a Hawksley microhaematocrit reader at the IDEAL project field laboratory in Busia, Kenya (Conradie van Wyk et al., 2012; Jain, 1993). In the PCV$_{TP}$ GWAS, the PCV at seroconversion to *T. parva* was used as the continuous response variable. Only calves which survived until the end of the study period and by definition had seroconverted to *T. parva* were included in the analysis.

Table 5.1 Cattle populations and breeds with high density SNP genotypes

Cattle Population	Cattle Breed	Sample Location	Number of Individuals Sampled
Asian Zebu	Gir	India	30
	Nellore	India	35
East African Zebu	Ankole	Uganda	25
	Karamojong Zebu	Uganda	16
	Nanda	Uganda	23
	Serere Zebu	Uganda	13
	Sheko	Ethiopia	18
West African Zebu	Adamawa Gudali	Nigeria	25
	Azawak	Nigeria	2
	Bunaji	Nigeria	23
	Red Bororo	Nigeria	22
	Sokoto Gudali	Nigeria	21
	Wadara	Nigeria	3
	Yakanaji	Nigeria	12
West African Taurine	Muturu	Nigeria	12
	N'dama	Nigeria	24
European Taurine	Holstein-Friesian	USA	63
	Jersey	UK	36

5.2.3 DNA genotyping, SNP quality control and construction of the genomic kinship matrix

All calves were genotyped using a 50K Illumina® BovineSNP50 beadchip v1. The beadchip contained 55777 SNPs before quality control, spread evenly throughout the genome with an average of 1895 SNPs on each autosome and 1362 SNPs on the X chromosome (Mbole-Kariuki et al., 2014; Murray et al., 2013). High density SNP genotypes from other cattle populations, which included both taurine and zebu breeds with varying degrees of exposure to *T. parva*, were obtained from Olivier Hanotte, Hussain Bahbahani (University of Nottingham, UK) and Tad Sonstegard (USDA-ARS, Maryland, USA). These populations and breeds are listed in Table 5.1.

Quality control was applied to all IDEAL calf genotypes prior to analysis using the *check.marker* function in GenABEL within R version 2.15.2 (Aulchenko et al., 2007). The quality control criterion used was the same as that used in Callaby et al. (2015) which was as follows: SNP call rate cut-off of 0.9; individual call rate of 0.9 and an identity by state (IBS) threshold cut-off of 0.9. The IBS threshold means that if a pair of individuals is estimated to be exceptionally highly related (e.g. identical twins) then one of the individuals would be removed. The minimum minor allele frequency for SNPs was set to 0.005, to include all SNPs where the minor allele count was 5 or

more. Any X chromosome markers which were inconsistent with the observed sex were set as missing. This quality control resulted in 42,119 autosomal and X markers (41419 autosomal markers plus 700 X markers) left for further analysis and 22 cases and 305 controls in the 50K dataset.

All autosome SNPs and calves which passed the quality control checks were used to construct an identity-by-state (IBS) matrix in GenABEL using the allele frequency weighted option, giving the genomic kinship matrix (Aulchenko et al., 2007). The average genomic estimate of kinship between calves as given by the IBS matrix ranged from -0.02 to 0.24. Three pairs of calves had a genomic estimate of relatedness greater than 0.2 and 6 pairs of calves had a genomic estimate of relatedness between 0.15 and 0.2.

5.2.4 GWAS and identification of candidate SNPs

Genome-wide association analysis (GWAS) was carried out using the *egscore* function in GenABEL (Aulchenko et al., 2007). The *egscore* function computes the fast score test for association between a trait and the genetic polymorphism, and adjusts for possible stratification in both the trait and genotypes by using eigenstratification to account for the first three principal components from Classical Multidimensional Scaling of the genomic kingship matrix (Aulchenko et al., 2007; Price et al., 2006). These first three principal components accounted for the largest proportion of the variance (Figure 5.1). Calf sex and the five agro-ecological zones were included as categorical fixed effects in the model. An alternative method utilizing the 'FASTA' (FAmily-based Score Test for Association) approach, using the *polygenic* and *mmscore* functions in GenABEL to account for the relatedness between individuals directly, instead of using principal components of the genetic structure, was tried and no difference in the top 20 SNPs was found, Table F.1 and Table F.2 (Aulchenko et al., 2007; Chen and Abecasis, 2007; Eu-ahsunthornwattana et al., 2014).

Previous research into the IDEAL study population by Mbole-Kariuki et al. (2014) has indicated that there is evidence for the presence of European Taurine (ET) introgression in 20% of the IDEAL calves. This is consistent with the crossing of EASZ with European breeds fewer than 5 generations ago and is believed to be a result of the breed improvement programs from the mid-90s, in which local cattle were crossed with European breeds such as Holstein-Friesian, Ayrshire, Jersey and Guernsey (Mbole-Kariuki, 2012). In addition, inbreeding is also present within this population (Murray et al., 2013). See Mbole-Kariuki et al. (2014), Murray et al. (2013) and Appendix E.2 for details on how percentage introgression and inbreeding were quantified. To evaluate the effect of inbreeding and introgression on the probability
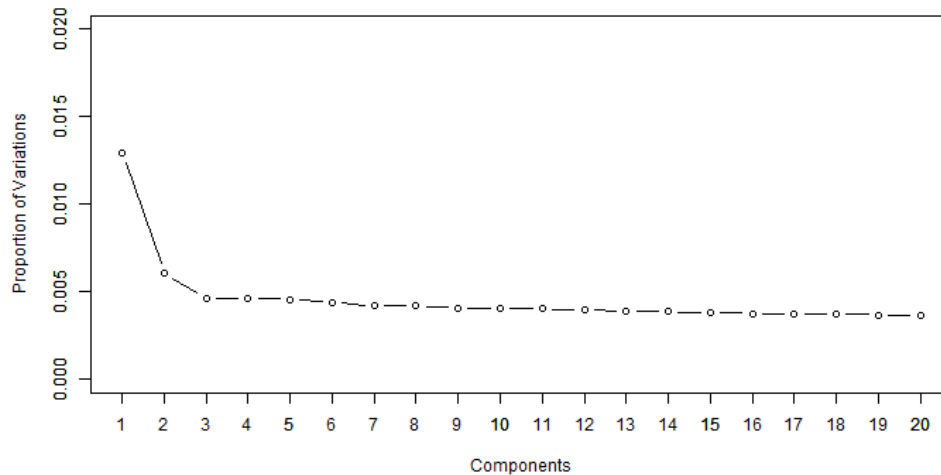
Figure 5.1 Proportion of variance explained by each of the principal components from classical multidimensional scaling of the genomic kinship matrix. I accounted for the first 3 principal components in the *egscore* GWAS analysis, which explain the majority of the variance.

of ECF death and on PCV_{TP} GWAS results, further GWAS were performed with either (i) SNP heterozygosity or (ii) percentage of ET introgression or (iii) both i and ii included as continuous fixed effects in the GWAS. Lastly, (iv) a GWAS was performed in which I excluded all substantially ET introgressed individuals from the analysis (including from the IBS matrix); details are in the Appendix F. Substantially ET introgressed individuals are individuals which have $\geq 12.5\%$ European taurine introgression (Mbole-Kariuki et al., 2014). It was found that these alternative models identified similar SNPs (Table F.3 - Table F.6). A generalized linear mixed model with binomial errors and logit link function was also used to model the risk of ECF death due to ET introgression.

Normally in GWAS analysis, the P values of the association between the SNP and trait are drawn on Manhattan plots in chromosome order to visualize SNPs associated with the trait on neighboring areas of the genome. SNPs which pass the Bonferroni correction of 0.05 divided by the number of SNPs used in the GWAS (in this case, the Bonferroni correction cut-off P value = 1.19×10^{-6}) are considered to be significantly associated with the trait. However, in human medicine, the Bonferroni correction is considered to be too conservative as it does not account for linkage disequilibrium between SNPs, therefore an arbitrary cut-off of $P < 10^{-4}$ is often used to identify SNPs associated with a trait (Johnson et al., 2010; Wu et al., 2010).

In the ECF death GWAS the 12 SNPs associated with a genomic control corrected P-value of $P < 10^{-4}$ were considered to be candidate SNPs and were used for further

analysis (Table 5.2). A genomic control corrected P value is obtained by dividing the observed P value for the association between the SNP and a trait by the genomic inflation factor (λ, Price et al. (2006)). The genomic inflation factor (λ) in GenABEL is given by the regression coefficient of the observed χ^2 test statistic onto the expected χ^2 test statistic (Aulchenko et al., 2007). P values in genetic association studies are likely to be inflated as observations are not independent, for example, infected individuals are more likely to be similar to each other than control individuals as it is hypothesized that they share a common genetic basis for a disorder (Devlin and Roeder, 1999). Since genetic association studies break the assumption about independence of observations, it causes extra variance to occur in the test statistics which need to be accounted for in the form of the genomic inflation factor (λ) and genomic control corrected P-values (Devlin and Roeder, 1999; Hirschhorn and Daly, 2005). In addition, the top 10 SNPs from the PCV_{TP} GWAS were classified as the candidate SNPs for the measure of PCV at the time of seroconversion to *T. parva* (Table 5.2).

To investigate the candidate SNPs more closely, high resolution Manhattan plots of the region surrounding the candidate SNP were plotted. In addition, the pairwise linkage disequilibrium (r^2) was calculated for pairs of SNPs within 500Kbp of the candidate SNP using the *r2fast* function in GenABEL (Aulchenko et al., 2007; Hao et al., 2007).

Furthermore, quantile-quantile (QQ) plots where constructed to show the distribution of the expected χ^2 test statistic (X-axis) compared to the observed χ^2 test statistic (Y-axis). In a QQ plot, the null line $x = y$ means that there is no difference between the observed and expected χ^2 test statistic implying that SNPs are having no effect and are unassociated with the trait. Since the observed values for each SNP are ordered from largest to smallest in the QQ plot, if a few points deviate from the null line at the end of the QQ plot then this suggests that these SNPs are significantly associated with the trait (Ehret, 2010). However, if there is an early deviation from the null line in the QQ plot then it is likely that population stratification is occurring and so there are systematic differences in the allele frequencies between cases and control which results in a larger number of very low P values (Ehret, 2010).

To investigate the effect of small sample sizes on GWAS results, Fisher's exact test was performed to test the association between an individual's genotype at each of the 12 candidate SNPs and its ECF death phenotype. By using Fisher's exact test, I am able to avoid making assumptions about the sample size being less than 5 which may be violated in the GWAS analysis.

Table 5.2 Candidate SNPs from the East Coast Fever death and packed cell volume at the time of seroconversion to *T. parva* GWASs

ECF death candidate SNPs			PCV$_{TP}$ candidate SNPs		
SNP Name	Chromosome	Position	SNP Name	Chromosome	Position
ARS-BFGL-BAC-18388	1	53489802	UA-IFASA-4025	13	33366445
ARS-BFGL-NGS-78606	24	48899795	BTB-01529431	20	50044555
ARS-BFGL-NGS-62008	5	43253141	BTB-01619529	26	29658789
BTA-79451-no-rs	7	60414842	BTB-01930671	5	76956329
BTA-25932-no-rs	12	54218542	BTB-00636189	16	34820781
BFGL-NGS-112172	23	51145823	BTB-01347410	26	1577113
BTB-01968129	5	43315469	ARS-BFGL-NGS-57388	10	4742149
ARS-BFGL-NGS-100189	1	53296440	ARS-BFGL-NGS-87466	10	7830003
ARS-BFGL-NGS-36574	13	40031719	Hapmap38486-BTA-30602	1	41379300
ARS-BFGL-NGS-24963	22	52056494	ARS-BFGL-NGS-42106	7	97772240
BTB-01477690	12	14358955			
ARS-BFGL-NGS-84142	18	18449746			

5.2.5   Simulation analysis

Randomly defining cases and controls

The usual conservative approach in GWAS is to use a Bonferroni correction threshold to identify SNPs that are statistically significant and are associated with a particular trait. Here, I use empirical simulations to explore whether the Bonferroni correction gives appropriate levels of significance. Therefore a simulation was performed to investigate the effect of choosing random cases and controls in varying proportions on obtaining SNPs greater than the Bonferroni correction value, even when there is no underlying association between SNP and trait. To run the simulation, the same quality control checks and GWAS model structure was used as that of the ECF death GWAS. Individuals who passed the quality control checks were randomly assigned to either a case or control group according to the proportions in Table 5.3. *Egscore* GWAS was then performed for 1000 runs of each proportion. The top 20 SNPs from each run were obtained and the cumulative frequency of P values was plotted.

An empirical significance threshold was generated using permutation. The case control status was randomly shuffled 1000 times (whilst keeping the proportion of cases and controls the same as that used in the ECF analysis) and a GWAS was performed for each permutation. The most extreme P value was recorded for each run. The top 1000 P values were then ranked in order, and the 95% most extreme P value was classified as the empirical significance threshold (Churchill and Doerge, 1994).

Table 5.3 Proportion of cases and controls included in the random cases and controls simulation, with the final row (indicated by *) representing a random sample of individuals in the exact same proportion as that used in the ECF death GWAS.

Percentage of individuals which were cases	Number of cases	Number of controls	Total number of individuals in GWAS
50%	165	165	330
40%	132	198	330
30%	99	231	330
20%	66	264	330
10%	33	297	330
5%	17	313	330
*ECF death proportion - (6.73%)	22	308	330

Complete association between SNP and trait

Finally, I took an alternative approach of considering what happens when a SNP is completely associated with the phenotype of interest. A new set of cases and controls was constructed with complete association between a SNP and the trait. Following quality control, every individual with the non-wildtype Chr.13.40031719 SNP (genotype=C/A or A/A, frequency = 11.93%) was classified as a case and all wildtype individuals were classified as controls (genotype= C/C). *Egscore* GWAS was then run on this definition and SNPs which formed a peak were identified.

5.2.6 Alternative approaches to investigate evidence for selection in the IDEAL population

Alternative approaches to investigate evidence for selection in the IDEAL calves has been performed by Bahbahani (2015). If the candidate SNPs identified in the GWAS were truly associated with ECF death or PCV_{TP}, then it is feasible to expect that regions containing these SNPs are under selection and it would therefore be possible to identify the same regions in Bahbahani's 2015 study.

Bahbahani (2015) investigated the evidence for signatures of selection in the non-ET introgressed IDEAL study population using indices based upon the fixation index and extended haplotype homozygosity (EHH) methods. The procedures and results used to calculate each index are fully described in Bahbahani (2015), so here I provide a brief summary of the relevant indices calculated.

The interpopulation genome-wide fixation index (F_{ST}) was calculated to identify adaptively differentiated regions for each chromosome. F_{ST} is the average reduction in heterozygosity of a subpopulation relative to the whole population (Wright, 1949). Bahbahani (2015) calculated the FST between EASZ and three different populations: European (Holstein-Friesian and Jersey), African (N'Dama) and Asian (Nellore) population. A 10 SNP sliding window with a 5 SNP overlap was used to calculate the F_{ST}. A genomic region was considered to be a candidate if it had at least two overlapping windows and if they both were in the upper 0.2% of the distribution (Bahbahani, 2015).

Extended haplotype homozygosity (EHH) based methods investigate the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical by descent for a defined interval; thus EHH detects transmission of an extended haplotype without recombination (Sabeti et al., 2002). A higher EHH value indicates evidence for selection of a particular allele.

The first intrapopulation based EHH statistic Bahbahani (2015) used was the

integrated haplotype score (*iHS*). *iHS* provides a measure of how unusual haplotypes are around a given SNP (Voight et al., 2006). iHS is calculated from the integral of the observed decay in EHH from a specified allele until EHH reaches 0.05. The integral is called the integrated EHH (iHH) and is used to compare the ancestral EHH (iHH$_A$) to the derived EHH (iHH$_D$). Therefore, the unstandardized *iHS* is given by the natural logarithm of iHH$_A$ divided by iHH$_D$ (unstandardized $iHS = ln(iHH_A/iHH_D)$ and this was used to calculate the *iHS* for EASZ (Bahbahani, 2015; Voight et al., 2006).

The second EHH based statistic Bahbahani (2015) calculated was *Rsb* (extended haplotype homozygosity across populations, Gautier and Naves (2011)) which allows pairwise comparison of EHH between the EASZ and the combined reference breeds (Holstein, Jersey, N'Dama and Nellore (Bahbahani, 2015)). This method compares the decay of haplotypes from a specific SNP site of a population as a function of distance (EHHS, extended haplotype homozygosity of a single nucleotide polymorphism site, Tang et al. (2007)). Similarly to *iHS*, *Rsb* is calculated from the log ratio of the integrated EHHS for each population or breed in this case e.g. $ln(Rsb)' = ln(iES_{population_1}/iES_{population_2})$ where iES is the integrated EHHS in each population (Bahbahani, 2015; Tang et al., 2007). In both the *iHS* and *Rsb* analyses, candidate regions were identified if two SNPs which were separated by $\leq$1Mb pass the threshold. Whilst in the between breeds Rsb analyses, a threshold of $log_{10}(1/P\ value) = 3$ was used to account for the smaller sample size used (Bahbahani, 2015).

Thirdly, Bahbahani (2015) calculated the interpopulation change in derived allele frequency (ΔDAF) for EASZ, zebu populations from Nigeria and Uganda as well as for each breed. ΔDAF "is designed to detect variants with high derived alleles in the selected population" relative to other populations (Grossman et al., 2010). It is calculated by subtracting the derived allele frequency in the selected population in this case the EASZ (D$_S$) from the mean derived allele frequency in the non-selected populations (D$_{NS}$), $\Delta DAF = D_S - D_{NS}$ (Bahbahani, 2015; Grossman et al., 2010).

Lastly, a composite analysis combined the *Rsb*, *iHS* and ΔDAF P values from each population (i.e. EASZ, Nigerian and Ugandan cattle). Bahbahani (2015) achieved this by transforming the P value from each index for each SNP to a Z-score and combining them using the following equation:

$$Z_i = \frac{Z_{Rsb} + Z_{iHS} + Z_{\Delta DAF}}{\sqrt{k}} \tag{5.1}$$

were *i* is the number of SNPs and *k* are the number of tests (Bahbahani, 2015). The Z-score was then transformed back to the normal distribution to obtain the combined P values. Bahbahani (2015) considered a region as a candidate for selection if 5 adjacent

SNPs were less than 500kb apart and passed the threshold of $log_{10}(1/P) = 4$. In addition, one of the adjacent SNPs had to pass the Bonferroni threshold of 0.05 divided by the total number of analysed SNPs. See Bahbahani (2015) for a more in depth description of how the composite analysis was computed.

I took the results from Bahbahani's 2015 indices (interpopulation F_{ST} between pure EASZ and pure reference breeds and the composite analysis of pure EASZ *vs.* pure reference breeds) and investigated if any of the candidate SNPs fell in or close to (within ±500000bp) of Bahbahani's candidate sweep regions, to see if there was evidence for selection for genes protecting the calves from death from ECF or associated with PCV_{TP}.

5.2.7 Identification of genes and QTLs associated with candidate SNPs

Since the alternative methods to detect regions of selection in the IDEAL calves highlighted a region containing my ECF death candidate SNP Chr.13.40031719 and the signatures of selection were also close to the candidate SNP Chr.5.76956329 for PCV_{TP} (see section 5.3.4), it suggests that I should investigate the candidate SNPs further. Therefore I identified genes and their functions associated with the candidate SNPs using the methods outlined below.

Regions of ±500000bp either side of the candidate SNP were run through the genome browser, *Ensembl* (release 78, Flicek et al. (2013)) and the associated genes were recorded. Functional annotation clustering for the genes associated with the candidate SNPs was carried out using DAVID bioinformatics resources with a medium level of stringency (Dennis et al., 2003). Functional annotation clustering groups the genes' annotation terms given by *Ensembl* together by their function. This is achieved by identifying gene-term enrichment in the study sample compared to that expected by chance for *B. taurus*, using a modified Fisher's exact test (EASE score, Huang et al. (2008)). Gene-term enrichment means "to identify the most over-represented and biologically relevant terms associated with a given gene list" (Dennis et al., 2003; Huang et al., 2008). To identify genes with a similar function, otherwise known as functional annotation clusters and rank them in order of importance, an enrichment score is used. The enrichment score is the geometric mean of all the EASE scores for each annotation term associated with the gene terms in the group. A higher enrichment score indicates that the gene members in the group are more enriched, and so a particular gene annotation terms occurs more often in a group then you would expect by chance. Furthermore, an enrichment score of 1.3 is equivalent to a non-log

scale P value of 0.05 and so groups with an enrichment score of >1.3 are of particular interest (Huang et al., 2008).

Quantitative trait loci (QTL) within ±5000000bp of the candidate SNPs were ascertained using the Cattle QTL database (Hu et al., 2013).

5.2.8 Comparing minor allele frequencies in different breeds

The minor allele frequency (MAF) was calculated for each of the ECF death and PCV_{TP} candidate SNPs. If the candidate SNP is under selection then neighboring SNPs which are in LD with the candidate SNP should show similar patterns in MAF. Therefore the MAF of all SNPs in a region ±500000bp of each candidate SNP was calculated for each cattle population. Chi squared tests were used to compare the MAF between the different cattle populations.

To remove the confounding effect of outbreeding with a high-risk population from my analysis, all comparisons of MAF are carried out on individuals which do not have ET introgression (estimated fraction of ET introgression is ≤1.56%). In order to specifically investigate the effect of ET introgression on allele frequencies, a generalized linear model was built in R version 2.15.2 to examine if the frequency of the minor allele was correlated with the level of ET introgression in the calves. The response variable was coded as a binary variable for the calf having the minor allele of the candidate SNP or not; and the explanatory variable is ET introgression as a categorical variable (≥12.5% ET introgression representing individuals with "substantial" introgression; 1.56% to 12.5% ET introgression representing the "moderate" introgressed; and the third category included calves with less than ≤1.56% ET introgression representing the "non-European introgressed or pure" calves (Mbole-Kariuki, 2012; Mbole-Kariuki et al., 2014)).

5.3 Results

5.3.1 GWAS of East Coast Fever death

22 cases and 305 controls passed quality control checks. *Egscore* GWAS in GenABEL identified two SNPs significantly associated with ECF death with a P value less than the Bonferroni correction of 1.19×10^{-6} (Chr.1.53489802, P = 2.93×10^{-7} and Chr.24.48899795, P = 8.34×10^{-7}, Figure 5.2). Ten other SNPs were associated with a P value less than 10^{-4} (Table 5.4). The QQ plot of the observed *vs.* expected P values from the GWAS suggests that there is a difference in the SNPs in cases and controls (Figure 5.3) despite the high resolution Manhattan plots (Figure 5.4) showing

no 'peaks' in association. The lack of peaks is presumably due to the low linkage disequilibrium between the candidate SNP and neighbouring SNPs (Figure 5.5). In addition, Table 5.5 shows the association between the candidate SNP and ECF death using Fisher's exact test to account for small sample sizes.

Accounting for inbreeding and outbreeding by including the level of heterozygosity and ET introgression as a fixed effect in the GWAS does not affect the candidate SNPs identified (Table F.3). Excluding substantially ET-introgressed calves from the IBS matrix and case/control population leaves 41359 SNPs, 21 cases and 296 controls for analysis and produces the same top 10 SNPs in a different order to the original analysis which included ET introgressed individuals (Table F.4). This is presumably due to the level of ET introgression having no effect on the risk of ECF death ($OR_{moderate\ ET\ introgression}$=0.51, 95% CI=0.11-2.29, P=0.377; $OR_{substantial\ ET\ ingression}$=1.68, 95% CI=0.36-7.91, P=0.513).
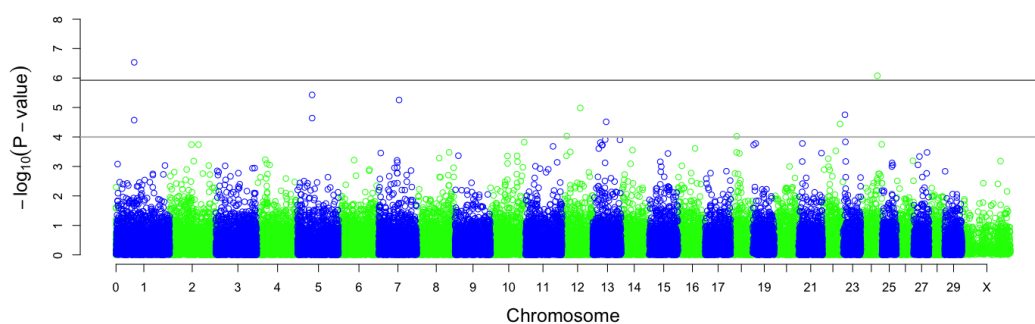
Figure 5.2 Manhattan plot showing SNPs associated with East Coast Fever death from the *egscore* GWAS. The black line is the Bonferroni cut-off (P value = 1.19×10^{-6}) and the grey line is the P value = 10^{-4} cut-off.
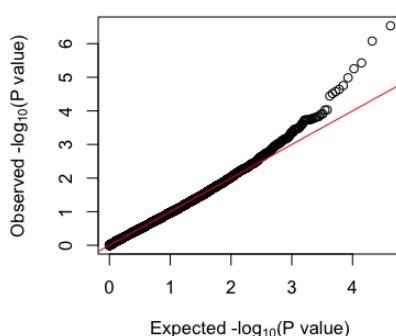
Figure 5.3 QQ plot showing the observed *vs.* expected P values of SNPs from the *egscore* GWAS of East Coast Fever death. The red line represents the null hypothesis of no difference between the observed and expected P values.

Table 5.4 SNPs associated with East Coast Fever death with a P value less than 10^{-4} in the *egscore* GWAS (without accounting for inbreeding or outbreeding). SNPs are ordered in terms of their significance. The Q value is the P value corrected for a false discovery rate (FDR) of 0.005 using the *qvaluebh95* function in GenABEL. The effect size is given by the effect estimate of that allele/genotype in the allelic/genotypic test.

SNP Name	Chromosome	Position	P value	Q value	Effect of the B allele	SE of the effect of the B allele	Effect of the AB genotype relative to AA	Effect of the BB genotype relative to AA
ARS-BFGL-BAC-18388	1	53489802	2.93×10^{-7}	5.86×10^{-6}	0.07	0.01	0.20	7.13×10^{-18}
ARS-BFGL-NGS-78606	24	48899795	8.34×10^{-7}	8.34×10^{-6}	0.10	0.02	0.30	7.13×10^{-18}
ARS-BFGL-NGS-62008	5	43253141	3.74×10^{-6}	2.49×10^{-5}	0.06	0.01	0.20	-1.11×10^{-3}
BTA-79451-no-rs	7	60414842	5.56×10^{-6}	2.78×10^{-5}	0.15	0.03	0.45	-2.77×10^{-4}
BTA-25932-no-rs	12	54218542	1.03×10^{-5}	4.13×10^{-5}	0.22	0.05	0.65	7.13×10^{-18}
BFGL-NGS-112172	23	5145823	1.76×10^{-5}	5.88×10^{-5}	-0.02	0.00	-0.02	-3.42×10^{-2}
BTB-01968129	5	43315469	2.29×10^{-5}	6.54×10^{-5}	0.03	0.01	0.10	7.13×10^{-18}
ARS-BFGL-NGS-100189	1	53296440	2.68×10^{-5}	6.71×10^{-5}	0.03	0.01	0.11	7.13×10^{-18}
ARS-BFGL-NGS-36574	13	40031719	3.06×10^{-5}	6.80×10^{-5}	0.05	0.01	0.16	7.13×10^{-18}
ARS-BFGL-NGS-24963	22	52056494	3.62×10^{-5}	7.23×10^{-5}	0.04	0.01	0.13	2.23×10^{-3}
BTB-01477690	12	14358955	9.36×10^{-5}	1.58×10^{-4}	0.08	0.02	0.23	7.13×10^{-18}
ARS-BFGL-NGS-84142	18	18449746	9.47×10^{-5}	1.58×10^{-4}	0.02	0.00	0.02	3.47×10^{-2}

(a) Chr.1.53296440 (red);
Chr.1.53489802 (green)

(b) Chr.5.43253141 (red);
Chr.5.43315469 (green)

(c) Chr.7. 60414842

(d) Chr.12.14358955 (red);
Chr.12.54218542 (green)

(e) Chr.13. 40031719

(f) Chr.18.18449746

(g) Chr.22.52056494

(h) Chr.23.5145823

(i) Chr.24.48899795
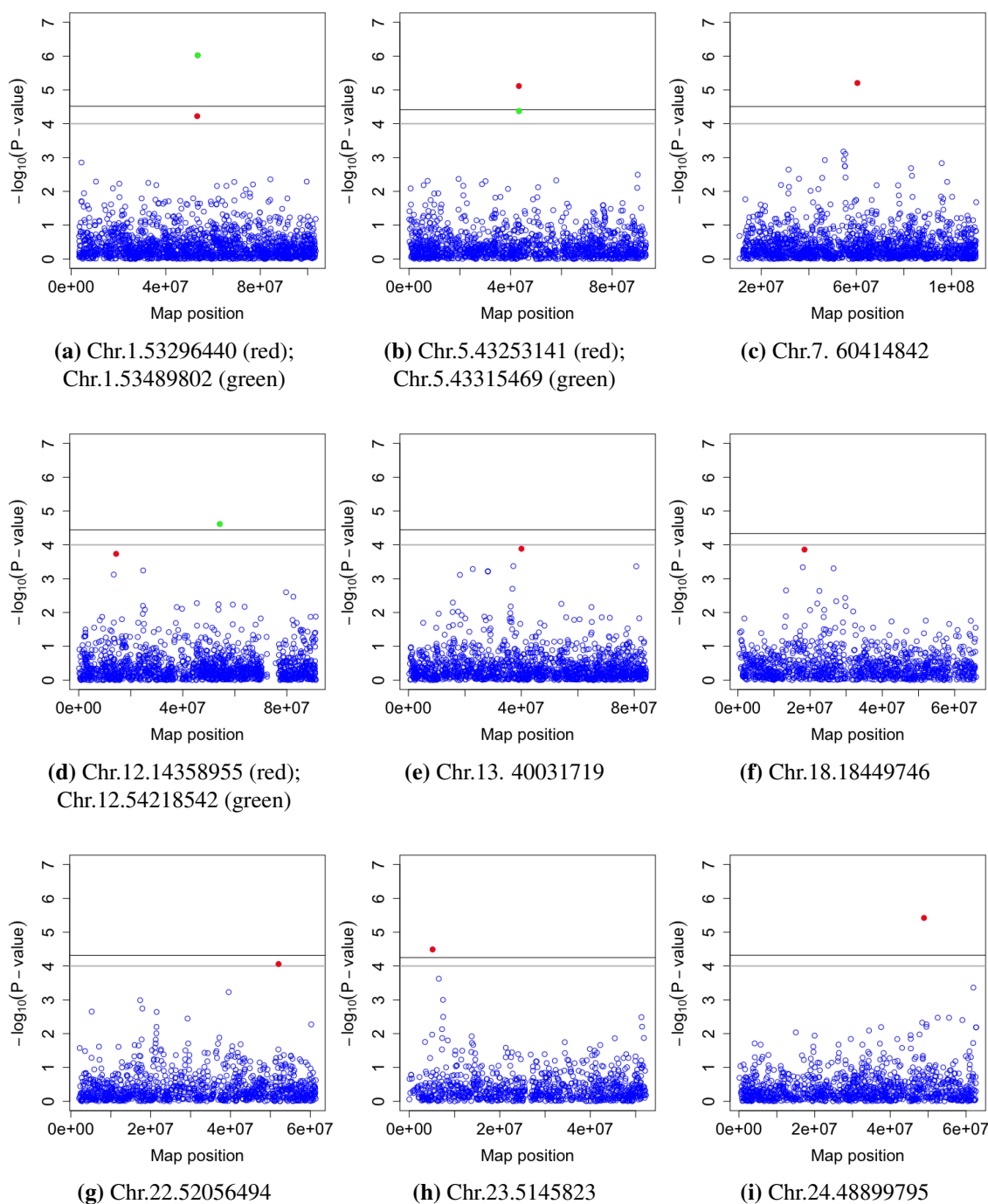
Figure 5.4 Fine scale resolution Manhattan plots from the *egscore* GWAS of East Coast Fever death. Candidate SNPs are highlighted in red or green if there were two candidate SNPs in close proximity to one another. The black line is the Bonferroni cut-off given by 0.05 divided by the number of SNPs in each region of the chromosome and the grey line is the $P = 10^{-4}$ cut-off.
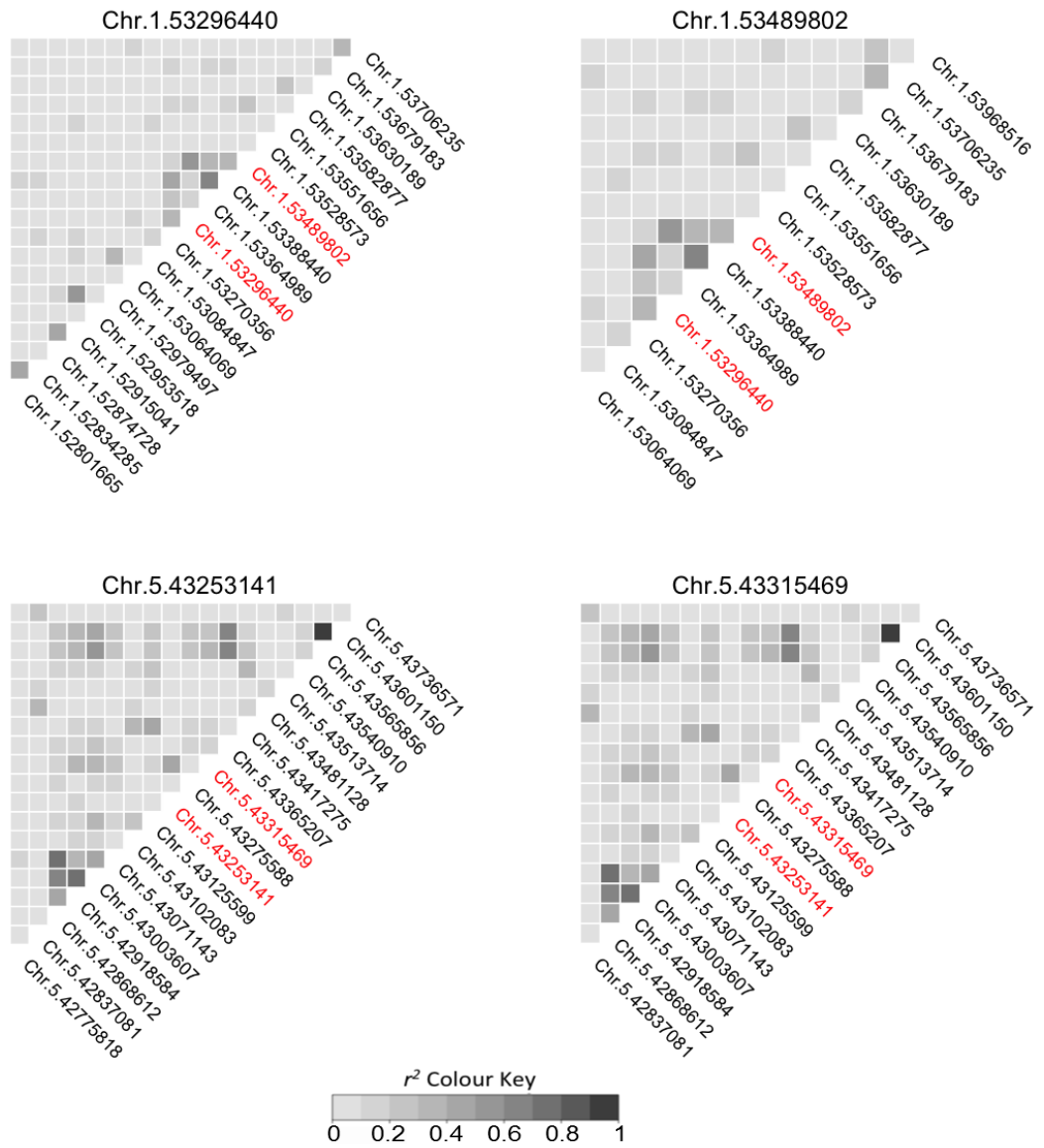
Figure 5.5 Continued on next page
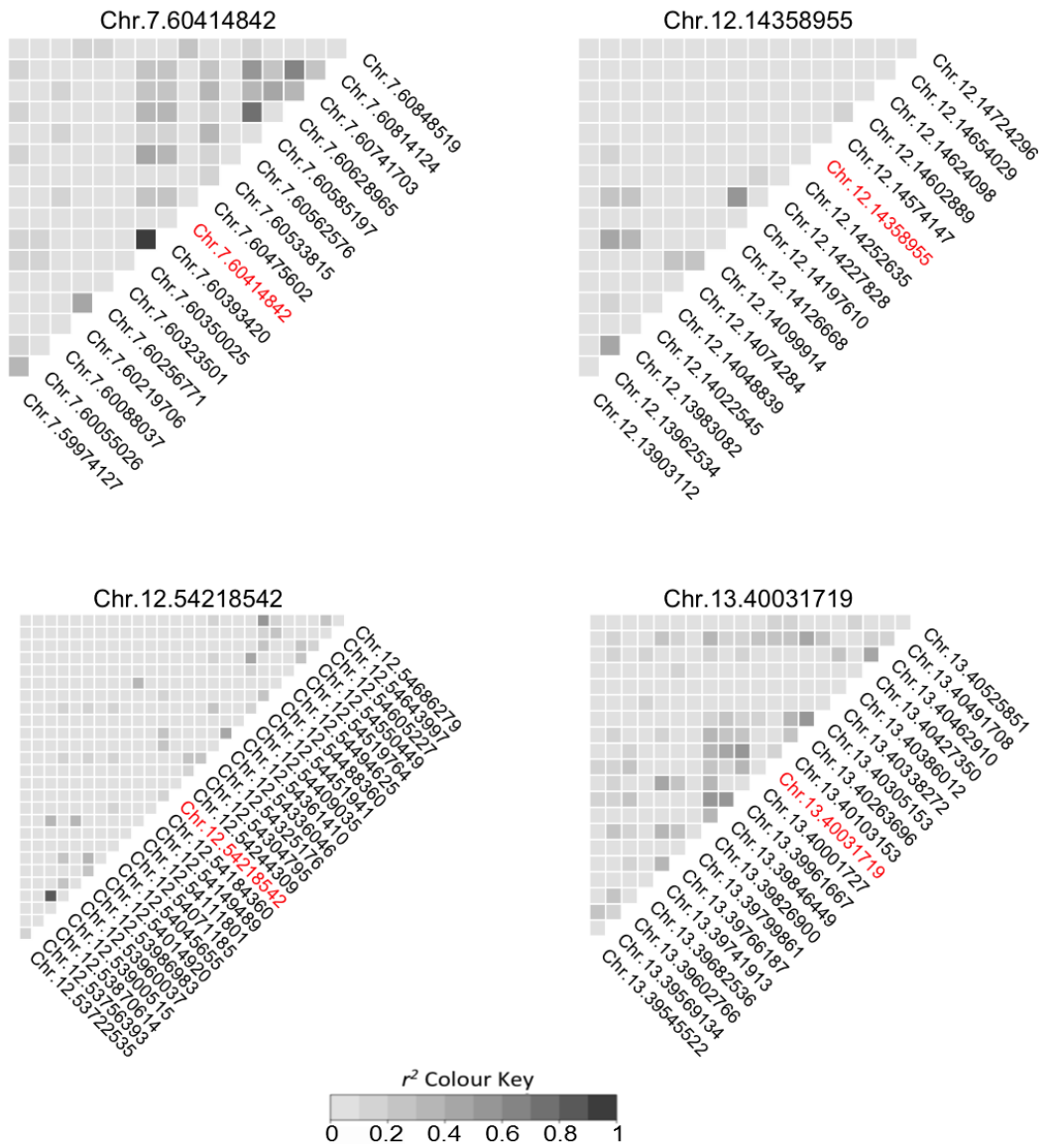
Figure 5.5 Continued on next page
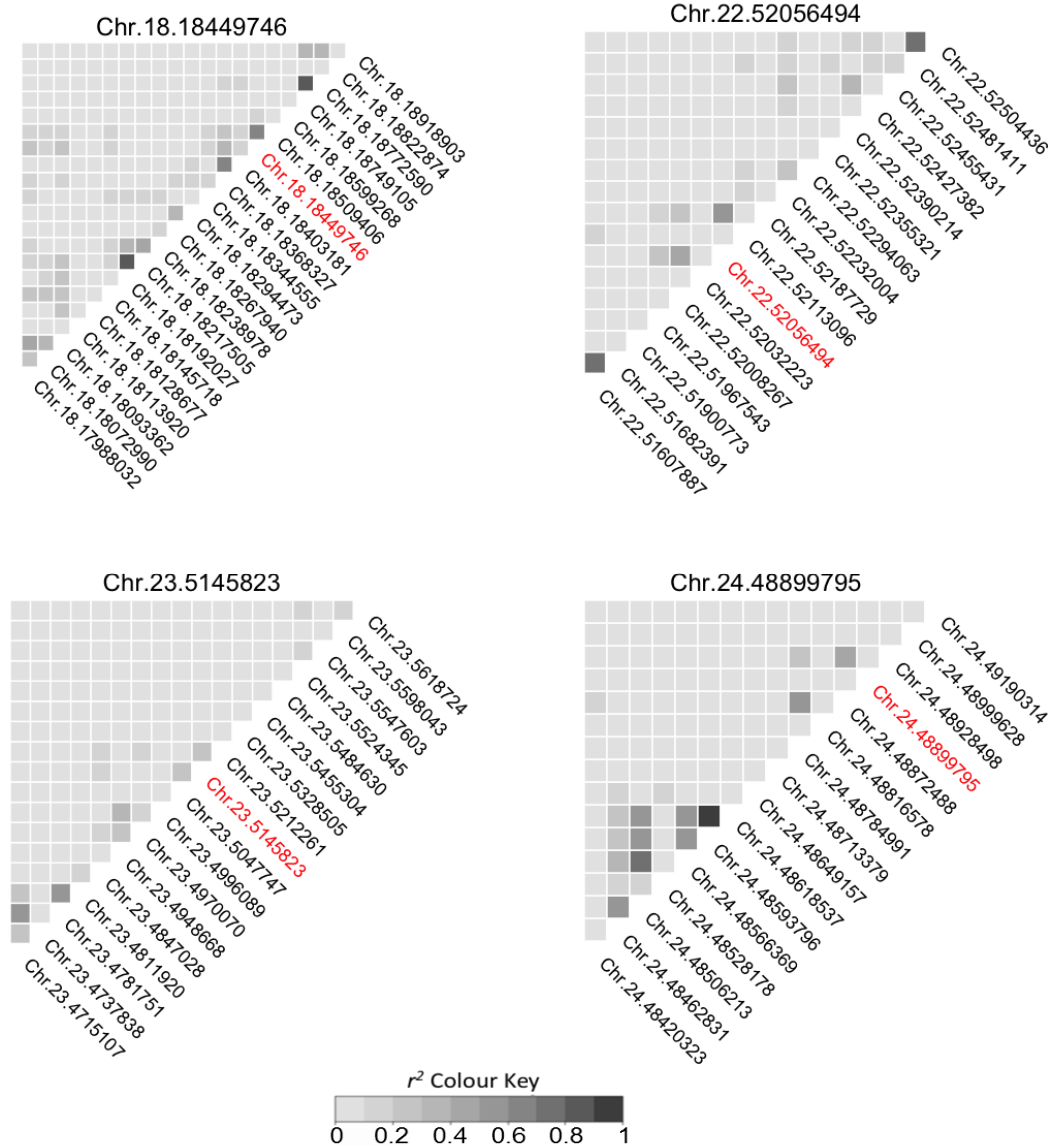
Figure 5.5 Heat map of pairwise linkage disequilibrium (LD, r^2) measurements between SNPs within ±500000bp of the twelve East Coast Fever death candidate SNPs. The candidate SNPs are indicated by the red label and the darker grey colour symbolises higher LD between pairs of SNPs.

Table 5.5 Frequency of each genotype for the candidate SNPs in East Coast Fever death cases (N_{Cases}) and controls ($N_{Controls}$). SNPs are ordered by their P value from the Fisher's exact test of association between the genotype of the candidate SNP and East Coast Fever death (final column).

SNP	Genotype	N_{Cases}	$N_{Control}$	Fisher exact P value
	C/C	20	127	
Chr.23.5145823	T/C	2	143	6.35×10^{-5}
	T/T	0	29	
	C/C	13	280	
Chr.1.53489802	T/C	9	25	9.42×10^{-5}
	T/T	0	0	
	A/A	0	0	
Chr.24.48899795	G/A	60	9	1.55×10^{-4}
	G/G	16	296	
	C/C	10	240	
Chr.5.43315469	T/C	11	65	2.07×10^{-4}
	T/T	1	0	
	C/C	3	2	
Chr.1.53296440	T/C	8	63	2.32×10^{-4}
	T/T	11	240	
	A/A	3	2	
Chr.22.52056494	T/A	6	44	2.74×10^{-4}
	T/T	12	257	
	C/C	2	3	
Chr.5.43253141	T/C	6	21	3.82×10^{-4}
	T/T	14	275	
	A/A	7	26	
Chr.18.18449746	G/A	10	105	5.26×10^{-4}
	G/G	5	173	
	A/A	1	1	
Chr.13.40031719	C/A	7	27	6.54×10^{-4}
	C/C	14	277	
	A/A	0	0	
Chr.12.14358955	G/A	5	11	2.31×10^{-3}
	G/G	17	294	
	C/C	0	0	
Chr.7.60414842	T/C	3	2	2.46×10^{-3}
	T/T	19	302	
	G/G	20	305	
Chr.12.54218542	G/T	2	0	4.33×10^{-3}
	T/T	0	0	

5.3.2 GWAS of packed cell volume at the time of seroconversion to *Theileria parva*

The GWAS for PCV_{TP} was carried out on 380 individuals, with a median PCV of 29% (Range: 11% - 50%). None of the SNPs were significantly associated with PCV_{TP} with a P value less than the Bonferroni correction of 1.19×10^{-6}; however, two SNPs were associated with a P value less than 10^{-4} (Figure 5.6 and Table 5.6). One of these SNPs was Chr.13.33366445 (P value = 2.57×10^{-5}) which is close to the ECF death GWAS candidate SNP Chr.13.40031719 (P value = 3.06×10^{-5}). However the QQ-plot suggests that quantitative traits are less likely to show deviations from the null line and that there is no association between the SNPs and PCV_{TP} (Figure 5.7). This is further illustrated by the lack of peaks above the threshold lines in the high resolution Manhattan plots (Figure 5.8). In addition, the heat maps indicate that there is little pairwise LD between the candidate SNP and its neighbouring SNPs (Figure 5.9).
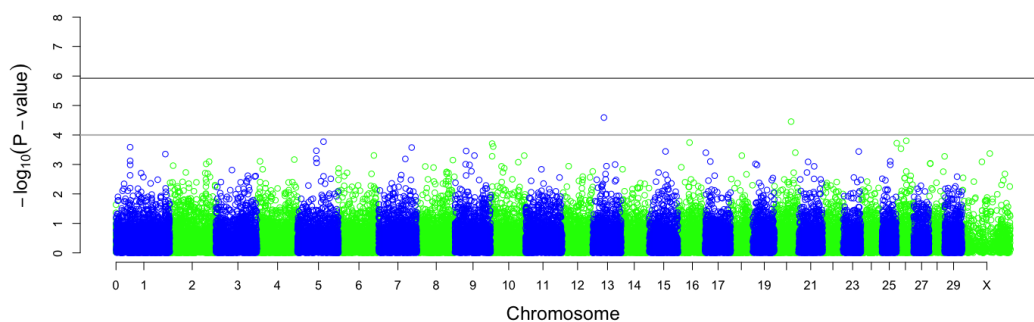
Figure 5.6 Manhattan plot showing SNPs associated with packed cell volume at the time of seroconversion to *T. parva* from the *egscore* GWAS. The black line is the Bonferroni cut-off (P value = 1.19×10^{-6}) and the grey line is the P value=10^{-4} cut-off.
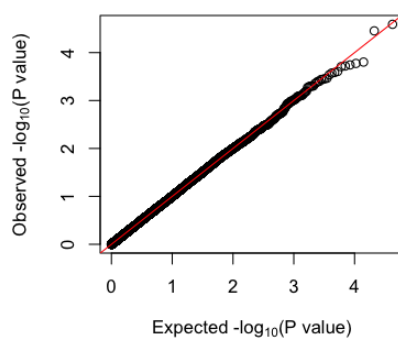
Figure 5.7 QQ plot showing the observed *vs.* expected P values of SNPs from the *egscore* GWAS of packed cell volume at the time of seroconversion to *T. parva*. The red line represents the null hypothesis of no difference between the observed and expected P values.

Table 5.6 Top 10 SNPs from the packed cell volume at the time of seroconversion to *T. parva* GWAS. SNPs are ordered in terms of their significance. The Q value is the P value corrected for a false discovery rate (FDR) of 0.005 using the *qvaluebh95* function in GenABEL. The effect size is given by the effect estimate of that allele/genotype in the allelic/genotype test.

SNP Name	Chromosome	Position	P value	Q value	Effect of the B allele	SE of the effect of the B allele	Effect of the AB genotype relative to AA	Effect of the BB genotype relative to AA
UA-IFASA-4025	13	33366445	2.57×10^{-5}	1.76×10^{-4}	0.25	0.06	0.25	5.06×10^{-1}
BTB-01529431	20	50044555	3.52×10^{-5}	1.76×10^{-4}	0.66	0.16	2.10	1.73×10^{-16}
BTB-01619529	26	29658789	1.58×10^{-4}	2.67×10^{-4}	0.34	0.09	0.34	6.87×10^{-1}
BTB-01930671	5	76956329	1.68×10^{-4}	2.67×10^{-4}	1.09	0.28	3.35	1.73×10^{-16}
BTB-00636189	16	34820781	1.81×10^{-4}	2.67×10^{-4}	0.29	0.07	0.29	5.72×10^{-1}
BTB-01347410	26	1577113	1.90×10^{-4}	2.67×10^{-4}	-0.22	0.06	-0.22	-4.32×10^{-1}
ARS-BFGL-NGS-57388	10	4742149	1.96×10^{-4}	2.67×10^{-4}	-0.35	0.09	-0.35	-7.00×10^{-1}
ARS-BFGL-NGS-87466	10	7830003	2.47×10^{-4}	2.67×10^{-4}	0.27	0.07	0.27	5.45×10^{-1}
Hapmap38486-BTA-30602	1	41379300	2.60×10^{-4}	2.67×10^{-4}	0.30	0.08	0.30	5.97×10^{-1}
ARS-BFGL-NGS-42106	7	97772240	2.67×10^{-4}	2.67×10^{-4}	1.06	0.28	3.19	1.73×10^{-16}

(a) Chr.1.41379300

(b) Chr.5.76956329

(c) Chr.7.97772240

(d) Chr.10.4742149 (red); Chr.10.7830003 (green)

(e) Chr.13.33366445

(f) Chr.16.34820781

(g) Chr.20.50044555

(h) Chr.26.1577113 (red); Chr.26.29658789 (green)
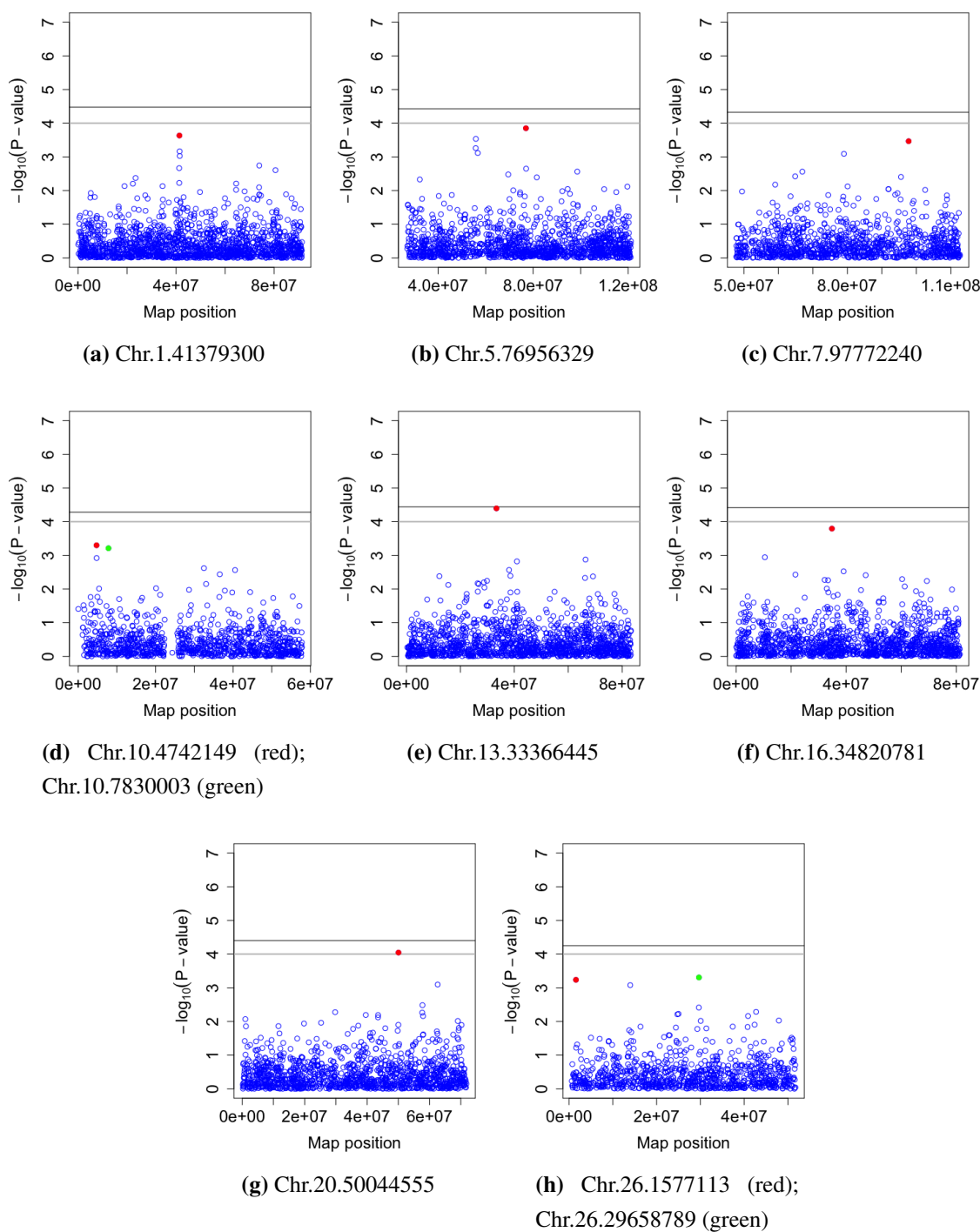
Figure 5.8 Fine scale resolution Manhattan plots from the *egscore* GWAS of packed cell volume at the time of seroconversion to *T. parva*. The candidate SNPs are highlighted in red or green or purple if there were two or three candidate SNPs in close proximity to each other. The black line is the Bonferroni cut-off is given by 0.05 divided by the number of SNPs in each region of the chromosome and the grey line is the P=10^{-4} cut-off.

125
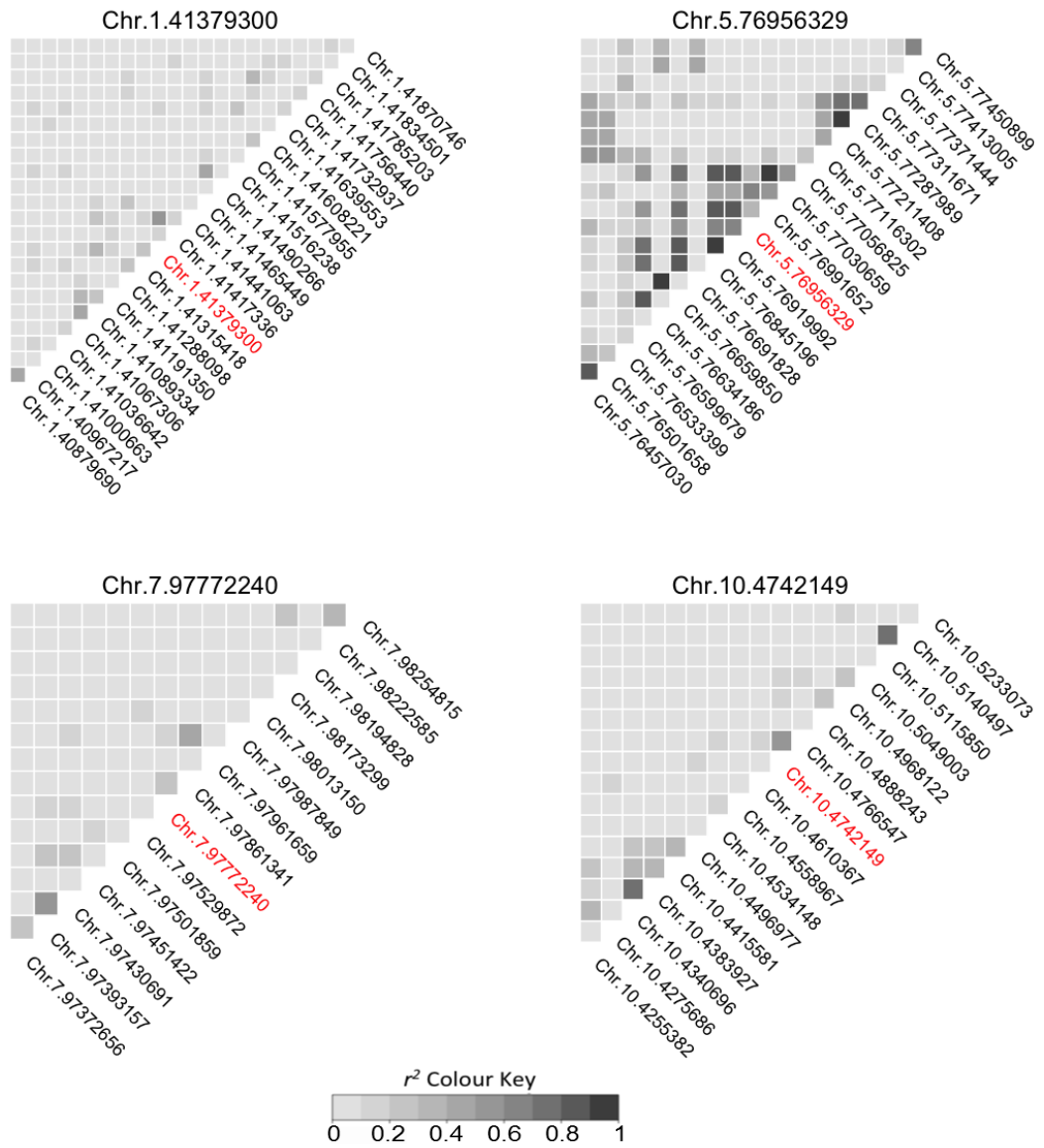
Figure 5.9 Continued on next page
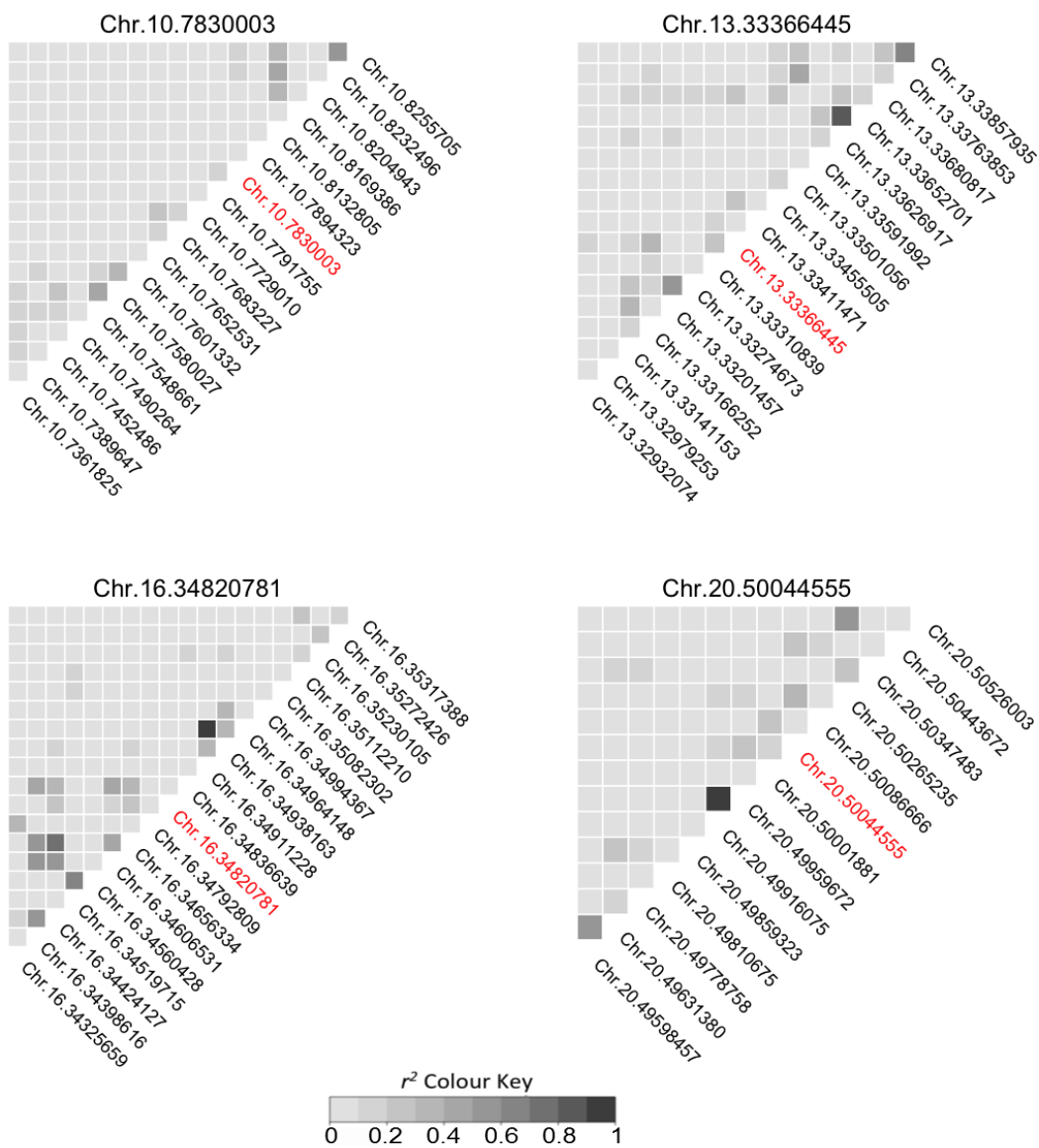
Figure 5.9 Continued on next page
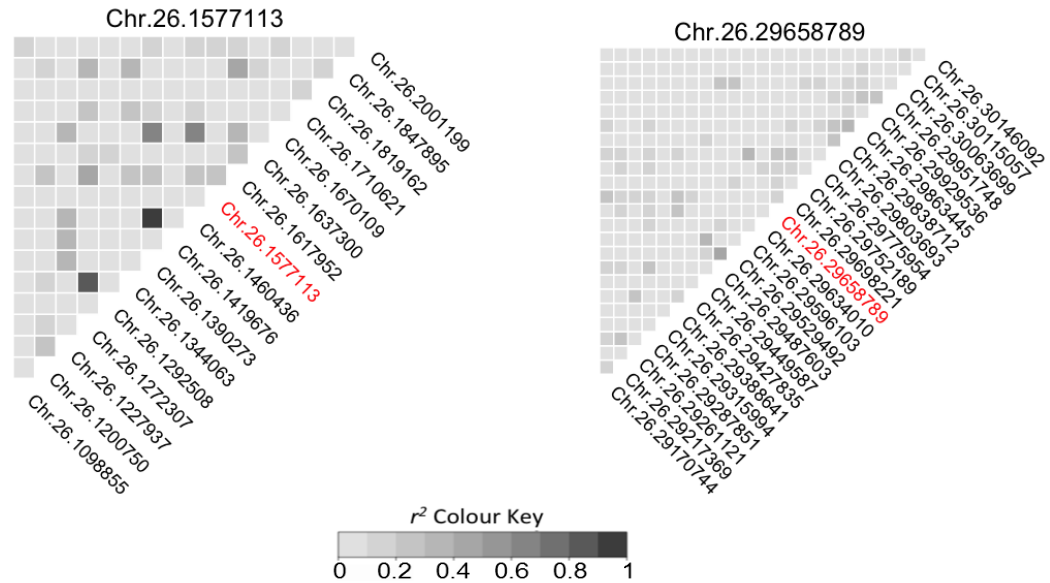
Figure 5.9 Heat map of pairwise linkage disequilibrium (LD, r^2) measurements between SNPs within ±500000bp of the candidate SNP for packed cell volume at the time of seroconversion to *T. parva*. The candidate SNP is indicated by the red label and the darker grey colour symbolises higher LD between pairs of SNPs.

5.3.3 Simulation analysis

Randomly defining cases and controls

Using my simulated sample in which cases and controls had been assigned randomly to different genotypes (i.e. no underlying association), at the sample size used in the ECF death GWAS, 1298 SNPs crossed the Bonferroni threshold of P=1.19x10^{-6} in 1000 runs of the simulation (N.B. the same SNP could be counted multiple times). An average of 1.3 significant SNPs occurred per run, with the maximum being 10 (range: 0-10, Figure 5.10). A one sample Kolmogorov-Smirnov test of the median chi square of the top 20 SNPs from each of the 1000 simulations and the ECF death GWAS suggest that the two samples come from the same distribution (Kolmogorov-Smirnov test, D=0.539, P value=0.923). This suggests that the ECF death GWAS does not identify any truly significant SNPs. Thus, the empirical significance threshold for the 50K tests is P=1.21x10^{-9} and is equivalent to the 95% most extreme P value from taking the lowest P value from each of the 1000 runs and ranking them in order (Churchill and Doerge, 1994).

Furthermore, Figure 5.11 shows that simulations using a lower proportion of cases compared to controls produced more false positive results (according to the number of SNPs which were above the Bonferroni cut-off of 1.19x10^{-6} than GWAS simulations

with a higher proportion of cases. Study populations where 30% of the individuals are defined as cases produce similar results to those that were 40 or 50% of individuals are defined as cases, Figure 5.11. In addition, GWAS with a lower proportion of cases are more likely to deviate from the null line in a QQ-plot then those analysis which include a higher proportion (>30%) of cases (Figure 5.12).
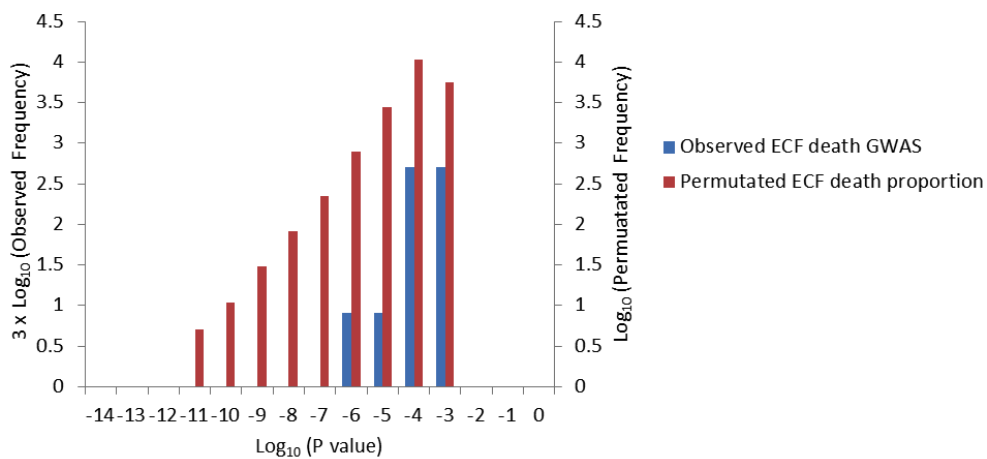
Figure 5.10 Frequency of P values from the top 20 SNPs observed in the East Coast Fever death GWAS and the permutated dataset with randomly-assigned cases and controls at the same proportion as the East Coast Fever death GWAS.
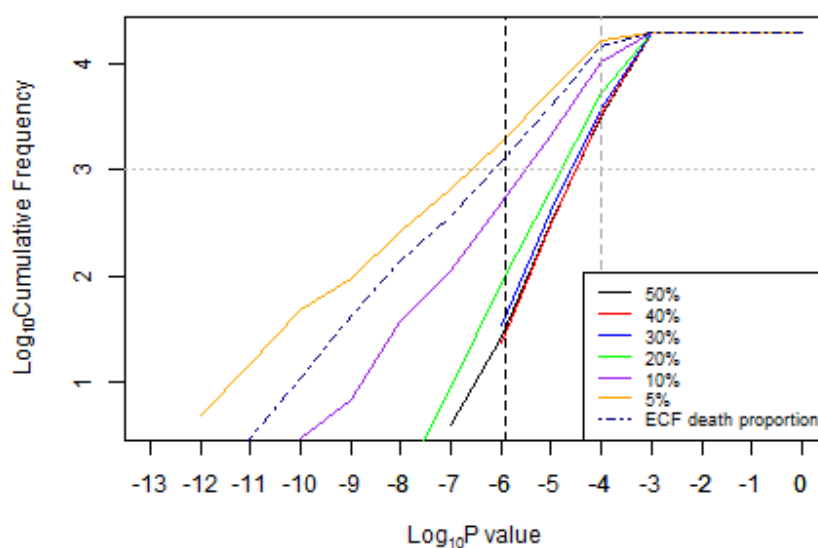
Figure 5.11 Cumulative frequency of P values from the top 20 SNPs of 1000 runs of *egscore* GWAS on randomly-assigned cases and controls at varying proportions. The vertical black line represents the Bonferroni correction of 1.19×10^{-6} and the vertical grey line represents the 10^{-4} cut-off. A $\log_{10}$ cumulative frequency of 10^{-3} is equivalent to having an average of 1 Bonferroni significant SNP per iteration of the GWAS analysis and is indicated by the horizontal grey dotted line.
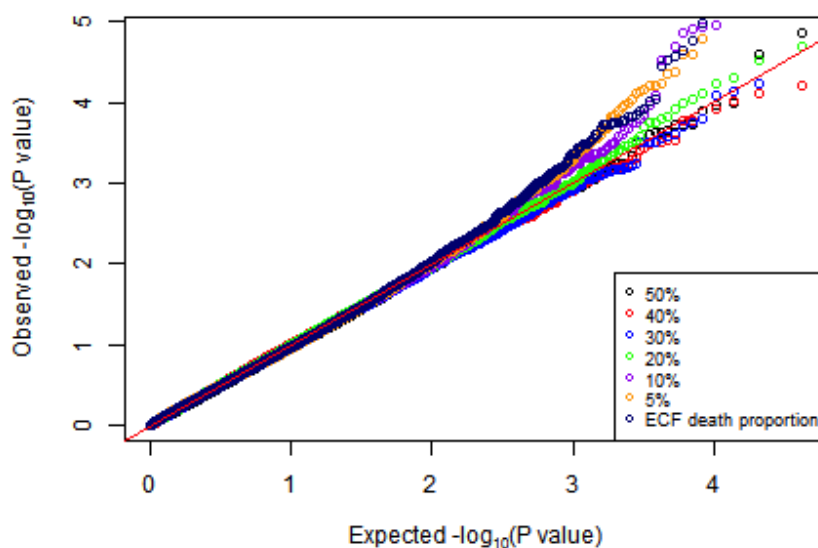
Figure 5.12 QQ plot showing the observed *vs.* expected P values of SNPs from the *egscore* GWAS with varying proportions of cases, but no underlying association between cases and SNPs. The red line represents the null hypothesis of no difference between the observed and expected P values.

Complete association between SNP and trait

When a SNP is fully associated with a simulated trait, the maximum possible level of significance achieved is P=4.41x10^{-93} and a clear peak in P values of associated SNPs is observed (Figure 5.13 and Table 5.7). The peak involves neighboring SNPs with a high pairwise LD and a similar MAF (Figure 5.14 and Table 5.7). However, it should be noted that not all the neighboring SNPs are included in the peak, Table 5.7.
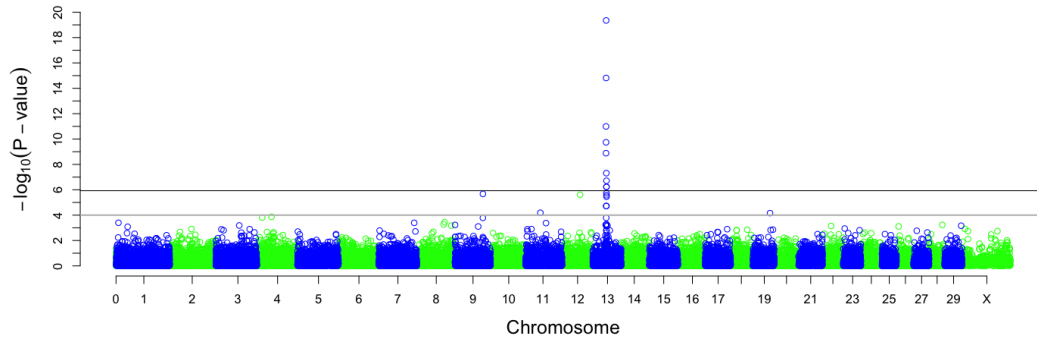
Figure 5.13 Manhattan plot showing SNPs associated with the simulated trait. The black line is the Bonferroni cut-off (P value = 1.19x10^{-6}) and the grey line is the P value = 10^{-4} cut-off. This plot is missing one point at 10^{-92} for Chr.13.40031719.
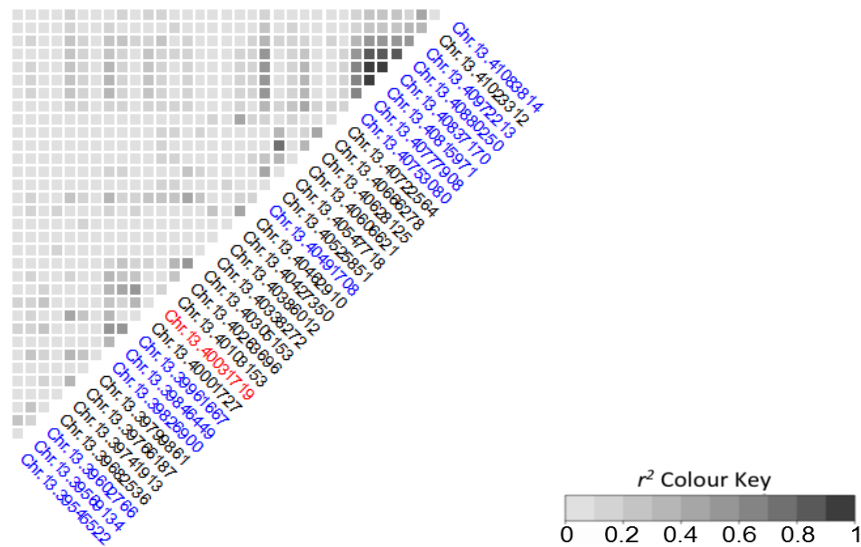
Figure 5.14 Heat map of pairwise linkage disequilibrium (LD, r^2) measurements between SNPs on chromosome 13 within the position range of those in the peak identified in Table 5.7. The East Coast Fever death candidate SNP Chr.13.40031719 is shown in red (N.B. it is the most statistically significant SNP in the peak), and the other SNPs in the peak are highlighted in blue. The darker grey colour symbolises higher LD between pairs of SNPs.

Table 5.7 Top 20 SNPs from the *egscore* GWAS of simulation of complete association between trait and SNP. Cases: Have the non-wild genotype (C/A or A/A, N=63). Controls: Have the wildtype genotype (C/C, N=465).

SNP Name	Chromosome	Position	P value	Q value	MAF Cases	MAF Controls
ARS-BFGL-NGS-36574	13	40031719	4.41×10^{-92}	8.82×10^{-91}	0.516	0.000
ARS-BFGL-NGS-108568	13	39961667	4.49×10^{-20}	4.49×10^{-19}	0.405	0.077
ARS-BFGL-NGS-65199	13	39826900	1.54×10^{-15}	1.02×10^{-14}	0.540	0.167
ARS-BFGL-NGS-4010	13	39545522	1.02×10^{-11}	5.08×10^{-11}	0.135	0.005
Hapmap41547-BTA-46171	13	39602766	1.81×10^{-10}	7.25×10^{-10}	0.214	0.041
ARS-BFGL-NGS-5494	13	39569134	1.31×10^{-9}	4.35×10^{-9}	0.421	0.152
ARS-BFGL-NGS-44116	13	40491708	4.90×10^{-8}	1.40×10^{-7}	0.413	0.152
BTB-00521804	13	40972213	1.95×10^{-7}	4.87×10^{-7}	0.242	0.058
ARS-BFGL-NGS-41919	13	40777908	5.90×10^{-7}	1.18×10^{-6}	0.357	0.123
ARS-BFGL-NGS-37354	13	40837170	5.90×10^{-7}	1.18×10^{-6}	0.357	0.123
ARS-BFGL-NGS-15975	13	40815971	1.86×10^{-6}	3.37×10^{-6}	0.347	0.122
Hapmap51958-BTA-84395	9	82972692	2.12×10^{-6}	3.53×10^{-6}	0.063	0.003
BTB-00496783	12	53654325	2.48×10^{-6}	3.81×10^{-6}	0.040	0.001
ARS-BFGL-NGS-58346	13	40880250	2.67×10^{-6}	3.81×10^{-6}	0.325	0.110
BFGL-NGS-113071	13	41083814	3.56×10^{-6}	4.74×10^{-6}	0.111	0.016
ARS-BFGL-NGS-21158	13	40753080	1.89×10^{-5}	2.25×10^{-5}	0.254	0.081
ARS-BFGL-NGS-8027	13	39846449	1.92×10^{-5}	2.25×10^{-5}	0.206	0.067
ARS-BFGL-NGS-90639	11	43296086	6.47×10^{-5}	7.19×10^{-5}	0.040	0.001
ARS-BFGL-NGS-90673	19	51581082	6.90×10^{-5}	7.27×10^{-5}	0.040	0.001
ARS-BFGL-NGS-90974	4	43710013	1.35×10^{-4}	1.35×10^{-4}	0.071	0.010

5.3.4 Alternative approaches to investigate evidence for selection in the IDEAL population

Despite the simulation analysis suggesting that the GWAS results are likely to be artifacts of a small sample size, the same few SNPs were identified in various preliminary analyses into identifying an association with ECF death. This could symbolize that there is something important occurring within these regions which is worth investigating further. In addition, alternative approaches to investigate evidence for selection in the IDEAL calves have been performed by Bahbahani (2015) and so if the candidate SNPs identified in the GWAS were truly associated with ECF death or PCV_{TP}, then it is feasible to expect that regions containing these SNPs are under selection.

Bahbahani's 2015 interpopulation F_{ST} identified a candidate region close to the ECF death candidate SNP Chr.13.40031719 at Chr.13.40606621-40972213. Within this region, there is a reduction in the level of heterozygosity of alleles in EASZ compared to Holstein, Jersey, N'Dama and Nellore breeds which suggests that selection has occurred within this area. No interpopulation F_{ST} candidate regions were close to or contained the PCV_{TP} candidate SNPs.

The composite analysis which combined the *Rsb*, *iHs* and *ΔDAF* P values identified a candidate region from Chr.13.39754654 - 41333451 which included the ECF death candidate SNP Chr.13.40031719. In addition, the composite analysis identified a region at Chr.5.57977594 close to the PCV_{TP} candidate SNP Chr.5.76956329. These regions suggest that the candidate SNPs are in areas of the genome in EASZ which have different allele frequencies to that of the other breeds investigated (Holstein, Jersey, N'Dama, Muturu, Nellore and Gir) and so it implies that the candidate regions have been selected for.

5.3.5 Identification of genes and QTLs associated with candidate SNPs

East Coast Fever death

Within the genomic region surrounding the 12 ECF death candidate SNPs described above, 97 genes were identified from the genome browser, *Ensembl*. Of these, 5 genes (IFT57, KCNMB4, CFAP61, HCRTR2 and CTIF) were of particular interest as they included some of the ECF death candidate SNPs. All of these were protein coding genes and their functions, as described on the *Ensembl* genome browser (release 78, Flicek et al. (2013)), are presented in Table 5.8.

Functional annotation cluster analysis in DAVID bioinformatics resources (Dennis et al., 2003) identified 8 functional clusters within the 97 genes, of which 1 cluster had an enrichment score greater than 1.3 (enrichment score = 3.81) and included the annotation terms for defense (immune) response, functional annotation terms are listed in Table 5.9. These functional annotation terms were found on genes in chromosome 1 and 22.

The QTLs associated with ±500000bp of the candidate SNPs and those which are relevant to *T. parva* tolerance are presented in Table 5.10. In the region surrounding candidate SNPs on chromosome 5 and 7, two QTL were associated with tick (*Riphicephalus (Boophilus) microplus*) resistance in cattle (Gasparin et al., 2007). In the SNPs surrounding Chr.13.40031719, particular QTL of interest included those for a decrease in PCV following trypanosome challenge and in the parasite detection rate (Hanotte et al., 2003).

Packed cell volume at the time of seroconversion to *Theileria parva*

I identified 60 genes in the region surrounding the PCV_{TP} candidate SNPs, of which 4 genes contained the SNP of interest (Table 5.8). These genes formed 4 functional clusters in DAVID; the cluster with the highest enrichment score had an enrichment score of 1.39 (functional annotation terms are listed in Table 5.9). In addition, already known QTLs associated with PCV on chromosomes 1 and 13 also contained my PCV_{TP} candidate SNPs (Table 5.10). These QTL were the same as those identified for ECF death and coincided with already known QTL associated with trypanosome tolerance and change in PCV following infection (Hanotte et al., 2003).

Table 5.8 Genes which contain the candidate SNPs from East Coast Fever death and packed cell volume at the time of seroconversion to *T. parva* GWAS. The functions of which are as that described on the *Ensembl* genome browser (release 78, Flicek et al. (2013)).

Gene Name	Associated candidate SNP	Gene start (bp)	Gene end (bp)	Description	% GC content	Status
East Coast Fever death						
IFT57	Chr.1.532296440	53295477	53368940	*Bos taurus* intraflagellar transport 57 homolog (*Chlamydomonas*) (IFT57), mRNA	37.82	Known
KCNMB4	Chr.5.43315469	43266425	43347032	*Bos taurus* potassium large conductance calcium-activated channel, subfamily M, beta member 4 (KCNMB4), mRNA	40.12	Known
CFAP61	Chr.13.40031719	39934251	40222551	Cilia and flagella associated protein 61	43.35	Known by projection
HCRTR2	Chr.23.5145823	5043872	5175679	*Bos taurus* hypocretin (orexin) receptor 2 (HCRTR2), mRNA	37.69	Known
CTIF	Chr.24.48899795	48769621	49013492	CBP80/20-dependent translation initiation factor	51.52	Known by projection
Packed cell volume at the time of seroconversion to *T. parva*						
EPHA6	Chr.1.41379300	40608729	41623778	EPH receptor A6	36.70	Known by projection
ATG12	Chr.10.4742149	4619383	4799774	*Bos taurus* ATG12 autophagy related 12 homolog (*S. cerevisiae*) (ATG12), mRNA	40.35	Known
IQGAP2	Chr.10.7830003	7554265	7867328	IQ motif containing GTPase activating protein 2	42.43	Known by projection
NSUN6	Chr.13.33366445	33291610	33369610	*Bos taurus* NOP2/Sun domain family, member 6	41.28	Known by projection

Table 5.9 The functional annotation cluster from DAVID with the highest enrichment score.

Term in the functional annotation cluster	Count of genes involved in term	Modified Fisher's exact test P value (EASE score)	Chromosomes in cluster
East Coast Fever death			
Cathelicidin, conserved site	5	3.20×10^{-9}	22
Cathelicidin	5	6.50×10^{-9}	22
Pyrrolidone carboxylic acid	6	7.70×10^{-7}	1, 22
Antibiotic	5	3.10×10^{-6}	22
Antimicrobial	5	7.10×10^{-6}	22
Defense response to bacterium	5	2.20×10^{-5}	22
Defense response	7	4.70×10^{-5}	22
Response to bacterium	5	1.40×10^{-4}	22
Extracellular region	8	2.70×10^{-2}	22
Signal peptide	7	4.10×10^{-2}	1,22
Disulfide bond	6	5.50×10^{-2}	1,22
Signal	7	9.80×10^{-2}	1,22
Disulfide bond	6	1.30×10^{-1}	1,22
Secreted	5	1.30×10^{-1}	22

Table 5.9 – continued from previous page

Term in the functional annotation cluster	Count of genes involved in term	Modified Fisher's exact test P value (EASE score)	Chromosomes in cluster
Packed cell volume at the time of seroconversion to *T. parva*			
Propeptide: Removed for receptor activation	3	2.00×10^{-5}	10
Thrombin receptor activity	3	2.10×10^{-5}	10
Protease-activated receptor	3	2.70×10^{-5}	10
Peptide binding	4	9.20×10^{-4}	10
Blood coagulation	3	7.10×10^{-3}	10
Coagulation	3	7.10×10^{-3}	10
Hemostasis	3	7.60×10^{-3}	10
Response to wounding	4	9.10×10^{-3}	10
Regulation of body fluid levels	3	1.00×10^{-2}	10
Peptide receptor activity	3	1.10×10^{-2}	10
Peptide receptor activity, G-protein coupled	3	1.10×10^{-2}	10
Wound healing	3	1.40×10^{-2}	10
Cell membrane	5	2.60×10^{-2}	10, 13
Plasma membrane part	6	3.90×10^{-2}	1,5,10,13
7TM GPCR, rhodopsin-like	3	4.90×10^{-2}	10
GPCR, rhodopsin-like superfamily	3	4.90×10^{-2}	10
G-protein coupled receptor	3	6.90×10^{-2}	10
Plasma membrane	7	8.90×10^{-2}	1,5,10,13
Transducer	3	9.00×10^{-2}	10

137

Table 5.9 – continued from previous page

Term in the functional annotation cluster	Count of genes involved in term	Modified Fisher's exact test P value (EASE score)	Chromosomes in cluster
Neuroactive ligand-receptor interaction	3	1.10×10^{-1}	10
Disulfide bond	5	1.30×10^{-1}	7, 10, 16, 18, X
Disulfide bond	5	1.40×10^{-1}	7, 10
Signal	5	2.00×10^{-1}	7, 10
Signal peptide	5	2.40×10^{-1}	7, 10
Glycosylation site: N-linked (GlcNAc...)	5	2.40×10^{-1}	7, 10
Glycoprotein	5	2.50×10^{-1}	7, 10
Receptor	3	3.40×10^{-1}	10
Membrane	7	3.40×10^{-1}	10
Topological domain: Extracellular	3	4.20×10^{-1}	10
Intrinsic to membrane	10	4.90×10^{-1}	1, 5, 10, 13, 26
Transmembrane	5	5.90×10^{-1}	10, 26
Topological domain: Cytoplasmic	3	6.00×10^{-1}	10
Integral to membrane	9	6.20×10^{-1}	1, 5, 10, 13, 26
Transmembrane region	4	6.70×10^{-1}	10, 26
Cell surface receptor linked signal transduction	4	7.30×10^{-1}	10, 1
G-protein coupled receptor protein signalling pathway	3	8.10×10^{-1}	10

138

Table 5.10 Identified QTL associated with health traits related to East Coast Fever deaths or packed cell volume at the time of seroconversion to *T. parva*.

Region (bp)	Trait	Breed identified in	Reference
Chr.1.24654824-60441319	Initial packed red blood cell volume	Boran, N'dama	Hanotte et al. (2003)
Chr.5.13744248-85115661	Tick resistance	Gir, Holstein-Friesian	Gasparin et al. (2007)
Chr.5.61698332-101042998	Tick resistance	Gir, Holstein-Friesian	Gasparin et al. (2007)
Chr.7.2733229-96066390	Tick resistance	Gir, Holstein-friesian	Gasparin et al. (2007)
Chr.13.28857706-71832245	Percentage decrease in packed cell volume up to100 and150 days after challenge with *Trypanosoma congolense* clone IL1180	Boran, N'dama	Hanotte et al. (2003)
Chr.13.28857706-71832245	*Trypanosoma congolense* detection rate	Boran, N'dama	Hanotte et al. (2003)

139

5.3.6   Comparing minor allele frequencies in different breeds

The MAFs for the candidate SNPs are presented in Figure 5.15 and Figure 5.16. A significant difference in MAF was observed between EASZ and European breeds for all ECF death candidate SNPs apart from Chr.1.53296440 (Figure 5.15). In general, ECF death cases have a higher MAF than controls and their MAF is more similar to the ET population than the controls, although Chr.23.5145823 was an exception (Figure 5.15). Since cases tend to have a MAF closer to taurine breeds than zebus, the more zebu a calf is, the better its chances of surviving *T. parva* infections relative to taurine breeds.

The MAF in the region surrounding the candidate SNP is presented in Figure F.1 and Figure F.2. There was a large variation in MAF in SNPs surrounding the candidate SNP, presumably as a result of the low LD and the long distance between genotyped loci (Figure 5.5 and Figure 5.9, Figure F.1 and Figure F.2). Therefore the effect of MAF is local and does not spill over to the neighboring SNPs.

Examining the MAF of the candidate SNPs more closely, I can establish that the MAF is moderately associated with the level of ET introgression in the ECF death candidate SNPs but not in the PCV_{TP} candidate SNPs (Table 5.11). For example, calves which had moderate levels of introgression have 2.18 times the odds of having the minor Chr.13.40031719 allele then pure bred individuals (95% CI=1.17-4.05, P=0.014, Table 5.11). In contrast substantially introgressed calves had 7.58 times the odds of having the minor allele in comparison to pure bred individuals (95% CI=2.68-21.47, P<0.001, Table 5.11). This pattern was observed in 6/12 of the ECF death candidate SNPs and none of the PCV_{TP} candidate SNPs; therefore, the ECF death MAF increases as a calf becomes more ET introgressed.
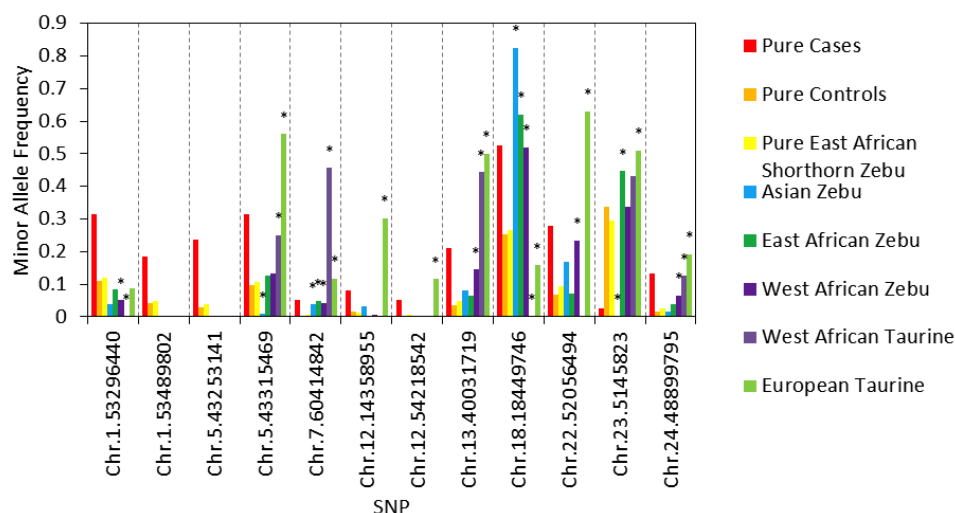
Figure 5.15 Minor allele frequencies (MAF) of the East Coast Fever death candidate SNPs in each of the cattle populations. The MAF for the pure East African Shorthorn zebu (EASZ) is the MAF from all IDEAL calves (both cases and controls) without substantial European taurine introgression. N.B. Chr.1.53489802 and Chr.5.43253141 are not in the high density dataset, so the MAF of these SNPs in the other breeds are unknown. *indicates a statistical significant difference in the allele frequency of each breed compared to the EASZ allele frequency.
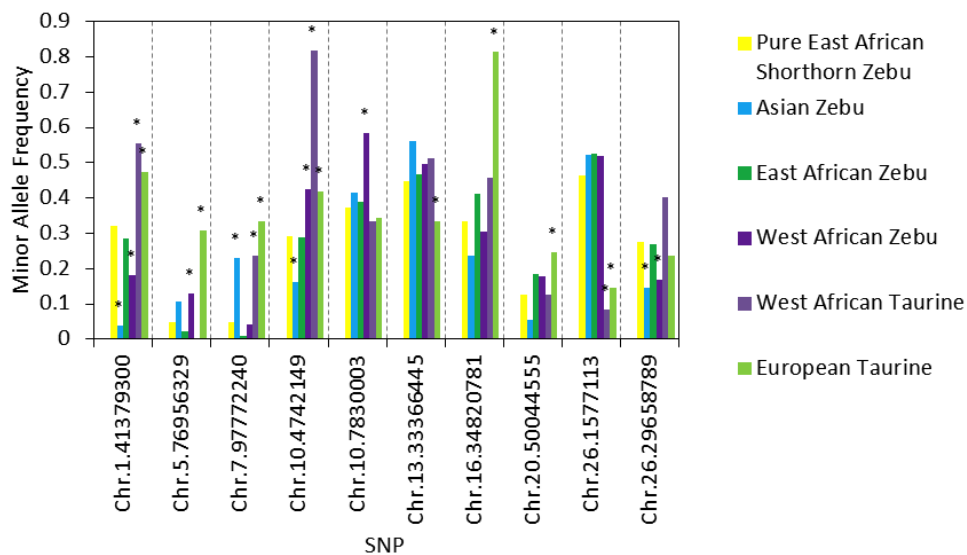
Figure 5.16 Minor allele frequencies (MAF) of the packed cell volume at the time of seroconversion to *T. parva* candidate SNPs in each of the cattle populations. The pure East African Shorthorn zebu is the MAF from all IDEAL calves (both cases and controls) without substantial European taurine introgression. *indicates a statistical significant difference in the allele frequency of each breed compared to the EASZ allele frequency.

141

Table 5.11 Association between level of European Taurine introgression and minor allele frequency of the candidate SNPs. The GLM compares the number of individuals with the minor allele (N_{with}) to those without the minor allele ($N_{without}$) in the pure, moderate and substantial European Taurine introgressed animals. Pure individuals are used as the base line group.

SNP	Level of European Taurine Introgression	Genotype Frequency			MAF	$N_{without}$	N_{with}	OR	95% CI	P value
East Coast Fever death										
Chr.1.53296440		T/T	T/C	C/C						
	Pure	327	85	7	0.118	327	92	-	-	-
	Moderate	70	22	1	0.129	70	23	1.17	0.69-1.97	0.562
	Substantial	12	4	0	0.125	12	4	1.18	0.37-3.76	0.774
Chr.1.53489802		C/C	C/T	T/T						
	Pure	380	37	1	0.047	380	38	-	-	-
	Moderate	83	9	0	0.049	83	9	1.08	0.5-2.33	0.836
	Substantial	15	1	0	0.031	15	1	0.67	0.09-5.19	0.698
Chr.5.43253141		T/T	T/C	C/C						
	Pure	384	24	4	0.039	384	28	-	-	-
	Moderate	81	11	1	0.07	81	12	2.03	0.99-4.16	0.053
	Substantial	12	4	0	0.125	12	4	4.57	1.38-15.1	0.013

142

Table 5.11 – continued from previous page

SNP	Level of European Taurine Introgression	Genotype Frequency			MAF	N$_{without}$	N$_{with}$	OR	95% CI	P value
Chr.5.43315469		C/C	C/T	T/T						
	Pure	333	84	2	0.105	333	86	-	-	-
	Moderate	69	24	0	0.129	69	24	1.35	0.8-2.27	0.263
	Substantial	9	7	0	0.219	9	7	3.01	1.09-8.32	0.033
Chr.7.60414842		T/T	T/C	C/C						
	Pure	414	5	0	0.006	414	5	-	-	-
	Moderate	91	1	0	0.005	91	1	0.91	0.11-7.88	0.932
	Substantial	14	2	0	0.063	14	2	11.83	2.11-66.33	0.005
Chr.12.14358955		G/G	G/A	A/A						
	Pure	407	12	0	0.014	407	12	-	-	-
	Moderate	89	4	0	0.022	89	4	1.52	0.48-4.84	0.474
	Substantial	13	3	0	0.094	13	3	7.83	1.97-31.13	0.003
Chr.12.54218542		G/G	G/T	T/T						
	Pure	413	6	0	0.007	413	6	-	-	-
	Moderate	91	2	0	0.011	91	2	1.51	0.3-7.62	0.616
	Substantial	16	0	0	0	16	0	5.95 x 10-7	0.00-Inf	0.993

143

Table 5.11 – continued from previous page

SNP	Level of European Taurine Introgression	Genotype Frequency			MAF	$N_{without}$	N_{with}	OR	95% CI	P value
Chr.13.40031719		C/C	C/A	A/A						
	Pure	380	37	2	0.049	380	39	-	-	-
	Moderate	76	17	0	0.091	76	17	2.18	1.17-4.05	0.014
	Substantial	9	7	0	0.219	9	7	7.58	2.68-21.47	<0.001
Chr.18.18449746		G/G	G/A	A/A						
	Pure	232	150	36	0.266	232	186	-	-	-
	Moderate	45	41	7	0.296	45	48	1.33	0.85-2.09	0.214
	Substantial	6	8	2	0.375	6	10	2.08	0.74-5.83	0.164
Chr.22.52056494		T/T	T/A	A/A						
	Pure	346	60	8	0.092	346	68	-	-	-
	Moderate	72	19	2	0.124	72	21	1.48	0.86-2.58	0.16
	Substantial	9	7	0	0.219	9	7	3.96	1.43-10.99	0.008
Chr.23.51145823		C/C	C/T	T/T						
	Pure	201	177	32	0.294	201	209	-	-	-
	Moderate	47	41	5	0.274	47	46	0.94	0.6-1.48	0.792
	Substantial	4	10	2	0.438	4	12	2.89	0.92-9.09	0.071
Chr.24.48899795		G/G	G/A	A/A						
	Pure	398	21	0	0.025	398	21	-	-	-
	Moderate	89	4	0	0.022	89	4	0.85	0.29-2.54	0.774
	Substantial	15	1	0	0.031	15	1	1.26	0.16-10.03	0.825

Table 5.11 – continued from previous page

Packed cell volume at the time of seroconversion to *T. parva*

SNP	Level of European Taurine Introgression	Genotype Frequency			MAF	N$_{without}$	N$_{with}$	OR	95% CI	P value
Chr.1.41379300		G/G	G/T	T/T						
	Pure	193	183	43	0.321	193	226	-	-	-
	Moderate	40	44	8	0.326	40	52	1.11	0.7-1.75	0.652
	Substantial	9	6	1	0.25	9	7	0.66	0.24-1.82	0.425
Chr.5.76956329		C/C	T/C	T/T						
	Pure	0	41	378	0.049	378	41	-	-	-
	Moderate	1	9	83	0.059	83	10	1.11	0.53-2.31	0.778
	Substantial	0	3	13	0.094	13	3	2.13	0.58-7.78	0.254
Chr.7.97772240		C/C	C/A	A/A						
	Pure	379	40	0	0.048	379	40	-	-	-
	Moderate	84	9	0	0.048	84	9	1.02	0.47-2.17	0.969
	Substantial	15	1	0	0.031	15	1	0.63	0.08-4.91	0.661
Chr.10.4742149		A/A	A/G	G/G						
	Pure	211	171	37	0.292	211	208	-	-	-
	Moderate	47	43	3	0.263	47	46	0.99	0.63-1.56	0.975
	Substantial	4	10	2	0.438	4	12	3.04	0.97-9.59	0.057

145

Table 5.11 – continued from previous page

SNP	Level of European Taurine Introgression	Genotype Frequency			MAF	N$_{without}$	N$_{with}$	OR	95% CI	P value
Chr.10.7830003		A/A	A/G	G/G						
	Pure	154	218	47	0.372	154	265	-	-	-
	Moderate	36	44	13	0.376	36	57	0.92	0.58-1.46	0.724
	Substantial	5	11	0	0.344	5	11	1.28	0.44-3.75	0.654
Chr.13.33366445		C/C	C/T	T/T						
	Pure	125	212	82	0.449	125	294	-	-	-
	Moderate	27	42	24	0.484	27	66	1.04	0.63-1.7	0.878
	Substantial	3	5	8	0.656	3	13	1.84	0.52-6.58	0.347
Chr.16.34820781		T/T	T/C	C/C						
	Pure	186	184	47	0.333	186	231	-	-	-
	Moderate	45	30	18	0.355	45	48	0.86	0.55-1.35	0.508
	Substantial	4	8	4	0.5	4	12	2.42	0.77-7.61	0.132
Chr.20.50044555		T/T	T/G	G/G						
	Pure	320	92	7	0.126	320	99	-	-	-
	Moderate	62	29	2	0.177	62	31	1.62	0.99-2.63	0.053
	Substantial	11	5	0	0.156	11	5	1.47	0.5-4.33	0.485
Chr.26.1577113		T/T	T/C	C/C						
	Pure	118	210	89	0.465	118	299	-	-	-
	Moderate	32	41	19	0.429	32	60	0.74	0.46-1.19	0.218
	Substantial	4	9	3	0.469	4	12	1.18	0.37-3.74	0.774

Table 5.11 – continued from previous page

SNP	Level of European Taurine Introgression	Genotype Frequency			MAF	N$_{without}$	N$_{with}$	OR	95% CI	P value
		G/G	G/A	A/A						
Chr.26.29658789										
	Pure	219	165	33	0.277	219	198	-	-	-
	Moderate	42	48	3	0.29	42	51	1.34	0.86-2.11	0.2
	Substantial	8	8	0	0.25	8	8	1.11	0.41-3	0.843

5.4 Discussion

5.4.1 GWAS analysis

My GWAS suggests that there is no robust evidence for a genetic predisposition to ECF death in the IDEAL population; the GWAS results follow the same distribution of that of the simulated null model. In addition, I found no evidence for a genetic association between PCV_{TP} and the genotyped loci. However, there are observed differences in the response to *T. parva* infection and MAF between breeds. Furthermore, alternative approaches to investigate selection in EASZ has identified that the candidate SNPs are in areas of the genome where there is a reduction in the level heterozygosity of alleles in EASZ compared to Holstein, Jersey, N'Dama and Nellore breeds suggesting that selection has occurred within this area (Bahbahani, 2015). Moreover, the candidate SNPs are close to genes and QTL with relevance to *T. parva* tolerance. Therefore with a larger sample size and a denser SNP chip which focuses on zebu polymorphisms rather than taurine ones it may be possible to identify a genetic predisposition to ECF death and PCV_{TP}. Below, I review the issues associated with the GWAS of ECF death in more detail before discussing the key findings of this chapter.

ECF death is a complex trait with many steps involved in producing the observed phenotype which may explain why I was unable to detect any genetic effect using GWAS. Previous research into the IDEAL population has shown that the order of exposure to different *Theileria* species and coinfection with other pathogens such as *Trypanosoma* spp. and strongyles influence the risk of ECF death in EASZ (Thumbi et al., 2014; Woolhouse et al., 2015). Given that all individuals in the GWAS were exposed to *T. parva*, it is possible that these external factors are having a greater influence over the phenotype then the genotype and are hiding the genetic effects.

It is also thought that different loci are responsible for controlling resistance (the ability to limit parasite burden), tolerance (the ability to limit harm) and survival (Råberg et al., 2009). For example, different loci were associated with resistance and survival from *Toxoplasma gondii* infection in laboratory mice (Johnson et al., 2002). This may explain why I observed different candidate SNPs associated with ECF death and PCV_{TP}.

In addition, as the sample size is small with few cases, it implies that the majority of animals in the IDEAL study are tolerant to ECF. The reason why I assume that the low number of cases is caused by tolerance of *T. parva* infection rather than low exposure to *T. parva* is because all individuals in this study where exposed to *T. parva* (due to the case and control definitions used) and the overall prevalence of *T. parva* in the IDEAL population is 77% (Kiara et al., 2014). Therefore exposure to *T. parva* is

said to be high in this population. If I was to increase the power of the study by using a larger sample size then it may be possible to identify the genetic basis to ECF death. However, even in large human medicine studies, the genetic basis to many common diseases has still not been identified and it may be unrealistic to assume that every disease has a genetic component which segregates within the population (Goldstein, 2009; Visscher et al., 2012).

To overcome the issues of small sample sizes in this chapter, I experimented with using Fisher's exact test, a non-parametric test to avoid making assumptions about sample sizes. As expected, I found that this method decreases the P values associated with each SNP. An alternative method to increase the strength of the GWAS in a small sample size is to use a continuous trait rather than a binary trait as continuous traits can provide extra information in the form of the variance. This is illustrated by the QQ plots; my QQ-plots of ECF death and PCV_{TP} show that continuous data are less likely to deviate from the null line than binary ones.

Moreover, GWAS may not be the most appropriate method for the analysis of genetic associations in this situation. Issues with GWAS has been raised, for example McClellan and King (2010) suggest that although stratification to account for population structure, as I have done, is appropriate and necessary it is not sufficient as individual outliers may still be influencing the results. The recommended method to correct this issue is to replicate the study with a larger sample size or to use family study designs (McClellan and King, 2010; Wang et al., 2005), both of which are not possible using data available from the IDEAL study. Accounting for population structure in the form of ET introgression and heterozygosity had little effect upon the associations observed, despite ET introgression being strongly associated with MAF for some SNPs. In addition, the FASTA method (Aulchenko et al., 2007; Chen and Abecasis, 2007; Eu-ahsunthornwattana et al., 2014), which directly accounts for the relatedness between individuals, identified similar SNPs to the *egscore* method.

Another issue with my analysis is that the coverage of the genome of the 50K chip is not optimal. It is based on the assessment of polymorphisms in taurine breeds, which may not be polymorphic in zebu. In addition, the SNPs are mainly for production-related traits rather than infection-related traits. Furthermore, the LD between SNPs on the 50K chip is low, especially in EASZ compared to other breeds (Mbole-Kariuki et al., 2014). Given the low density of SNPs, poor coverage of the zebu genome and few SNPs surrounding the candidate SNP, the loci in my analysis are virtually independent which may explain why I only see a single significant association between isolated SNPs and no peaks.

5.4.2 Simulation analysis

The simulation results suggest that the Bonferroni correction is not stringent enough; I observe apparently 'significant' SNPs above the Bonferroni cut-off, even when a random sample of individuals is used. The probability of a 'significant' SNP being identified is a reflection of the proportion of cases in the study. In addition, as I have used a 5% Bonferroni threshold, I would expect to observe 5% of SNPs passing the cut-off. Thus, when analysing data from a small sample size with a low density SNP chip, as in the ECF death situation, the empirical threshold is a more suitable significance threshold than the Bonferroni correction because it reflects the distribution of the data in terms of the number of SNPs and the sample size (proportion of cases and controls).

Human GWAS studies often use a lower cut-off, as low as $P<10^{-4}$ (although the standard cut-off is $P<10^{-7}$ or $P<10^{-8}$), instead of the Bonferroni correction. The explanation for using the lower cut-off is that the Bonferroni correction is too conservative since it does not account for LD between SNPs (Johnson et al., 2010; Wu et al., 2010). Even though human GWAS studies tend to have a larger sample size and a denser SNP chip with a higher LD, given the results of my simulations, I think that these studies are still at risk of false positive results as SNPs with a low MAF may yield sporadic significant results. Therefore human GWAS studies should still be careful about the cut-offs they use.

The simulations show that GWAS with a lower proportion of cases are more likely to deviate from the null line in a QQ-plot then those analysis which include a higher proportion (>30%) of cases, implying that in GWASs with a lower proportion of cases, SNPs are more likely to be statistically associated with the trait. Furthermore, QQ-plots assume the sample distribution of the test statistic (χ^2). However, when a small sample size is used the assumptions of the test statistic are broken. Therefore, I observe more SNPs deviating from the null line then expected by chance given a larger sample size.

Another factor which I did not have the data available to explore is the effect of the total number of individuals on the case-control dataset. It would be interesting to repeat the simulation analysis on a larger sample size to investigate the effect of total sample size on the rates of obtaining false positive SNPs in addition to varying the proportion of cases.

In the second simulation, which defined cases as "all individuals which had the minor allele" a clear peak was produced. This peak is due to having a highly significant SNP. The SNPs within the peak had a similar allele frequency to those used to define the cases. Subsequently, as the loci are in LD with each other it caused the neighbouring SNPs to be drawn up into a peak. It is likely that the allele frequency in

this small population is having a large effect, since the MAF determines the power to detect genotypes associated with a particular phenotype (Guthery et al., 2007). Therefore, when designing GWAS studies it is important to consider the frequency of the trait, the MAF and the density of the SNP chip in addition to the sample size and the proportion of cases and controls.

5.4.3 Alternative approaches to investigate evidence for selection in the IDEAL population

Since alternative approaches to look for evidence for selection in the IDEAL calves were already being carried out, I compared my candidate SNPs to the regions identified as being under selection by Bahbahani (2015). A region (Chr.13.40606621-40972213) was identified using interpopulation F_{ST} which contained the ECF death candidate SNP Chr.13.40031719. This region had a reduction in the level of heterozygosity of alleles in EASZ compared to the other breeds suggesting that selection has occurred within the region (Bahbahani, 2015). One explanation as to why not all of the candidate SNPs overlapped with the selection analysis is because GWAS is looking for alleles which are at a high frequency in the calves with a rare trait. In contrast, Bahbahani (2015) selection analysis is looking for segregating alleles that are different between different populations.

The composite analysis identified a candidate region from Chr.13.39754654-41333451 which included the ECF death candidate SNP Chr.13.40031719 and a further region at Chr.5.57977594 which was close to the PCV$_{TP}$ candidate SNPs Chr.5.76956329. These regions overlap with a region identified by O'Brien et al. (2014) which shows a signature of selection in *B. indicus* Gir cattle when compared to *B. taurus* Brown Swiss cattle. This strengthens the evidence for selection and decreases the likelihood that other forces such as genetic drift are acting upon this region and generating the population differences observed. It also suggests that I should investigate this region further.

5.4.4 Genes and QTLs

Looking specifically at the genes associated with the ECF death candidate SNPs, I observed that the genes formed a cluster in which the functional annotation was related to infection defence (Table 5.9). For example, the functional annotation term cathelicidins had an EASE score of 9.7×10^{-9}. Cathelicidins are precursor sequences of antimicrobial peptides which are stored in the cytoplasmic granules of neutrophil, leukocytes and are released upon leukocyte activation (Zanetti et al., 1995). Likewise,

the annotation terms antibiotics and antimicrobial could mean that the ECF death candidate SNPs are in regions of the genome which have deleterious effects upon infecting microbes such as protozoa and bacteria. The annotation terms in Table 5.9 also included defence and response, meaning that a reaction is triggered in response to infection, to protect the cell or organism, resulting in a change in state or activity (Binns et al., 2009). Whereas the annotation term 'signal' means that a protein is secreted into the cell surroundings (Binns et al., 2009). It is easily plausible that all these functions are likely to be useful in the defence against *T. parva* infection.

In comparison, the PCV_{TP} candidate SNPs were associated with genes which had functional annotation terms for receptors and binding, which imply a change in the activity of the cell (Binns et al., 2009). This suggests that an active process is required to maintain PCV_{TP} following infection and seroconversion to *T. parva*.

QTLs surrounding the candidate SNPs were identified which have been previously linked to tick resistance (Gasparin et al., 2007). Furthermore, a QTL previously associated with parasite detection (number of times an individual was detected to be infected with *Trypanosoma conglonse*) was also identified in the area surrounding the candidate SNPs (Hanotte et al., 2003). In addition, one candidate SNP from the ECF death and PCV_{TP} GWAS were in close proximity to each other (ECF death GWAS: Chr.13.40031719, P value = 0.218; PCV_{TP} GWAS: Chr.13.33366445, P value = 1.53 x 10^{-5}). Since, Hanotte et al. (2003) also identified QTL in the area of the PCV_{TP} candidate SNP Chr.13.33366445 associated with change in PCV following trypanosome challenge, this furthers the idea that this region is important in parasite response. Whilst there is not a high enough prevalence of *Trypanosoma* spp. within this population for a direct comparison, in the case of ECF, it is known that PCV is negatively associated with *T. parva* infection (Conradie van Wyk et al., 2014). In addition, PCV is associated with EASZ calf mortality (Thumbi, 2012) and calves which died during the IDEAL study had a lower PCV than those which survived until the end of the study period (Woolhouse et al., 2015). Therefore, further work is required to determine the mechanism which allows individuals to maintain their PCV and tolerate *T. parva* infection.

5.4.5 Comparing minor allele frequencies in different breeds

All of the candidate SNP alleles are rare and are most commonly found in heterozygotes. In addition, cases generally had a higher MAF than controls. Moreover, cases have a more similar MAF to the European breeds than the controls. European taurine breeds such as Holstein-Friesians and Jersey cattle have not been exposed to the selection pressure exhibited by ECF and so they can exist with a higher MAF.

In contrast, the *B. indicus* (zebu) breeds originated from Asia where *T. annulata* is present, and therefore zebu cattle evolved under a similar selection pressure and it is likely that they brought the tolerance allele with them to Africa (Coetzer and Tustin, 2004; Decker et al., 2014; Hanotte et al., 2002).

Differences in MAF in other populations exposed to different selection pressures is well known, for example resistance to malaria in humans as described in the introduction (Durand and Coetzer, 2008). In cattle, gene expression differences in macrophages have been identified in *B. inidicus* from the endemic areas and exotic *B. taurus* breeds in response to *T. annulata* infection (Glass and Jensen, 2007).

5.4.6 Conclusion

In conclusion, I have found no conclusive evidence for a genetic predisposition to ECF death or association with PCV_{TP}. However, evidence for selection exists in EASZ (Bahbahani, 2015) and the MAF differs between EASZ and European breeds, suggesting that EASZ are adapted to an environment with a high natural selection pressure such as ECF. The difference between the results is likely to be due to the sample size needed to power the analysis. Therefore if this study was repeated with a larger sample size and a denser SNP chip which focuses on zebu polymorphisms rather than taurine ones it may be possible to identify a genetic predisposition to ECF death and PCV_{TP} using GWAS.

The effect of the number of cases on GWAS and significance thresholds has been investigated further in this chapter. I have shown that the number of cases is important in case-control GWAS, as the probability of a SNP being associated with a phenotypic trait is reflected by the proportion of cases. Therefore, the empirical threshold is a more suitable significance threshold than the Bonferroni correction, when analysing data from small sample sizes since it reflects the distribution of the data in terms of the number of SNPs and the proportion of cases and controls.

Chapter 6

General Discussion

The aim of this thesis was to investigate the evidence for coinfections and to generate a better understanding of the genetic and phenotypic factors affecting coinfections and their consequences for East African shorthorn zebu (EASZ) in Western Kenya using data gathered by the Infectious Diseases in East African Livestock (IDEAL) project. I have shown that coinfection is complex, with both positive and negative associations occurring between pairs of parasites. In addition, there are environmental, seasonal and host level factors which affect infection risk and the consequence of infection as well as influencing the likelihood of coinfection.

This discussion is split into three sections. The first section emphasises the key results and draws attention to new findings which were not known prior to commencing this thesis. The second section discusses the implications of these findings in the context of EASZ in Western Kenya and treatment to improve cattle health and productivity in this region, whilst simultaneously reviewing my findings in consideration of other research carried out using the IDEAL data. Lastly, in the third section, I highlight the questions raised by this thesis and provide suggestions for the direction of future work before drawing the thesis to a conclusion.

6.1 Key results

Following the structure of the literature review in Chapter 1, here I present the key results of the thesis by summarising the evidence for coinfections in the IDEAL study before portraying the effect of the host's history of exposure on coinfection in addition to the effect of environmental, seasonal, genetic and phenotypic variation on infection. Lastly, in this section I also describe the impacts of infection upon the host.

a) The occurrence of coinfections

Coinfection was common within the IDEAL population. Focusing on the most prevalent, routinely-sampled parasites *Theileria* spp., *Coccidia* spp., *Strongyloides* spp., strongyles and *Calicophoron* spp. I found evidence for both positive and negative associations between pairs of parasites. For example, concurrent infection with *Strongyloides* spp. was associated with an increase in the risk of strongyle infection. Conversely, in other cases being infected with one parasite decreased the calf's risk of infection with another parasite: for example concurrent infection with *Strongyloides* spp. was associated with a decrease in the risk of *Calicophoron* spp. infection.

Similarly, positive associations were identified between the respiratory viruses infectious bovine rhinotracheitis (IBR), bovine parainfluenza virus Type 3 (PIV3) and bovine viral diarrhoea virus (BVDV), which suggests that IBR, PIV3 and BVDV are co-distributed. This result suggests that there is an increased risk of an individual being seropositive for one of these viruses if the calf is also seropositive for one or both of the other viruses. Furthermore, coinfection with any of the other parasites in the IDEAL project does not increase the risk of being IBR or PIV3 seropositive. However, I detected a previously unidentified association between seroconversion to *Babesia bigemina* and BVDV: *B. bigemina* seropositive individuals were more likely to be BVDV seropositive then seronegative individuals.

b) Effect of host history of exposure

Coinfection in the IDEAL population occurred between parasites at concurrent time points. Interspecific associations between different parasites at previous time points were rare: I only identified one interspecific association which involved a time lag between *Theileria* spp. 5 weeks prior to the current visit (T_{-5}) and *Coccidia* spp. at the current visit, (T_0), with *Theileria* spp. T_{-5} increasing the risk of infection with *Coccidia* spp at T_0. However, this effect was no longer apparent once environmental and seasonal variation was accounted for.

c) Effect of environmental and seasonal variation

Inclusion of indicators of calf condition, seasonal and environmental variation in the parasite-parasite association models improved model fit. However, the majority of associations between parasites in Chapter 2 remained unchanged after accounting for these variables. In addition, the inclusion of environmental confounders into the virus models in Chapter 3 did not affect the relationship observed between the serostatus of the three viruses.

d) Effect of host phenotype

This thesis has shown that the main phenotypic effect on infection risk and outcome in EASZ is through the recruitment weight of the calf. Heavier calves at recruitment were less likely to be infected with *Strongyloides* spp. than lighter calves and had a lower estimate of strongyle eggs per gramme of faeces (EPG) later in life (at aged 16-51 weeks old). In addition, heavier calves at recruitment were marginally less likely to experience a clinical episode than lighter calves. However, in contrast, heavier calves at recruitment were at *increased* risk of being IBR seropositive.

e) Effect of host genotype

Host genotype can also determine an individual's infection risk. I found that additive genetic variance contributed the most (after residual variance, and compared to permanent environmental variance and sublocation variance) to the overall phenotypic variance in strongyle EPG, which resulted in heritable variation in strongyle EPG in EASZ (h^2=23.9%, SE=11.8%). Likewise, I identified significant additive genetic variance in white blood cell counts, giving a heritability of 27.6% (SE=10.6%).

I used genome wide association studies (GWAS) to investigate genetic predisposition to East Coast Fever (ECF) death and to evaluate the genetic basis of the packed cell volume at the time of seroconversion to *T. parva* (PCV_{TP}). The GWAS suggested that there is no robust evidence for a genetic predisposition to ECF death in the IDEAL population; the GWAS results followed the same distribution of that of a simulated null model. However, the minor allele frequency differed between EASZ and European breeds, and Bahbahani's 2015 selection analysis suggests that EASZ are adapted to an environment with a high natural selection pressure such as ECF. I found no evidence for a genetic association between packed cell volume at the time of seroconversion to *Theileria parva* (PCV_{TP}) and the genotyped loci.

f) Consequence of infection

Whilst other authors, e.g. Conradie van Wyk et al. (2014) and Thumbi et al. (2013b, 2014), have investigated the effect of coinfection in EASZ on growth, mortality and haematological parameters, this thesis investigates the consequences of infection on the host by looking at the impact of IBR, PIV3 and BVDV serostatus on the average daily weight gain (ADWG) of the calf and clinical illness. I showed that there was no association between IBR, PIV3 and BVDV serostatus and weight gain. However, IBR seropositive calves were less likely to experience any kind of clinical episode than IBR seronegative calves. This effect was independent of the other two viruses serostatus

and the relationship was not confounded by environmental variation or coinfection with pathogenic parasites such as *Theileria parva* and *Babesia bigemina*.

In addition to looking at the effect of exposure to multiple viruses on the host, I also investigated the impact of strongyle worm burden on the calf. Calves with a higher strongyle EPG at a given age tended to have a lower red blood cell count (RBC), total serum protein (TSP), and absolute eosinophil count (EO) than those with a lower strongyle EPG. Furthermore, calves with a high strongyle EPG were also lighter than those with a lower EPG, which indicates that an increase in strongyle EPG was associated with a decrease in weight.

6.2 Implications of the thesis

The main practical implication of my thesis is its relevance to livestock health campaigns. Since cattle in the IDEAL study region are minimally managed and there is very limited use of vaccination or other preventative measures against infectious diseases, the data analysed in this thesis can be used to provide ideas for setting treatment priorities and schemes which could be efficient in removing parasites. This could be achieved by selecting a treatment plan to target one parasite which was positively associated with several other parasites, and so the removal of one coinfecting species might result in the decrease in prevalence of several other parasites. In turn, knowledge of coinfecting parasites could also help avoid the situation where an livestock health campaign could have net negative consequences if the removal of one species inadvertently promotes another species which is negatively associated with the treated parasite (Lafferty, 2010). Therefore this section examines the potential treatments which could be used to improve cattle health in Western Kenya.

6.2.1 Parasite treatment and control

My thesis indicates that helminth control would be beneficial in this population. Calves with a high strongyle EPG were lighter than those with a lower EPG. By lowering the worm burden an individual experiences, it may be possible to increase the weight gain of the calf and thus increase its productivity. Most helminth control is non-specific, therefore treatment with anthelminthics such as albendazole or ivermectin will remove developing larvae and adult worms of many different species (Aiello and Moses, 2012; Urquhart et al., 1996). Since I identified a negative association between *Coccidia* spp. and *Calicophoron* spp. before the addition of environmental variables, there is the possibility, if this is a causal relationship, that the removal of helminths will increase

the burden of *Coccidia* spp. in this population. Experimental treatment of wild mice with ivermectin has shown a similar shift in the parasite community of the host. Mice treated with ivermectin had a reduced nematode infection and an increased *Coccidia* spp. infection (Pedersen and Antonovics, 2013), however the mechanism behind this association remains unclear. Following the results of Chapter 2, I would not expect the promotion of any other parasites, apart from *Coccidia* spp., following the treatment with anthelminthics.

Theileria spp. were not associated with any other parasites analyzed in Chapter 2, after accounting for indicators of calf condition, environmental and seasonal variation. Therefore the removal of *Theileria* spp. should not have any adverse consequences upon the other parasites studied by inadvertently promoting them through removing a competing species. Moreover, it has been shown that *Theileria* spp. infection has a strong impact upon growth rates and mortality, therefore its removal should increase the productivity in this region (Thumbi et al., 2013b, 2014). The traditional method to control ticks is through dipping cattle in acaricide preparations. The topical application of acaricides to control tick infections will result in the removal of all tick species, not just the *Rhipicephalus appendiculatus* and *Amblyomma* spp. which transmit *Theileria* spp. but also other ticks such as *Boophilius* spp., which transmit *Babesia bigemina*. Since Chapter 3 identified a positive association between *B. bigemina* and BVDV, removal of *B. bigemina* could, if this is a causal relationship, result in a decrease in BVDV seroprevalence, which may in turn result in a decrease in the seroprevalence of IBR and PIV3 as these viruses are positively associated with BVDV.

However, there are many problems associated with the topical application of acaricides. Their application is not environmentally friendly and ticks can develop resistance to them, plus in areas of high tick challenge, such as the IDEAL study region, cattle need to be treated two or three times a week, which is not feasible (Babo Martins et al., 2010; de Castro, 1997). Furthermore, the application of intensive treatments could upset the endemic stability to all tick-borne diseases within an area as it disrupts the acquirement of immunity, which can result in an increase in tick-borne disease if the treatment is suddenly stopped (Coetzer and Tustin, 2004; Jonsson et al., 2012; Rynkiewicz et al., 2015).

Kobylinski et al. (2014) has suggested that a combined treatment for malaria and soil-transmitted helminths such as *Ascaris lumbricoides* would work as a cost-effective treatment in humans, as both parasites occur in the same region and helminths increase malaria infection risk. Although I found no association between *Theileria* spp. and helminths in Chapter 2, it is possible that if I had used the presence/absence of specific *Theileria* species such as *T. parva* at each visit (instead of all species combined) then I

would have obtained a different result.

In order to achieve a combined *Theileria*-helminth treatment scheme, a more specific treatment directed to *T. parva* (which has the most severe consequences for EASZ) such as using chemotherapy to treat clinically apparent ECF or inoculating cattle using a live vaccine would be required. Both the chemotherapy and vaccination method have been combined into an effective ECF control called the 'Infection and Treatment Method' (ITM); cattle are inoculated with a vaccine combining several different live strains of *T. parva*, known as a Muguga Cocktail, with the concurrent administration of an antimicrobial, which results in long lasting acquired immune response, however immunized individuals will become carriers (Di Giulio et al., 2009; Uilenberg, 1999). A field trial carried out in Tanzania by Babo Martins et al. (2010) showed that the ITM method resulted in an increase in the profitability of vaccinated cattle and that ITM was a better control method than natural infection and subsequent treatment. Therefore, by adding an anthelminthic into the ITM treatment mixture it could act as a cost-effective treatment to *Theileria* spp. and helminth control.

In order to treat IBR, PIV3 and BVDV a combined vaccine is best to have optimal chance of removal of all three viruses. A combined vaccine has already been designed and is currently licensed for use in cattle in the UK (Rispoval 4, Zoetis UK Limited). However, treatment of these three viruses in this population is nonessential as I found no association between IBR, PIV3 and BVDV serostatus and calf growth rate or PIV3 and BVDV serostatus and clinical illness. Moreover, assuming a causal relationship, seroconversion to IBR is protecting the IDEAL calves against clinical illness. In addition, there are other parasites such as *T. parva* and strongyles which are having more severe effects upon the calf.

In summary, given the limited resources devoted to animal health in this setting, selection of parasites for treatment to increase production should be based upon parasites which are having most effect upon the calf. Therefore, I would support suggestions to prioritize of the treatment of *T. parva* and helminths.

6.2.2 Increasing recruitment weight

A interesting finding of this thesis is that recruitment weight is important: heavier calves at recruitment (when they are 3-7 days old) have decreased infection risk (apart from IBR) and experience fewer clinical episodes then lighter calves. The mechanism by which recruitment weight is having an effect is unknown, as recruitment weight could be reflecting many different things. For example is it the weight of the calf itself that is important? Or is recruitment weight acting as an indicator of the general condition of the calf or the dam or of a good environment?

A calf with a low recruitment weight could be in overall poor condition and so more likely to become infected as a result of being in poor condition, i.e. Beldomenico and Begon's (2010) vicious circle. In addition, heavier calves are more likely to be born to larger dams or dams in better condition, and it is possible that better condition dams could be providing more resources in the form of higher quality colostrum which contains maternal antibodies to help fight infection at an early age.

By being in a good environment the calf could have more resources around it, so the calf will be heavier as a result and better at fighting infection. In the IDEAL population these resources could be access to water at the homestead, supplementary feeding or access to a stall which could affect calf recruitment weight through the dam's condition. In contrast, heavier calves at recruitment may also be born in a region that has a lower infectious disease burden; therefore they would be less likely to experience infection later in life. However, I found that the occurrence of different parasites was fairly homogenous across the study region so this is unlikely. In addition, the effect of recruitment weight on infection risk still held when other factors were included in the model, so the association is not completely confounded by the environment.

If there is a causal effect of recruitment weight upon infection risk then one possible method to treat cattle in this region is to provide supplementary feeding to pregnant cows or lightweight calves and therefore reduce health problems later on in a calf's life.

6.2.3   Genetic selection

An alternative method to disease control is using selection. I have identified heritable variation in strongyle EPG therefore it should be possible to selectively breed for individuals with a low worm burden. In addition, selective breeding for both disease and worm resistance has been achieved in experimental sheep flocks (Bishop and Morris, 2007; Kemper et al., 2010); the same result might be possible for EASZ. However, breeding for parasite resistance is likely to be a long process with many challenges along the way. These challenges have been reviewed by Bishop (2012) but they include the uncertainty in the outcome of selection if parasite resistance is correlated with other traits. Therefore, genetic selection is not a short-term or guaranteed answer in the treatment of infection within this population.

EASZ often experience insignificant or subclinical effects of infections with pathogenic parasites such as *T. parva* in comparisons to European breeds which are severely affected by ECF (Coetzer and Tustin, 2004; Kariuki et al., 1995; Moll et al., 1984; Mukhebi et al., 1992). Therefore, genetic selection in this population should be managed carefully so that the diversity which contributes to the yet undiscovered

resistant genes to *T. parva* is not lost (Hanotte et al., 2010).

6.2.4 How do my results fit with the other IDEAL project findings?

Other PhD theses (Conradie van Wyk, 2012; Jennings, 2013; Thumbi, 2012) using data gathered by the IDEAL project have reviewed the consequences of coinfection or single infection on the host (see Appendix A). My thesis has turned this question around to ask what causes hosts to be infected with multiple parasites in the first place and to identify parasites which are associated with each other. I have identified roles of various aspects of the host's genotype and phenotype in addition to particular environmental and seasonal factors which contribute to the causes of coinfection. Furthermore, I have identified associations occurring between many parasites infecting EASZ which were unknown in this population before I started my thesis.

6.3 Questions raised by the thesis and future studies

Since the IDEAL project only followed individuals over the first year of life, only the short-term effects of infection and coinfection can be investigated. In this time period, I have shown that infection is having an impact upon the calves: for example I have observed that calves with a high strongyle EPG were lighter than those with a lower EPG. It would therefore be interesting to extend the IDEAL project for longer and follow the cohort for the whole of their lifetime to see the long-term carry over effects of infection at a young age on fitness and life history traits. In addition, the use of a long-term study design could allow the maternal effects of the dam on the calf to be investigated in more detail, which would help to disentangle the effect of birth weight upon later infection risk.

With a 5 week interval between observations I could be missing the impact of acute infections within this population. I am curious to know what pattern of associations would be identified if the IDEAL project was repeated with a shorter time interval between visits. In addition, it would be easier to understand the associations occurring between the respiratory viruses, other parasites and clinical episodes if I knew the age at which individuals were exposed to the respiratory viruses. Therefore, further work could involve testing the stored blood samples, from the biobank created during the IDEAL project, for the respiratory viruses at every routine visit.

Since I have identified both positive and negative associations occurring between parasites in the IDEAL project, follow-up studies could be designed to investigate the mechanisms of association occurring between parasites. For example, this could

include experiments which treat individuals with anthelminthics to see what effect the removal of helminths has on other parasites in the study system c.f. Pedersen and Antonovics (2013). Following on from this study, *T. parva* could also be treated to evaluate the effect of a combined helminth-*Theileria* treatment plan on the calf. During both of these experiments the immunological response to coinfection could also be investigated with e.g. cytokine levels such as IFN-γ, IL-10 and IL-12 being compared in the single infection and coinfected state. Another *in vivo* study could involve experimental infection with BVDV and *Babesia bigemina* to determine the association between these two parasites. Lastly, *in vivo* experiments could be designed to investigate the mechanism for the protective effect of IBR serostatus on clinical episodes.

In addition, there are many other parasites such as the protozoa *Trypanosoma* spp. and the rickettsia *Ehrlichia ruminantium* (Heartwarter) which have severe consequences for the calves and could be having a strong effect upon the other parasites identified in the study system. For example, both Kaufmann et al. (1992) and Dwinger et al. (1994) found that N'Dama cattle infected with *Trypanosoma* spp. have higher strongyle egg counts then cattle which are not infected with *Trypanosoma* spp. and that coinfected cattle had a reduced ability to mount a normal response to infection with *Haemonchus contortus*. It is hypothesised that this association occurs via immunosuppression and Th1-Th2 trade-offs, whereby the Th1 immune response elicited by trypanosome infection would adversely influence the Th2 immune response to helminth infection (Chiejina et al., 2003). However, since both *Trypanosoma* spp. and *E. ruminantium* have a low prevalence (<10%) within this population, which is below the parasite selection criteria I used, I did not include them in my analysis. A more focussed study could be designed to evaluate coinfection with *Trypanosoma* spp. and *E. ruminantium* in more detail.

This thesis focused on the binary association between the presence/absence of parasites. It should be possible to extend my analysis, using data already available from the IDEAL project, to look at the effect of coinfection on other aspects of infection biology. One such question could be: does infection with one parasite alter the length of infection or clinical severity of a second parasite? In addition, further work could look at the effect of parasite abundance or diversity of infections rather than presence/absence of a parasite on coinfection and its consequence for the host.

Furthermore, there is lots of potential to extend the genetic component of this thesis. With a higher-density SNP chip which focuses on zebu infection polymorphisms and a much larger sample size to increase the study power, it may be possible to identify a genetic predisposition to East Coast Fever death and to evaluate

the genetic basis to coinfection.

6.4   Conclusion

To conclude, there are many different parasites infecting EASZ, and coinfections occur amongst them. This thesis provides an understanding of these parasite-parasite associations, and of their causes and consequences. I also discuss these findings in terms of their implications for disease control in EASZ in Western Kenya. I suggest that the most feasible treatment to improve the health and production of EASZ include the integrated control of worm and *T. parva* by the use of anthelminthics and ITM. In addition, it may also be possible to use supplementary feeding of lightweight individuals or genetic selection for parasite resistance to improve the health of EASZ.

References

Abbas, A., Murphy, K., and Sher, A. (1996). Functional diversity of helper T lymphocytes. *Nature*, 383(6603):787–793.

Aiello, S. and Moses, M. A. (2012). *The Merck veterinary manual online*. Merck Publishing Group, Whitehouse Station, N.J., U.S.A.

Ameni, G., Aseffa, A., Engers, H., Young, D., Gordon, S., Hewinson, G., and Vordermeier, M. (2007). High prevalence and increased severity of pathology of bovine tuberculosis in Holsteins compared to Zebu breeds under field cattle husbandry in Central Ethiopia. *Clinical and Vaccine Immunology*, 14(10):1356–1361.

Anderson, R. M., Truscott, J. E., Pullan, R. L., Brooker, S. J., and Hollingsworth, T. D. (2013). How effective is school-based deworming for the community-wide control of soil-transmitted helminths. *PLoS Neglected Tropical Diseases*, 7(2):e2027.

Antia, R., Regoes, R. R., Koella, J. C., and Bergstrom, C. T. (2003). The role of evolution in the emergence of infectious diseases. *Nature*, 426(6967):658–661.

Aulchenko, Y. S., Ripke, S., Isaacs, A., and Van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23(10):1294–1296.

Austrialian Society for Parasitology Inc. (2015). Para-site. www.parasite.org.au/para-site.

Babo Martins, S., Di Giulio, G., Lynen, G., Peters, A., and Rushton, J. (2010). Assessing the impact of East Coast Fever immunisation by the infection and treatment method in Tanzanian pastoralist systems. *Preventive Veterinary Medicine*, 97(3):175–182.

Bahbahani, H. (2015). *Genome-wide identification of signatures of positive selection in African admixed zebu cattle*. PhD thesis, School of Life Sciences, University of Nottingham.

Barlow, N. (1996). The ecology of wildlife disease control: simple models revisited. *Journal of Applied Ecology*, pages 303–314.

Batchelor, N. A., Atkinson, P. M., Gething, P. W., Picozzi, K., Fèvre, E. M., Kakembo, A. S. L., and Welburn, S. C. (2009). Spatial predictions of Rhodesian Human African Trypanosomiasis (Sleeping Sickness) prevalence in Kaberamaido and Dokolo, two newly affected districts of Uganda. *PLoS Neglected Tropical Diseases*, 3(12):e563.

Bates, D., Martin, M., Ben, B., and Walker, S. (2014). lme4: Linear mixed effects models using Eigen and S4.(r package v. 1.0–6).

Behnke, J. M., Gilbert, F. S., Abu-Madi, M. A., and Lewis, J. W. (2005). Do the helminth parasites of wood mice interact? *Journal of Animal Ecology*, 74(5):982–993.

Beldomenico, P. M. and Begon, M. (2010). Disease spread, susceptibility and infection intensity: vicious circles? *Trends in Ecology & Evolution*, 25(1):21–27.

Beldomenico, P. M., Telfer, S., Gebert, S., Lukomski, L., Bennett, M., and Begon, M. (2008). Poor condition and infection: a vicious circle in natural populations. *Proceedings of the Royal Society B: Biological Sciences*, 275(1644):1753–1759.

Beldomenico, P. M., Telfer, S., Lukomski, L., Gebert, S., Bennett, M., and Begon, M. (2009). Host condition and individual risk of cowpox virus infection in natural animal populations: cause or effect? *Epidemiology and Infection*, 137(9):1295–1301.

Beraldi, D., McRae, A. F., Gratten, J., Pilkington, J. G., Slate, J., Visscher, P. M., and Pemberton, J. M. (2007). Quantitative trait loci (QTL) mapping of resistance to strongyles and coccidia in the free-living Soay sheep (*Ovis aries*). *International Journal for Parasitology*, 37(1):121–129.

Bèrènos, C., Ellis, P. A., Pilkington, J. G., and Pemberton, J. M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular Ecology*, 23:3434–3451.

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, 2(3):47–53.

Bermingham, M. L., Bishop, S. C., Woolliams, J. A., Pong-Wong, R., Allen, A. R., McBride, S. H., Ryder, J. J., Wright, D. M., Skuce, R. A., McDowell, S. W., and Glass, E. J. (2014). Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity*, 112:543–551.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., and Apweiler, R. (2009). QuickGo: a web-based tool for geneontology searching. *Bioinformatics*, 25(22):3045–3046.

Bishop, S. C. (2012). Possibilities to breed for resistance to nematode parasite infections in small ruminants in tropical production systems. *Animal*, 6(5):741–747.

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

Bishop, S. C. and Morris, C. A. (2007). Genetics of disease resistance in sheep and goats. *Small Ruminant Research*, 70(1):48–59.

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

Bishop, S. C. and Woolliams, J. A. (2014). Genomics and disease resistance studies in livestock. *Livestock Science*, 166:190–198.

Bock, R. E., de Vos, A. J., Kingston, T. G., and McLellan, D. J. (1997). Effect of breed of cattle on innate resistance to infection with *Babesia bovis*, *B. bigemina* and *Anaplasma marginale*. *Australian Veterinary Journal*, 75(5):337–340.

Booth, M., Vennervald, B. J., Kenty, L., Butterworth, A. E., Kariuki, H. C., Kadzo, H., Ireri, E., Amaganga, C., Kimani, G., Mwatha, J. K., et al. (2004). Micro-geographical variation in exposure to *Schistosoma mansoni* and malaria, and exacerbation of splenomegaly in Kenyan school-aged children. *BMC Infectious Diseases*, 4(1):13.

Bowland, S. L. and Shewen, P. E. (2000). Bovine respiratory disease: commercial vaccines currently available in Canada. *The Canadian Veterinary Journal*, 41(1):33–48.

Box, G., Jenkins, G., and Reinsel, G. (1994). *Time series analysis: forecasting and control*. Holden-Day, 3rd edition.

Bradshaw, B. J. and Edwards, S. (1996). Antibody isotype responses to experimental infection with bovine herpesvirus 1 in calves with colostrally derived antibody. *Veterinary Microbiology*, 53(1):143–151.

Bricarello, P. A., Zaros, L. G., Coutinho, L. L., Rocha, R. A., Kooyman, F. N. J., De Vries, E., Gonçalves, J. R. S., Lima, L. G., Pires, A. V., and Amarante, A. F. T. (2007). Field study on nematode resistance in Nelore-breed cattle. *Veterinary Parasitology*, 148:272–278.

Brock, K. V. (2003). The persistence of bovine viral diarrhea virus. *Biologicals*, 31(2):133–135.

Bronsvoort, B. M. d. C., Tanya, V., Kitching, R., Nfon, C., Hamman, S., and Morgan, K. (2003). Foot and mouth disease and livestock husbandry practices in the Adamawa province of Cameroon. *Tropical Animal Health and Production*, 35(6):491–507.

Bronsvoort, B. M. d. C., Thumbi, S., Poole, E., Kiara, H., Auguet, O., Handel, I., Jennings, A., Conradie, I., Mbole-Kariuki, M., Toye, P., Hanotte, O., Coetzer, J., and Woolhouse, M. (2013). Design and descriptive epidemiology of the Infectious Diseases of East African Livestock (IDEAL) project, a longitudinal calf cohort study in Western Kenya. *BMC Veterinary Research*, 9(1):171.

Brooker, S. (2010). Estimating the global distribution and disease burden of intestinal nematode infections: adding up the numbers-a review. *International Journal for Parasitology*, 40(10):1137–1144.

Brooker, S. and Clements, A. C. A. (2009). Spatial heterogeneity of parasite co-infection: Determinants and geostatistical prediction at regional scales. *International Journal for Parasitology*, 39(5):591–597.

Brooker, S. J., Pullan, R. L., Gitonga, C. W., Ashton, R. A., Kolaczinski, J. H., Kabatereine, N. B., and Snow, R. W. (2012). Plasmodium-helminth coinfection and its sources of heterogeneity across East Africa. *Journal of Infectious Diseases*, 205(5):841–852.

Brun-Hansen, H., Christensson, D., Eide, D., and Grønstøl, H. (1998). Experimental infection with *Babesia divergens* in cattle persistently infected with bovine virus diarrhoea virus. *Journal of Veterinary Medicine, Series B*, 45(1-10):269–277.

Bryan, R. P. and Kerr, J. D. (1989). The relation between the natural worm burden of steers and the faecal egg count differentiated to species. *Veterinary Parasitology*, 30(4):327–334.

Callaby, R., Hanotte, O., Conradie van Wyk, I., Kiara, H., Toye, P., Mbole-Kariuki, M. N., Jennings, A., Thumbi, S. M., Coetzer, J. A. W., Bronsvoort, B. M. d. C., Knott, S. A., Woolhouse, M. E. J., and Kruuk, L. E. B. (2015). Variation and covariation in strongyle infection in East African shorthorn zebu calves. *Parasitology*, 142(03):499–511.

Campbell, E., Scadding, J. G., and Roberts, R. (1979). The concept of disease. *British Medical Journal*, 2(6193):757–762.

Casadevall, A. and Pirofski, L.-a. (1999). Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infection and Immunity*, 67(8):3703–3713.

Casadevall, A. and Pirofski, L.-a. (2000). Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infection and Immunity*, 68(12):6511–6518.

Cattadori, I., Boag, B., and Hudson, P. (2008). Parasite co-infection and interaction as drivers of host heterogeneity. *International Journal for Parasitology*, 38(3):371–380.

CDC website (2015). www.cdc.gov/parasites.

Chatrath, R., Saili, A., Jain, M., and Dutta, A. (1997). Immune status of full-term small-for-gestational age neonates in India. *Journal of Tropical Pediatrics*, 43(6):345–348.

Chen, W.-M. and Abecasis, G. R. (2007). Family-based association tests for genomewide association scans. *The American Journal of Human Genetics*, 81(5):913–926.

Chiejina, S., Wakelin, D., and Goyal, P. (2003). Trypanosome-induced modulation of responses to concurrent helminth infection. *Research in Veterinary Science*, 74(1):47–53.

Chongsuvivatwong, V. (2012). Epicalc: Epidemiological calculator.

Christensen, N., Nansen, P., Fagbemi, B. O., and Monrad, J. (1987). Heterologous antagonistic and synergistic interactions between helminths and between helminths and protozoans in concurrent experimental infection of mammalian hosts. *Parasitology Research*, 73(5):387–410.

Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, 138(3):963–971.

Cizauskas, C. A. (2013). *The ecology of anthrax and coinfection trade-offs from an immunological perspective: Seasonal aspects of host susceptibility*. PhD thesis, Environmental Science, Policy and Management, UC Berkeley.

Cizauskas, C. A., Turner, W. C., Wagner, B., Küsters, M., Vance, R. E., and Getz, W. M. (2014). Gastrointestinal helminths may affect host susceptibility to anthrax through seasonal immune trade-offs. *BMC Ecology*, 14(1):27.

Clapperton, M., Diack, A., Matika, O., Glass, E., Gladney, C., Mellencamp, M., Hoste, A., and Bishop, S. (2009). Traits associated with innate and adaptive immunity in pigs: heritability and associations with performance under different health status conditions. *Genetics Selection Evolution*, 41(1):1–11.

Clements, M. N., Clutton-Brock, T. H., Guinness, F. E., Pemberton, J. M., and Kruuk, L. E. B. (2011). Variances and covariances of phenological traits in a wild mammal population. *Evolution*, 65(3):788–801.

Coelingh, K., Winter, C., Murphy, B., Rice, J., Kimball, P., Olmsted, R., and Collins, P. (1986). Conserved epitopes on the hemagglutinin-neuraminidase proteins of human and bovine parainfluenza type 3 viruses: nucleotide sequence analysis of variants selected with monoclonal antibodies. *Journal of Virology*, 60(1):90–96.

Coetzer, J. A. W. and Tustin, R. C. (2004). *Infectious diseases of livestock.* Oxford University Press, Cape Town, Southern Africa.

Coltman, D. W., Pilkington, J., Kruuk, L. E. B., Wilson, K., and Pemberton, J. M. (2001). Positive genetic correlation between parasite resistance and body size in a free-living ungulate population. *Evolution*, 55(10):2116–2125.

Coltman, D. W., Pilkington, J. G., Smith, J. A., and Pemberton, J. M. (1999). Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution*, 53(4):1259–1267.

Conradie van Wyk, I. (2012). *Anaemia in East African short-horn zebu calves: field diagnosis, infectious causes and pathogen interactions.* PhD thesis, University of Pretoria.

Conradie van Wyk, I., Goddard, A., Bronsvoort, B. M. d. C., Coetzer, J., Booth, C., Hanotte, O., Jennings, A., Kiara, H., Mashego, P., Muller, C., Pretorius, G., Poole, E., Thumbi, S., Toye, P., Woolhouse, M., and Penzhorn, B. (2012). Hematological profile of East African short-horn zebu calves from birth to 51 weeks of age. *Comparative Clinical Pathology*, 22:1029–1036.

Conradie van Wyk, I., Goddard, A., Bronsvoort, B. M. d. C., Coetzer, J., Handel, I., Hanotte, O., Jennings, A., Lesosky, M., Kiara, H., Thumbi, S., Toye, P., Woolhouse, M., and Penzhorn, B. (2014). The impact of co-infections on the haematological profile of East African short-horn zebu calves. *Parasitology*, 141(3):374–388.

Cox, F. E. G. (2001). Concomitant infections, parasites and immune responses. *Parasitology*, 122(SupplementS1):S23–S38.

Craig, B. H., Tempest, L. J., Pilkington, J. G., and Pemberton, J. M. (2008). Metazoan-protozoan parasite co-infections and host body weight in St Kilda Soay sheep. *Parasitology*, 135(4):433–441.

Crawford, A., Paterson, K., Dodds, K., Diez Tascon, C., Williamson, P., Roberts Thomson, M., Bisset, S., Beattie, A., Greer, G., Green, R., Wheeler, R., Shaw, R., Knowler, K., and McEwan, J. (2006). Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: i. analysis of outcross pedigrees. *BMC Genomics*, 7(1):178.

Cross, D. S., Ivacic, L. C., Stefanski, E. L., and McCarty, C. A. (2010). Population based allele frequencies of disease associated polymorphisms in the Personalized Medicine Research Project. *BMC Genetics*, 11(1):51.

Davies, T. J. and Pedersen, A. B. (2008). Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proceedings of the Royal Society B: Biological Sciences*, 275(1643):1695–1701.

Davis, L. R., Bowman, G. W., and Boughton, D. C. (1957). The endogenous development of *Eimeria alabamensis* Christensen, 1941, an intranuclear coccidium of cattle. *The Journal of Protozoology*, 4(4):219–225.

de Castro, J. J. (1997). Sustainable tick and tickborne disease control in livestock improvement in developing countries. *Veterinary Parasitology*, 71(2):77–97.

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., Praharani, L., Babar, M. E., Correia de Almeida Regitano, L., Yildiz, M. A., Heaton, M. P., Liu, W.-S., Lei, C.-Z., Reecy, J. M., Saif-Ur-Rehman, M., Schnabel, R. D., and Taylor, J. F. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genetics*, 10(3):e1004254.

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., Lempicki, R. A., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biology*, 4(5):P3.

Devlin, B. and Roeder, K. (1999). Genomic control for association studies. *Biometrics*, 55(4):997–1004.

Di Giulio, G., Lynen, G., Morzaria, S., Oura, C., and Bishop, R. (2009). Live immunization against East Coast fever-current status. *Trends in Parasitology*, 25(2):85–92.

Díaz, P., Lomba, C., Pedreira, J., Arias, M., Sánchez-Andrade, R., Suárez, J., Díez-Baños, P., Morrondo, P., and Paz-Silva, A. (2006). Analysis of the IgG antibody response against *Paramphistomidae trematoda* in naturally infected cattle: application to serological surveys. *Veterinary Parasitology*, 140(3):281–288.

Durand, P. M. and Coetzer, T. L. (2008). Hereditary red cell disorders and malaria resistance. *Haematologica*, 93(7):961–963.

Durham, P. J. and Hassard, L. E. (1990). Prevalence of antibodies to infectious bovine rhinotracheitis, parainfluenza 3, bovine respiratory syncytial, and bovine viral diarrhea viruses in cattle in Saskatchewan and Alberta. *Canadian Veterinary Journal*, 31(12):815–820.

Dwinger, R., Agyemang, K., Kaufmann, J., Grieve, A., and Bah, M. (1994). Effects of trypanosome and helminth infections on health and production parameters of village N'Dama cattle in the Cambia. *Veterinary Parasitology*, 54(4):353–365.

Ehret, G. B. (2010). Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension. *Current Hypertension Reports*, 12(1):17–25.

Eu-ahsunthornwattana, J., Miller, E. N., Fakiola, M., Jeronimo, S. M., Blackwell, J. M., Cordell, H. J., 2, W. T. C. C. C., et al. (2014). Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genetics*, 10(7):e1004445.

Ezenwa, V. O. and Jolles, A. E. (2011). From host immunity to pathogen invasion: the effects of helminth coinfection on the dynamics of microparasites. *Integrative and Comparative Biology*, 51(4):540–551.

Ezenwa, V. O. and Jolles, A. E. (2015). Opposite effects of anthelmintic treatment on microbial infection at individual versus population scales. *Science*, 347(6218):175–177.

Falconer, D. and Mackay, T. (1996). *Introduction to quantitiative genetics*. Harlow: Longman, 4th edition.

Fenner, F., Henderson, D. A., Arita, I., Jezek, Z., Ladnyi, I. D., and Organization, W. H. (1988). *Smallpox and its eradication*. World Health Organization, Geneva.

Fenton, A., Knowles, S. C., Petchey, O. L., and Pedersen, A. B. (2014). The reliability of observational approaches for detecting interspecific parasite interactions: comparison with experimental results. *International Journal for Parasitology*, 44(7):437–445.

Fenton, A. and Perkins, S. E. (2010). Applying predator-prey theory to modelling immune-mediated, within-host interspecific parasite interactions. *Parasitology*, 137(06):1027–1038.

Fenton, A., Viney, M. E., and Lello, J. (2010). Detecting interspecific macroparasite interactions from ecological data: patterns and process. *Ecology Letters*, 13(5):606–615.

Figueroa-Chávez, D., Segura-Correa, J. C., García-Márquez, L. J., Pescador-Rubio, A., and Valdivia-Flores, A. G. (2012). Detection of antibodies and risk factors for infection with bovine respiratory syncytial virus and parainfluenza virus 3 in dual-purpose farms in Colima, Mexico. *Tropical Animal Health and Production*, 44(7):1417–1421.

Fine, P. E. (1995). Variation in protection by BCG: implications of and for heterologous immunity. *The Lancet*, 346(8986):1339–1345.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kähäri, A. K., Keenan, S., Kulesha, E., Martin, F. J., Maurel, T., McLaren, W. M., Murphy, D. N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H. S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S. J., Vullo, A., Wilder, S. P., Wilson, M., Zadissa, A., Aken, B. L., Birney, E., Cunningham, F., Harrow, J., Herrero, J., Hubbard, T. J., Kinsella, R., Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D. R., and Searle, S. M. (2013). Ensembl 2014. *Nucleic Acids Research*.

Fulton, R. W., Purdy, C. W., Confer, A. W., Saliki, J. T., Loan, R. W., Briggs, R. E., and Burge, L. J. (2000). Bovine viral diarrhea viral infections in feeder calves with respiratory disease: interactions with pasteurella spp., parainfluenza-3 virus,

and bovine respiratory syncytial virus. *Canadian Journal of Veterinary Research*, 64(3):151–159.

Gardner, B., Dolezal, H., Bryant, L., Owens, F., and Smith, R. (1999). Health of finishing steers: effects on performance, carcass traits, and meat tenderness. *Journal of Animal Science*, 77(12):3168–3175.

Garner, C., Tatu, T., Reittie, J., Littlewood, T., Darley, J., Cervino, S., Farrall, M., Kelly, P., Spector, T. D., and Thein, S. L. (2000). Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood*, 95(1):342–346.

Gasparin, G., Miyata, M., Coutinho, L., Martinez, M., Teodoro, R., Furlong, J., Machado, M., Silva, M., Sonstegard, T., and Regitano, L. (2007). Mapping of quantitative trait loci controlling tick *Riphicephalus (Boophilus) microplus* resistance on bovine chromosomes 5, 7 and 14. *Animal Genetics*, 38(5):453–459.

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, 20(15):3128–3143.

Gay, E. and Barnouin, J. (2009). A nation-wide epidemiological study of acute bovine respiratory disease in France. *Preventive Veterinary Medicine*, 89(3):265–271.

Ghirotti, M., Semproni, G., Meneghi, D., Mungaba, F. N., Nannini, D., Calzetta, G., and Paganico, G. (1991). Sero-prevalences of selected cattle diseases in the Kafue flats of Zambia. *Veterinary Research Communications*, 15(1):25–36.

Ghram, A., Reddy, P., Morrill, J., Blecha, F., and Minocha, H. (1989). Bovine herpesvirus-1 and parainfluenza-3 virus interactions: clinical and immunological response in calves. *Canadian Journal of Veterinary Research*, 53(1):62–67.

Gilbert, A. T., Fooks, A. R., Hayman, D. T., Horton, D. L., Müller, T., Plowright, R., Peel, A. J., Bowen, R., Wood, J. L., Mills, J., et al. (2013). Deciphering serology to understand the ecology of infectious diseases in wildlife. *Ecohealth*, 10(3):298–313.

Gilmore, A. R., Gogel, B. J., Cullis, B. R., and Thompson, R. (2006). *ASReml user guide release 3*. VSN International Ltd., Hemel Hamstead.

Glass, E. J. and Jensen, K. (2007). Resistance and susceptibility to a protozoan parasite of cattle-gene expression differences in macrophages from different breeds of cattle. *Veterinary Immunology and Immunopathology*, 120(1):20–30.

Glass, E. J., Preston, P. M., Springbett, A., Craigmile, S., Kirvar, E., Wilkie, G., and Brown, C. D. (2005). *Bos taurus* and *Bos indicus* (Sahiwal) calves respond differently to infection with *Theileria annulata* and produce markedly different levels of acute phase proteins. *International Journal for Parasitology*, 35(3):337–347.

Goddard, M. E. and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6):381–391.

Goldstein, D. B. (2009). Common genetic variation and human traits. *New England Journal of Medicine*, 360(17):1696.

Gorsich, E. E., Ezenwa, V. O., and Jolles, A. E. (2014). Nematode-coccidia parasite co-infections in African buffalo: epidemiology and associations with host condition and pregnancy. *International Journal for Parasitology: Parasites and Wildlife*, 3(2):124–134.

Graham, A. L. (2008). Ecological rules governing helminth-microparasite coinfection. *Proceedings of the National Academy of Sciences*, 105(2):566–570.

Graham, A. L., Cattadori, I. M., Lloyd-Smith, J. O., Ferrari, M. J., and Bjørnstad, O. N. (2007). Transmission consequences of coinfection: cytokines writ large? *Trends in Parasitology*, 23(6):284–291.

Graham, A. L., Lamb, T. J., Read, A. F., and Allen, J. E. (2005). Malaria-filaria coinfection in mice makes malarial disease more severe unless filarial infection achieves patency. *Journal of Infectious Diseases*, 191(3):410–421.

Griffin, D., Chengappa, M., Kuszak, J., and McVey, D. S. (2010). Bacterial pathogens of the bovine respiratory disease complex. *Veterinary Clinics of North America: Food Animal Practice*, 26(2):381–394.

Griffiths, E. C., Pedersen, A. B., Fenton, A., and Petchey, O. L. (2014). Analysis of a summary network of co-infection in humans reveals that parasites interact most via shared resources. *Proceedings of the Royal Society B: Biological Sciences*, 281(1782):20132286.

Grossman, S. R., Shylakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., Lander, E. S., Schaffner, S. F., and Sabeti, P. C. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*, 327(5967):883–886.

Guthery, S. L., Salisbury, B. A., Pungliya, M. S., Stephens, J. C., and Bamshad, M. (2007). The structure of common genetic variation in United States populations. *The American Journal of Human Genetics*, 81(6):1221–1231.

Hammond, D. M., Bowman, G. W., Davis, L. R., and Simms, B. (1946). The endogenous phase of the life cycle of *Eimeria bovis*. *The Journal of Parasitology*, 32(4):409–427.

Hampson, K., Lembo, T., Bessell, P., Auty, H., Packer, C., Halliday, J., Beesley, C. A., Fyumagwa, R., Hoare, R., Ernest, E., et al. (2011). Predictability of anthrax infection in the Serengeti, Tanzania. *Journal of Applied Ecology*, 48(6):1333–1344.

Handel, I. G., Willoughby, K., Land, F., Koterwas, B., Morgan, K. L., Tanya, V. N., and Bronsvoort, B. M. d. C. (2011). Seroepidemiology of bovine viral diarrhoea virus (BVDV) in the Adamawa region of Cameroon and use of the SPOT test to identify herds with PI calves. *PLoS One*, 6(7):e21620.

Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., and Rege, J. E. O. (2002). African pastoralism: genetic imprints of origins and migrations. *Science*, 296(5566):336–339.

Hanotte, O., Dessie, T., and Kemp, S. (2010). Time to tap Africa's livestock genomes. *Science*, 328(5986):1640–1641.

Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., Sugimoto, Y., Kemp, S., Gibson, J., Korol, A., Soller, M., and Teale, A. (2003). Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proceedings of the National Academy of Sciences*, 100(13):7443–7448.

Hansen, J. and Perry, B. (1994). *The epidemiology, diagnosis and control of helminth parasites of ruminants: A handbook*. ILRAD, Nairobi, Kenya, 2nd edition.

Hao, K., Di, X., and Cawley, S. (2007). LdCompare: rapid computation of single-and multiple-marker r^2 and genetic coverage. *Bioinformatics*, 23(2):252–254.

Hartgers, F. and Yazdanbakhsh, M. (2006). Co-infection of helminths and malaria: modulation of the immune responses to malaria. *Parasite immunology*, 28(10):497–506.

Hawley, D. M. and Altizer, S. M. (2011). Disease ecology meets ecological immunology: understanding the links between organismal immunity and infection dynamics in natural populations. *Functional Ecology*, 25(1):48–60.

Hayward, A. D., Pilkington, J. G., Pemberton, J. M., and Kruuk, L. E. B. (2010). Maternal effects and early-life performance are associated with parasite resistance across life in free-living Soay sheep. *Parasitology*, 137(08):1261–1273.

Hayward, A. D., Wilson, A. J., Pilkington, J. G., Pemberton, J. M., and Kruuk, L. E. B. (2009). Ageing in a variable habitat: environmental stress affects senescence in parasite resistance in St Kilda Soay sheep. *Philosphical Transactions of the Royal Society B*, 276(1672):3477–3485.

Hedrick, P. W. and Kim, T. J. (2000). Genetics of complex polymorphisms: parasites and maintenance of the major histocompatibility complex variation. *Evolutionary Genetics: from molecules to morphology*, pages 204–234.

Hellard, E., Pontier, D., Sauvage, F., Poulet, H., and Fouchet, D. (2012). True versus false parasite interactions: a robust method to take risk factors into account and its application to feline viruses. *PLoS One*, 7(1):e29618.

Hirschhorn, J. N. and Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, 6(2):95–108.

Homewood, K., Trench, P., Randall, S., Lynen, G., and Bishop, B. (2006). Livestock health and socio-economic impacts of a veterinary intervention in Maasailand: infection-and-treatment vaccine against East Coast fever. *Agricultural Systems*, 89(2):248–271.

Hong, Y. H., Lillehoj, H. S., Lillehoj, E. P., and Lee, S. H. (2006). Changes in immune-related gene expression and intestinal lymphocyte subpopulations following *Eimeria maxima* infection of chickens. *Veterinary Immunology and Immunopathology*, 114(3):259–272.

Hooper, L. V. and Gordon, J. I. (2001). Commensal host-bacterial relationships in the gut. *Science*, 292(5519):1115–1118.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2000). Model-building strategies and methods for logistic regression. *Applied Logistic Regression, Third Edition*, pages 89–151.

Hotchkiss, E., Dagleish, M., Willoughby, K., McKendrick, I., Finlayson, J., Zadoks, R., Newsome, E., Brulisauer, F., Gunn, G., and Hodgson, J. (2010). Prevalence of *Pasteurella multocida* and other respiratory pathogens in the nasal tract of Scottish calves. *Veterinary Record*, 167(15):555–560.

Hoverman, J. T., Hoye, B. J., and Johnson, P. T. J. (2013). Does timing matter? how priority effects influence the outcome of parasite interactions within hosts. *Oecologia*, 173(4):1471–1480.

Hu, Z.-L., Park, C. A., Wu, X.-L., and Reecy, J. M. (2013). Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Research*, 41(D1):D871–D879.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1):44–57.

Hutchings, D. L., Campos, M., Qualtiere, L., and Babiuk, L. (1990). Inhibition of antigen-induced and interleukin-2-induced proliferation of bovine peripheral blood leukocytes by inactivated bovine herpes virus 1. *Journal of Virology*, 64(9):4146–4151.

Hviid, A. and Melbye, M. (2007). The impact of birth weight on infectious disease hospitalization in childhood. *American Journal of Epidemiology*, 165(7):756–761.

Jackson, J. A., Pleass, R. J., Cable, J., Bradley, J. E., and Tinsley, R. C. (2006). Heterogenous interspecific interactions in a host-parasite system. *International Journal for Parasitology*, 36(13):1341–1349.

Jain, N. C. (1993). *Essentials of veterinary hematology*. Wiley, Philadelphia.

Jamieson, A. M., Yu, S., Annicelli, C. H., and Medzhitov, R. (2010). Influenza virus-induced glucocorticoids compromise innate host defense against a secondary bacterial infection. *Cell Host & Microbe*, 7(2):103–114.

Janeway, C., Travers, P., Walport, M., and Shlomchik, M. (2005). *Immunobiology: The Immune System in Health and Disease*. Garland Science, New York.

Jennings, A. (2013). *The epidemiology and clinical outcomes associated with Theileria Parva in a cohort of East African short horn zebu calves*. PhD thesis, University of Edinburgh.

Johnson, J., Suzuki, Y., Mack, D., Mui, E., Estes, R., David, C., Skamene, E., Forman, J., and McLeod, R. (2002). Genetic analysis of influences on survival following *Toxoplasma gondii* infection. *International Journal for Parasitology*, 32(2):179–185.

Johnson, R. C., Nelson, G. W., Troyer, J. L., Lautenberger, J. A., Kessing, B. D., Winkler, C. A., and O'Brien, S. J. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*, 11(1):724.

Jolles, A. E., Ezenwa, V. O., Etienne, R. S., Turner, W. C., and Olff, H. (2008). Interactions between macroparasites and microparasites drive infection patterns in free-ranging African buffalo. *Ecology*, 89(8):2239–2250.

Jonsson, N. N., Bock, R. E., Jorgensen, W. K., Morton, J. M., and Stear, M. J. (2012). Is endemic stability of tick-borne disease in cattle a useful concept? *Trends in Parasitology*, 28(3):85–89.

Kaashoek, M. J., Straver, P. H., Van Rooij, E. M. A., Quak, J., and Van Oirschot, J. T. (1996). Virulence, immunogenicity and reactivation of seven bovine herpesvirus 1.1 strains: clinical and virological aspects. *Veterinary Record*, 139:416–421.

Kanyari, P. W. N., Kagira, J. M., and Mhoma, J. R. L. (2010). Prevalence of endoparasites in cattle within urban and peri-urban areas of Lake Victoria Basin, Kenya with special reference to zoonotic potential. *Scientia Parasitologica*, 11:171–178.

Kariuki, D., Young, A., Morzaria, S., Lesan, A., Mining, S., Omwoyo, P., Wafula, J., and Molyneux, D. (1995). *Theileria parva* carrier state in naturally infected and artificially immunised cattle. *Tropical Animal Health and Production*, 27(1):15–25.

Kaufmann, J., Dwinger, R. H., Hallebeek, A., van Dijk, B., and Pfister, K. (1992). The interaction of *Trypanosoma congolense* and *Haemochus contortus* infections in trypanotolerant N'dama cattle. *Veterinary Parasitology*, 43(3):157–170.

Kemper, K., Palmer, D., Liu, S., Greeff, J., Bishop, S., and Karlsson, L. (2010). Reduction of faecal worm egg count, worm numbers and worm fecundity in sheep selected for worm resistance following artificial infection with *Teladorsagia circumcincta* and *Trichostrongylus colubriformis*. *Veterinary Parasitology*, 171(3-4):238–246.

Keyyu, J. D., Kassuku, A. A., Kyvsgaard, N. C., and Willingham, A. L. (2003). Gastrointestinal nematodes in indigenous zebu cattle under pastoral and nomadic management systems in the lower plain of the Southern Highlands of Tanzania. *Veterinary Research Communications*, 27(5):371–380.

Kiara, H., Jennings, A., Bronsvoort, B. M. d. C., Handel, I., Mwangi, S. T., Mbole-Kariuki Ndila, M., Conradie van Wyk, I., Poole, E. J., Hanotte, O., Coetzer, J. A. W., Woolhouse, M., and Toye, P. (2014). A longitudinal assessment of the serological response to *Theileria parva* and other tick-borne parasites from birth to one year in a cohort of indigenous calves in Western Kenya. *Parasitology*, 141:1289–1298.

Knowles, S. C. L. (2011). The effect of helminth co-infection on malaria in mice: a meta-analysis. *International Journal for Parasitology*, 41(10):1041–1051.

Knowles, S. C. L., Fenton, A., Petchey, O. L., Jones, T. R., Barber, R., and Pedersen, A. B. (2013). Stability of within-host-parasite communities in a wild mammal system. *Proceedings of the Royal Society B: Biological Sciences*, 280(1762).

Kobylinski, K. C., Alout, H., Foy, B. D., Clements, A., Adisakwattana, P., Swierczewski, B. E., and Richardson, J. H. (2014). Rationale for the coadministration of albendazole and ivermectin to humans for malaria parasite transmission control. *The American Journal of Tropical Medicine and Hygiene*, 91(4):655–662.

Koppers-Lalic, D., Rijsewijk, F. A. M., Verschuren, S. B. E., van Gaans-van den Brink, J. A. M., Neisig, A., Ressing, M. E., Neefjes, J., and Wiertz, E. J. H. J. (2001). The UL41-encoded virion host shutoff (vhs) protein and vhs-independent mechanisms are responsible for down-regulation of MHC class I molecules by bovine herpesvirus 1. *Journal of General Virology*, 82(9):2071–2081.

Krause, P. J., Telford, I. I. I. S. R., Spielman, A., Sikand, V., Ryan, R., Christianson, D., Burke, G., Brassard, P., Pollack, R., Peck, J., and Persing, D. H. (1996). Concurrent lyme disease and babesiosis: evidence for increased severity and duration of illness. *JAMA: The Journal of the American Medical Association*, 275(21):1657–1660.

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1446):873–890.

Kruuk, L. E. B. and Hadfield, J. D. (2007). How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology*, 20(5):1890–1903.

Lafferty, K. D. (2010). Interacting parasites. *Science*, 330(6001):187–188.

Leighton, E. A., Murrell, K. D., and Gasbarre, L. C. (1989). Evidence for genetic-control of nematode egg-shedding rates in calves. *Journal of Parasitology*, 75(4):498–504.

Lello, J., Boag, B., Fenton, A., Stevenson, I. R., and Hudson, P. J. (2004). Competition and mutualism among the gut helminths of a mammalian host. *Nature*, 428(6985):840–844.

Lello, J., Norman, R., Boag, B., Hudson, P., and Fenton, A. (2008). Pathogen interactions, population cycles, and phase shifts. *The American Naturalist*, 171(2):176–182.

Lemaire, M., Weynants, V., Godfroid, J., Schynts, F., Meyer, G., Letesson, J.-J., and Thiry, E. (2000). Effects of bovine herpesvirus type 1 infection in calves with maternal antibodies on immune response and virus latency. *Journal of Clinical Microbiology*, 38(5):1885–1894.

Lemon, K. P., Armitage, G. C., Relman, D. A., and Fischbach, M. A. (2012). Microbiota-targeted therapies: an ecological perspective. *Science Translational Medicine*, 4(137):137rv5–137rv5.

Lesosky, M., Dumas, S., Conradie van Wyk, Ilana, I., Handel, I., Jennings, A., Thumbi, S., Toye, P., and Bronsvoort, B. M. d. C. (2013). A live weight-heart girth relationship for accurate dosing of East African shorthorn zebu cattle. *Tropical Animal Health and Production*, 45:311–316.

Lessard, P., l'Eplattenier, R., Norval, R., Kundert, K., Dolan, T., Croze, H., Walker, J., Irvin, A., and Perry, B. (1990). Geographical information systems for studying the epidemiology of cattle diseases caused by *Theileria parva*. *The Veterinary Record*, 126(11):255–262.

Levin, S. A., Carpenter, S. R., Godfray, H. C. J., Kinzig, A. P., Loreau, M., Losos, J. B., Walker, B., and Wilcove, D. S. (2009). *The Princeton Guide to Ecology*. Princeton University Press.

Levings, R. L. and Roth, J. A. (2013). Immunity to bovine herpesvirus 1: II. adaptive immunity and vaccinology. *Animal Health Research Reviews*, 14(01):103–123.

Lotfy, W. M., Brant, S. V., Ashmawy, K. I., Devkota, R., Mkoji, G. M., and Loker, E. S. (2010). A molecular approach for identification of paramphistomes from Africa and Asia. *Veterinary Parasitology*, 174(3):234–240.

Lu, Y. P., Zeng, D. Y., Chen, Y. P., Liang, X. J., Xu, J. P., Huang, S. M., Lai, Z. W., Wen, W. R., Von Websky, K., and Hocher, B. (2013). Low birth weight is associated with lower respiratory tract infections in children with hand, foot, and mouth disease. *Clinical Laboratory*, 59(9-10):985–992.

Lumley, T. (2012). Analysis of complex survey samples.

Lyaku, J., Nettleton, P., and Marsden, H. (1992). A comparison of serological relationships among five ruminant alphaherpesviruses by ELISA. *Archives of Virology*, 124(3-4):333–341.

Lynch, M. and Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer Associates Inc.,U.S.

Machado, M., S Azevedo, A., Teodoro, R., Pires, M., CD Peixoto, M., de Freitas, C., Prata, M., Furlong, J., da Silva, M., Guimaraes, S., Regitano, L., Coutinho, L., Gasparin, G., and Verneque, R. (2010). Genome wide scan for quantitative trait loci affecting tick resistance in cattle (*Bos taurus* x *Bos indicus*). *BMC Genomics*, 11(1):280.

Mainar-Jaime, R., Berzal-Herranz, B., Arias, P., and Rojo-Vázquez, F. (2001). Epidemiological pattern and risk factors associated with bovine viral-diarrhoea virus (BVDV) infection in a non-vaccinated dairy-cattle population from the Asturias region of Spain. *Preventive Veterinary Medicine*, 52(1):63–73.

Mangen, M.-J. J., Plass, D., Havelaar, A. H., Gibbons, C. L., Cassini, A., Mühlberger, N., van Lier, A., Haagsma, J. A., Brooke, R. J., Lai, T., et al. (2013). The pathogen-and incidence-based DALY approach: an appropriated methodology for estimating the burden of infectious diseases. *PLoS ONE*.

Margolis, L., Esch, G., Holmes, J., Kuris, A., and Schad, G. (1982). The use of ecological terms in parasitology (report of an ad hoc committee of the american society of parasitologists). *The Journal of Parasitology*, pages 131–133.

Marshall, D. M. (1994). Breed differences and genetic parameters for body composition traits in beef cattle. *Journal of Animal Science*, 72(10):2745–2755.

Martin, S. and Bohac, J. (1986). The association between serological titers in infectious bovine rhinotracheitis virus, bovine virus diarrhea virus, parainfluenza-3 virus, respiratory syncytial virus and treatment for respiratory disease in Ontario feedlot calves. *Canadian Journal of Veterinary Research*, 50(3):351–358.

Martin, S. W., Nagy, E., Armstrong, D., and Rosendal, S. (1999). The associations of viral and mycoplasmal antibody titers with respiratory disease and weight gain in feedlot calves. *The Canadian Veterinary Journal*, 40(8):560–570.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P., Sonstegard, T. S., and Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*, 4(4):e5350.

Mbole-Kariuki, M. N. (2012). *Genomic diversity of East African shorthorn zebu of Western Kenya*. PhD thesis, School of Life Sciences.

Mbole-Kariuki, M. N., Sonstegard, T., Orth, A., Thumbi, S. M., Bronsvoort, B. M. d. C., Kiara, H., Toye, P., Conradie van Wyk, I., Jennings, A., Coetzer, K., Woolhouse, M. E. J., Hanotte, O., and Tapio, M. (2014). Genome-wide analysis reveals the ancient and recent admixture history of East African shorthorn zebu from Western Kenya. *Heredity*, 113:297–305.

McClellan, J. and King, M. (2010). Genetic heterogeneity in human disease. *Cell*, 141(2):210–217.

McClelland, E. E., Penn, D. J., and Potts, W. K. (2003). Major histocompatibility complex heterozygote superiority during coinfection. *Infection and Immunity*, 71(4):2079–2086.

McCune, B. and Mefford, M. J. (2011). *PC-ORD: multivariate analysis of ecological data*. MjM Software, Gleneden Beach, Oregon, U.S.A.

McDermott, J., Kadohira, M., O'Callaghan, C., and Shoukri, M. (1997). A comparison of different models for assessing variations in the sero-prevalence of infectious bovine rhinotracheitis by farm, area and district in Kenya. *Preventive Veterinary Medicine*, 32(3):219–234.

Medzhitov, R., Schneider, D. S., and Soares, M. P. (2012). Disease tolerance as a defense strategy. *Science*, 335(6071):936–941.

Méthot, P.-O. and Alizon, S. (2014). What is a pathogen? Toward a process view of host-parasite interactions. *Virulence*, 5(8):775–785.

Ministry of Planning and National Development (2003). *GOK: Geographical Dimensions of Well-Being in Kenya. Where the Poor Live.* Regal Press.

Moll, G., Lohding, A., and Young, A. S. (1984). Epidemiology of theilerioses in the Trans-Mara division, Kenya: husbandry and disease background and preliminary investigations on theilerioses in calves. *Preventive Veterinary Medicine*, 2(6):801–831.

Moore, S. L. and Wilson, K. (2002). Parasites as a viability cost of sexual selection in natural populations of mammals. *Science*, 297:2015–2018.

Moreno, P. G., Eberhardt, M. A. T., Lamattina, D., Previtali, M. A., and Beldomenico, P. M. (2013). Intra-phylum and inter-phyla associations among gastrointestinal parasites in two wild mammal species. *Parasitology Research*, 112(9):3295–3304.

Mosmann, T. R. and Sad, S. (1996). The expanding universe of T-cell subsets: Th1, Th2 and more. *Immunology Today*, 17(3):138–146.

Mukhebi, A., Perry, B., and Kruska, R. (1992). Estimated economics of theileriosis control in Africa. *Preventive Veterinary Medicine*, 12(1):73–85.

Murphy, P. M. (1996). Chapter 11 Phagocytes in immunity and inflammation. In Bittar, E. and Bittar, N., editors, *Immunobiology*, volume 6 of *Principles of Medical Biology*, pages 197 – 229. Elsevier.

Murray, G. G., Woolhouse, M. E. J., Tapio, M., Mbole-Kariuki, M. N., Sonstegard, T. S., Thumbi, S. M., Jennings, A. E., van Wyk, I. C., Chase-Topping, M., and Kiara, H. (2013). Genetic susceptibility to infectious disease in East African shorthorn zebu: a genome-wide analysis of the effect of heterozygosity and exotic introgression. *BMC Evolutionary Biology*, 13(1):246.

Murray, M., Trail, J., Davis, C., and Black, S. (1984). Genetic resistance to African trypanosomiasis. *Journal of Infectious Diseases*, 149(3):311–319.

Muylkens, B., Thiry, J., Kirten, P., Schynts, F., and Thiry, E. (2007). Bovine herpesvirus 1 infection and infectious bovine rhinotracheitis. *Veterinary Research*, 38(2):181–209.

Nakagawa, S. and Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, 85:935–956.

Njiiri, N. E., Bronsvoort, B. M. d. C., Collins, N. E., Steyn, H. C., Troskie, M., Vorster, I., Thumbi, S. M., Sibeko, K. P., Jennings, A., van Wyk, I. C., Mbole-Kariuki, M., Kiara, H., Poole, E. J., Hanotte, O., Coetzer, K., Oosthuizen, M. C., Woolhouse, M., and Toye, P. (2015). The epidemiology of tick-borne haemoparasites as determined by the reverse line blot hybridization assay in an intensively studied cohort of calves in Western Kenya. *Veterinary Parasitology*, In Press, Corrected Proof.

Norris, K. and Evans, M. R. (2000). Ecological immunology: life history trade-offs and immune defense in birds. *Behavioral Ecology*, 11(1):19–26.

Norval, R., Perry, B., Gebreab, F., and Lessard, P. (1991). East Coast fever: a problem of the future for the horn of Africa? *Preventive Veterinary Medicine*, 10(3):163–172.

Noyes, H., Brass, A., Obara, I., Anderson, S., Archibald, A. L., Bradley, D. G., Fisher, P., Freeman, A., Gibson, J., Gicheru, M., Hall, L., Hanotte, O., Hulme, H., McKeever, D., Murray, C., Oh, S. J., Tate, C., Smith, K., Tapio, M., Wambugu, J., Williams, D. J., Agaba, M., and Kemp, S. J. (2011). Genetic and expression analysis of cattle identifies candidate genes in pathways responding to *Trypanosoma congolense* infection. *Proceedings of the National Academy of Sciences*, 108(22):9304–9309.

O'Brien, A. M. P., Utsunomiya, Y. T., Mészáros, G., Bickhart, D. M., Liu, G. E., Van Tassell, C. P., Sonstegard, T. S., Da Silva, M. V., Garcia, J. F., and Sölkner, J. (2014). Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genetics Selection Evolution*, 46(1):19.

Paling, R., Moloo, S., Scott, J., McOdimba, F., Enfrey, L. L.-H., Murray, M., and Williams, D. (1991). Susceptibility of N'Dama and Boran cattle to tsetse-transmitted primary and rechallenge infections with a homologous serodeme of *Trypanosoma congolense*. *Parasite Immunologymmunology*, 13(4):413–425.

Pankow, J. S., Folsom, A. R., Cushman, M., Borecki, I. B., Hopkins, P. N., Eckfeldt, J. H., and Tracy, R. P. (2001). Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis*, 154(3):681–689.

Pedersen, A. B. and Antonovics, J. (2013). Anthelmintic treatment alters the parasite community in a wild mouse host. *Biology Letters*, 9(4):20130205.

Pedersen, A. B. and Fenton, A. (2007). Emphasizing the ecology in parasite community ecology. *Trends in Ecology & Evolution*, 22(3):133–139.

Pemberton, J. M. (2008). Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences*, 275:613–621.

Periasamy, K., Pichler, R., Poli, M., Cristel, S., Cetrá, B., Medus, D., Basar, M., A. K., T., Ramasamy, S., Ellahi, M. B., Mohammed, F., Teneva, A., Shamsuddin, M., Podesta, M. G., and Diallo, A. (2014). Candidate gene approach for parasite resistance in sheep – variation in immune pathway genes and association with fecal egg count. *PLoS One*, 9(2):e88337.

Perry, B. and Sones, K. (2007). Science for development: poverty reduction though animal health. *Science*, 315:333–334.

Peters, A., Thevasagayam, S., Wiseman, A., and Salt, J. (2004). Duration of immunity of a quadrivalent vaccine against respiratory diseases caused by BHV-1, PI3V, BVDV, and BRSV in experimentally infected calves. *Preventive Veterinary Medicine*, 66(1):63–77.

Phiri, B., Benschop, J., and French, N. (2010). Systematic review of causes and factors associated with morbidity and mortality on smallholder dairy farms in Eastern and Southern Africa. *Preventive Veterinary Medicine*, 94(1):1–8.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed effects models in S and S plus*. Springer Verlag, New York.

Pirofski, L.-a. and Casadevall, A. (2012). Q&a: What is a pathogen? A question that begs the point. *BMC Biology*, 10(1):6.

Piyathilake, C. J., Henao, O. L., Macaluso, M., Cornwell, P. E., Meleth, S., Heimburger, D. C., and Partridge, E. E. (2004). Folate is associated with the natural history of high-risk human papillomaviruses. *Cancer Research*, 64(23):8788–8793.

Poletto, C., Meloni, S., Van Metre, A., Colizza, V., Moreno, Y., and Vespignani, A. (2015). Characterising two-pathogen competition in spatially structured environments. *Scientific Reports*, 5.

Poulin, R. (2011). *Evolutionary ecology of parasites*. Princeton University Press.

Powell, J. E., Visscher, P. M., and Goddard, M. E. (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics*, 11(11):800–805.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8):904–909.

Pullan, R. L., Bethony, J. M., Geiger, S. M., Correa-Oliveira, R., Brooker, S., and Quinnell, R. J. (2010). Human helminth co-infection: no evidence of common genetic control of hookworm and *Schistosoma mansoni* infection intensity in a Brazilian community. *International Journal for Parasitology*, 40(3):299–306.

Råberg, L., Graham, A. L., and Read, A. F. (2009). Decomposing health: tolerance and resistance to parasites in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1513):37–49.

Raqib, R., Alam, D. S., Sarker, P., Ahmad, S. M., Ara, G., Yunus, M., Moore, S. E., and Fuchs, G. (2007). Low birth weight is associated with altered immune function in rural Bangladeshi children: a birth cohort study. *The American Journal of Clinical Nutrition*, 85(3):845–852.

Read, J. S., Clemens, J. D., and Klebanoff, M. A. (1994). Moderate low birth weight and infectious disease mortality during infancy and childhood. *American Journal of Epidemiology*, 140(8):721–733.

Riggio, V., Matika, O., Pong-Wong, R., Stear, M. J., and Bishop, S. C. (2013). Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity*, 110(5):420–429.

Robinson, M. R., Santure, A. W., DeCauwer, I., Sheldon, B. C., and Slate, J. (2013). Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology*, 22(15):3963–3980.

Rodriguez, M., Terrazas, L. I., Marquez, R., and Bojalil, R. (1999). Susceptibility to *Trypanosoma cruzi* is modified by a previous non-related infection. *Parasite Immunology*, 21(4):177–185.

Roth, J., Bolin, S., and Frank, D. (1986). Lymphocyte blastogenesis and neutrophil function in cattle persistently infected with bovine viral diarrhea virus. *American Journal of Veterinary Research*, 47(5):1139–1141.

Rowlands, G. J., Mulatu, W., Nagda, S. M., Dolan, R. B., and Dieteren, G. D. M. (1995). Genetic-variation in packed red-cell volume and frequency of parasitemia in East African zebu cattle exposed to drug-resistant trypanosomes. *Livestock Production Science*, 43(1):75–84.

Rynkiewicz, E. C., Pedersen, A. B., and Fenton, A. (2015). An ecosystem approach to understanding and managing within-host parasite community dynamics. *Trends in Parasitology*, pages In Press, Corrected Proof.

Saathoff, E., Olsen, A., Kvalsvig, J. D., Appleton, C. C., Sharp, B., and Kleinschmidt, I. (2005). Ecological covariates of *Ascaris lumbricoides* infection in schoolchildren from rural KwaZulu-Natal, South Africa. *Tropical Medicine & International Health*, 10(5):412–422.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., and Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909):832–837.

Sanabria, R. and Romero, J. (2008). Review and update of paramphistomosis. *Helminthologia*, 45(2):64–68.

Schmid-Hempel, P. (2011). *Evolutionary parasitology: the integrated study of infections, immunology, ecology, and genetics*. Oxford University Press New York.

Schneider, D. S. and Ayres, J. S. (2008). Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases. *Nature Reviews Immunology*, 8(11):889–895.

Schneider, M. J., Tait, R. G., Busby, W. D., and Reecy, J. M. (2009). An evaluation of bovine respiratory disease complex in feedlot cattle: impact on performance and carcass traits using treatment records and lung lesion scores. *Journal of Animal Science*, 87(5):1821–1827.

Schur, N., Gosoniu, L., Raso, G., Utzinger, J., and Vounatsou, P. (2011). Modelling the geographical distribution of co-infection risk from single-disease surveys. *Statistics in Medicine*, 30(14):1761–1776.

Shaner, G., Stromberg, E. L., Lacy, G. H., Barker, K. R., and Pirone, T. P. (1992). Nomenclature and concepts of pathogenicity and virulence. *Annual Review of Phytopathology*, 30(1):47–66.

Shapiro-Ilan, D. I., Fuxa, J. R., Lacey, L. A., Onstad, D. W., and Kaya, H. K. (2005). Definitions of pathogenicity and virulence in invertebrate pathology. *Journal of Invertebrate Pathology*, 88(1):1–7.

Sin, Y. W., Annavi, G., Dugdale, H. L., Newman, C., Burke, T., and MacDonald, D. W. (2014). Pathogen burden, co-infection and major histocompatibility complex variability in the European badger (*Meles meles*). *Molecular Ecology*, 23(20):5072–5088.

Snowder, G. D., Van Vleck, L. D., Cundiff, L. V., and Bennett, G. L. (2006). Bovine respiratory disease in feedlot cattle: environmental, genetic, and economic factors. *Journal of Animal Science*, 84(8):1999–2008.

Stear, M., Bishop, S., Doligalska, M., Duncan, J., Holmes, P., Irvine, J., McCririe, L., McKellar, Q., Sinski, E., and Murray, M. (1995). Regulation of egg production, worm burden, worm length and worm fecundity by host responses in sheep infected with *Ostertagia circumcincta*. *Parasite Immunology*, 17(12):643–652.

Stear, M. J., Hetzel, D. J. S., Brown, S. C., Gershwin, L. J., Mackinnon, M. J., and Nicholas, F. W. (1990). The relationships among ecto- and endoparasite levels, class I antigens of the bovine major histocompatibility system, immunoglobulin E levels and weight gain. *Veterinary Parasitology*, 34(4):303 – 321.

Stear, M. J., Park, M., and Bishop, S. C. (1996). The key components of resistance to *Ostertagia circumcincta* in lambs. *Parasitology Today*, 12(11):438–441.

Stear, M. J., Tierney, T. J., Baldock, F. C., Brown, S. C., Nicholas, F. W., and Rudderh, T. H. (1988). Failure to find an association between class I antigens of the bovine major histocompatibility system and faecal worm egg counts. *International Journal for Parasitology*, 18(6):859–861.

Stockdale, P. (1976). Schizogony and gametogony of *Eimeria zuernii* (Rivolta, 1878) Martin, 1909. *Veterinary Parasitology*, 1(4):367–376.

Suchdev, P. S., Davis, S. M., Bartoces, M., Ruth, L. J., Worrell, C. M., Kanyi, H., Odero, K., Wiegand, R. E., Njenga, S. M., Montgomery, J. M., et al. (2014). Soil-transmitted helminth infection and nutritional status among urban slum children in Kenya. *The American Journal of Tropical Medicine and Hygiene*, 90(2):299–305.

Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology*, 5(7):e171.

Taylor, J. D., Fulton, R. W., Lehenbauer, T. W., Step, D. L., and Confer, A. W. (2010). The epidemiology of bovine respiratory disease: what is the evidence for predisposing factors? *The Canadian Veterinary Journal*, 51(10):1095.

Telfer, S., Birtles, R., Bennett, M., Lambin, X., Paterson, S., and Begon, M. (2008). Parasite interactions in natural populations: insights from longitudinal data. *Parasitology*, 135(Special Issue 07):767–781.

Telfer, S., Lambin, X., Birtles, R., Beldomenico, P., Burthe, S., Paterson, S., and Begon, M. (2010). Species interactions in a parasite community drive infection risk in a wildlife population. *Science*, 330(6001):243–246.

Temple, L. K., McLeod, R. S., Gallinger, S., and Wright, J. G. (2001). Defining disease in the genomics era. *Science*, 293(5531):807–808.

Thomas, M. B., Watson, E. L., and Valverde-Garcia, P. (2003). Mixed infections and insect-pathogen interactions. *Ecology Letters*, 6(3):183–188.

Thomas, S. R. and Elkinton, J. S. (2004). Pathogenicity and virulence. *Journal of Invertebrate Pathology*, 85(3):146–151.

Thrusfield, M. (2005). *Veterinary epidemiology*. Blackwell Publishing, third edition.

Thumbi, S. (2012). *Epidemiology of infections and co-infections: impact on survival and growth of zebu cattle under one year*. PhD thesis, University of Edinburgh.

Thumbi, S., Bronsvoort, B. M. d. C., Kiara, H., Toye, P., Poole, J., Ndila, M., Conradie, I., Jennings, A., Handel, I., Coetzer, J., Steyl, J., Hanotte, O., and Woolhouse, M. (2013a). Mortality in East African shorthorn zebu cattle under one year: predictors of infectious-disease mortality. *BMC Veterinary Research*, 9(1):175.

Thumbi, S. M., Bronsvoort, B. M. d. C., Poole, E. J., Kiara, H., Toye, P., Ndila, M., Conradie, I., Jennings, A., Handel, I. G., Coetzer, J. A. W., Hanotte, O., and Woolhouse, M. E. J. (2013b). Parasite co-infections show synergistic and antagonistic interactions on growth performance of East African zebu cattle under one year. *Parasitology*, 140:1789–1798.

Thumbi, S. M., Bronsvoort, B. M. d. C., Poole, E. J., Kiara, H., Toye, P. G., Mbole-Kariuki, M. N., Conradie, I., Jennings, A., Handel, I. G., Coetzer, J. A. W., Steyl, J. C. A., Hanotte, O., and Woolhouse, M. E. J. (2014). Parasite co-infections and their impact on survival of indigenous cattle. *PLoS One*, 9(2):e76324.

Tikkinen, K. A., Leinonen, J. S., Guyatt, G. H., Ebrahim, S., and Järvinen, T. L. (2012). What is a disease? Perspectives of the public, health professionals and legislators. *BMJ Open*, 2(6):e001632.

Tollenaere, C., Bryja, J., Galan, M., Cadet, P., Deter, J., Y., C., Berthier, K., Ribas Salvador, A., Voutilainen, L., Laakkonen, J., Henttonen, H., Cosson, J.-F., and Charbonnel, N. (2008). Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. *Journal of Evolutionary Biology*, 21(5):1307–1320.

Tompkins, D. M., Dunn, A. M., Smith, M. J., and Telfer, S. (2011). Wildlife diseases: from individuals to ecosystems. *Journal of Animal Ecology*, 80(1):19–38.

Toye, P., Handel, I., Gray, J., Kiara, H., Thumbi, S., Jennings, A., van Wyk, I. C., Ndila, M., Hanotte, O., Coetzer, K., Woolhouse, M., and Bronsvoort, B. M. d. C. (2013a). Maternal antibody uptake, duration and influence on survival and growth rate in a cohort of indigenous calves in a smallholder farming system in Western Kenya. *Veterinary Immunology and Immunopathology*, 155(1-2):129–134.

Toye, P. G., Batten, C. A., Kiara, H., Henstock, M. R., Edwards, L., Thumbi, S., Poole, E. J., Handel, I. G., Bronsvoort, B. M. d. C., Hanotte, O., Coetzer, J. A. W., Woolhouse, M. E. J., and Oura, C. A. L. (2013b). Bluetongue and epizootic haemorrhagic disease virus in local breeds of cattle in Kenya. *Research in Veterinary Science*, 94(3):769–773.

Trail, J., D'Ieteren, G., and Teale, A. (1989). Trypanotolerance and the value of conserving livestock genetic resources. *Genome*, 31(2):805–812.

Turner, W. C. and Getz, W. M. (2010). Seasonal and demographic factors influencing gastrointestinal parasitism in ungulates of Etosha National Park. *Journal of Wildlife Diseases*, 46(4):1108.

Uilenberg, G. (1999). Immunization against diseases caused by *Theileria parva*: a review. *Tropical Medicine & International Health*, 4(9):A12–A20.

Urquhart, G. M., Armour, J., Duncan, J. L., Dunn, A. M., and Jennings, F. W. (1996). *Veterinary parasitology*. Blackwell Science Ltd, Oxford, 2nd edition.

Van der Poel, W., Kramps, J., Quak, J., Brand, A., and Van Oirschot, J. (1995). Persistence of bovine herpesvirus-1-specific antibodies in cattle after intranasal vaccination with a live virus vaccine. *The Veterinary Record*, 137(14):347–348.

van Riet, E., Hartgers, F. C., and Yazdanbakhsh, M. (2007). Chronic helminth infections induce immunomodulation: consequences and mechanisms. *Immunobiology*, 212(6):475–490.

van Schaik, G., Dijkhuizen, A. A., Huirne, R. B., Schukken, Y. H., Nielen, M., and Hage, H. J. (1998). Risk factors for existence of bovine herpes virus 1 antibodies on nonvaccinating Dutch dairy farms. *Preventive Veterinary Medicine*, 34(2):125–136.

Vanimisetti, H. B., Andrew, S. L., Zajac, A. M., and Notter, D. R. (2004). Inheritance of fecal egg count and packed cell volume and their relationship with production traits in sheep infected with *Haemonchus contortus*. *Journal of Animal Science*, 82(6):1602–1611.

Viney, M. E. and Graham, A. L. (2013). *Patterns and processes in parasite co-Infection*, volume 82 of *Advances in Parasitology*, pages 321–369. Elsevier Academic Press Inc, San Diego.

Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, 90(1):7–24.

Visscher, P. M., Hemani, G., Vinkhuyzen, A. A., Chen, G.-B., Lee, S. H., Wray, N. R., Goddard, M. E., and Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genetics*, 10(4):e1004269.

Visscher, P. M., Hill, W. G., and Wray, N. R. (2008). Heritability in the genomics era - concepts and misconceptions. *Nature Reviews Genetics*, 9(4):255–266.

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, 4(3):e72.

Walling, C. A., Nussey, D. H., Morris, A., Clutton-Brock, T. H., Kruuk, L. E., and Pemberton, J. M. (2011). Inbreeding depression in red deer calves. *BMC Evolutionary Biology*, 11(1):318.

Wang, W. Y., Barratt, B. J., Clayton, D. G., and Todd, J. A. (2005). Genome-wide association studies: theoretical and practical concerns. *Nature Reviews Genetics*, 6(2):109–118.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Wardrop, N. A., Atkinson, P. M., Gething, P. W., Fèvre, E. M., Picozzi, K., Kakembo, A. S. L., and Welburn, S. C. (2010). Bayesian geostatistical analysis and prediction of Rhodesian Human African Trypanosomiasis. *PLoS Neglected Tropical Diseases*, 4(12):e914.

Waruiru, R., Kyvsgaard, N., Thamsborg, S., Nansen, P., Bøgh, H., Munyua, W., and Gathuma, J. (2000). The prevalence and intensity of helminth and coccidial infections in dairy cattle in central Kenya. *Veterinary Research Communications*, 24(1):39–53.

Wearing, H. J. and Rohani, P. (2006). Ecological and immunological determinants of dengue epidemics. *Proceedings of the National Academy of Sciences*, 103(31):11802–11807.

Wellenberg, G., Van Der Poel, W., and Van Oirschot, J. (2002). Viral infections and bovine mastitis: a review. *Veterinary Microbiology*, 88(1):27–45.

White, B. J. and Renter, D. G. (2009). Bayesian estimation of the performance of using clinical observations and harvest lung lesions for diagnosing bovine respiratory disease in post-weaned beef calves. *Journal of Veterinary Diagnostic Investigation*, 21(4):446–453.

White, E., Armstrong, B. K., and Saracci, R. (2008). *Principles of exposure measurement in epidemiology: collecting, evaluating and improving measures of disease risk factors*. Oxford University Press.

Williams, T. N., Maitland, K., Bennett, S., Ganczakowski, M., Peto, T. E. A., Newbold, C. I., Bowden, D. K., Weatherall, D. J., and Clegg, J. B. (1996). High incidence of malaria in [alpha]-thalassaemic children. *Nature*, 383(6600):522–525.

Wilson, A. J. (2008). Why h^2 does not always equal V_A/V_P? *Journal of Evolutionary Biology*, 21:647–650.

Wilson, A. J., Pemberton, J. M., Pilkington, J. G., Clutton-Brock, T. H., Coltman, D. W., and Kruuk, L. E. B. (2007). Quantitative genetics of growth and cryptic evolution of body size in an island population. *Evolutionary Ecology*, 21(3):337–356.

Woolhouse, M. E. J., Haydon, D. T., and Antia, R. (2005). Emerging pathogens: the epidemiology and evolution of species jumps. *Trends in Ecology & Evolution*, 20(5):238–244.

Woolhouse, M. E. J., Thumbi, S. M., Jennings, A., Chase-Topping, M., Callaby, R., Kiara, H., Oosthuizen, M. C., Mbole-Kariuki, M. N., Conradie, I., Handel, I. G., Poole, E. J., Njiiri, E., Collins, N. E., Murray, G., Tapio, M., Auguet, O. T., Weir, W., Morrison, W. I., Kruuk, L. E. B., Bronsvoort, B. M. d. C., Hanotte, O., Coetzer, K., and Toye, P. G. (2015). Co-infections determine patterns of mortality in a population exposed to parasite infection. *Science Advances*, 1(2):e1400026.

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, 15(1):323–354.

Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J., and Lin, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *The American Journal of Human Genetics*, 86(6):929–942.

Wymann, M. N., Bonfoh, B., Traore, K., Tembely, S., and Zinsstag, J. (2007). Species diversity and acquisition of gastrointestinal parasites in calves aged 0-13 months in periurban livestock production in Mali. *Veterinary Parasitology*, 143(1):67–73.

Wymann, M. N., Traore, K., Bonfoh, B., Tembely, S., Tembely, S., and Zinsstag, J. (2008). Gastrointestinal parasite egg excretion in young calves in periurban livestock production in Mali. *Research in Veterinary Science*, 84(2):225–231.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., and Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42(7):565–569.

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M. G., Hill, W. G., Landi, M. T., Alonso, A., Lettre, G., Lin, P., Ling, H., Lowe, W., Mathias, R. A., Melbye, M., Pugh, E., Cornelis, M. C., Weir, B. S., Goddard, M. E., and Visscher, P. M. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, 43(6):519–525.

Yates, W. (1982). A review of infectious bovine rhinotracheitis, shipping fever pneumonia and viral-bacterial synergism in respiratory disease of cattle. *Canadian Journal of Comparative Medicine*, 46(3):225.

Zanetti, M., Gennaro, R., and Romeo, D. (1995). Cathelicidins: a novel protein family with a common proregion and a variable C-terminal antimicrobial domain. *FEBS Letters*, 374(1):1–5.

Zinsstag, J., Ankers, P., Njie, M., Smith, T., Pandey, V. S., Pfister, K., and Tanner, M. (2000). Heritability of gastrointestinal nematode faecal egg counts in West African village N'dama cattle and its relation to age. *Veterinary Parasitology*, 89(1):71–78.

Appendix A

The IDEAL project

My thesis is based upon data gathered by the Infectious Diseases in East African Livestock (IDEAL) project. In this appendix, with permission from the authors, I adapt Bronsvoort et al. (2013) manuscript to provide a detailed description of the IDEAL project.

The IDEAL project was a Welcome Trust funded (grant number 079445) epidemiological study of cattle health in Western Kenya which ran between 2007 and 2010. The aim of the IDEAL project was to provide baseline epidemiological data and to investigate the impact of different infections and coinfections on growth, mortality and morbidity, in addition, to the effect of management and genetic factors.

A.1   The study area

The study laboratory was based in the town of Busia, on the Kenya/Uganda border, therefore due to logistical reasons the study area consisted of a 45km radium semi-circle from Busia town (Figure A.1). The study area covered 5 agro-ecological zones, which is a way to describe the type of land, its topography, soil, climate and suitability for different crops. 280 sublocations, which is the smallest administrative region within Kenya were included in the study area. Each sublocation is about 10-20km^2 and consists of 60-90 households per km^2. Approximately 60% of households within the study region own cattle, which are grazed communally. The IDEAL project focused on sedentary mixed crop-livestock small holdings, a system commonly found in Kenya and surrounding countries which have a typical income of less than US\$15 per month (Ministry of Planning and National Development, 2003).

In the IDEAL study population, 86% of owners reported farming as their only source of income. It is common for land to be sub-divided amongst sons as inheritance and ownership rights, therefore the majority of farmers (96%) own the land they farm,
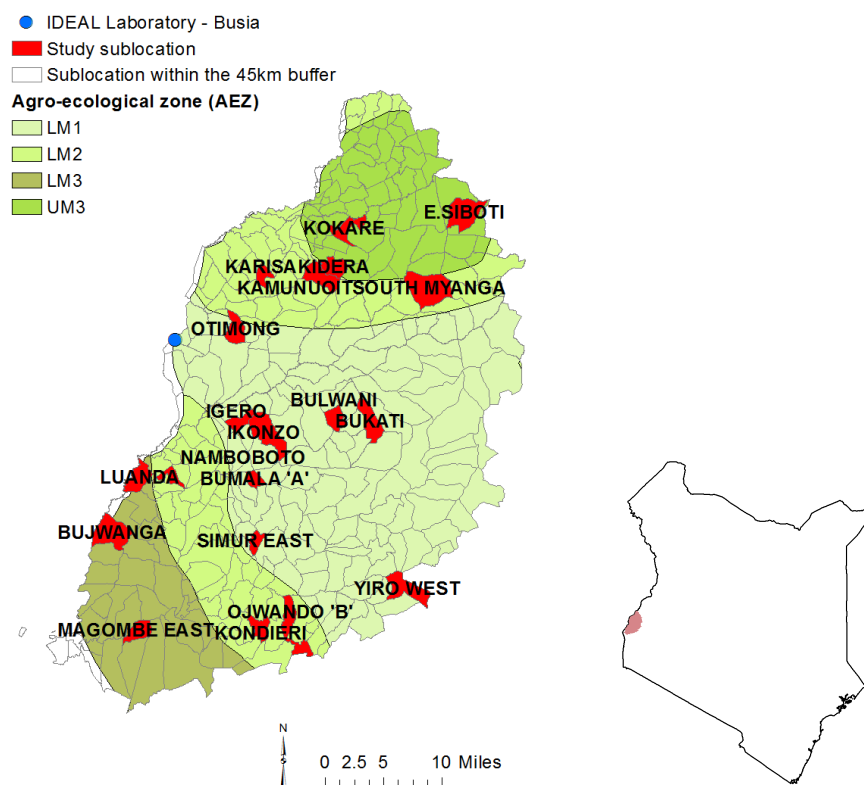
Figure A.1 Map showing the IDEAL project study area in western Kenya, highlighting the selected sublocations in red, the five agro-ecological zones in green and study laboratory in Busia (the blue point). The small insert map shows the study area in relation to the whole of Kenya. Figure is taken from Bronsvoort et al. (2013).

which on average is 1.98 (± 0.1 SE) hectares (range 0.1-23.1ha). All farmers grew food crops in addition to keeping cattle and other livestock species for multiple purposes. The average herd size is 5 (range 1-131) cattle, with indigenous shorthorn zebu being the predominant breed kept.

In 94% of farms, calves were not allowed to graze with adults until after they had weaned. Calves were allowed to suckle from the dam after the farmer had obtained their share of the milk. 60% of farmers provided housing for their livestock in the form of an open yard/kraal surrounded by a fence with no roof; those farmers which did not provided any form of housing left the cattle free or tethered them in the homestead at night. Water was provided for the cattle in 49% of homesteads, the rest drove their animals to a water source 0-10km away.

A.2   Study design

Calves were selected using a stratified two-stage random cluster study design. In the first stage, 20 sublocations (the smallest administrative unit in Kenya) were randomly selected from 5 agro-ecological zones within a radius of 45km from the town of Busia. During the second stage, 28 calves were recruited at less than one week old from each sublocation. A 5 week cycle was used for the recruitment with 4 out of 20 sublocations being visited each week to ensure an even distribution of calves across space and season. Recruitment lasted for two years.

The selection criteria for calf enrolment into the study were as follows: (1) the calf had to be between 3 and 7 days old at recruitment; (2) the calf was not as a result of artificial insemination; (3) the dam was not managed under zero-grazing conditions. This criterion meant that the IDEAL project could avoid the recruitment of exotic cattle and capture new-borne calves. Calves were excluded from the study if they had any congenital deformities. Only one calf per dam could be recruited and a farmer could only have one calf enrolled into the study at one time. Following permission from the farmer to access the calf, and given their willingness to report clinical episodes and not to "self-treat", the calf was enrolled into the IDEAL project. A flat rate of compensation was agreed with the local veterinary office for this. If the calf became ill during the study period, then it was asked that the farmer called the IDEAL project office and a veterinary surgeon would then examine the calf. If the calf was seriously ill or welfare was an issue then the calf would be treated. Calves were censored from the study following treatment.

A.3   Data collection

At the recruitment visit, a questionnaire was completed by the calf's owner. This questionnaire collected information about the farm and farmer, other livestock, water sources and animal husbandry practices. Calf locations were geo-referenced using hand-held GPS devices (Garmin 12, Garmin Kansas, USA).

Calves received routine visits every 5 weeks until they were 51 weeks old. During these visits, individuals were given a systematic physical examination. This included checking the body for lesions and discharges, the presence of ticks and other ecotoparasites, observation of the animal at rest, posture, alertness, rectal temperature, weight, girth, mucus membrane colour and skin elasticity. In addition, a questionnaire was used to update the IDEAL project about other activities on the farm such as illness or treatment of the other livestock as well as animal transactions. During each

routine visit, a standard set of biological samples were taken for laboratory analysis, hematological profiling and parasite identification. The tests and parasites screened for by the IDEAL project are described in Table A.1.

If the calf became unwell (for example it had a temperature >40.5°C, anorexia, diarrhoea, respiratory distress or lameness etc.), then the IDEAL project veterinary surgeon was contacted. The calf then received a full clinical examination and additional samples were collected based on the syndrome observed. A list of possible tests and parasites screened for by the IDEAL project following a clinical episode visit is presented in Table A.1.

Following the death or euthanasia of a calf, a full gross post mortem examination was carried out using standard veterinary procedures. The definitive and aetiological cause of death was identified and samples specific to the aetiology were collected.

The calf's dam was also examined at each visit. A full clinical examination was performed at recruitment and in subsequent visits until the calf was weaned, the girth was measured and body condition was scored and the udder was examined. Phenotypic measurements of coat colour and pattern, horn length and shape, ear shape, size of hump and dewlap was also recorded at recruitment.

There were two periods during the IDEAL project when sampling and recruitment was suspended. In the first, field work was suspended for 6 weeks in 2008 due to political unrest which resulted in a small number of calves missing one or two 5 weekly visits. The second was due to an extended holiday period in 2009/2010. All the data gathered by the IDEAL project is stored in a MySQL database and biological samples are biobanked at the International Livestock Research Institute (ILRL) Nairobi, Kenya.

Table A.1 Table showing the tests and timing of parasites screened for during the IDEAL study. Table taken from Bronsvoort et al. (2013). All data collection and testing was performed by other members of the IDEAL project at the IDEAL project study laboratory in Busia, the International Livestock Research Institute (ILRI), Nairobi, Kenya and Onderstepoort Veterinary Research Institute, the University of Pretoria, South Africa. RB = routine bacteriology; RLB = reverse line blot; FM = faecal examination by McMaster's technique; FC = faecal culture; MIC = routine microscopy; SNP = skin snip and culture; ZN = Ziehl-Neelsen stain; DG = dark ground microscopy; HCT = haematocrit; PCR = polymerase chain reaction; Ab = antibody; Ag = antigen; CE = clinical episode; CL = clinical examination; 7D = recruitment visit; 5W = routine 5 weekly visit; Y = final year visit

Parasite	Test	Time Period
Bovine Diarrhoea virus (Ag)	ELISA (Svanova)	7D
Amblyomma variegatum	CL	7D, 5W, Y

Parasite	Test	Time Period
Boophilus spp.	CL	7D, 5W, Y
Hyalomma spp.	CL	7D, 5W, Y
Hypoderma bovis	CL	7D, 5W, Y
Riphicephalus appendiculatus	CL	7D, 5W, Y
Anaplasma marginale (Ab)	ELISA (ILRI in house)	7D, 5W, Y
Babesia bigemina (Ab)	ELISA (ILRI in house)	7D, 5W, Y
Theileria mutans (Ab)	ELISA (ILRI in house)	7D, 5W, Y
Theileria parva (Ab)	ELISA (ILRI in house)	7D, 5W, Y
Dictyocaulus viviparus (L1)	FB	7D, 5W, Y
Bunostomum trigonocephalum	FM	7D, 5W, Y
Ostertagia ostertagi	FM, FC	7D, 5W, Y
Calicophoron spp.	FM, FC	7D, 5W, Y
Chabertia ovina	FM, FC	7D, 5W, Y
Coccidia spp.	FM, FC	7D, 5W, Y
Cooperia spp.	FM, FC	7D, 5W, Y
Haemonchus placei	FM, FC	7D, 5W, Y
Moniezia spp.	FM, FC	7D, 5W, Y
Nematodirus spp.	FM, FC	7D, 5W, Y
Oesophagostomum radratium	FM, FC	7D, 5W, Y
Toxocara vitulorum	FM, FC	7D, 5W, Y
Trichostrongylus axei	FM, FC	7D, 5W, Y
Trichuris spp.	FM, FC	7D, 5W, Y
Eimeria alabamensis	FM, MIC	7D, 5W, Y
Eimeria auburnensis	FM, MIC	7D, 5W, Y
Eimeria bovis	FM, MIC	7D, 5W, Y
Eimeria cylindrica	FM, MIC	7D, 5W, Y
Eimeria ellipsoidalis	FM, MIC	7D, 5W, Y
Eimeria subspherica	FM, MIC	7D, 5W, Y
Eimeria zuernii	F MIC	7D, 5W, Y
Trypanosoma brucei	HCT, DG, PCR	7D, 5W, Y
Trypanosoma congolense	HCT, DG, PCR	7D, 5W, Y
Trypanosoma spp.	HCT, DG, PCR	7D, 5W, Y
Trypanosoma theileri	HCT, DG, PCR	7D, 5W, Y
Trypanosoma vivax	HCT, DG ,PCR	7D, 5W, Y
Cryptosporidium spp.	ZN, MIC	7D, 5W, Y
Theileria spp.	MIC, (RLB)	7D, 5W, Y, CE

Parasite	Test	Time Period
Rotavirus	ELISA	CE
Trichophyton spp.	MIC	CE
Lumpy skin disease	PCR	CE
Actinomyces sp.	RB	CE
Actinomycetes	RB	CE
Arcanobacterium pyogenes	RB	CE
Bacillus spp.	RB	CE
Clostridium spp.	RB	CE
Coccobacillary	RB	CE
Corynebacterium spp.	RB	CE
Dermatophilus congolensis	RB	CE
E. coli	RB	CE
Klebsiella ozaenae	RB	CE
Klebsiella pneumoniae	RB	CE
Listeria spp.	RB	CE
Micrococcus spp.	RB	CE
Non-pathogenic *Staphylococci*	RB	CE
Pasteurella multocida	RB	CE
Salmonella spp.	RB	CE
Staphylococcus aureus	RB	CE
Staphylococcus epidermicus	RB	CE
Staphylococcus epidermidis	RB	CE
Staphylococcus spp.	RB	CE
Streptococcus bovis	RB	CE
Streptococcus spp.	RB	CE
Weksella zoohelcum	RB	CE
Brucella spp. (Ab)	ELISA (IDEXX)	Dam 7D
Leptospira hardjo	ELISA (Linnodee)	Dam 7D
Sarcocystis spp.	HIS	PM
Bacillus anthracis	RB	PM
M. bovis (Ab)	Bovigam ELISA (Prionics)	Y
Bluetongue virus (Ab)	ELISA (Pirbright Institute)	Y
Epizootic Haemorrhagic Disease virus (Ab)	ELISA (Pirbright Institute)	Y
Bovine Diarrhoea virus (Ab)	ELISA (Svanova)	Y
Bovine Respiratory Syncitial virus (Ab)	ELISA (Svanova)	Y
Infectious Bovine Rhinotracheitis virus (Ab)	ELISA (Svanova)	Y

Table A.1 – continued from previous page

Parasite	Test	Time Period
Neospora caninum (Ab)	ELISA (Svanova)	Y
Parainfluenza Type 3 virus (Ab)	ELISA (Svanova)	Y
Akabane Disease virus (Ab)	ELISA (University of Pretoria)	Y
Palyam group (Ab)	ELISA (University of Pretoria)	Y
Anaplasma bovis	RLB	Y
Anaplasma centrale	RLB	Y
Anaplasma marginale	RLB	Y
Anaplasma ovis	RLB	Y
Anaplasma phagocytophilum	RLB	Y
Babesia bicornis	RLB	Y
*Babesia bigemin*a	RLB	Y
Babesia bovis	RLB	Y
Babesia caballi	RLB	Y
Babesia canis	RLB	Y
Babesia divergens	RLB	Y
Babesia felis	RLB	Y
Babesia gibsoni (Japan)	RLB	Y
Babesia microti	RLB	Y
Babesia motasi	RLB	Y
Babesia odocoilei	RLB	Y
Babesia ovis	RLB	Y
Babesia rossi	RLB	Y
Babesia vogeli	RLB	Y
Borrelia afzelii	RLB	Y
Borrelia burgdorferi s. lao	RLB	Y
Borrelia burgdorferi s. stricto	RLB	Y
Borrelia garinii	RLB	Y
Borrelia valaisiana	RLB	Y
Ehrlichia chaffeensis	RLB	Y
Ehrlichia spp. (Omatjenne)	RLB	Y
Hepatozoon spp. catch-all	RLB	Y
Rickettsia spp. (DnS14) raoultii	RLB	Y
Rickettsia spp. catch-all	RLB	Y
Theileria annae	RLB	Y
Theileria annulata	RLB	Y
Theileria bicornis	RLB	Y

Continued on next page

193

Parasite	Test	Time Period
Theileria buffeli	RLB	Y
Theileria cervi	RLB	Y
Theileria equi	RLB	Y
Theileria equi-like	RLB	Y
Theileria lestoquardi	RLB	Y
Theileria mutans	RLB	Y
Theileria orientalis1	RLB	Y
Theileria spp. (buffalo)	RLB	Y
Theileria spp. (duiker)	RLB	Y
Theileria spp. (kudu)	RLB	Y
Theileria spp. (sable)	RLB	Y
Theileria taurotragi	RLB	Y
Theileria velifera	RLB	Y
Theileria parva	RLB, PCR	Y
Onchocerca spp.	SNP, MIC	Y
M. avium paratuberculosis	ZN	Y
Bluetongue virus	PCR	Y, CE
Epizootic haemorrhagic disease	PCR	Y, CE
Ehrlichia ruminantium	RLB, MIC, PCR	Y, CE

A.4   Morbidity and mortality

548 calves were recruited into the IDEAL project, of which 88 died before 51 weeks of age. Figure A.2, taken from Bronsvoort et al. (2013), shows the life line for each calf, and highlights when calves experienced clinical episodes, died or were euthanized (15 calves). Post mortem was carried out on all but 6 individuals for logistical reasons. Of the remaining 82 deaths, 8 individuals died of non-infectious causes such as cassava poisoning, mis-mothering and trauma etc (Table A.2). Out of the infectious causes of death, East Coast Fever (ECF) was the most common cause with 32 calves dying as a result of it. Other infectious causes of death included Haemonchosis and Heartwater (Table A.2).

275 calves experienced clinical episodes, with 307 occurring at routine visits and 179/216 non-routine (clinical episode) visits were classified as the calf being clinically ill by the visiting veterinarian.
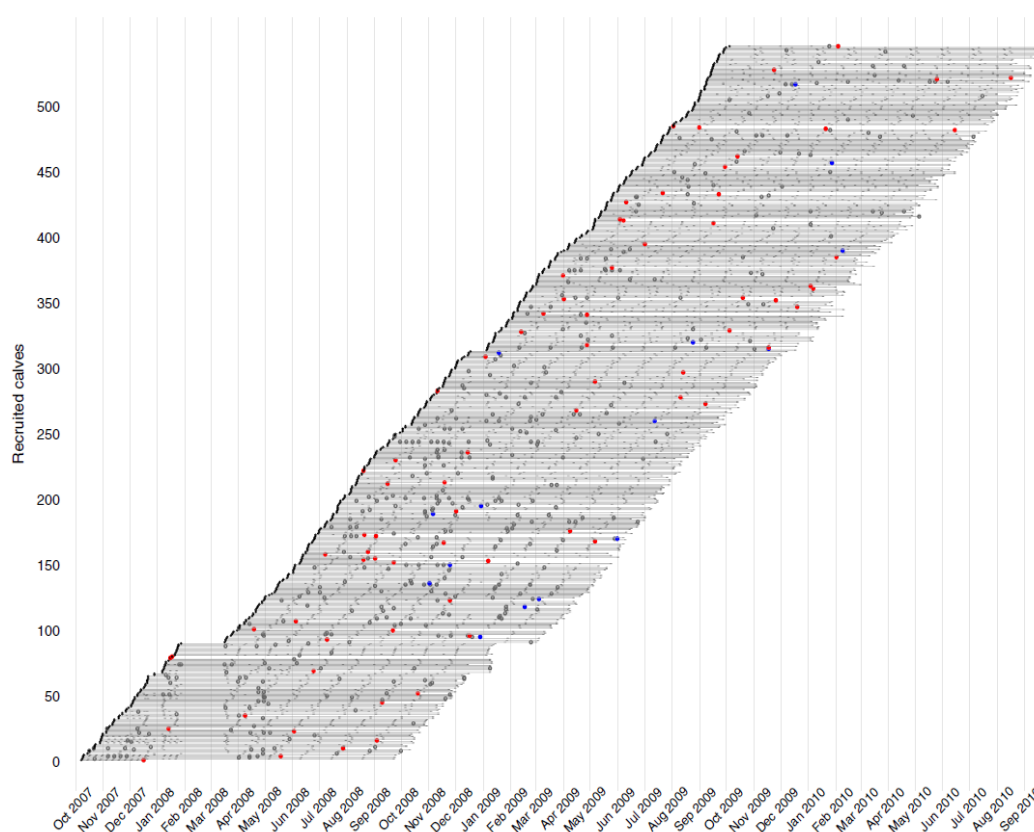
Figure A.2 Life line for each calf showing the time of recruitment (black dot), routine examinations (grey bar), clinical episodes (grey circle) or deaths (red dot = died, blue dot = euthanized) over the 3 years of the IDEAL project. Figure taken from Bronsvoort et al. (2013).

Table A.2 Counts of the primary cause of deaths attributed by expert committee. N.B. 2 additional calves were considered to have died with ECF as a secondary contributing cause, one with heartwater and one with black quarter. Table adapted from Bronsvoort et al. (2013).

Cause of Death	Number of calves
East Coast Fever	32
Unknown	20
Heamonchosis	9
Heartwater	6
Trauma	3
Actiomyces pyogenes, Babesiosis, Bacterial pneumonia, Black Quarter, Cassava, Foreign Body, Mis-mothering, Rabies, Salmonellosis, Trypansomosiasis, Turning sickness, Virual pneumonia	12 (one individual with each cause)
No post-mortem carried out	6
Total	88

A.5 Parasites and exposures

Figure A.3 (also taken from Bronsvoort et al. (2013)) shows that the calves in the IDEAL project were infected with over 50 different parasites. Some parasites such as *Theileria* spp. appear multiple times as several techniques were used to identify them. In addition, some assays, such as microscopy, do not distinguish between species. Only a few parasites where found in the majority of calves, and these main parasites were helminths and protozoan heamoparasites such as such as *T. parva*, *A. marginale*, *B. bigemina* and *H. placei*.
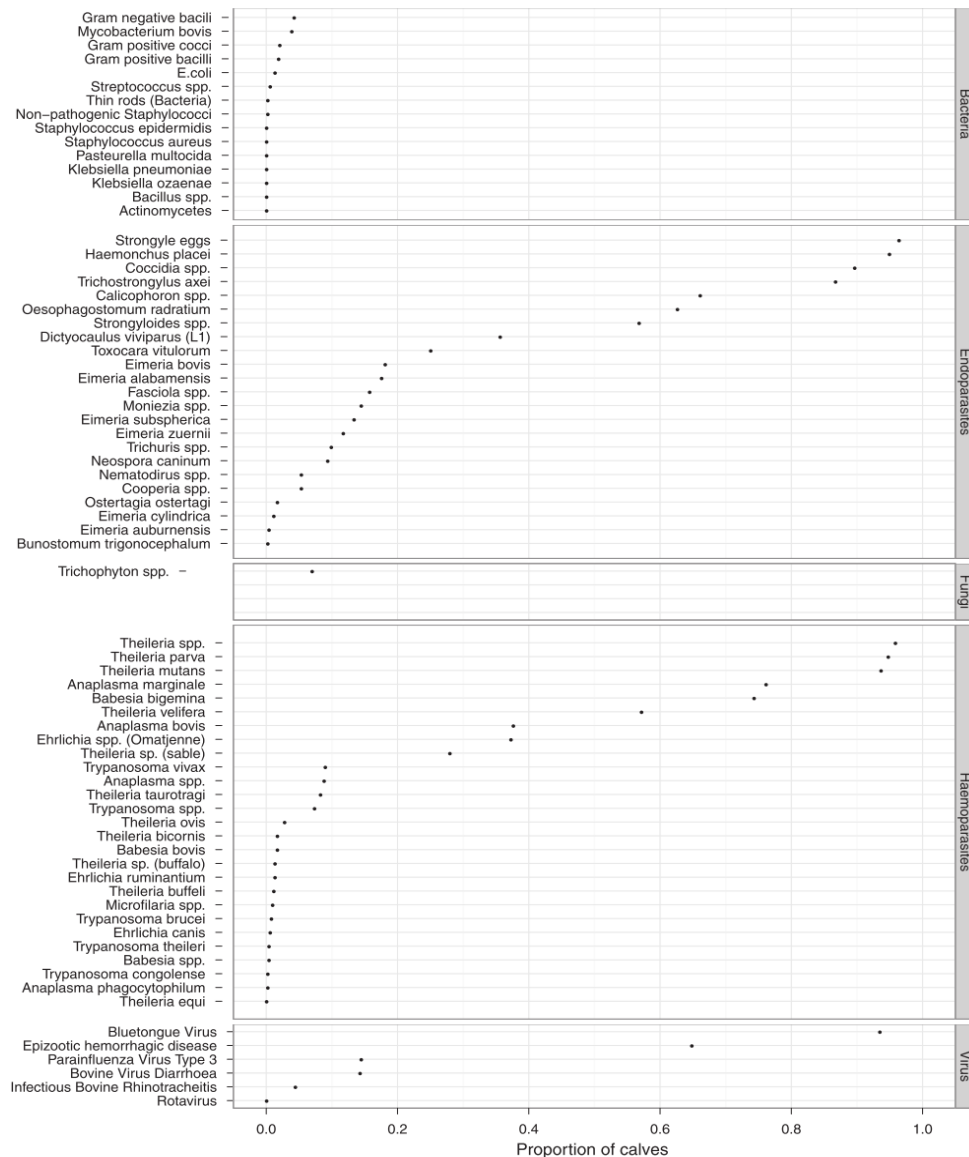
Figure A.3 The proportion of animals which were positive for a given parasite/test combination at any time during their 51 weeks of enrolment in the IDEAL project. Figure is taken from Bronsvoort et al. (2013)

A.6 Previous research into coinfections using data from the IDEAL project

Other manuscripts and PhD theses (Conradie van Wyk, 2012; Jennings, 2013; Thumbi, 2012) which used data gathered by the IDEAL project have reviewed the consequences of coinfection or single infection on the host. For instance, as described in the introduction, individuals concurrently infected with *T. mutans* experience an 89% reduction in mortality associated with *T. parva* (which causes East Coast Fever, ECF), this is thought to correspond to a greater than 40% net reduction in ECF mortality within this population (Jennings, 2013; Woolhouse et al., 2015). Coinfection can also increase the risk of mortality, for example coinfection with *Trypanosoma* spp. increased the calves hazard of death from ECF by 6 times (Thumbi et al., 2014). Independent to the effect of *Trypanosoma* spp., calves coinfected with strongyle eggs experienced a 1.5 increase in the hazard of ECF death per 1000 strongyle eggs (Thumbi et al., 2014).

Coinfection can also effect average daily weight gain of the calf. Calves coinfected with *T. parva* and *A. marginale* experience lower growth rates whereas calves had relatively higher growth rates when they were coinfected with *T. parva* and *T. mutans*, compared to infection with *T. parva* only (Thumbi et al., 2013b).

Conradie van Wyk et al. (2014) investigated the impact of coinfections on the haematological profile of the IDEAL calves. Coinfection with trypanosomes and strongyles was associated with a significant decrease in packed cell volume and platelet counts. In addition, coinfection with *A. marginale* and *T. parva* resulted with a decrease in the platelet counts (Conradie van Wyk et al., 2014).

The evidence for coinfection amongst tick-borne haemoparasites was examined further by Njiiri et al. (2015) using test results from reverse line blots (RLB). Njiiri et al. (2015) found strong positive associations amongst different *Theileria* spp. Two species, *Theileria* sp. (sable) and *T. taurotragi*, were never detected as single infections, they were only detected in the presence of one or more haemoparasites. It is not expected that all *Theileria* species should coinfect since they are transmitted by different tick species. *T. mutans* and *T. velifera* are transmitted by *Amblyomma* spp. whereas *T. parva* and *T. taurotragi* are transmitted by *Rhipicephalus appendiculatus*.

In summary, the calves enrolled in the IDEAL project are infected with many different parasites and coinfections can have both positive and negative consequences for the host. My thesis turns the focus around to answer questions on what causes hosts to be infected with multiple parasites in the first place and to identify parasites which are associated with each other.

Appendix B

Published Manuscript

The published version of Chapter 4. The supplementary materials referred to in the manuscript can be found in Appendix E. This manuscript has been reproduced with the publishers permission.

Variation and covariation in strongyle infection in East African shorthorn zebu calves

R. CALLABY[1,2,3]*, O. HANOTTE[4], I. CONRADIE VAN WYK[5], H. KIARA[6], P. TOYE[6], M. N. MBOLE-KARIUKI[7], A. JENNINGS[8], S. M. THUMBI[2]†, J. A. W. COETZER[5], B. M. de C. BRONSVOORT[9], S. A. KNOTT[1], M. E. J. WOOLHOUSE[1,2] *and* L. E. B. KRUUK[10,1]

[1] *Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Kings Buildings, West Mains Road, Edinburgh EH9 3JT, UK*
[2] *Centre for Immunity, Infection and Evolution, University of Edinburgh, Ashworth Laboratories, Kings Buildings, West Mains Road, Edinburgh EH9 3JT, UK*
[3] *James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, UK*
[4] *School of Life Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK*
[5] *Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private bag X04, Onderstepoort, South Africa*
[6] *International Livestock Research Institute, P.O. Box 30709, Nairobi 00100, Kenya*
[7] *African Union – Interafrican Bureau for Animal Resources (AU-IBAR), Kenindia Business Park building, Museum Hill, Westlands Road, P.O Box 30786, 00100 Nairobi, Kenya*
[8] *The Farm Animal Practice, The University of Edinburgh, Easter Bush, Edinburgh EH25 9RG, UK*
[9] *The Roslin Institute, Easter Bush, University of Edinburgh, Roslin, Midlothian EH25 9RG, UK*
[10] *Division of Evolution, Ecology & Genetics, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia*

SUMMARY

Parasite burden varies widely between individuals within a population, and can covary with multiple aspects of individual phenotype. Here we investigate the sources of variation in faecal strongyle eggs counts, and its association with body weight and a suite of haematological measures, in a cohort of indigenous zebu calves in Western Kenya, using relatedness matrices reconstructed from single nucleotide polymorphism (SNP) genotypes. Strongyle egg count was heritable ($h^2 = 23 \cdot 9\%$, S.E. $= 11 \cdot 8\%$) and we also found heritability of white blood cell counts (WBC) ($h^2 = 27 \cdot 6\%$, S.E. $= 10 \cdot 6\%$). All the traits investigated showed negative phenotypic covariances with strongyle egg count throughout the first year: high worm counts were associated with low values of WBC, red blood cell count, total serum protein and absolute eosinophil count. Furthermore, calf body weight at 1 week old was a significant predictor of strongyle EPG at 16–51 weeks, with smaller calves having a higher strongyle egg count later in life. Our results indicate a genetic basis to strongyle EPG in this population, and also reveal consistently strong negative associations between strongyle infection and other important aspects of the multivariate phenotype.

Key words: gastrointestinal parasite infection, strongyle, indigenous cattle, Kenya, heritability, haematology.

INTRODUCTION

Gastrointestinal parasite infections of livestock are responsible for large economic losses in pastoral systems (Keyyu *et al.* 2003). They reduce weight gain and fertility, and may even cause direct losses through mortality (Wymann *et al.* 2008). Reduction of gastrointestinal parasite infections would therefore improve animal health and remove some of the constraints on livestock enterprises in developing

countries, thereby reducing poverty (Perry and Sones, 2007). However, management of parasite infection requires an understanding of the causes of variation in parasite burdens, variation which can be substantial even between individuals within a population. For example, in indigenous East African Shorthorn Zebu (*Bos indicus*, EASZ) calves in Western Kenya, most individuals experience an apparently low level of strongyle worm infection, whilst others experience a high worm burden and suffer severe consequences (Thumbi *et al.* 2013*a*). In this paper, we explore possible causes of this variation, and quantify its covariation with other variables.

Strongyles are a group of nematode gut worms which produce morphologically similar eggs. Strongyle-type eggs include the species *Haemonchus placei*, *Trichostrongylus axei* and *Oesophagostomum*

* Corresponding author. Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Kings Buildings, West Mains Road, Edinburgh EH9 3JT, UK. E-mail: rebecca.callaby@ed.ac.uk
† Current address: Paul G Allen School for Global Animal Health, Washington State University, Pullman, Washington 99164-7079, USA

radiatum (Urquhart *et al*. 1996). The most common method used to quantify worm burden is a count of the number of strongyle eggs per gramme of faeces (EPG), a non-invasive, relatively easily measured variable. It has been shown that faecal egg counts (FECs) are a good index of parasite burden in Australian cattle, although the relationship between the two may not be exactly linear (Bryan and Kerr, 1989). Variation in strongyle FEC can be due to a variation in susceptibility, resistance, tolerance or exposure to infection by strongyle worms. Evidence from other domestic ungulates suggests that variation in strongyle FEC frequently has a heritable genetic basis: for example, FEC has a heritability of 18% (95% CI = 0·10–0·25) in West African N'Dama cattle (Zinsstag *et al*. 2000), and the heritability is approximately 30% in many other cattle breeds (Stear *et al*. 1988, 1990; Leighton *et al*. 1989). Similarly, strongyle EPG in Scottish Blackface sheep lambs has a heritability of 32% (Riggio *et al*. (2013); see also Beraldi *et al*. (2007); Crawford *et al*. (2006)).

In addition to additive genetic effects, there may also be consistent environmental-based causes of variation in parasite burden between individuals. These 'permanent environmental effects' comprise all variance of non-(additive) genetic origin that persist throughout an individual's life-time, and so for example may include long-running effects of maternal environment or of how an individual was raised and housed: for instance, in a feral Soay sheep population, lambs born as twins or born to very young or old mothers have higher parasite burdens than those born as singletons or to prime-age mothers (Hayward *et al*. 2010). Stear *et al*. (1996) also found higher parasite burdens in Scottish Blackface sheep twins in comparison to singletons. The physical environment that an individual resides in will also be important for determining its exposure to a particular pathogen, which in turn can affect the burden of infection observed (e.g. Batchelor *et al*. 2009). Finally, there may be variation between measures made on an individual at different time points, due to, for example, effects of ageing, immediate climatic effects or simply stochastic variation and measurement error.

Variation in parasite burden may also have implications for the expression of other important traits, especially if parasite resistance is costly and may therefore be traded off against investment in other traits (Norris and Evans, 2000). Such associations can be quantified within individuals by looking at the covariation of parasite burden and other traits, for example morphological variables such as growth rates or weight, or physiological variables, such as haematological parameters, to test for any costs associated with high parasite burdens (e.g. Coltman *et al*. 2001). In particular, one of the strongyle species, *H. placei,* is an important cause of anaemia in ruminants (Kaufmann *et al*. 1992): Conradie van Wyk *et al*. (2013) and Vanimisetti *et al*. (2004) have shown

negative correlations between parasite burden and various haematological parameters in EASZ and sheep. Finally, it is possible that an individual's phenotype at birth may influence their infection risk later in life. For example, in humans, babies that have a lower birth weight are more likely to develop lower respiratory tract infections when they are coinfected with hand, foot and mouth disease (Lu *et al*. 2013). Likewise, Read *et al*. (1994) showed there is a higher risk of childhood infectious disease mortality in lower birth weight babies than heavier ones.

Traditionally, pedigree information has been used to estimate quantitative genetic parameters such as the heritability of a trait (Falconer and Mackay, 1996). More recently, the development of high density SNP beadchips means that novel alternative approaches can be used without reference to pedigree records (Yang *et al*. 2010; Visscher *et al*. 2014). This has reduced previous constraints faced during estimation of heritability in wild populations due to the lack of accuracy and completeness of the pedigree (Pemberton 2008). Bérénos *et al*. (2014) compared heritability estimates produced from using both pedigrees and SNPs from related Soay sheep and demonstrated that heritability estimates obtained from dense SNP data are in correspondence with pedigree estimates.

The Infectious Diseases of East African Livestock (IDEAL) project (Bronsvoort *et al*. 2013) provides a unique opportunity to study natural variation and covariation in strongyle EPG in indigenous EASZ from Western Kenya. Cattle in this region are minimally managed and there is very limited use of vaccination or other preventative measures against infectious diseases. Therefore the study population is similar to a wild population in that, unlike other estimates of genetic variation in FEC in domestic populations (e.g. Bishop *et al*. (1996)), animals have not been treated for anthelmintics (those individuals which were treated with anthelmintics were retrospectively removed from the cohort as part of the IDEAL study design); variation therefore reflects natural diversity in parasite burden. Calves were enrolled in the study at birth and their infectious disease burden, haematological profiles and growth were tracked for the first year of life (Conradie van Wyk *et al*. 2012; Bronsvoort *et al*. 2013). Strongyle worm burdens (assessed via EPG) have a major impact upon the calves in the study population: for example, an increase in strongyle EPG by a count of 1000 eggs is associated with a 3·3% reduction in weight gain over the first year (Thumbi *et al*. 2013b), and an increase in the hazard of death by 1·5 (95% CI = 1·4–1·7, $P < 0·001$; Thumbi *et al*. (2013a)). Moreover, genome-wide genetic information is available in the form of SNPs as each calf enrolled in the IDEAL project was genotyped with a 50 K Illumina® BovineSNP50 beadchip (Murray *et al*. 2013; Mbole-Kariuki *et al*. 2014), providing the

opportunity to exploit this information to estimate a relatedness matrix and thereby derive estimates of variance components, including additive genetic variance of different traits.

Our aim in this study is to dissect the potential genetic and non-genetic sources of between- and within-individual level variation in strongyle EPG. We present a multivariate analysis of associations between strongyle EPG, body size and a suite of haematological measures. We quantified the variance components of five physiological traits and their covariation with strongyle EPG. Finally we investigated whether the characteristics of newborn calves could be used to predict subsequent EPG levels, by looking at the association between weight at birth and strongyle EPG later in life.

MATERIALS AND METHODS

Study population

Five hundred and forty-eight free-grazing indigenous EASZ calves in Western Kenya were selected using a stratified two-stage random cluster study design. In the first stage, 20 sublocations (the smallest administrative unit in Kenya) were selected from five agro-ecological zones, across an area of roughly 45 × 90 km. Around 28 3–7-day-old calves were recruited from each sublocation, all from different mothers and different farms; see Bronsvoort *et al*. (2013) for a detailed description of the study design. Recruited calves were followed for their first year of life. They were visited every 5 weeks for a clinical examination at which they were weighed and blood and faecal samples were taken for parasite identification and haematological profiling. A total of 446 calves that survived to 51 weeks of age (and had passed the SNP quality control checks, see SNP quality control section below) were included in this analysis, giving a total of 4727 observations and an average of 10·6 visits per calf.

Data collection

The McMaster counting technique (Hansen and Perry, 1994) was performed on the faecal samples from each visit to each calf to quantify the number of strongyle eggs per gramme of faeces (EPG) present. We refer to our measurement of strongyle faecal egg count as EPG (eggs per gramme); though note that this may also be referred to as FEC in the literature.

The other traits considered in this study were: white blood cell count (WBC), red blood cell count (RBC), total serum protein (TSP), absolute eosinophil count (EO) and body weight. Blood cell analysis was automatically performed using the pocH-100iV Diff (Sysmex® Europe GMBG); see Conradie van Wyk *et al*. (2012) for more details. Haematological profiles were produced for the total WBC and RBC. TSP was determined using a refractometer and

EO was quantified by differential counts from thin EDTA blood smears stained with Diff Quick. Previous studies have shown that higher RBC and heavier body weights are associated with lower FECs (Conradie van Wyk *et al*. 2013; Thumbi *et al*. 2013*b*; Vanimisetti *et al*. 2004).

Calves were weighed (in kilogrammes, measured to the nearest 500 g) at recruitment, then again every 5 weeks until 31 weeks of age, and once again at a last visit at 51 weeks. The number of observations for each trait is presented in Table 1.

SNP quality control and construction of the relationship matrix

All calves were genotyped using a 50 K Illumina® BovineSNP50 beadchip v.1. The beadchip contained 55 777 SNPs before quality control, spread evenly throughout the genome with an average of 1895 SNPs on each autosome and 1362 SNPs on the X chromosome (Murray *et al*. 2013). Quality control was applied to all SNP data prior to analysis using GenABEL (Aulchenko *et al*. 2007), with the following criteria: SNP call rate cut-off of 0·9; individual call rate of 0·9 and an identity by state (IBS) threshold cut-off of 0·9. The IBS threshold means that if a pair of individuals is estimated to be exceptionally highly related (e.g. identical twins) then one of the individuals would be removed. The minimum minor allele frequency for SNPs was set to 0·005, to include all SNPs with a minor allele count of 5 or more. Any X chromosome genotypes that were inconsistent with the phenotype were removed. This quality control resulted in 42 119 autosomal and X markers (41 419 autosomal markers plus 700 X markers) and 446 calves for analysis. We explored the effect of varying the quality control parameters and the number of SNPs included in the IBS matrix on the resulting estimates of heritability; details are given in Supplementary Tables 1 and 2; in general, estimates of heritability for strongyle EPG increased with increasing marker density. Plots of the distribution of the minor allele frequencies at SNP markers and the association between linkage disequilibrium and the distance between pairs of SNPs are presented in Supplementary Figure 1.

All SNPs and calves which passed the quality control checks were then used to construct an identity-by-state matrix in GenABEL (Aulchenko *et al*. 2007) using the allele frequency weighted option, giving the kinship coefficients for use in the variance component and heritability analyses described below. The average genomic estimate of kinship between calves as given by the IBS matrix ranged from −0·02 to 0·24. Three pairs of calves had a genomic estimate of relatedness greater than 0·2 and 6 pairs of calves had a genomic estimate of relatedness between 0·15 and 0·2.

Approximately 20% of the calves in the IDEAL study cohort were shown to have some level of

Table 1. Variance components ($\pm$ s.e.) for all traits considered in univariate repeated measures models which include calf age and sex as fixed effects and V_A, V_{PE} and V_{SL} as a random effect

Trait	Number of observations	Mean ($\pm$ s.d.)	V_{SL}	V_A	V_{PE}	V_{RES}	h^2 (%)	V_{PE}/V_P (%)	r^2 (%)	Sex effect estimate
Strongyle EPG	4032	542·72±15·79	0·06±0·03	0·45±0·23	0·08±0·22	1·30±0·03	23·92±11·83	4·32±11·53	31·44±2·16	0·23±0·08
WBC (×10^3 μL^{-1})	4693	11·08±3·44	0·22±0·15	3·15±1·22	0·86±1·17	7·17±0·16	27·63±10·56	7·54±10·28	37·17±1·97	−0·15±0·21
RBC (×10^6 μL^{-1})	4281	8·79±2·03	0·12±0·06	0·57±0·41	0·60±0·04	1·95±0·04	17·60±12·46	18·50±12·36	39·68±2·10	−0·26±0·11
TSP (gdL^{-1})	4721	7·98±0·90	2·56×10^{-3}±3·15×10^{-3}	0·07±0·04	0·03±0·04	0·38±0·01	14·50±8·20	7·05±8·09	22·08±1·63	−0·04±0·04
EO	3690	0·14±0·14	1·71e^{-4}±1·12e^{-4}	9·44e^{-5}±8·78e^{-4}	1·74e^{-3}±9·14e^{-4}	1·57e^{-2}±3·89e^{-4}	0·53±4·97	9·84±5·15	11·34±1·44	−0·01±0·01
Weight	3338	1·55±0·18	4·19e^{-4}±2·41e^{-4}	1·91e^{-3}±1·87e^{-3}	4·46e^{-3}±1·87e^{-5}	2·92e^{-3}±7·69e^{-5}	19·64±19·22	45·93±19·06	69·89±1·67	0·02±0·01

The proportion of total variance (V_P) explained by the permanent environment variance (V_{PE}) is also presented. The total number of calves for each trait is 446. EO, transformed EO (×10^3 μL^{-1}, log$_{10}$(EO+1)); weight, transformed body weight (kg, log$_{10}$(weight)); V_{SL}, sublocation variance; V_A, additive genetic variance; V_{PE}, permanent environment variance; V_{RES}, residual variance; h^2, heritability; V_{PE}/V_P (%), proportion of the total phenotypic variance explained by the permanent environment variance expressed as a percentage; r^2, repeatability; sex effect estimate, the effect estimate of being male.

introgression from European taurine (ET) cattle, although calves that were first generation offspring from ET were explicitly excluded from the study (Bronsvoort *et al.* 2013; Mbole-Kariuki *et al.* 2014). These calves with lower levels of ET introgression were included in our study since the aim of the study was to describe the components of variation in strongyle EPG in the population. The effect of excluding the introgressed calves on the heritability estimates is presented in the supplementary materials (Supplementary Table 3).

Statistical analysis

Trait distributions. In order to account for the distribution of the strongyle EPG counts, we used generalized linear mixed models (GLMMs) with a negative binomial distribution and log link function; as observations of strongyle EPG were in multiples of 50, they were first divided by 50 so that the data resemble typical count data. Note that estimates of variance components for EPG are therefore on a latent scale rather than on the original data scale (Nakagawa and Schielzeth, 2010). All other variables were analysed assuming Gaussian distributions. Body weight was first transformed to log$_{10}$ (weight) and EO to log$_{10}$ (EO + 1) to account for their slightly skewed distribution.

A significant increase in RBC was found between the calves aged 1 *vs* aged 6 weeks old, followed by a general decreasing trend in calves aged 6–51 weeks (Supplementary Figure 2 and Conradie van Wyk *et al.* 2012). We therefore focused our analysis of RBC on calves aged 6–51 weeks old for RBC. Removal of the records from 1-week-old calves did not affect the direction of associations observed and only resulted in small changes to the variance and heritability estimates.

Random effects and variance components estimation. We used an animal model to estimate the variance components of each trait (Lynch and Walsh, 1998; Kruuk 2004). Animal models are a form of mixed model, with fixed and random effects, that can break phenotypic variation down into the different components via a model of the form:

$$y = Xb + Za + Pc + Sd + e$$

where y is the phenotype of interest and b is a vector of fixed effects that are unknown constants that affect the mean of the distribution. The random effects, which determine the variance of the trait, were additive genetic (a), permanent environment (c), sublocation (d) and residual effects (e). In particular, a is a vector associated with the identity-by-state matrix (see Visscher *et al.* (2008) and Powell *et al.* (2010) for more details on calculating heritabilities using identity-by-state matrices rather than pedigrees) and is derived from the principle that if a trait has a high

degree of genetic variance relative to its other components of variance, pairs of relatives will have high phenotypic similarity. $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{P}$ and $\boldsymbol{S}$ are all design matrixes corresponding to the appropriate fixed or random effects. Permanent environmental effects are measurable because of the repeated observations on the same individual; this between-individual variation is likely to result from long-term environmental or non-additive genetic effects, and in this case will probably incorporate most of any maternal effects (Kruuk and Hadfield, 2007). The total phenotypic variance (V_P) for a trait was therefore broken down into the additive genetic variance (V_A), permanent environmental variance (V_{PE}), sublocation variance (V_{SL}) and residual variance (V_R):

$$V_P = V_A + V_{PE} + V_{SL} + V_R$$

The narrow-sense heritability of a trait (h^2) is defined as the proportion of phenotypic variance (V_P) explained by the additive genetic variance (V_A), $h^2 = V_A/V_P$. It describes the extent to which differences between individuals are determined by additive genetic effects (Falconer and Mackay, 1996). We also report the repeatability (r^2) of each trait, defined as the proportion of the phenotypic variance due to consistent differences between individuals and is given by the ratio of the between individual variance to the total variance, $r^2 = (V_A + V_{PE} + V_{SL})/V_P$.

The covariances between traits can be investigated using multivariate models. By extending the above approach of variance partitioning to multiple traits, and linking them through a covariance term in the random effects, we can ask how much of the phenotypic covariance (COV_P) between traits is due to covariance of the different random effects described above, for example covariance in the permanent environment effects (COV_{PE}).

All statistical analyses were carried out in ASReml version 3.0.5 (Gilmore *et al.* 2006).

Components of variation in strongyle EPG.
Estimation of the components of variance of strongyle EPG at each visit indicated that there was insufficient statistical power to analyse measures at every visit separately. In order to overcome this, we used a univariate animal model fitted with a negative binomial distribution to estimate the heritability of strongyle EPG across all ages. Age (as a multi-level factor) was fitted as a fixed effect to account for changes across visits in mean EPG with age. Sex was also included in this model as a fixed effect and V_A, V_{PE} and V_{SL} were fitted as random effects. Unlike other studies which have estimated genetic variation in FEC in domestic animals (e.g. Bishop *et al.* (1996)), individuals in this study population have not been treated with anthelmintics, and so represent natural levels of variation. Repeated observations on individuals are therefore not necessarily independent assessments of resistance, because nematodes might

persist between sample dates. However our mixed models account for the repeated measures structure of the data by fitting a permanent environment effect, defining the number of individuals as the appropriate number of independent observations (Kruuk and Hadfield 2007).

The significance of V_A was evaluated by comparing the component estimate to the standard error, as it is not advisable to carry out likelihood ratio tests (LRTs) for GLMMs with negative binomial errors in ASReml (Gilmore *et al.* 2006). Finally, for comparison with previous studies which have analysed FECs assuming Gaussian errors (Stear *et al.* 1990; Bishop *et al.* 1996; Coltman *et al.* 2001; Beraldi *et al.* 2007), we also present analyses of linear mixed models assuming a normal distribution of $\log_{10}$ (strongyle EPG + 50). These results are presented in the supplementary materials.

Components of variation in physiological traits. The components of variance in the physiological traits were examined by constructing a univariate Gaussian repeated measures animal model for each trait. As above, age and sex were included as fixed effects, and V_A, V_{PE} and V_{SL} were fitted as random effects in all models. The significance of V_A for each trait was assessed with a LRT comparing the full animal model to one in which the additive genetic variance was set to zero.

Associations between strongyle infection and physiological traits. We assessed associations between strongyle infection and the physiological traits (and body size) in three different ways, by: (1) testing whether infection affected mean levels of the physiological traits; (2) testing whether size at birth predicted levels of strongyle infection later in life; and (3) assessing components of covariance between all traits.

The effects of strongyle infection on the physiological traits were therefore first quantified by univariate animal models with the trait as the response variable and explanatory variables of age at visit, calf sex and strongyle EPG classified into two categories of 'high' and 'low' EPG. A 'high' strongyle EPG was defined as a value above the median strongyle EPG across all visits (200 EPG), and a 'low' strongyle EPG one below the median. This categorization was chosen to reflect the non-linearity in effect of strongyle EPG estimate of effect (Conradie van Wyk *et al.* 2013). All of the explanatory variables were coded as factors and V_A, V_{PE} and V_{SL} were fitted as random effects.

Secondly, we tested whether a calf's phenotype very early in life was an informative predictor of our index of infection burden, EPG, later in life, and specifically whether the calf's recruitment weight predicted strongyle EPG later in the first year of life. This was achieved by constructing a univariate animal model with a negative binomial distribution
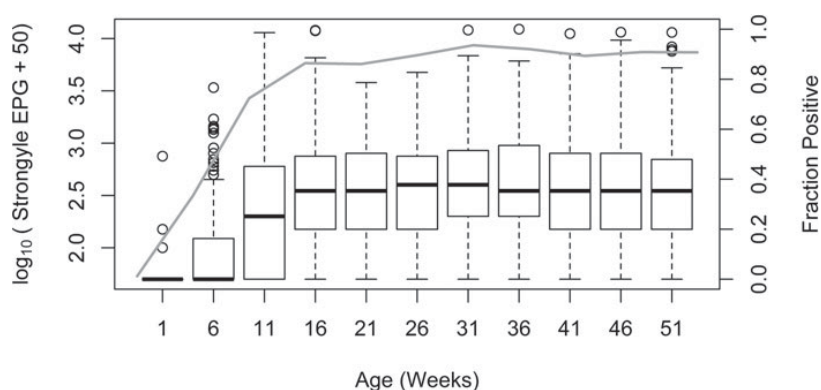
Fig. 1. Distribution of strongyle EPG (box plots, left-hand axis) and the fraction of calves which tested positive at each age (right-hand axis). The black heavy solid lines in each box are the median EPG at each age group, the bottom and top of the box represent the 25th and 95th percentiles, respectively, and the whiskers represent 1·5 times the interquartile range. Points beyond the whiskers are outliers. Strongyle EPG is transformed as log_{10}(strongyle EPG + 50). The solid grey line represents the fraction of tested calves positive for strongyle eggs at each age.

to evaluate the effect of calf weight at recruitment (when the calf is less than 1 week old) on strongyle EPG in older calves (aged 16–51 weeks, following a plateau in median strongyle EPG after 16 weeks, Figure 1). This model includes calf age and sex as fixed effects and V_A, V_{PE} and V_{SL} as random effects. The magnitude and directionality of association between the trait and strongyle EPG is given by the parameter estimate, whilst its significance was assessed using Wald F statistics.

Thirdly, the covariances and correlations between strongyle EPG and the physiological traits were assessed by constructing a multivariate model of all six traits (strongyle EPG, WBC, RBC, TSP, EO and weight), using measures across the whole year for all traits. Calf age and sex were included as fixed effects and strongyle EPG was fitted with a negative binomial error distribution, whilst the other traits were fitted with a Gaussian error distribution. The resulting six-trait multivariate model was computationally much more demanding than the univariate models described above, due to the much greater number of parameters (an extra 80 parameters) being estimated. We therefore had to take several steps to facilitate reliable convergence. Firstly, we were unable to separate between-individual differences into genetic *vs* permanent environment effects, so we restricted the analysis to separating between- *vs* within-individual-level variances and covariances, omitting the genetic relationship matrix from the model. By only including calf identity as a random effect, we obtained estimates of the individual- (phenotypic-) level variance, which reflects consistent differences between individuals; similarly, the model partitions the total phenotypic *covariance* between two traits into that due to between-individual *vs* within-individual (residual) components. Secondly, we were unable to fit sublocation as a random effect in the models, so it was omitted from

the multivariate analysis. Note however that sub-location was never significant in any of the univariate models (Table 1), and its effects will be included in the permanent environment effect (co)variance. Since LRTs are not advisable with GLMMs with negative binomial errors in ASReml (Gilmore *et al*. 2006), significance of estimates was assessed based on their magnitude relative to the standard error.

RESULTS

Summary statistics

Out of the 4032 visits with faecal samples taken from the 446 live calves that passed the genetic quality control checks, strongyle eggs were detected in 3071 (76·2%) visits using the McMasters technique. The overall median number of strongyle EPG of faeces was 200 EPG (range: 0–12250 EPG). All calves were infected with strongyle eggs at some point during their 51 weeks of inclusion in the study. Infection rates increased up to 16 weeks of age, and then levelled off afterwards, with an average of 89·8% of visits showing non-zero EPG between the ages of 16–51 weeks, and a median strongyle EPG of 300 EPG (range: 0–12250 EPG). The median strongyle EPG and the fraction of calves positive at each age are shown in Figure 1.

Components of variation in strongyle EPG

Additive genetic variance contributed the most (after residual variance) to the overall variance in strongyle EPG, resulting in heritable variation in strongyle EPG in EASZ calves ($h^2 = 23·9\%$, s.e. = 11·8%, Table 1). In contrast, the contribution of permanent environmental effects to the overall variance was relatively low (4·3%, s.e. = 11·5%, Table 1 and Figure 2). Strongyle EPG had a repeatability of 31·4%
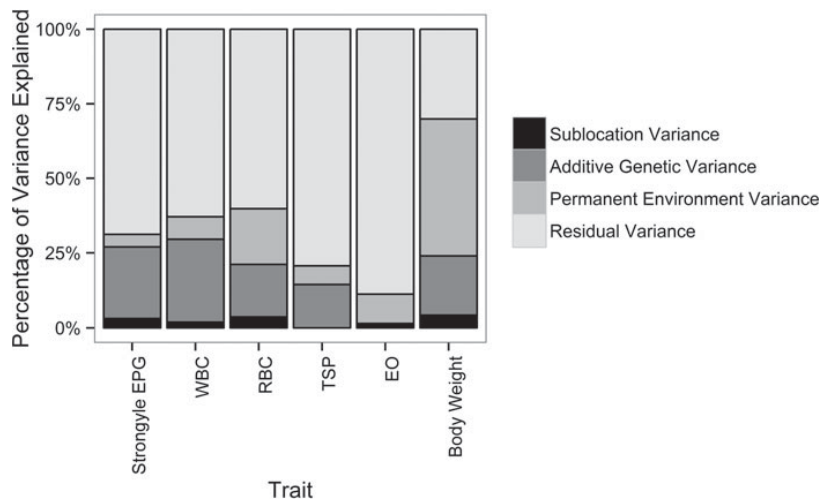
Fig. 2. Percentage of variance explained by each component for all traits investigated. Details of the full model (including SEs on variance component estimates) are given in Table 1. Strongyle EPG uses an animal model fitted with a negative binomial distribution model. WBC, white blood cell count ($\times 10^3 \mu L^{-1}$); RBC, red blood cell count ($\times 10^6 \mu L^{-1}$); TSP, total serum protein ($g\,dL^{-1}$); EO, transformed absolute eosinophil count ($\times 10^3 \mu L^{-1}$, $\log_{10}(EO+1)$); body weight, transformed body weight (kg, $\log_{10}(weight)$).

(S.E. = 2·2%). In addition, male calves had a higher strongyle EPG than female calves (effect estimate = 0·23, S.E. = 0·08, *P* value = 0·01).

Complete removal of the 'introgressed' calves from the study resulted in a lower heritability estimate and larger standard errors, whilst inclusion of the ET introgression as a fixed effect did not alter the heritability estimate (with ET introgressed calves included $h^2 = 23\cdot9\%$, S.E. = 11·8%, *N* calves = 446; with ET introgression included as a fixed effect, $h^2 = 25\cdot7\%$, S.E. = 11·9%, *N* calves = 446; with ET introgressed calves excluded $h^2 = 13\cdot3\%$, S.E. = 13·4%, *N* calves = 353; see Supplementary Tables 1 and 3).

For comparison of the negative binomial errors model with models assuming Gaussian errors, we present analyses of linear mixed models assuming a normal distribution of $\log_{10}$ (strongyle EPG + 50) in Supplementary Table 4. Both methods produce similar estimates of heritability of strongyle EPG although, notably, the standard errors are much larger with the GLMM.

Components of variation in physiological traits

The age-related profiles for the physiological traits are shown in Supplementary Figure 2 (split according to whether the calf had high or low EPG at the time). WBC, EO and weight all increased with age, as expected. However, RBC increased rapidly until 6 weeks old and then declined sharply. A decline from birth in TSP was observed until 21 weeks of age when TSP started to increase again. These distributions and the effect of coinfections on WBC are

discussed in Conradie van Wyk *et al.* (2012) and (2013), respectively.

Estimates of the variance components and the heritability of each trait are shown in Table 1. WBC was the only physiological trait to show evidence for significant V_A (LRT: $\chi^2 = 8\cdot8$, D.F. = 1, *P* = 0·003, Table 1; $h^2 = 27\cdot6\%$, S.E. = 10·6%). There were large differences between traits in the proportion of the total variance (V_P) explained by each variance component: for example, permanent environment effects explained most (45·9%, S.E. = 19·1%) of the total variance in body weight, but only a relatively small proportion of the variance in the other parameters (7·1–18·5%; Table 1). Weight had the highest repeatability of 69·9% (S.E. = 1·7%); repeatability otherwise ranged from 11·3% (S.E. = 1·4%) for EO to 37·2% (S.E. = 2·0%) for WBC.

Associations between strongyle infection and physiological traits

Effect of strongyle infection on physiological traits.
We found significant effects of strongyle infection on all the physiological traits considered. The impact of strongyle EPG on every trait at each age is illustrated in Supplementary Figure 2 and quantified in Table 2. Table 2 shows that calves with a higher strongyle EPG at a given age tended to have a lower RBC, TSP and EO than those with a lower strongyle EPG. Furthermore, calves with a high strongyle EPG were also lighter than those with a lower EPG (by −0·02 $\log_{10}$ (kg); S.E. = 0·03 on average, Table 2). Similar results were observed when a continuous measure of EPG ($\log_{10}$ (strongyleEPG + 50)) rather than a binary measure was used as an explanatory variable.

205

Table 2. The effect of a high or low strongyle EPG at a given age on the trait of interest, using univariate animal models

Trait	WBC				RBC				TSP			
	Estimate	S.E.	F	P value	Estimate	S.E.	F	P value	Estimate	S.E.	F	P value
Intercept	9·676	0·217	8049·50	<0·001	10·560	0·124	10999·24	<0·001	9·690	0·041	200000·00	<0·001
High strongyle EPG	−0·471	0·102	0·35	0·550	−0·362	0·054	236·11	<0·001	−0·090	0·023	646·29	<0·001
Sex (effect of being male)	−0·046	0·210	0·05	0·819	−0·224	0·111	4·04	0·047	−0·034	0·036	0·88	0·349
Calf age (11 factor levels)	NA	NA	25·94	<0·001	NA	NA	184·00	<0·001	NA	NA	269·40	<0·001

Trait	EO				Weight			
	Estimate	S.E.	F	P value	Estimate	S.E.	F	P value
Intercept	0·055	0·009	1166·54	<0·001	1·267	0·008	7274·08	<0·001
High strongyle EPG	−0·007	0·005	10·96	0·001	−0·015	0·003	3579·83	<0·001
Sex (effect of being male)	−0·009	0·006	2·22	0·14	0·018	0·008	5·44	0·021
Calf age (11 factor levels)	NA	NA	37·15	<0·001	NA	NA	2682·21	<0·001

A high EPG is defined as being above the median strongyle EPG whilst a low EPG is defined as being below the median strongyle EPG. The median is the overall median taken across all visits. The significance is given by the Wald F statistic. NA, not applicable, as multiple factor level estimates are not reported.

Table 3. Association between strongyle EPG in older calves (aged 16–51 weeks old) and the calf's weight at the recruitment visit (calf aged <1 week), using a univariate animal model.

Variable	Estimate	S.E.	P value
Calf weight at recruitment (kg)	−0·030	0·012	0·011
Calf sex (effect of being male)	0·276	0·087	0·006
Calf age (7 factor levels)	NA	NA	0·185

Does weight at first visit predict strongyle infection in older calves? Weight at the recruitment visit (when the calf was less than a week old) was significantly associated with later strongyle EPG: calves that were lighter at the first visit had a higher strongyle EPG when aged 16–51 weeks old than calves that were heavier (Table 3). As above, males also had higher levels of EPG.

Covariances between strongyle EPG and physiological traits. The individual-level and residual covariances between strongyle EPG and the physiological traits of interest are shown in Table 4. All traits had a negative individual-level covariance with strongyle EPG whilst positive covariances were found amongst all the blood parameters and weight. This indicates that an increase in strongyle EPG was associated with a decrease in blood parameters and weight, whilst an increase in weight, etc. was associated with an increase in blood parameters and *vice versa*. Comparison of the between-individual *vs* residual (within-individual) variance showed that both follow the same pattern, but that there were higher levels of between-individual than residual (within-individual) level correlations.

DISCUSSION

Our analyses of data from zebu calves in Western Kenya quantified several sources of variation: firstly, in strongyle worm burdens, and secondly in body size and a suite of haematological parameters that we anticipated might be affected by strongyle infection. Measures of associations between strongyle EPG and the physiological traits were consistently negative, suggesting a possible cost of increased parasite burdens. Below, we discuss each of these aspects of our results in turn.

Components of variation in strongyle EPG

Our results show, firstly, substantial changes with age in median levels of strongyle EPG in EASZ calves. The difference in median strongyle EPG in young

Table 4. Covariance/variance/correlation matrix for the between-individual and residual (within-individual) level variance between strongyle EPG and trait

	Strongyle EPG	WBC	RBC	TSP	EO	Body weight
Individual level variance						
Strongyle EPG	*0·620 (0·065)*	−0·120 (0·068)	−0·386 (0·062)	−0·192 (0·075)	−0·140 (0·098)	−0·182 (0·060)
WBC	−0·204 (0·118)	*4·678 (0·426)*	0·133 (0·064)	0·284 (0·068)	0·468 (0·079)	0·280 (0·054)
RBC	−0·355 (0·067)	0·336 (0·168)	*1·365 (0·128)*	0·251 (0·069)	0·138 (0·092)	0·399 (0·050)
TSP	−0·046 (0·019)	0·188 (0·049)	0·090 (0·027)	*0·094 (0·011)*	0·030 (0·104)	0·273 (0·061)
EO	−0·005 (0·004)	0·048 (0·010)	0·008 (0·005)	0·0004 (0·001)	*0·002 (0·0004)*	0·368 (0·077)
Body weight	−0·013 (0·004)	0·053 (0·011)	0·041 (0·006)	0·007 (0·002)	0·002 (0·0003)	*0·008 (0·001)*
Residual level variance						
Strongyle EPG	*1·124 (0·041)*	−0·095 (0·025)	−0·140 (0·025)	−0·124 (0·025)	−0·022 (0·026)	−0·137 (0·025)
WBC	−0·255 (0·069)	*6·472 (0·239)*	0·345 (0·023)	0·152 (0·025)	0·246 (0·024)	0·126 (0·026)
RBC	−0·214 (0·040)	1·262 (0·101)	*2·065 (0·046)*	0·265 (0·024)	0·082 (0·026)	0·209 (0·025)
TSP	−0·067 (0·014)	0·199 (0·034)	0·196 (0·020)	*0·265 (0·010)*	0·032 (0·026)	0·182 (0·025)
EO	−0·003 (0·003)	0·078 (0·009)	0·015 (0·005)	0·002 (0·002)	*0·016 (0·001)*	0·018 (0·026)
Body weight	−0·007 (0·001)	0·016 (0·003)	0·015 (0·002)	0·005 (0·001)	0·0001 (0·0002)	*0·002 (0·0001)*

Covariances are shown below the diagonal (in italics), the associated correlations above the diagonal and variances on the diagonal. Standard errors are in brackets. WBC, white blood cell count ($-10^3 \mu L^{-1}$); RBC, red blood cell count ($\times 10^6 \mu L^{-1}$); TSP, total serum protein (gdL^{-1}); EO, transformed absolute eosinophil count ($\times 10^3 \mu L^{-1}$, $\log_{10}(EO + 1)$); body weight, transformed body weight (kg, $\log_{10}(weight)$).

(age 1–11 weeks) and old (age 16–51 week) calves is possibly due to weaning, with calves moving more once they are weaned and so older calves being at higher risk of becoming infected due to sampling more areas. We observe lower median FECs then might normally be expected for *Haemonchus* infections (e.g. compare to Hansen and Perry (1994)). However, Kanyari *et al*.'s (2010) study of cattle from a peri-urban area in a neighbouring area of Western Kenya (which included exotic breeds) observed a similar prevalence and mean strongyle EPG (mean = 296, range = 0–8300 EPG (Kanyari *et al*. 2010) and Fig. 1 for comparison). Secondly, as in other studies for example (Hayward *et al*. (2009) and Moore and Wilson (2002)); male calves have a higher strongyle EPG then female calves.

Thirdly, our analyses indicated that strongyle EPG was heritable ($\underline{h^2} = 23 \cdot 9\%$, s.e. = 11·8%). Similar heritabilities have been observed in feral Soay sheep lambs on St Kilda ($h^2 = 26\%$, s.e. = 12%, Beraldi *et al*. (2007)) and in Scottish Blackface sheep ewes ($h^2 = 23\%$, s.e. = 9%, (Bishop and Stear, 2001)). These estimates are from models which included the same fixed effects of age and sex as used in our analysis, but they also included additional fixed effects such as weight and twin status, so direct comparisons of heritability need to treated cautiously (Wilson, 2008). As we have found evidence for the presence of heritable variation in strongyle EPG, it may therefore be possible for selection for parasite resistance to occur. Quantitative trait loci and SNPs associated with strongyle FEC have been identified in Soay sheep (Beraldi *et al*. 2007) and Blackface lambs (Riggio *et al*. 2013), but so far have not yet been tested for in indigenous African cattle.

Lastly, complete removal of the 'introgressed' calves from the study resulted in a lower heritability estimate and larger standard errors. The decrease in heritability is possibly due to European introgressed calves having a higher genetic variance whilst the larger standards errors are likely to be due to a decrease in sample size. Inclusion of the ET introgression as a fixed effect did not alter the heritability estimate. However, as the focus of the aim of this study is to describe the components of variation in strongyle EPG in the study cohort we wish to include as much variance in the population as possible in the dataset. Furthermore, the level of ET introgression is on a continuous scale and the cut-off to determine what level of introgression should be excluded is somewhat arbitrary.

Components of variation in physiological traits

We found evidence for significant additive genetic variation in WBC in our study population (WBC, $h^2 = 27 \cdot 6\%$, s.e. = 10·6%; $V_A = 3 \cdot 1$, s.e. = 1·2). This is in accordance with other analyses of WBC count, which have found it to be heritable in both humans

and pigs ($h^2 = 35\%$, s.e. = 9%, Pankow *et al*. (2001); $h^2 = 29\%$, s.e. = 10%, Clapperton *et al*. (2009), respectively). None of the other traits investigated showed evidence for significant additive genetic variation. However, Rowlands *et al*. (1995) showed that packed red-cell volume was heritable in zebu ($h^2 = 32\%$, s.e. = 7%, sample size = 936) and body weight is known to be highly heritable in many other species, including in a much larger study of beef cattle ($h^2 = 41\%$; Marshall (1994)). More generally, haematological parameters are highly heritable in humans, for example haemoglobin levels, RBC, WBC and platelet numbers have heritability estimates of 37, 42, 62 and 57%, respectively (Garner *et al*. 2000). The difference with our results may reflect limited statistical power. In addition, age may be playing an important role in determining the overall (co)variance seen, as heritability (of for example weight and hindleg length in Soay sheep) changes with age (Wilson *et al*. 2007). Furthermore, all of these traits are likely to be polygenic, and so are influenced by many loci of small effect (Goddard and Hayes, 2009), and so it is unlikely that all of the causal loci were detected given the low linkage disequilibrium in EASZ (see below).

Possible biases in heritability estimation

Using our SNP data, we have demonstrated here that it is possible to estimate the heritability of select traits without the need for pedigree information or even the presence of close relatives. We found evidence of heritable variation in strongyle EPG and in WBC. However, it is worth noting that our estimates may be slightly lower than the true heritability because of the ascertainment bias of the SNP chip (Matukumalli *et al*. 2009). Additionally, in the absence of close relatives (such as in our study sample, as all the calves had different mothers and the average genomic relatedness from the IBS matrix ranged from −0·02 to 0·24, and only 9 pairs of calves out of the 446 individuals had a genomic estimate of relatedness greater than 0·15), the heritability estimated is determined by the variance explained by causal variants that are in linkage disequilibrium with the genotyped SNPs (Yang *et al*. 2010). Mbole-Kariuki *et al*. (2014) showed that EASZ have lower levels of average linkage disequilibrium between adjacent SNP pairs on the SNP chip than other cattle breeds (Nelore and N'dama cattle). Therefore the residual relatedness (i.e. between two 'unrelated' individuals) is low; consequently unrelated individuals (by known pedigree) will only share very short proportions of the genome. Furthermore, as marker density increases, our estimate of heritability also increased (Supplementary Table 2). These factors suggest that our estimates of heritability may be lower than those which would be estimated using more closely related individuals and more dense markers (Yang *et al*. 2011; Bérénos *et al*. 2014). Similarly, Robinson *et al*.

(2013) found marker-based estimates to be as low as 60% of the value of pedigree-based estimates of heritability of wing length in a wild bird population.

As such, the estimates presented here should be taken as lower limits on the true estimates of heritability of the different traits in this population, which may also explain why we did not find significant heritability for body weight ($h^2 = 19\cdot6\%$ s.e. = 19·2%), a trait which is commonly found to have significant additive variance. However, conversely, use of known relatives can result in an overestimation of the true heritability as relatives may share non-additive effects such as dominance, epistasis and shared environmental conditions, which may then confound estimates of similarity due to genetic effects if not adequately accounted for (Kruuk and Hadfield, 2007). Since our study does not include close relatives, our estimates will not be affected by this issue.

Care needs to be taken in distinguishing additive genetic effects from other sources of variance in this analysis as maternal or shared environment effects may be important. The IDEAL dataset has information on only one calf per mother; therefore we cannot estimate maternal effects explicitly. However, this data structure also means that maternal effects are less likely to confound estimates of additive genetic variance, as the most usual scenario is that covariance between full-sibs or maternal half-sibs due to maternal effects is mistaken for additive genetic effects (Kruuk and Hadfield, 2007). Any maternal effects are most likely to be contained in the permanent environment effect variance; however, there is also the possibility that if the maternal effects themselves are to any extent genetically based and if related mothers are in the same sublocation, they may also contribute to the sublocation variance. Note however that all calves were from different farms, so very immediate local effects will not be generating any covariance between individuals.

It is also worth pointing out that our estimates had relatively large standard errors, especially for the parameters associated with additive genetic effects. This may be a result of the relatively small sample size (446 individuals) and a lack of relatedness structure between calves, though our sample sizes are relatively standard for similar analyses on wild animal populations (e.g. sample sizes are between 306–576 Soay sheep (Coltman *et al*. 2001) and 333–634 red deer (Clements *et al*. 2011) for some heritability estimates on wild mammal populations).

Associations between strongyle infection and physiological traits

Previous work on this study population has also found associations between EPG and other key components of individual phenotypes, specifically survival rates and body size (Thumbi *et al*. 2013*a*, *b*).

Thus strongyle EPG has a major impact on life history in this population. We have added to this information the contribution of the different components of variance in each of these traits, and the observation that birth weight predicts subsequent worm infection.

Calves with a higher strongyle EPG tended to have lower mean EO, WBC, RBC and TSP than those with fewer eggs: these associations applied both to average values across all observations on a calf (the 'individual-level' covariances in Table 4), and within each visit ('residual' covariances in Table 4). Some strongyle species, such as *H. placei*, are important causes of anaemia in cattle (Kaufmann *et al*. 1992). Since anaemia is defined as an erythrocyte count, haemoglobin concentration or packed cell volume below the reference value for that species (Jain, 1993), it is expected that RBC will decrease in association with strongyle infection, as we observed in this study. Furthermore, as some strongyles such as *H. placei* are blood sucking parasites, a reduction in all blood parameters at the same time is likely to be due to total blood loss in calves with high burdens. The loss in TSP will probably also contribute to the reduction in weight. Meanwhile, the negative association between EO and strongyle EPG could be explained by EO having been implicated in the resistance to infection in ruminants. For example, Bricarello *et al*. (2007) found a negative association between nematode FEC and blood eosinophil counts in Nelore-breed cattle.

Calves that were lighter weight at less than 1 week old had a higher strongyle EPG than heavier calves when they are aged 16–51 weeks old. In a study of humans, Raqib *et al*. (2007) observed altered immune function in low birth weight babies which may increase vulnerability to infection later in life. Alternatively, the association could be generated by correlations of both early weight and subsequent strongyle infection with some other unmeasured aspect of individual condition, without requiring any causal component. It is also possible that lighter calves may be eating less and therefore might be expected to have lower intensities of infection, due to sampling fewer areas, but we observe the opposite direction of effect, with lighter calves having a higher strongyle EPG. However we did not monitor the calves' consumption of food during the study so cannot investigate this further.

Concluding remarks

To conclude, in this study we have used relationship matrices reconstructed from SNP genotypes to demonstrate evidence for heritable variation in strongyle EPG in EASZ. We also found significant additive genetic variation in WBC. All additional traits investigated showed negative phenotypic covariances with strongyle EPG throughout the first year: high strongyle EPG was associated with low WBC, RBC, TSP, and EO. Weight at 1 week old was significantly associated with strongyle EPG at 16–51 weeks: smaller calves had a higher strongyle EPG later in life. Our results indicate that additive genetic variation in strongyle EPG is present in this population, and that strongyle EPG is associated with variation in other important variables. Further investigation is needed to understand the physiological mechanisms of the interactions between strongyle EPG and haematological parameters that allow EASZ calves to tolerate a high strongyle EPG.

ETHICAL AND REGULATORY GUIDELINES

The IDEAL project received approval by the University of Edinburgh Ethics Committee (reference number OS 03–06), and the Animal Care and Use Committee of the ILRI. All participating farmers gave informed consent in their native language prior to recruiting their animals into the study.

REFERENCES

Aulchenko, Y. S., Ripke, S., Isaacs, A. and Van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296.
Batchelor, N. A., Atkinson, P. M., Gething, P. W., Picozzi, K., Fèvre, E. M., Kakembo, A. S. L. and Welburn, S. C. (2009). Spatial predictions of Rhodesian Human African Trypanosomiasis (Sleeping Sickness) prevalence in Kaberamaido and Dokolo, two newly affected districts of Uganda. *PLoS Neglected Tropical Diseases* **3**, e563.
Beraldi, D., McRae, A. F., Gratten, J., Pilkington, J. G., Slate, J., Visscher, P. M. and Pemberton, J. M. (2007). Quantitative trait loci (QTL) mapping of resistance to strongyles and coccidia in the free-living Soay sheep (*Ovis aries*). *International Journal for Parasitology* **37**, 121–129.
Bérénos, C., Ellis, P. A., Pilkington, J. G. and Pemberton, J. M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular Ecology* **23**, 3434–3451.

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

Bishop, S. C., Bairden, K., McKellar, Q. A., Park, M. 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.

Bricarello, P. A., Zaros, L. G., Coutinho, L. L., Rocha, R. A., Kooyman, F. N. J., De Vries, E., Gonçalves, J. R. S., Lima, L. G., Pires, A. V. and Amarante, A. F. T. (2007). Field study on nematode resistance in Nelore-breed cattle. *Veterinary Parasitology* **148**, 272–278.

Bronsvoort, B. d. C., Thumbi, S., Poole, E., Kiara, H., Auguet, O., Handel, I., Jennings, A., Conradie, I., Mbole-Kariuki, M., Toye, P., Hanotte, O., Coetzer, J. and Woolhouse, M. (2013). Design and descriptive epidemiology of the Infectious Diseases of East African Livestock (IDEAL) project, a longitudinal calf cohort study in western Kenya. *BMC Veterinary Research* **9**, 171.

Bryan, R. P. and Kerr, J. D. (1989). The relation between the natural worm burden of steers and the faecal egg count differentiated to species. *Veterinary Parasitology* **30**, 327–334.

Clapperton, M., Diack, A., Matika, O., Glass, E., Gladney, C., Mellencamp, M., Hoste, A. and Bishop, S. (2009). Traits associated with innate and adaptive immunity in pigs: heritability and associations with performance under different health status conditions. *Genetics Selection Evolution* **41**, 1–11.

Clements, M. N., Clutton-Brock, T. H., Guinness, F. E., Pemberton, J. M. and Kruuk, L. E. B. (2011). Variances and covariances of phenological traits in a wild mammal population. *Evolution* **65**, 788–801.

Coltman, D. W., Pilkington, J., Kruuk, L. E. B., Wilson, K. and Pemberton, J. M. (2001). Positive genetic correlation between parasite resistance and body size in a free-living ungulate population. *Evolution* **55**, 2116–2125.

Conradie van Wyk, I., Goddard, A., Bronsvoort, B. M. de C., Coetzer, J., Booth, C., Hanotte, O., Jennings, A., Kiara, H., Mashego, P., Muller, C., Pretorius, G., Poole, E., Thumbi, S., Toye, P., Woolhouse, M. and Penzhorn, B. (2012). Hematological profile of East African short-horn zebu calves from birth to 51 weeks of age. *Comparative Clinical Pathology* **22**, 1029–1036.

Conradie van Wyk, I., Goddard, A., Bronsvoort, B. M. de C., Coetzer, J., Handel, I., Hanotte, O., Jennings, A., Lesosky, M., Kiara, H., Thumbi, S., Toye, P., Woolhouse, M. and Penzhorn, B. (2014). The impact of co-infections on the haematological profile of East African Short-horn Zebu calves. *Parasitology* **141**, 374–388.

Crawford, A., Paterson, K., Dodds, K., Diez Tascon, C., Williamson, P., Roberts Thomson, M., Bisset, S., Beattie, A., Greer, G., Green, R., Wheeler, R., Shaw, R., Knowler, K. and McEwan, J. (2006). Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: I. Analysis of outcross pedigrees. *BMC Genomics* **7**, 178.

Falconer, D. and Mackay, T. (1996). *Introduction to Quantitiative Genetics*, 4th Edn. Longman, Harlow.

Garner, C., Tatu, T., Reittie, J. E., Littlewood, T., Darley, J., Cervino, S., Farrall, M., Kelly, P., Spector, T. D. and Thein, S. L. (2000). Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood* **95**, 342–346.

Gilmore, A. R., Gogel, B. J., Cullis, B. R. and Thompson, R. (2006). *Asreml User Guide Release 3*. VSN International Ltd., Hemel Hempsted.

Goddard, M. E. and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**, 381–391.

Hansen, J. and Perry, B. (1994). *The Epidemiology, Diagnosis and Control of Helminth Parasites of Ruminants: A Handbook*, 2nd Edn. ILRAD, Nairobi, Kenya.

Hayward, A. D., Wilson, A. J., Pilkington, J. G., Pemberton, J. M. and Kruuk, L. E. B. (2009). Ageing in a variable habitat: environmental stress affects senescence in parasite resistance in St Kilda Soay sheep. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **276**, 3477–3485.

Hayward, A. D., Pilkington, J. G., Pemberton, J. M. and Kruuk, L. E. B. (2010). Maternal effects and early-life performance are associated with parasite resistance across life in free-living Soay sheep. *Parasitology* **137**, 1261–1273.

Jain, N. C. (1993). *Essentials of Veterinary Hematology*. Wiley, Philadelphia.

Kanyari, P. W. N., Kagira, J. M. and Mhoma, J. R. L. (2010). Prevalence of endoparasites in cattle within urban and peri-urban areas of Lake Victoria Basin, Kenya with special reference to zoonotic potential. *Scientia Parasitologica* **11**, 171–178.

Kaufmann, J., Dwinger, R. H., Hallebeek, A., van Dijk, B. and Pfister, K. (1992). The interaction of *Trypanosoma congolense* and *Haemonchus contortus* infections in trypanotolerant N'Dama cattle. *Veterinary Parasitology* **43**, 157–170.

Keyyu, J. D., Kassuku, A. A., Kyvsgaard, N. C. and Willingham, A. L., III (2003). Gastrointestinal nematodes in indigenous zebu cattle under pastoral and nomadic management systems in the lower plain of the Southern Highlands of Tanzania. *Veterinary Research Communications* **27**, 371–380.

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**, 873–890.

Kruuk, L. E. B. and Hadfield, J. D. (2007). How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology* **20**, 1890–1903.

Leighton, E. A., Murrell, K. D. and Gasbarre, L. C. (1989). Evidence for genetic-control of nematode egg-shedding rates in calves. *Journal of Parasitology* **75**, 498–504.

Lu, Y. P., Zeng, D. Y., Chen, Y. P., Liang, X. J., Xu, J. P., Huang, S. M., Lai, Z. W., Wen, W. R., Von Websky, K. and Hocher, B. (2013). Low birth weight is associated with lower respiratory tract infections in children with hand, foot, and mouth disease. *Clinical Laboratory* **59**, 985–992.

Lynch, M. and Walsh, B. (1998). *Genetics and analysis of quantitative traits*, Sinauer Associates.

Marshall, D. M. (1994). Breed differences and genetic parameters for body composition traits in beef cattle. *Journal of Animal Science* **72**, 2745–2755.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P. L., Sonstegard, T. S. and Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *Plos One* **4**, e5350.

Mbole-Kariuki, M. N., Sonstegard, T., Orth, A., Thumbi, S. M., Bronsvoort, B. M. d. C., Kiara, H., Toye, P., Conradie, I., Jennings, A., Coetzer, K., Woolhouse, M. E. J., Hanotte, O. and Tapio, M. (2014). Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity*. doi: 10.1038/hdy.2014.31.

Moore, S. L. and Wilson, K. (2002). Parasites as a viability cost of sexual selection in natural populations of mammals. *Science* **297**, 2015–2018.

Murray, G. G., Woolhouse, M. E., Tapio, M., Mbole-Kariuki, M. N., Sonstegard, T. S., Thumbi, S. M., Jennings, A. E., Conradie van Wyk, I., Chase-Topping, M. and Kiara, H. (2013). Genetic susceptibility to infectious disease in East African Shorthorn Zebu: a genome-wide analysis of the effect of heterozygosity and exotic introgression. *BMC Evolutionary Biology* **13**, 246.

Nakagawa, S. and Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews* **85**, 935–956.

Norris, K. and Evans, M. R. (2000). Ecological immunology: life history trade-offs and immune defense in birds. *Behavioral Ecology* **11**, 19–26.

Pankow, J. S., Folsom, A. R., Cushman, M., Borecki, I. B., Hopkins, P. N., Eckfeldt, J. H. and Tracy, R. P. (2001). Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis* **154**, 681–689.

Pemberton, J. M. (2008). Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences* **275**, 613–621.

Perry, B. and Sones, K. (2007). Science for development: poverty reduction though animal health. *Science* **315**, 333–334.

Powell, J. E., Visscher, P. M. and Goddard, M. E. (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics* **11**, 800–805.

Raqib, R., Alam, D. S., Sarker, P., Ahmad, S. M., Ara, G., Yunus, M., Moore, S. E. and Fuchs, G. (2007). Low birth weight is associated with altered immune function in rural Bangladeshi children: a birth cohort study. *The American Journal of Clinical Nutrition* **85**, 845–852.

Read, J. S., Clemens, J. D. and Klebanoff, M. A. (1994). Moderate low birth weight and infectious disease mortality during infancy and childhood. *American Journal of Epidemiology* **140**, 721–733.

Riggio, V., Matika, O., Pong-Wong, R., Stear, M. J. and Bishop, S. C. (2013). Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity* **110**, 420–429.

Robinson, M. R., Santure, A. W., DeCauwer, I., Sheldon, B. C. and Slate, J. (2013). Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology* **22**, 3963–3980.

Rowlands, G. J., Mulatu, W., Nagda, S. M., Dolan, R. B. and Dieteren, G. D. M. (1995). Genetic-variation in packed red-cell

volume and frequency of parasitemia in East African zebu cattle exposed to drug-resistant trypanosomes. *Livestock Production Science* **43**, 75–84.

Stear, M. J., Tierney, T. J., Baldock, F. C., Brown, S. C., Nicholas, F. W. and Rudderh, T. H. (1988). Failure to find an association between class I antigens of the bovine major histocompatibility system and faecal worm egg counts. *International Journal for Parasitology* **18**, 859–861.

Stear, M. J., Hetzel, D. J. S., Brown, S. C., Gershwin, L. J., Mackinnon, M. J. and Nicholas, F. W. (1990). The relationships among ecto- and endoparasite levels, class I antigens of the bovine major histocompatibility system, immunoglobulin E levels and weight gain. *Veterinary Parasitology* **34**, 303–321.

Stear, M. J., Park, M. and Bishop, S. C. (1996). The key components of resistance to *Ostertagia circumcincta* in lambs. *Parasitology Today* **12**, 438–441.

Thumbi, S., Bronsvoort, M., Kiara, H., Toye, P., Poole, J., Ndila, M., Conradie, I., Jennings, A., Handel, I., Coetzer, J., Steyl, J., Hanotte, O. and Woolhouse, M. (2013*a*). Mortality in East African shorthorn zebu cattle under one year: predictors of infectious-disease mortality. *BMC Veterinary Research* **9**, 175.

Thumbi, S. M., Bronsvoort, B. M. d. C., Poole, E. J., Kiara, H., Toye, P., Ndila, M., Conradie, I., Jennings, A., Handel, I. G., Coetzer, J. A., Hanotte, O. and Woolhouse, M. E. (2013*b*). Parasite co-infections show synergistic and antagonistic interactions on growth performance of East African zebu cattle under one year. *Parasitology* **140**, 1789–1798.

Urquhart, G., Armour, J., Duncan, J., Dunn, A. and Jennings, F. (1996). *Veterinary Parasitology*, 2nd Edn. Blackwell Science Ltd, Oxford.

Vanimisetti, H. B., Andrew, S. L., Zajac, A. M. and Notter, D. R. (2004). Inheritance of faecal egg count and packed cell volume and their relationship with production traits in sheep infected with *Haemonchus contortus*. *Journal of Animal Science* **82**, 1602–1611.

Visscher, P. M., Hill, W. G. and Wray, N. R. (2008). Heritability in the genomics era – concepts and misconceptions. *Nature Reviews Genetics* **9**, 255–266.

Visscher, P. M., Hemani, G., Vinkhuyzen, A. A. E., Chen, G.-B., Lee, S. H., Wray, N. R., Goddard, M. E. and Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genetics* **10**, e1004269.

Wilson, A. J. (2008). Why h^2 does not always equal V_A/V_P? *Journal of Evolutionary Biology* **21**, 647–650.

Wilson, A. J., Pemberton, J. M., Pilkington, J. G., Clutton-Brock, T. H., Coltman, D. W. and Kruuk, L. E. B. (2007). Quantitative genetics of growth and cryptic evolution of body size in an island population. *Evolutionary Ecology* **21**, 337–356.

Wymann, M. N., Traore, K., Bonfoh, B., Tembely, S., Tembely, S. and Zinsstag, J. (2008). Gastrointestinal parasite egg excretion in young calves in periurban livestock production in Mali. *Research in Veterinary Science* **84**, 225–231.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E. and Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565–569.

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M. G., Hill, W. G., Landi, M. T., Alonso, A., Lettre, G., Lin, P., Ling, H., Lowe, W., Mathias, R. A., Melbye, M., Pugh, E., Cornelis, M. C., Weir, B. S., Goddard, M. E. and Visscher, P. M. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43**, 519–525.

Zinsstag, J., Ankers, P., Njie, M., Smith, T., Pandey, V. S., Pfister, K. and Tanner, M. (2000). Heritability of gastrointestinal nematode faecal egg counts in West African village N'Dama cattle and its relation to age. *Veterinary Parasitology* **89**, 71–78.

Appendix C

Evidence for parasite-parasite associations in East African shorthorn zebu calves

C.1 Parasite characteristics

Table C.1 Summary information on the characteristics of each of the seven parasites included in this analysis which were tested at every 5 weekly visit with a prevalence of greater than 10% across the whole of the IDEAL study period.

	Strongyloides **spp.**	*Oesophagostomum radiatum*
Type	Nematode	Nematode
	Non-strongyle type helminth	Strongyle type helminth
Method of Identification	McMasters quantitative (counting) technique	Faecal culture
Site of Infection	Small intestine	Caecum and colon
Prepatent period	8-14 days (2 weeks)	45 days (6 weeks)
Clinical Symptoms	Diarrhoea, anorexia, dullness, weight loss, reduced growth weight but symptoms are usually only observed in very young animals	Acute infections: Serve dark green diarrhoea, rapid weight loss, sometimes submandibular oedema. Chronic infections: loss of appetite, emaciation with intermittent diarrhoea and anaemia
Mode of Transmission	Faecal-oral	Faecal-oral
References	Urquhart et al. (1996)	Urquhart et al. (1996)

Table C.1 – continued from previous page

	Calicophoron **spp.**	*Trichostrongylus axei*
Type	Trematode	Nematode
	Non-strongyle type helminth	Strongyle type helminth
Method of Identification	Sedimentation	Faecal culture
Site of Infection	Juvenile parasites are found in the small intestine and feed on lining of the intestine. Adults are found in the rumen and reticulum	abomasum
Prepatent period	37 days (in the intermediate snail host)	2-3 weeks
Clinical Symptoms	Loss of appetite, resulting in diarrhoea, anorexia, anaemia and weakness. Heavy infections may damage the mucus of the rumen	Heavy levels of infection: rapid weight loss, diarrhoea. Low levels of infection: loss of appetite, poor growth rates, soft faeces
Mode of Transmission	Faecal-oral	Faecal-oral
References	Díaz et al. (2006); Lotfy et al. (2010); Sanabria and Romero (2008)	Urquhart et al. (1996)

Continued on next page

214

Table C.1 – continued from previous page

	Coccidia spp.	*Haemonchus placei*
Type	Protozoa	Nematode
		Strongyle type helminth
Method of Identification	McMasters quantitative (counting) technique	Faecal culture
Site of Infection	Depends on the species, they may be found in the epithelial cells of the small intestine or the caecum and colon	Abomasum
Prepatent period	17–18 days depending on species	4 weeks
Clinical Symptoms	Depends on species but in heavy levels of infection: severe blood-stained dysentery accompanied by tenesmus, severe enteritis and diarrhoea	Acute cases: anaemia, oedema, lethargy, dark coloured faeces. Chronic cases: progressive weight loss and weakness, no severe anaemia or gross oedema present
Mode of Transmission	Faecal-oral	Faecal-oral
References	Urquhart et al. (1996)	Urquhart et al. (1996)

215

Table C.1 – continued from previous page

	Theileria spp.
Type	Protozoa
Method of Identification	Microscopy
Site of Infection	White and red blood cells
Prepatent period	NA
Clinical Symptoms	Clinical symptoms vary depending upon the species. *Theileria parva* causes East Coast fever, which is characterized by high fever, swelling of the lymph nodes, dyspnea, and high mortality. Other species such as *T. mutans* and *T. velifera* are less pathogenic and do not have as severe clinical symptoms
Mode of Transmission	Tick-borne
References	Aiello and Moses (2012); Coetzer and Tustin (2004)

216

C.2   Data structure

Table C.2 Example subset of data structure for a single calf from the current (T_0), 5 (T_{-5}) and 10 (T_{-10}) week visits for *Coccidia* spp. The status of the calf at the present visit is given by *Coccidia* spp. T_0, where 0 means that *Coccidia* spp. was not found at that particular visit, 1 means that *Coccidia* spp. was identified at that visit and NA that no measure was available.

Calf Age	*Coccidia* spp. T_0	*Coccidia* spp. T_{-5}	*Coccidia* spp. T_{-10}
1	0	NA	NA
6	1	0	NA
11	0	1	0
16	0	0	1

C.3 Non-infectious risk factors

Table C.3 A list of all the potential non-infectious risk factors examined to see if the parasite-parasite associations observed were generated by indicators of calf condition, seasonal or environmental variation.

Farm factors	Farmers age, gender, education level (Whether the farmer have at least primary school level education? (Yes/No)), main occupation (two levels a) no salaried income (farmer) b) salaried income (employed)), land size, number of cattle owned, distance to water, water provided at the homestead or not, whether or not the calf was housed with adults
'Farmer quality' index	As many of the environmental criteria are correlated, I decided to create a farmer quality index. The farmer quality index is a continuous scale from 0 to 6 based upon the number of requirements on the following criteria a farmer met. 0 means that the farmer met none of the criteria, whereas 6 means the farmer met all 6 requirements. The criteria used to form the farmer quality index where: a) the farmer has technical training; b) housing was provided in the form of a kraal or stall-shed or yard; c) the farmer used supplementary feeding mainly in the form of napier grass, as stated during the recruitment questionnaire; d) the farmer had access to veterinary services; e) the farmer has knowledge of diseases prevalent on the farm; and f) the farmer used veterinary interventions at some point during the study period in other members of the herd but NOT in the calf (due to the study requirements (Bronsvoort et al., 2013)). Veterinary interventions included the use of antibiotics, antiprotozoals, anthelminthics, insecticides/acaracides, traditional medicine, trypanocidials or vaccines.
Environmental factors	Farm altitude (elevation, ranges from 1127 to 1446 meters), northing from the geo-referenced location of the calf, agro-ecological zone (a five level factor, see (Bronsvoort et al., 2013)), normalized difference vegetation index (NDVI) at the current visit, NDVI 5 weeks prior to current visit and NDVI 10 weeks prior to current visit

Continued on next page

218

Table C.3 – continued from previous page

Seasonal factors	Sample month fitted as a factor, sample year fitted as a factor (2009 is the baseline category, before 2009 and after 2009)
Dam factors	Dam girth, body condition score (a method of assessing the body condition of the dam using a standard 10 point score).
Calf factors	Calf sex, heterozygosity and level of European Taurine introgression (the level of introgression consistent with crossing with European breeds less than 5 generations ago, see Mbole-Kariuki et al. (2014) for more details), weight at recruitment, weight at visit, haematological parameters: red blood cell count, white blood cell count, total serum protein, packed cell volume, absolute eosinophil count at each visit (see (Conradie van Wyk et al., 2012) for a detailed description of each one)
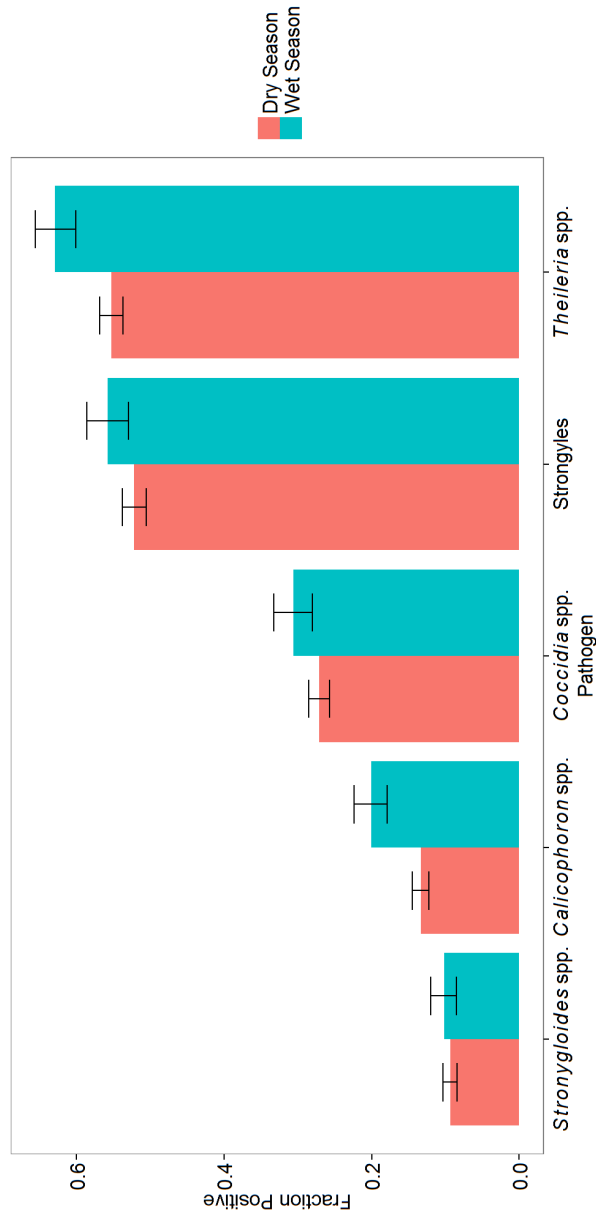
Figure C.1 Fraction of calves positive for the parasite of interest during visits which occurred in the dry or wet season. The wet season is traditionally thought to occur between March and May. The error bars represent the 95% confidence intervals.

C.4 Effect of including calf, environmental and seasonal variation in the parasite-parasite association model

Odds ratios for factors

Table C.4 Associations between multiple time lags and parasites identified from the multivariate analyses of the parasites present at the current visit (T_0) or at a 5 and 10 week lag (T_{-5} and T_{-10}, respectively) a) before (parasite only model) and b) after accounting for indicators of calf condition, environmental and seasonal variation. Parasites with a $P > 0.01$ are presented in the calf, seasonal and environmental variation model results for comparisons of odds ratios.

Response Parasite (T_0)	Explanatory Variable	(a) Parasite Only Model						(b) Calf, Seasonal and Environmental Variation Model					
		OR	95% CI	P value	AIC	N_O	N_C	OR	95% CI	P value	AIC	N_O	N_C
Strongyloides spp.	Strongyles T_0	2.56	1.85-3.56	<0.001	1687	3114	454	2.55	1.83-3.54	<0.001	1626	3105	453
	Calicophoron spp. T_0	0.47	0.29-0.74	0.001				0.53	0.33-0.85	0.008			
	Coccidia spp. T_0	1.44	1.11-1.87	0.006				1.53	1.17-2.00	0.002			
	Strongyloides spp. T_{-5}	3.31	2.4-4.56	<0.001				2.55	1.84-3.55	<0.001			
	Recruitment weight of the calf							0.95	0.91-0.98	0.003			
	Sampled before 2009							2.44	1.84-3.24	<0.001			
	Sampled after 2009							0.58	0.3-1.12	0.103			
	Random effects:	Calf variance = 0.10; Sublocation variance = 0.12						Calf variance = 0.00; Sublocation variance= 0.14					

C.4 Effect of including calf, environmental and seasonal variation in the parasite-parasite association model

Table C.4 – continued from previous page

Response Parasite (T_0)	Explanatory Variable	(a) Parasite Only Model						(b) Calf, Seasonal and Environmental Variation Model					
		OR	95% CI	P value	AIC	N_O	N_C	OR	95% CI	P value	AIC	N_O	N_C
Calicophoron spp.	*Strongyloides* spp. T_0	0.41	0.23-0.73	0.002	2233	2339	445	0.38	0.21-0.69	0.001	2100	2339	445
	Coccidia spp. T_0	0.71	0.57-0.89	0.003				0.79	0.62-1.00	0.052			
	Calicophoron spp. T_{-5}	1.88	1.47-2.41	<0.001				1.36	1.04-1.78	0.025			
	Calicophoron spp. T_{-10}	1.86	1.43-2.43	<0.001				1.55	1.16-2.08	0.003			
	NDVI 10 weeks prior to current visit							18.66	3.00-116.03	0.002			
	Sampled in February							1.96	1.14-3.37	0.015			
	Sampled in March							1.55	0.86-2.81	0.145			
	Sampled in April							1.28	0.71-2.31	0.414			
	Sampled in May							0.66	0.37-1.19	0.167			
	Sampled in June							0.28	0.16-0.5	<0.001			
	Sampled in July							0.30	0.17-0.53	<0.001			
	Sampled in August							0.23	0.13-0.41	<0.001			
	Sampled in September							0.39	0.22-0.70	0.002			
	Sampled in October							0.44	0.24-0.79	0.006			
	Sampled in November							0.28	0.15-0.50	<0.001			
	Sampled in December							0.24	0.13-0.44	<0.001			
	Sampled before 2009							0.78	0.56-1.08	0.139			
	Sampled after 2009							0.14	0.09-0.21	<0.001			
	Random effects:	Calf variance = 0.14; Sublocation variance = 0.32						Calf variance = 0.20; Sublocation variance= 0.10					

Table C.4 – continued from previous page

C.4 Effect of including calf, environmental and seasonal variation in the parasite-parasite association model

Response Parasite (T_0)	Explanatory Variable	(a) Parasite Only Model						(b) Calf, Seasonal and Environmental Variation Model					
		OR	95% CI	P value	AIC	N_O	N_C	OR	95% CI	P value	AIC	N_O	N_C
Coccidia spp.	*Strongyloides* spp. T_0	1.52	1.2-1.92	0.001	4730	3623	455	1.61	1.26-2.05	<0.001	4668	3623	455
	Calicophoron spp. T_0	0.78	0.65-0.93	0.006				0.85	0.70-1.02	0.088			
	Theileria spp. T_{-5}	1.25	1.06-1.47	0.008				1.13	0.96-1.34	0.142			
	Northing							0.15	0.05-0.47	0.001			
	AEZ: LM2 middle							0.56	0.40-0.78	0.001			
	AEZ: LM1							0.61	0.40-0.91	0.017			
	AEZ: LM2 south							0.37	0.20-0.67	0.001			
	AEZ: LM3							0.35	0.20-0.61	<0.001			
	Sampled in February							1.18	0.8-1.73	0.408			
	Sampled in March							1.63	1.14-2.33	0.007			
	Sampled in April							1.72	1.21-2.44	0.002			
	Sampled in May							1.79	1.26-2.55	0.001			
	Sampled in June							1.91	1.34-2.71	<0.001			
	Sampled in July							2.29	1.61-3.25	<0.001			
	Sampled in August							1.78	1.24-2.54	0.002			
	Sampled in September							2.66	1.86-3.81	0.000			
	Sampled in October							0.99	0.67-1.46	0.957			
	Sampled in November							1.60	1.10-2.33	0.014			
	Sampled in December							1.67	1.14-2.44	0.009			

Table C.4 – continued from previous page

C.4 Effect of including calf, environmental and seasonal variation in the parasite-parasite association model

Response Parasite (T_0)	Explanatory Variable	(a) Parasite Only Model						(b) Calf, Seasonal and Environmental Variation Model					
		OR	95% CI	P value	AIC	N_O	N_C	OR	95% CI	P value	AIC	N_O	N_C
	Sampled before 2009							0.80	0.67-0.97	0.022			
	Sampled after 2009							1.58	1.26-1.98	<0.001			
	Random effects:	Calf variance = 0.10; Sublocation variance = 0.07						Calf variance = 0.09; Sublocation variance= 0.01					
Strongyles	*Strongyloides* spp. T_0	2.59	1.89-3.57	<0.001	3420	3114	454	2.67	1.94-3.68	<0.001	3414	3114	454
	Strongyles T_{-5}	1.37	1.13-1.66	0.002				1.36	1.12-1.65	0.002			
	Northing							2.59	1.37-4.91	0.003			
	Random effects:	Calf variance = 0.06; Sublocation variance = 0.05						Calf variance = 0.06; Sublocation variance = 0.02					
Theileria spp.	*Theileria* spp. T_{-5}	1.84	1.56-2.16	<0.001	4054	3608	455	1.63	1.37-1.93	<0.001	3838	3576	451
	Theileria spp. T_{-10}	1.49	1.26-1.78	<0.001				1.29	1.07-1.54	0.007			
	Red blood cell count at current visit							0.91	0.87-0.95	<0.001			
	Total serum protein at current visit							1.23	1.07-1.42	0.005			
	Total number of cattle on farm is greater than 5							1.43	1.19-1.72	<0.001			
	Sampled in February							1.75	1.13-2.71	0.013			
	Sampled in March							2.26	1.42-3.62	0.001			
	Sampled in April							4.13	2.57-6.63	<0.001			
	Sampled in May							1.42	0.95-2.1	0.085			
	Sampled in June							2.07	1.40-3.06	<0.001			
	Sampled in July							1.16	0.8-1.69	0.440			
	Sampled in August							1.35	0.92-1.99	0.123			

Table C.4 – continued from previous page

Response Parasite (T_0)	Explanatory Variable	(a) Parasite Only Model						(b) Calf, Seasonal and Environmental Variation Model					
		OR	95% CI	P value	AIC	N_O	N_C	OR	95% CI	P value	AIC	N_O	N_C
	Sampled in September							0.98	0.67-1.43	0.898			
	Sampled in October							1.96	1.32-2.9	0.001			
	Sampled in November							2.02	1.36-3.02	0.001			
	Sampled in December							1.33	0.89-1.98	0.158			
	Sampled before 2009							0.82	0.68–1.00	0.048			
	Sampled after 2009							2.82	2.12-3.75	<0.001			
	Random effects:	Calf variance = 0.05; Sublocation variance = 0.08						Calf variance = 0.07; Sublocation variance= 0.07					

224

Strongyloides spp.

Table C.5 Effect of including calf, environmental and seasonal variation in the *Strongyloides* spp.-parasite association model. Three different models are shown to illustrate the individual effect of calf, environment or seasonal variation in turn upon the *Strongyloides* spp.-parasite associations. Variation in the dam's condition had no effect upon infection status. *Summary P value from the ANOVA F ratio of the GLMM. N_O = Number of observations; N_C = number of calves.

Explanatory Variable	Indicators of Calf Condition			Environmental Variation			Seasonal Variation		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Strongyles T_0	2.57	1.85-3.57	<0.001	2.56	1.85-3.56	<0.001	2.55	1.83-3.54	<0.001
Calicophoron spp. T_0	0.47	0.3-0.75	0.002	0.47	0.29-0.74	0.001	0.53	0.33-0.85	0.008
Coccidia spp. T_0	1.46	1.12-1.9	0.005	1.44	1.11-1.87	0.006	1.52	1.16-1.98	0.002
Strongyloides spp. T-5	3.11	2.25-4.3	<0.001	3.31	2.4-4.56	<0.001	2.71	1.95-3.75	<0.001
Recruitment weight	0.93	0.89-0.96	<0.001						
Sample Year							NA	NA	<0.001
Sampled before 2009							2.60	1.96-3.43	<0.001
Sampled after 2009							0.56	0.29-1.08	0.083
	Model Fit			**Model Fit**			**Model Fit**		
AIC	1671			1687			1634		
N_O	3105			3114			3114		
N_C	453			454			454		
	Random Effect Variance			**Random Effect Variance**			**Random Effect Variance**		
Calf	0.07			0.10			0.00		
Sublocation	0.12			0.12			0.13		

225

Calicophoron spp.

Table C.6 Effect of including calf, environmental and seasonal variation in the *Calicophoron* spp.-parasite association model. Three different models are shown to illustrate the individual effect of calf, environment or seasonal variation in turn upon the *Calicophoron* spp.-parasite associations. Variation in the damâ€™s condition had no effect upon infection status. *Summary P value from the ANOVA F ratio of the GLMM. N_O = Number of observations; N_C = number of calves.

Explanatory Variable	Indicators of Calf Condition			Environmental Variation			Seasonal Variation		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Strongyloides spp. T_0	0.41	0.23-0.73	0.002	0.42	0.24-0.76	0.004	0.37	0.2-0.67	0.001
Coccidia spp. T_0	0.71	0.57-0.89	0.003	0.71	0.57-0.9	0.004	0.79	0.62-1.00	0.054
Calicophoron spp. T_{-5}	1.88	1.47-2.41	<0.001	1.79	1.4-2.3	<0.001	1.32	1.01-1.73	0.041
Calicophoron spp. T_{-10}	1.86	1.43-2.43	<0.001	1.83	1.4-2.39	<0.001	1.53	1.14-2.04	0.004
NDVI 5 weeks prior to current visit				0.05	0.01-0.15	<0.001			
Sample Month*							NA	NA	<0.001
February							1.60	0.95-2.69	0.079
March							1.12	0.65-1.94	0.686
April							0.83	0.49-1.39	0.473
May							0.47	0.27-0.80	0.005
June							0.27	0.16-0.48	<0.001
July							0.35	0.2-0.61	<0.001
August							0.25	0.14-0.44	<0.001
September							0.33	0.19-0.58	<0.001

Continued on next page

226

Table C.6 – continued from previous page

Explanatory Variable	Indicators of Calf Condition			Environmental Variation			Seasonal Variation		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
October							0.34	0.19-0.61	<0.001
November							0.24	0.13-0.44	<0.001
December							0.25	0.13-0.46	<0.001
Sample Year*									
Sampled before 2009							NA	NA	<0.001
Sampled after 2009							0.86	0.63-1.20	0.378
							0.17	0.12-0.25	<0.001
	Model Fit			**Model Fit**			**Model Fit**		
AIC	2233			2208			2109		
NO	2339			2339			2339		
NC	445			445			455		
	Random Effect Variance			**Random Effect Variance**			**Random Effect Variance**		
Calf	0.14			0.16			0.24		
Sublocation	0.10			0.15			0.14		

Coccidia **spp.**

Table C.7 Effect of including calf, environmental and seasonal variation in the *Coccidia* spp.-parasite association model. Three different models are shown to illustrate the individual effect of calf, environment or seasonal variation in turn upon the *Coccidia* spp.-parasite associations. Variation in the dam's condition had no effect upon infection status. *Summary P value from the ANOVA F ratio of the GLMM. N_O = Number of observations; N_C = number of calves.

Explanatory Variable	Indicators of Calf Condition			Environmental Variation			Seasonal Variation		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Strongyloides spp. T_0	1.54	1.22-1.95	<0.001	1.53	1.21-1.93	<0.001	1.59	1.25-2.03	<0.001
Calicophoron spp. T_0	0.78	0.65-0.94	0.008	0.77	0.64-0.92	0.005	0.85	0.7-1.03	0.102
Theileria spp. T_{-5}	1.25	1.06-1.47	0.008	1.24	1.06-1.46	0.009	1.14	0.96-1.35	0.128
Packed cell volume count at current visit	1.03	1.01-1.04	<0.001						
Northing				0.14	0.04-0.45	0.001			
Agro-ecological zone				NA	NA	0.001			
LM2 middle				0.56	0.4-0.78	0.001			
LM1				0.60	0.40-0.91	0.015			
LM2 south				0.36	0.20-0.66	0.001			
LM3				0.36	0.21-0.62	<0.001			
Sample Month*							NA	NA	<0.001
February							1.16	0.79-1.7	0.447
March							1.63	1.14-2.32	0.008
April							1.72	1.21-2.44	0.002

Continued on next page

228

Table C.7 – continued from previous page

Explanatory Variable	Indicators of Calf Condition			Environmental Variation			Seasonal Variation		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
May							1.80	1.27-2.56	0.001
June							1.91	1.34-2.71	<0.001
July							2.26	1.59-3.21	<0.001
August							1.79	1.25-2.56	0.002
September							2.68	1.87-3.84	<0.001
October							0.99	0.67-1.46	0.955
November							1.59	1.09-2.31	0.016
December							1.68	1.15-2.46	0.008
Sample Year*							NA	NA	<0.001
Sampled before 2009							0.8	0.67-0.97	0.02
Sampled after 2009							1.56	1.25-1.95	<0.001
Model Fit	**Model Fit**			**Model Fit**			**Model Fit**		
AIC	4712			4722			4722		
NO	3619			3623			3623		
NC	455			455			455		
Random Effect Variance	**Random Effect Variance**			**Random Effect Variance**			**Random Effect Variance**		
Calf	0.09			0.10			0.09		
Sublocation	0.07			0.01			0.07		

229

Strongyles

Table C.8 Effect of including calf, environmental and seasonal variation in the Strongyle-parasite association model. Three different models are shown to illustrate the individual effect of calf, environment or seasonal variation in turn upon the Strongyle-parasite associations. Variation in the dam's condition had no effect upon infection status N_O = Number of observations; N_C = number of calves.

Explanatory Variable	Indicators of Calf Condition			Environmental Variation			Seasonal Variation		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Strongyloides spp. T_0	2.59	1.89-3.57	<0.001	2.67	1.94-3.68	<0.001	2.59	1.89-3.57	<0.001
Strongyles T_{-5}	1.37	1.13-1.66	0.002	1.36	1.12-1.65	0.002	1.37	1.13-1.66	0.002
Northing				2.59	1.37-4.91	0.003			
Model Fit									
AIC	3420			3414			3420		
N_O	3114			3114			3114		
N_C	454			454			454		
Random Effect Variance									
Calf	0.06			0.06			0.06		
Sublocation	0.05			0.02			0.05		

Theileria spp.

Table C.9 Effect of including calf, environmental and seasonal variation in the *Theileria* spp.-parasite association model. Three different models are shown to illustrate the individual effect of calf, environment or seasonal variation in turn upon the *Theileria* spp.-parasite associations. Variation in the dam's condition had no effect upon infection status. *Summary P value from the ANOVA F ratio of the GLMM. N_O = Number of observations; N_C = number of calves.

Explanatory Variable	Indicators of Calf Condition			Environmental Variation			Seasonal Variation		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Theileria spp. T_{-5}	1.8	1.53-2.12	<0.001	1.81	1.53-2.13	<0.001	1.69	1.43-2.00	<0.001
Theileria spp. T_{-10}	1.51	1.26-1.79	<0.001	1.51	1.26-1.79	<0.001	1.30	1.09-1.56	0.004
Red blood cell count at current visit	0.92	0.88-0.97	0.001						
Total serum protein at current visit	1.29	1.12-1.49	<0.001						
NDVI at current visit				4.73	1.96-11.38	0.001			
NDVI 10 weeks prior to current visit				0.26	0.11-0.62	0.002			
Farmer is greater than 60 years old				0.74	0.61-0.90	0.002			
Total number of cattle on farm is greater than 5				1.43	1.2-1.71	<0.001			
Sample Month*							NA	NA	<0.001
February							1.85	1.20-2.87	0.006
March							2.40	1.50-3.82	<0.001
April							4.33	2.7-6.92	<0.001
May							1.47	0.99-2.17	0.055
June							2.07	1.41-3.06	<0.001

Continued on next page

Table C.9 – continued from previous page

Explanatory Variable	Indicators of Calf Condition			Environmental Variation			Seasonal Variation		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
July							1.15	0.8-1.68	0.45
August							1.38	0.94-2.01	0.098
September							1.01	0.70-1.48	0.94
October							2.11	1.43-3.11	<0.001
November							2.21	1.49-3.29	<0.001
December							1.40	0.95-2.07	0.092
Sample Year*							NA	NA	<0.001
Sampled before 2009							0.86	0.71-1.04	0.11
Sampled after 2009							2.75	2.07-3.63	<0.001
	Model Fit			**Model Fit**			**Model Fit**		
AIC	4030			3956			3905		
N_O	3601			3563			3608		
N_C	455			447			455		
	Random Effect Variance			**Random Effect Variance**			**Random Effect Variance**		
Calf	0.06			0.03			0.06		
Sublocation	0.07			0.07			0.09		

232

Appendix D

Seroprevalence of respiratory viral pathogens of indigenous calves in Western Kenya

D.1 Description of viruses

Table D.1 Description of the aetiology, epidemiology and pathogenesis of IBR, PIV3 and BVDV

	Infectious Bovine Rhinotracheitis (IBR)	Bovine Parainfluenza Virus Type 3 (PIV3)
Aetiology	Genus: Varicellavirus Family: Herpeviridae	Genus: Paramyxovirus Family: Paramyxoviridae
Epidemiology	Spread by aerosols and secretions. Latent infection can occur.	Spread by aerosols. No persistently infected individuals reported.
Pathogenesis	Local lesions in mucus membranes. Virus multiples in the epithelial cells of the upper respiratory tract and then spreads to the lower respiratory tract in uncomplicated infections. The replicating virus causes destruction of the epithelial cells and alteration of the mucociliary clearance mechanisms.	Primarily infected cells are the epithelial cells of the trachea, bronchi and alveoli. Necrosis of ciliated epithelium interferes with the mucociliary clearance of the airways. No evidence being immunosuppressive. Predisposes lung tissue to bacterial invasion. Replicates in alveolar macrophages.
Clinical signs	Fever, increased respiratory rate, inappetence, coughing, nasal discharge, drop in milk production, hyperaemia and small foci of necrosis of the nasal passage, excessive salivation. Course of disease is variable among individuals and secondary bacterial infection will increase the severity and duration of clinical signs.	Mild or subclinical. Development of disease depends upon interaction with other infectious and environmental factors.
References	Coetzer and Tustin (2004)	Coetzer and Tustin (2004)

234

Continued on next page

Table D.1 – continued from previous page

	Bovine Viral Diarrhoea Virus (BVDV)
Aetiology	Genus: Pestivirus Family: Flavividae
Epidemiology	Spread by secretions, horizontal and vertical transmission can occur. Immuno-tolerant persistently infected individuals reported
Pathogenesis	Replicates in the nasal mucosa and the tonsils. Spreads to regional lymph nodes and the rest of the body. Induces immunosuppression.
Clinical signs	Depends on the host, environment, epidemiological and viral factors. May result in one of three disease syndromes: Bovine viral diarrhoea, mucosal disease or embryonal/foetal disease. Some of the clinical signs include abortions, fever, inappentance, diarrhoea, and ulceration of the gastrointestinal tract.
References	Coetzer and Tustin (2004)

235

D.2 Cut-offs and interpretation of the ELISA percent positivity values

Table D.2 Manufacturers cut-offs for the interpretation of ELISA percent positivity (PP) values for antibodies to IBR, PIV3 and BVDV and manufacturer cut-off for the interpration of sample to positive control percentage (S/P%) values for BVDV antigen

Test	Manufacturer	Manufacturers Cut-offs	
		PP	Interpretation
Infectious bovine rhinotracheitis antibody	SVANOVIR	< 8	Negative
		8-11	Doubtful
		> 12	Positive
Bovine parainfluenza virus type 3	SVANOVIR	< 10	Negative
		> 10	Positive
Bovine virus diarrhoea virus antibody	SVANOVIR	< 14	Negative
		> 14	Positive
Bovine virus diarrhoea virus antigen	IDEXX	< 0.300*	Negative
		> 0.300*	Positive

*Sample to positive control percentage (S/P%) given rather than PP

D.3 Case-case analysis of IBR doubtful and seropositive calves

Since the manufacturer's cut-off for IBR classifies individuals into 'seronegative', 'seropositive' or 'doubtful' (Table D.2), it is important to check that there are no differences in outcomes of interest between the seropositive and doubtful individuals if I am to combine them into a binary seropositive/seronegative serostatus variable. A case-case analysis was used to achieve this. The case-case analysis used chi-squared tests to investigate differences between IBR doubtful and seropositive calves in the number which experienced clinical episodes, or were seropositive for BVDV or PIV3. t-tests were used to test for a difference in the mean live body weight at recruitment and at one year old, mean daily weight gain of IBR seropositive and doubtful individuals. This whole process was then repeated to compare IBR doubtful and seronegative calves.

Comparison of IBR seropositive and doubtful individuals revealed that there was no statistically significant difference between the two categories. Both have a similar number of clinical episodes, BVDV and PIV3 seropositive, as well as having a similar mean daily weight gain (Table D.3 and Table D.4). In contrast, comparison of the IBR seronegative and doubtful individuals revealed that there was a significant difference between the two categories. IBR doubtful individuals were more likely to be BVDV and PIV3 seropositive than IBR seronegative individuals, in addition IBR seronegative calves were lighter than IBR doubtful individuals (Table D.5 and Table D.6). These findings imply that the IBR doubtful calves are more similar to the IBR seropositive calves than to the IBR seronegative ones; therefore it is reasonable to combine the IBR seropositive and doubtful categories into one group.

Table D.3 Comparison of categorical traits for IBR *doubtful* (N=66) and IBR *seropositive* (N=25) calves using chi-squared tests.

Categorical Trait of Interest	IBR doubtful		IBR seropositive		IBR doubtful vs. IBR seropositive		
	N	%	N	%	χ^2	d.f	P value
BVDV seronegative	42	46.15	16	17.58	0.001	1	0.974
BVDV seropositive	24	26.37	9	9.89			
PIV3 seronegative	42	46.15	16	17.58	0.001	1	0.974
PIV3 seropositive	24	26.37	9	9.89			
Did not experience a clinical episode	48	52.75	15	16.48	1.38	1	0.240
Experienced a clinical episode	18	19.78	10	10.99			
Did not experience a respiratory clinical episode	63	69.23	22	24.18	1.64	1	0.200
Experienced a respiratory clinical episode	3	3.30	3	3.30			

Table D.4 Comparison of continuous traits for IBR *doubtful* (N=66) and IBR *seropositive* calves (N=25) using t-tests.

Continuous Trait of Interest	IBR doubtful		IBR seropositive		T	d.f.	P value
	Mean	SE	Mean	SE			
Live body weight at recruitment	20.78	0.45	18.8	0.92	1.94	36	0.060
Live body weight at one year	69.89	2.44	63.1	3.79	1.51	45	0.138
Average daily weight gain	0.14	0.01	0.13	0.01	0.96	46	0.342

Table D.5 Comparison of categorical traits for IBR *doubtful* (N=66) and IBR *seronegative* (N=364) calves using chi-squared tests.

Categorical Trait of Interest	IBR doubtful		IBR seronegative		IBR doubtful vs. IBR seronegative		
	N	%	N	%	χ^2	d.f.	P value
BVDV seronegative	42	9.77	318	73.95	23.08	1	<0.001
BVDV seropositive	24	5.58	46	10.7			
PIV3 seronegative	42	9.79	316	73.66	22.17	1	<0.001
PIV3 seropositive	24	5.59	47	10.96			
Did not experience a clinical episode	48	11.16	186	43.26	10.54	1	0.001
Experienced a clinical episode	18	4.19	178	41.4			
Did not experience a respiratory clinical episode	63	14.65	328	76.28	1.94	1	0.164
Experienced a respiratory clinical episode	3	0.7	36	8.37			

Table D.6 Comparison of continuous traits for IBR *doubtful* (N=66) and IBR *seronegative* (N=364) calves using t-tests.

Continuous Trait of Interest	IBR doubtful		IBR seronegative		T	d.f.	P value
	Mean	SE	Mean	SE			
Live body weight at recruitment	20.78	0.45	18.95	0.19	3.77	89	<0.001
Live body weight at one year	69.89	2.44	64.45	0.90	2.09	83	0.039
Average daily weight gain	0.14	0.01	0.130	0.002	1.40	83	0.165

239

D.4 Associations between viruses analyzed as a continuous variable

Initially the association between viruses was assessed using the percentage positivity (PP) values as a continuous measure in a generalised linear mixed model (GLMM) fitted with a Laplace approximation to the maximum likelihood estimation in R v.2.15.2 using the *lme4* package (Bates et al., 2014). PP values were $\log_{10}$ transformed and outlying PPs were removed to normalize the data (Figure D.1). I investigate the association between PP values of multiple viruses at 51 weeks old with the PP value of virus A at 51 weeks using the following equation:

$$PP_{Virus\ A_i} = \alpha + \beta_1 PP_{Virus\ B_i{}_i} + \beta_2 PP_{Virus\ C_i}$$
$$+ b_{1_i} Sublocation_i + \varepsilon_i \tag{D.1}$$

where α is the intercept and ε is the residual variation. Fixed effects are symbolized by β. $PP_{Virus\ X}$ is the PP value of the calf (i) at 51 weeks of age. Sublocation (*Sublocation*, 20 levels) is included in the model as a random effect (b) to account for the study design and environmental similarity between calves clustered into each sublocation. A separate multivariable model was constructed for each virus and backwards-stepwise selection was used to remove viruses. The most parsimonious model was chosen as the final model.

The statistical analysis of the continuous PP values indicated that IBR, PIV3 and BVDV are co-distributed (Table D.7).

Table D.7 Parameter estimates for the association between the PP values of IBR, PIV3 and BVDV.

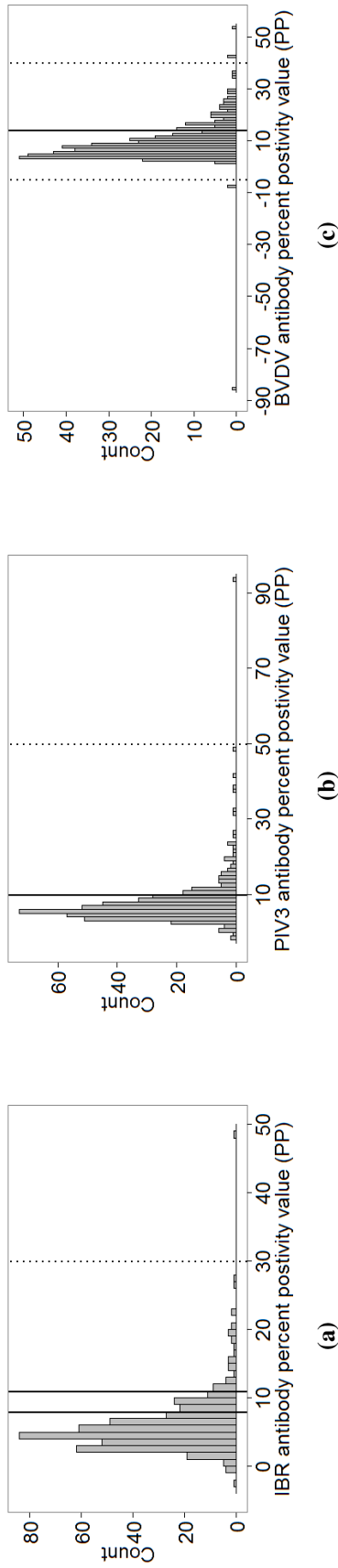| Explanatory Virus	Response Virus	Estimate	SE	T value	OR	95% CI	P value
IBR	PIV3	0.15	0.04	3.65	1.16	1.07-1.26	<0.001
	BVDV	0.21	0.04	4.81	1.23	1.13-1.34	<0.001
PIV3	IBR	0.2	0.05	3.66	1.22	1.10-1.36	<0.001
	BVDV	0.45	0.05	9.49	1.55	1.42-1.70	<0.001
BVDV	IBR	0.24	0.05	4.81	1.28	1.15-1.41	<0.001
	PIV3	0.39	0.04	9.46	1.48	1.36-1.60	<0.001

Figure D.1 Histogram showing the number of individuals against their percentage positivity values (PP) for the IBR, PIV3 and BVDV antibody ELISA tests. The solid vertical line represents the manufacturer's cut-off, below the dashed line indicates seronegative individuals and above the dashed line represents seropositive individuals. Individuals who fall between the two solid lines in the IBR results are inconclusive. The dotted line represents outliers which were removed in the continuous analysis.

241

D.5 Environmental variables

Table D.8 A list of all the non-infectious variables which were examined to investigate if the virus associations observed were robust to calf, dam and environmental variation

Farm factors	Farmer's age, gender, education level, main occupation, land size, number of cattle owned, distance to water, water provision, whether or not the calf was housed with adult cattle
'Farmer quality' index	As many of the environmental criteria are correlated, I decided to create a farmer quality index. The farmer quality index is a continuous scale from 0 to 6 based upon the number of requirements on the following criteria a farmer met. 0 means that the farmer met none of the criteria, whereas 6 means the farmer met all 6 requirements. The criteria used to form the farmer quality index where: a) The farmer has technical training; b) housing is provided in the form of a Kraal or Stall-Shed or Yard; c) the farmer used supplementary feeding mainly in the form of napier grass, as stated during the recruitment questionnaire; d) the farmer has access to veterinary services; e) the farmer has knowledge of diseases prevalent on the farm; and f) the farmer uses veterinary interventions at some point during the study period in other members of the herd but NOT in the calf (due to the study requirements (Bronsvoort et al. 2013)). Veterinary interventions included the use of antibiotics, antiprotozoals, anthelminthics, insecticides/acaracides, traditional medicine, trypanocidials or vaccines.
Environmental factors	Farm altitude (elevation, as categorical variable: <1198m, 1199-1238m, 1239-1269m, >1269m), agro-ecological zone
Dam factors	Dam girth (cm), body condition score (a method of assessing the body condition of the dam using a standard 10 point score).
Calf factors	Calf sex, heterozygosity and level of European Taurine introgression (the level of introgression consistent with crossing with European breeds less than 5 generations ago, see Mbole-Kariuki et al. (2014) for more details)

242

Table D.9 Association between viruses after accounting for environmental, dam and calf variation. Each line shows a separate model, account for one environmental factor at a time. Inclusion of environmental confounders in the virus-only models did not affect the relationship observed between the serostatus of the three viruses.

Explanatory Variable	Response Virus: IBR			Response Virus: PIV3			Response Virus: BVDV		
	OR	95% CI	P value	OR	95% CI	P Value	OR	95% CI	P Value
Farm Factors									
IBR seropositive	-	-	-	2.72	1.50-4.91	0.001	2.76	1.55-4.91	0.001
PIV3 seropositive	2.74	1.51-4.98	0.001	-	-	-	5.94	3.38-10.45	<0.001
BVDV seropositive	2.78	1.54-5.03	0.001	5.83	3.27-10.39	<0.001	-	-	-
Farmer age (effect of being greater than 60 years old)	1.50	0.84-2.66	0.170	1.10	0.58-2.08	0.774	0.89	0.47-1.69	0.731
IBR seropositive	-	-	-	2.77	1.53-5.01	0.001	2.77	1.56-4.93	0.001
PIV3 seropositive	2.77	1.53-5.02	0.001	-	-	-	5.98	3.40-10.51	<0.001
BVDV seropositive	2.79	1.55-5.02	0.001	5.87	3.30-10.46	<0.001	-	-	-
Farmer sex (effect of being male)	0.65	0.39-1.09	0.101	1.04	0.58-1.86	0.896	1.02	0.58-1.79	0.955
IBR seropositive	-	-	-	2.74	1.52-4.95	0.001	2.75	1.55-4.88	0.001
PIV3 seropositive	2.77	1.52-5.02	0.001	-	-	-	5.93	3.37-10.43	<0.001
BVDV seropositive	2.76	1.53-4.98	0.001	5.82	3.26-10.36	<0.001	-	-	-
Farmer education level (effect of not having at least primary school level education)	1.10	0.54-2.21	0.794	0.84	0.38-1.86	0.663	0.91	0.42-1.97	0.816

243

Table D.9 – continued from previous page

Explanatory Variable	Response Virus: IBR			Response Virus: PIV3			Response Virus: BVDV		
	OR	95% CI	P value	OR	95% CI	P Value	OR	95% CI	P Value
IBR seropositive	-	-	-	2.81	1.55-5.10	0.001	2.70	1.51-4.81	0.001
PIV3 seropositive	2.81	1.55-5.12	0.001	-	-	-	6.00	3.40-10.58	<0.001
BVDV seropositive	2.70	1.50-4.88	0.001	5.90	3.31-10.53	<0.001	-	-	-
Main occupation (effect of not being a farmer)	1.66	0.85-3.22	0.137	0.69	0.30-1.56	0.372	1.28	0.61-2.71	0.513
IBR seropositive	-	-	-	2.68	1.47-4.88	0.001	2.98	1.65-5.38	0.000
PIV3 seropositive	2.71	1.48-4.97	0.001	-	-	-	6.24	3.47-11.23	<0.001
BVDV seropositive	2.98	1.62-5.48	<0.001	6.18	3.39-11.27	<0.001	-	-	-
Total area of land used (acres)	0.74	0.28-1.95	0.545	0.54	0.18-1.61	0.271	1.74	0.62-4.92	0.293
IBR seropositive	-	-	-	2.79	1.55-5.02	0.001	2.65	1.48-4.75	0.001
PIV3 seropositive	2.80	1.54-5.07	0.001	-	-	-	6.15	3.48-10.87	<0.001
BVDV seropositive	2.67	1.47-4.86	0.001	6.07	3.39-10.84	<0.001	-	-	-
Total number of cattle on farm (greater than 5)	0.81	0.48-1.36	0.416	0.89	0.51-1.57	0.695	1.47	0.85-2.53	0.164
IBR seropositive	-	-	-	2.78	1.54-5.01	0.001	2.72	1.52-4.86	0.001
PIV3 seropositive	2.79	1.53-5.07	0.001	-	-	-	6.51	3.64-11.63	<0.001
BVDV seropositive	2.75	1.51-4.98	0.001	6.47	3.59-11.67	<0.001	-	-	-
Distance to water <1km	0.93	0.47-1.85	0.843	1.72	0.74-3.98	0.205	0.65	0.32-1.28	0.211
Distance to water 1-10Km	0.93	0.42-2.04	0.847	2.49	0.99-6.26	0.053	0.36	0.15-0.84	0.019

244

Table D.9 – continued from previous page

| Explanatory Variable	Response Virus: IBR			Response Virus: PIV3			Response Virus: BVDV		
	OR	95% CI	P value	OR	95% CI	P Value	OR	95% CI	P Value
IBR seropositive	-	-	-	2.81	1.56-5.07	0.001	2.67	1.49-4.78	0.001
PIV3 seropositive	2.82	1.56-5.09	0.001	-	-	-	6.11	3.46-10.8	<0.001
BVDV seropositive	2.67	1.47-4.83	0.001	6.02	3.37-10.75	<0.001	-	-	-
Water provisioning (animals go to water)	1.12	0.68-1.85	0.656	1.18	0.69-2.04	0.548	0.69	0.40-1.19	0.184
IBR seropositive	-	-	-	2.26	1.05-4.85	0.036	2.69	1.29-5.63	0.008
PIV3 seropositive	2.33	1.05-5.19	0.039	-	-	-	6.49	3.18-13.26	<0.001
BVDV seropositive	2.69	1.24-5.84	0.012	6.50	3.18-13.28	<0.001	-	-	-
Calves housed with adults	0.72	0.29-1.78	0.473	1.51	0.65-3.53	0.337	1.29	0.56-2.97	0.548
Good or bad farmer indicator scale									
IBR seropositive	-	-	-	2.58	1.4-4.74	0.002	2.85	1.58-5.12	<0.001
PIV3 seropositive	2.58	1.4-4.76	0.002	-	-	-	5.72	3.20-10.23	<0.001
BVDV seropositive	2.82	1.55-5.13	0.001	5.56	3.06-10.09	<0.001	-	-	-
Good/bad farmer indicator scale	1.10	0.88-1.38	0.403	0.93	0.73-1.18	0.566	1.20	0.94-1.54	0.151
Environmental factors									
IBR seropositive	-	-	-	2.77	1.53-5.02	0.001	2.89	1.61-5.18	<0.001
PIV3 seropositive	2.84	1.56-5.18	0.001	-	-	-	5.97	3.34-10.67	<0.001
BVDV seropositive	2.89	1.59-5.24	0.001	5.91	3.29-10.62	<0.001	-	-	-
Elevation - 1199-1238m	1.14	0.53-2.44	0.741	0.80	0.38-1.69	0.562	0.50	0.24-1.05	0.066
Elevation - 1239-1269m	1.47	0.68-3.18	0.326	0.41	0.18-0.93	0.033	0.77	0.38-1.57	0.480
Elevation - >1269m	1.66	0.76-3.64	0.203	0.83	0.37-1.85	0.645	0.42	0.19-0.91	0.028

Table D.9 – continued from previous page

Explanatory Variable	Response Virus: IBR			Response Virus: PIV3			Response Virus: BVDV		
	OR	95% CI	P value	OR	95% CI	P Value	OR	95% CI	P Value
IBR seropositive	-	-	-	2.54	1.40-4.63	0.002	2.92	1.62-5.27	<0.001
PIV3 seropositive	2.52	1.39-4.55	0.002	-	-	-	5.61	3.15-10.00	<0.001
BVDV seropositive	2.89	1.59-5.23	0.000	5.62	3.14-10.04	<0.001	-	-	-
AEZ - LM2 middle	0.75	0.25-2.25	0.608	0.68	0.21-2.16	0.514	1.75	0.6-5.10	0.304
AEZ - LM1	2.40	1.04-5.51	0.039	1.32	0.54-3.25	0.543	1.15	0.45-2.94	0.777
AEZ - LM2 south	1.69	0.64-4.48	0.291	1.08	0.37-3.14	0.890	1.79	0.62-5.14	0.279
AEZ - LM3	1.44	0.53-3.89	0.470	2.46	0.90-6.74	0.080	2.65	0.95-7.36	0.061
Dam factors									
IBR seropositive	-	-	-	2.63	1.44-4.82	0.002	2.83	1.58-5.05	<0.001
PIV3 seropositive	2.64	1.44-4.84	0.002	-	-	-	5.88	3.32-10.40	<0.001
BVDV seropositive	2.84	1.57-5.15	0.001	5.81	3.25-10.38	<0.001	-	-	-
Dam girth at recruitment	420.03	0.01 -1.34×10^7	0.253	0.001	0.00 -57.95	0.210	0.01	0-295.40	0.349
IBR seropositive	-	-	-	2.69	1.49-4.87	0.001	2.90	1.62-5.18	<0.001
PIV3 seropositive	2.69	1.48-4.89	0.001	-	-	-	6.04	3.42-10.67	<0.001
BVDV seropositive	2.88	1.59-5.22	<0.001	5.97	3.33-10.68	<0.001	-	-	-
Dam condition score at recruitment	1.26	0.98-1.63	0.074	1.13	0.85-1.50	0.398	0.81	0.62-1.07	0.136

Table D.9 – continued from previous page

Explanatory Variable	Response Virus: IBR			Response Virus: PIV3			Response Virus: BVDV		
	OR	95% CI	P value	OR	95% CI	P Value	OR	95% CI	P Value
Calf factors									
IBR seropositive	-	-	-	2.75	1.53-4.97	0.001	2.75	1.55-4.89	0.001
PIV3 seropositive	2.78	1.53-5.05	0.001	-	-	-	6.01	3.42-10.59	<0.001
BVDV seropositive	2.76	1.53-4.98	0.001	5.90	3.31-10.52	<0.001	-	-	-
Calf sex (effect of being male)	1.09	0.67-1.78	0.731	0.90	0.52-1.53	0.688	1.18	0.69-2.01	0.549
IBR seropositive	-	-	-	2.80	1.55-5.07	0.001	2.90	1.63-5.17	<0.001
PIV3 seropositive	2.80	1.54-5.09	0.001	-	-	-	5.92	3.34-10.48	<0.001
BVDV seropositive	2.88	1.59-5.22	<0.001	5.79	3.22-10.41	<0.001	-	-	-
European taurine introgression (moderate)	0.64	0.31-1.34	0.241	1.00	0.46-2.16	0.995	1.12	0.54-2.32	0.761
European taurine introgression (substantial)	0.73	0.22-2.37	0.599	0.55	0.14-2.17	0.398	2.36	0.84-6.62	0.104
IBR seropositive	-	-	-	2.79	1.54-5.03	0.001	2.86	1.61-5.09	<0.001
PIV3 seropositive	2.79	1.54-5.06	0.001	-	-	-	5.78	3.27-10.21	<0.001
BVDV seropositive	2.82	1.56-5.08	0.001	5.61	3.13-10.05	<0.001	-	-	-
Heterozygosity	0.22	0.00	0.826	0.10	0.00	0.774	3004.86	0.00	0.298
Heterozygosity		-1.85×10^5			-6.25×10^5			-1.08×10^{10}	
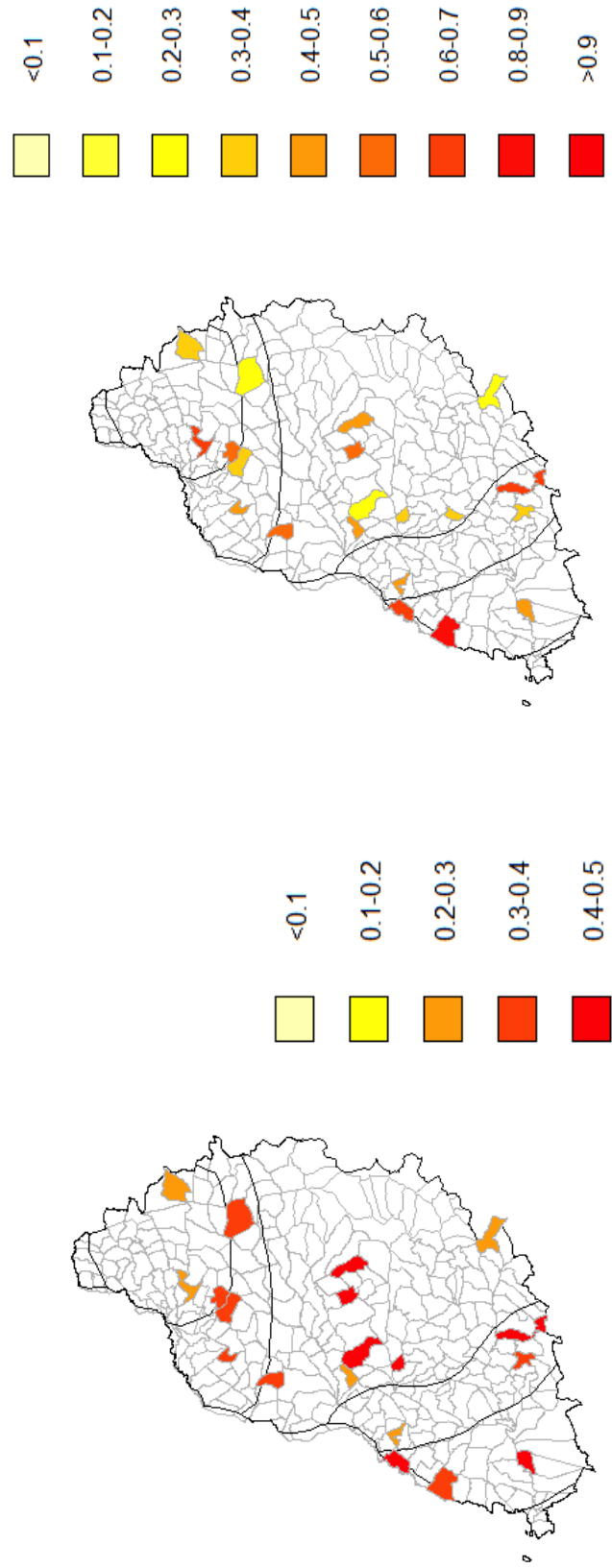
(a) *Babesia bigemina*

(b) Ever experienced any type of clinical episode

Figure D.2 Map showing the fraction of calves which are a) seropositive for *Babesia bigemina* or b) ever experienced any type of clinical episode in each sublocation. The map also shows the agro-ecological zones and the 20 sublocations in the study area.

248

D.6 Other parasites identified by the IDEAL project

Table D.10 The proportion of calves which survived until the end of the study period which were ever positive for each parasite at any time point during their 51 weeks of enrolment in the IDEAL project. Table adapted from Bronsvoort et al. (2013). The column 'Test' shows the test used to identify the virus: DB=direct Baermann; DG=dark ground microscopy; HCT=Haematocrit; ELISA=enzyme linked immunosorbent assay; FM=faecal examination by McMaster's technique; FC=faecal culture; MIC=routine microscopy; FS=faecal sedimentation; RB=routine bacteriology; RLB=reverse line blot. The column 'Visit Tested' shows the visit at which the test took place: CE=clinical examination; 7D=recruitment visit; 5W=routine 5 weekly visit; Y=final visit at 51 weeks. The column 'Laboratory' referrers to the laboratory where the testing was carried out: Busia= IDEAL project field laboratory in Busia, Kenya; IRLI=International Livestock Research Institute, Nairobi, Kenya; Pretoria=Onderstepoort Veterinary Research Institute, the University of Pretoria, South Africa. *Doubtful individuals classified as seropositive.

Parasite	Test	Visits Tested	Laboratory	Prevalence (%)
Endoparasites				
Strongyle eggs	FM	7D, 5W, Y	Busia	100.00
Haemonchus placei	FM, FC	7D, 5W, Y	Busia	99.56
Coccidia spp.	FM, FC	7D, 5W, Y	Busia	95.38
Trichostrongylus axei	FM, FC	7D, 5W, Y	Busia	92.31
Calicophoron spp.	FM, FC	7D, 5W, Y	Busia	73.85
Oesophagostomum radiatum	FM, FC	7D, 5W, Y	Busia	68.13
Strongyloides spp.	MIC	7D, 5W, Y	Busia	57.80
Dictyocaulus viviparus (L1)	DB	7D, 5W, Y	Busia	36.70
Toxocara vitulorum	FM, FC	7D, 5W, Y	Busia	26.37
Eimeria bovis	FM, MIC	7D, 5W, Y	Busia	19.56
Eimeria alabamensis	FM, MIC	7D, 5W, Y	Busia	19.12
Fasciola spp.	FS, MIC	7D, 5W, Y	Busia	18.24
Moniezia spp.	FM, FC	7D, 5W, Y	Busia	15.82
Eimeria subspherica	FS, MIC	7D, 5W, Y	Busia	14.29
Neospora caninum	ELISA	Y	Busia	13.85
Eimeria zuernii	FM, MIC	7D, 5W, Y	Busia	12.09
Trichuris spp.	FM, FC	7D, 5W, Y	Busia	10.77
Nematodirus spp.	FM, FC	7D, 5W, Y	Busia	5.71
Cooperia spp.	FM, FC	7D, 5W, Y	Busia	5.71
Ostertagia ostertagi	FM, FC	7D, 5W, Y	Busia	1.98
Eimeria cylindrica	FM, MIC	7D, 5W, Y	Busia	1.54

Table D.10 – continued from previous page

Parasite	Test	Visits Tested	Laboratory	Prevalence (%)
Eimeria auburnensis	FM, MIC	7D, 5W, Y	Busia	0.66
Bunostomum trigonocephalum	FM	7D, 5W, Y	Busia	0.44
Bacteria				
Gram negative bacilli	RB	CE	Busia	3.96
Gram positive cocci	RB	CE	Busia	2.20
Gram positive bacilli	RB	CE	Busia	1.98
E. *coli*	RB	CE	Busia	1.32
Streptococcus spp.	RB	CE	Busia	0.88
Thin rods (Bacteria)	RB	CE	Busia	0.44
Non-pathogenic *Staphylococci*	RB	CE	Busia	0.44
Staphylococcus epidermidis	RB	CE	Busia	0.22
Staphylococcus aureus	RB	CE	Busia	0.22
Klebsiella pneumoniae	RB	CE	Busia	0.22
Klebsiella ozaenae	RB	CE	Busia	0.22
Bacillus spp.	RB	CE	Busia	0.22
Actinomycetes spp.	RB	CE	Busia	0.22
Pasteurella multocida	RB	CE	Busia	0.00
Viruses				
Bluetongue Virus	ELISA	Y, CE	ILRI	94.73*
Epizootic haemorrhagic disease	ELISA	Y	ILRI	66.37*
Rotavirus	ELISA	CE	ILRI	0.00
Fungi				
Trichophyton spp.	MIC	CE	Busia	7.91
Haemoparasites				
Theileria spp.	MIC	7D,5W,Y,CE	Busia	100.00
Theileria parva	ELISA	7D, 5W, Y	ILRI	85.27
Theileria mutans	ELISA	7D, 5W, Y	ILRI	87.69
Anaplasma marginale	ELISA	7D, 5W, Y	ILRI	55.60
Theileria velifera	RLB	Y	Pretoria	64.14
Ehrlichia spp. (Omatjenne)	RLB	Y	Pretoria	45.21
Anaplasma bovis	RLB	Y	Pretoria	42.09
Babesia bigemina	ELISA	7D, 5W, Y	ILRI	37.80
Theileria sp. (sable)	RLB	Y	Pretoria	33.41
Theileria taurotragi	RLB	Y	Pretoria	14.25
Anaplasma spp.	MIC	7D, 5W, Y	Busia	10.33
Trypanosoma vivax	DG, HCT	7D, 5W, Y	Busia	9.67

Table D.10 – continued from previous page

Parasite	Test	Visits Tested	Laboratory	Prevalence (%)
Trypanosoma spp.	DG, HCT	7D, 5W, Y	Busia	8.13
Theileria ovis	RLB	Y	Pretoria	3.79
Babesia bovis	RLB	Y	Pretoria	3.12
Theileria bicornis	RLB	Y	Pretoria	2.23
Anaplasma phagocytophilum	RLB	Y	Pretoria	1.34
Microfilaria spp.	DG, HCT	7D, 5W, Y	Busia	1.32
Theileria sp. (buffalo)	RLB	Y	Pretoria	0.89
Ehrlichia ruminantium	RLB	Y	Pretoria	0.89
Babesia spp.	MIC	7D, 5W, Y	Busia	0.66
Trypanosoma theileri	DG, HCT	7D, 5W, Y	Busia	0.66
Ehrlichia canis	RLB	Y	Pretoria	0.45
Trypanosoma brucei	DG, HCT	7D, 5W, Y	Busia	0.44
Trypanosoma congolense	DG, HCT	7D, 5W, Y	Busia	0.44
Theileria equi	RLB	Y	Pretoria	0.22
Theileria buffeli	RLB	Y	Pretoria	0.00

Table D.11 Results of association between IBR, PIV3 or BVDV serostatus and ever being infected with another parasite. Models were fitted with either IBR, PIV3 or BVDV serostatus as a response variable and ever being infected with the parasite of interest during the calves 51 weeks of observation in the IDEAL project fitted as a binary fixed effect or median strongyle EPG fitted as a continuous fixed effect. The serostatus of the other viruses (IBR, PIV3 or BVDV depending on the virus of interest) was also fitted as a binary explanatory variable. Sublocation was included in the model as a random effect.

	Response Virus								
	IBR			PIV3			BVDV		
Explanatory Variable: Ever Had Parasite	**OR**	**95% CI**	**P Value**	**OR**	**95% CI**	**P value**	**OR**	**95% CI**	**P Value**
Anaplasma bovis	1.05	0.63-1.75	0.848	0.63	0.36-1.1	0.105	0.99	0.58-1.72	0.985
Anaplasma marginale	1.83	1.09-3.07	0.023	0.99	0.57-1.72	0.979	1.30	0.75-2.24	0.353
Anaplasma phagocytophilum	0.99	0.14-6.97	0.993	4.82	0.67-34.92	0.119	1.12	0.15-8.46	0.914
Anaplasma spp.	1.06	0.48-2.34	0.879	1.08	0.45-2.59	0.863	0.92	0.39-2.21	0.860
Babesia bigemina	1.18	0.71-1.96	0.529	0.94	0.53-1.66	0.828	2.88	1.67-4.97	<0.001
Babesia bovis	0.53	0.1-2.71	0.443	0.64	0.12-3.41	0.600	2.56	0.7-9.44	0.157
Bluetongue virus	0.63	0.23-1.74	0.370	0.97	0.29-3.25	0.962	1.70	0.44-6.50	0.440
Calicophoron spp.	0.98	0.56-1.73	0.958	0.68	0.38-1.22	0.199	0.65	0.37-1.15	0.142
Coccidia spp.	0.78	0.25-2.44	0.669	0.88	0.24-3.15	0.840	1.49	0.38-5.85	0.570
Cooperia spp.	1.12	0.39-3.22	0.829	1.07	0.32-3.51	0.914	0.37	0.08-1.72	0.206
Dictyocaulus viviparus (L1)	0.83	0.49-1.4	0.487	1.06	0.6-1.85	0.849	1.16	0.67-2.01	0.584
Ehrlichia spp. (Omatjenne)	1.63	0.97-2.74	0.066	0.97	0.56-1.69	0.917	1.33	0.78-2.26	0.300
Eimeria alabamensis	0.94	0.5-1.76	0.851	0.91	0.46-1.79	0.775	1.61	0.85-3.05	0.142
Eimeria bovis	1.11	0.61-2.03	0.729	0.87	0.44-1.71	0.684	1.53	0.81-2.88	0.192

Continued on next page.

Table D.11 – continued from previous page

Explanatory Variable: Ever Had Parasite	Response Virus								
	IBR			PIV3			BVDV		
	OR	95% CI	P Value	OR	95% CI	P value	OR	95% CI	P Value
Eimeria cylindrical	0.51	0.05-4.92	0.559	2.44	0.37-16.32	0.357	0.56	0.05-6.07	0.634
Eimeria subspherica	0.50	0.22-1.11	0.090	1.19	0.57-2.51	0.645	1.44	0.69-2.99	0.330
Eimeria zuernii	0.93	0.44-1.98	0.849	0.61	0.26-1.46	0.266	2.14	1.02-4.48	0.044
Epizootic haemorrhagic disease virus	1.51	0.87-2.62	0.146	0.72	0.40-1.29	0.269	1.41	0.78-2.53	0.255
Fasciola spp.	1.36	0.74-2.52	0.321	0.78	0.38-1.60	0.495	1.15	0.59-2.26	0.677
Microfilaria spp.	0.58	0.06-6.05	0.651	6.54	1.04-41.02	0.045	0.46	0.04-4.81	0.513
Moniezia spp.	1.44	0.75-2.78	0.273	1.08	0.51-2.28	0.840	0.42	0.17-1.01	0.052
Nematodirus spp.	0.32	0.07-1.46	0.140	0.81	0.22-3.02	0.756	0.86	0.24-3.13	0.825
Neospora caninum	1.57	0.81-3.02	0.178	0.82	0.38-1.78	0.614	1.40	0.68-2.87	0.364
Oesophagostomum radiatum	0.71	0.42-1.18	0.188	1.02	0.57-1.82	0.954	1.07	0.60-1.89	0.821
Strongyloides spp.	1.10	0.66-1.82	0.716	1.20	0.69-2.10	0.522	0.59	0.35-1.01	0.056
Theileria bicornis	1.74	0.39-7.84	0.468	0.95	0.15-5.86	0.955	1.01	0.18-5.67	0.995
Theileria mutans	1.49	0.66-3.38	0.336	1.02	0.44-2.35	0.968	0.99	0.44-2.27	0.988
Theileria ovis	0.26	0.03-2.08	0.202	1.32	0.34-5.15	0.689	1.27	0.33-4.90	0.726
Theileria parva	1.21	0.57-2.57	0.620	1.38	0.57-3.35	0.472	2.63	0.96-7.22	0.060
Theileria sp. (sable)	1.04	0.6-1.78	0.896	0.94	0.53-1.68	0.836	0.92	0.52-1.63	0.783
Theileria taurotragi	0.92	0.45-1.87	0.808	1.62	0.8-3.29	0.182	1.22	0.6-2.49	0.580
Theileria velifera	1.04	0.62-1.75	0.876	0.84	0.48-1.46	0.543	0.85	0.49-1.47	0.565
Toxocara vitulorum	1.01	0.57-1.77	0.975	0.90	0.48-1.69	0.751	0.85	0.46-1.59	0.616

Table D.11 – continued from previous page

Explanatory Variable: Ever Had Parasite	Response Virus								
	IBR			PIV3			BVDV		
	OR	95% CI	P Value	OR	95% CI	P value	OR	95% CI	P Value
Trichostrongylus axei	1.11	0.44-2.78	0.826	1.04	0.4-2.74	0.935	0.42	0.18-0.98	0.046
Trichuris spp.	1.01	0.45-2.25	0.981	1.18	0.5-2.77	0.704	0.46	0.16-1.32	0.149
Trypanosoma brucei	2.29	0.09-58.77	0.616	2.37	0.09-65	0.611	2.50	0.08-81.76	0.607
Trypanosoma spp.	1.19	0.5-2.81	0.695	1.43	0.58-3.55	0.442	1.49	0.61-3.64	0.381
Trypanosoma vivax	1.64	0.77-3.49	0.200	1.43	0.62-3.26	0.400	1.16	0.5-2.71	0.731
Median strongyle EPG	0.76	0.52-1.10	0.144	1.05	0.68-1.63	0.830	0.99	1.66-1.50	0.979

Table D.12 Results of pairwise analysis for the effect of IBR, PIV3 or BVDV serostatus on ever having parasite. Models were fitted with ever being infected with the parasite of interest during the calf's 51 weeks of observation in the IDEAL project fitted as a binary response or in the case of median strongyle EPG as a continuous value using a GLMM with a negative binomial error distribution. Either IBR, PIV3 or BVDV serostatus was fitted as a binary explanatory variable. Sublocation was included in the model as a random effect.

| | Explanatory Virus | | | | | | | | |
| | IBR | | | PIV3 | | | BVDV | | |
Response Parasite	OR	95% CI	P Value	OR	95% CI	P value	OR	95% CI	P Value
Anaplasma bovis	0.94	0.58-1.53	0.812	0.61	0.36-1.03	0.065	0.85	0.51-1.41	0.526
Anaplasma marginale	1.82	1.12-2.95	0.015	1.25	0.76-2.04	0.373	1.47	0.89-2.42	0.132
Anaplasma phagocytophilum	1.98	0.30-12.90	0.474	4.98	0.93-26.64	0.060	2.35	0.41-13.28	0.335
Anaplasma spp.	1.09	0.52-2.29	0.817	1.12	0.52-2.42	0.772	0.97	0.44-2.17	0.948
Babesia bigemina	1.46	0.92-2.32	0.112	1.45	0.89-2.36	0.137	2.95	1.79-4.85	<0.001
Babesia bovis	0.65	0.14-3.04	0.588	0.80	0.17-3.71	0.770	1.98	0.59-6.63	0.269
Bluetongue virus	0.70	0.26-1.87	0.474	1.04	0.33-3.26	0.940	1.54	0.44-5.45	0.503
Calicophoron spp.	0.79	0.45-1.36	0.389	0.57	0.33-1.00	0.051	0.56	0.32-0.98	0.042
Coccidia spp.	0.81	0.28-2.35	0.694	0.89	0.28-2.86	0.850	1.35	0.37-4.88	0.644
Cooperia spp.	0.95	0.35-2.59	0.920	0.84	0.28-2.51	0.758	0.38	0.09-1.65	0.196
Dictyocaulus viviparus (L1)	0.88	0.53-1.46	0.626	1.02	0.60-1.72	0.948	1.06	0.63-1.78	0.833
Ehrlichia spp. (Omatjenne)	1.72	1.05-2.83	0.031	1.16	0.70-1.93	0.562	1.42	0.86-2.36	0.173
Eimeria alabamensis	1.04	0.58-1.88	0.894	1.08	0.58-2.00	0.817	1.63	0.90-2.93	0.104
Eimeria bovis	1.20	0.68-2.12	0.530	1.04	0.56-1.92	0.904	1.53	0.86-2.74	0.147
Eimeria cylindrical	0.66	0.08-5.58	0.705	1.89	0.36-9.93	0.451	0.79	0.09-6.66	0.829

Continued on next page.

Table D.12 – continued from previous page

Response Parasite	Explanatory Virus								
	IBR			PIV3			BVDV		
	OR	95% CI	P Value	OR	95% CI	P value	OR	95% CI	P Value
Eimeria subspherica	0.60	0.28-1.27	0.180	1.20	0.61-2.35	0.593	1.39	0.72-2.69	0.321
Eimeria zuernii	0.99	0.48-2.01	0.974	0.77	0.35-1.71	0.517	1.79	0.92-3.51	0.088
Epizootic haemorrhagic disease	1.42	0.83-2.44	0.201	0.84	0.49-1.45	0.531	1.30	0.74-2.29	0.353
Fasciola spp.	1.35	0.75-2.41	0.314	0.88	0.46-1.69	0.699	1.10	0.59-2.06	0.760
Microfilaria spp.	0.72	0.07-7.57	0.787	2.11	0.33-13.4	0.430	0.32	0.03-3.49	0.353
Moniezia spp.	1.19	0.64-2.20	0.586	0.91	0.46-1.79	0.777	0.46	0.2-1.04	0.063
Nematodirus spp.	0.29	0.07-1.28	0.103	0.62	0.18-2.13	0.446	0.59	0.17-2.02	0.398
Neospora caninum	1.59	0.86-2.93	0.138	0.99	0.49-1.99	0.971	1.44	0.75-2.76	0.275
Oesophagostomum radiatum	0.73	0.44-1.21	0.220	0.99	0.58-1.71	0.981	0.99	0.58-1.70	0.971
Strongyloides spp.	0.99	0.61-1.63	0.982	0.93	0.55-1.58	0.794	0.58	0.34-0.97	0.038
Theileria bicornis	1.78	0.45-7.05	0.408	1.15	0.24-5.54	0.858	1.18	0.24-5.64	0.840
Theileria mutans	1.58	0.72-3.46	0.257	1.13	0.53-2.42	0.745	1.11	0.52-2.37	0.785
Theileria ovis	0.26	0.03-2.05	0.202	1.04	0.28-3.80	0.952	1.06	0.29-3.86	0.930
Theileria parva	1.50	0.72-3.12	0.282	1.78	0.77-4.14	0.179	2.74	1.06-7.09	0.037
Theileria sp. (sable)	0.92	0.54-1.58	0.775	0.90	0.51-1.57	0.705	0.95	0.54-1.68	0.871
Theileria taurotragi	1.06	0.54-2.09	0.873	1.60	0.82-3.09	0.165	1.35	0.69-2.65	0.381
Theileria velifera	0.97	0.6-1.58	0.912	0.86	0.52-1.42	0.563	0.84	0.51-1.38	0.490
Toxocara vitulorum	0.95	0.56-1.63	0.857	0.86	0.49-1.54	0.619	0.83	0.47-1.48	0.538
Trichostrongylus axei	0.84	0.36-1.97	0.690	0.74	0.31-1.75	0.494	0.41	0.19-0.90	0.026

Table D.12 – continued from previous page

| Response Parasite | Explanatory Virus | | | | | | | | | | | |
	IBR			PIV3			BVDV					
	OR	95% CI	P Value	OR	95% CI	P value	OR	95% CI	P Value			
Trichuris spp.	0.89	0.41-1.92	0.766	0.90	0.40-2.02	0.805	0.51	0.19-1.33	0.169			
Trypanosoma brucei	4.03	0.25-65.11	0.326	4.72	0.29-76.29	0.274	4.81	0.3-77.69	0.269			
Trypanosoma spp.	1.36	0.61-3.03	0.452	1.79	0.79-4.06	0.163	1.71	0.76-3.85	0.192			
Trypanosoma vivax	1.82	0.88-3.73	0.104	1.68	0.79-3.58	0.180	1.54	0.72-3.33	0.267			
Median strongyle EPG	0.95	0.74-1.21	0.677	1.15	089-1.49	0.278	1.00	0.77-1.30	0.994			

D.7 Association with clinical signs
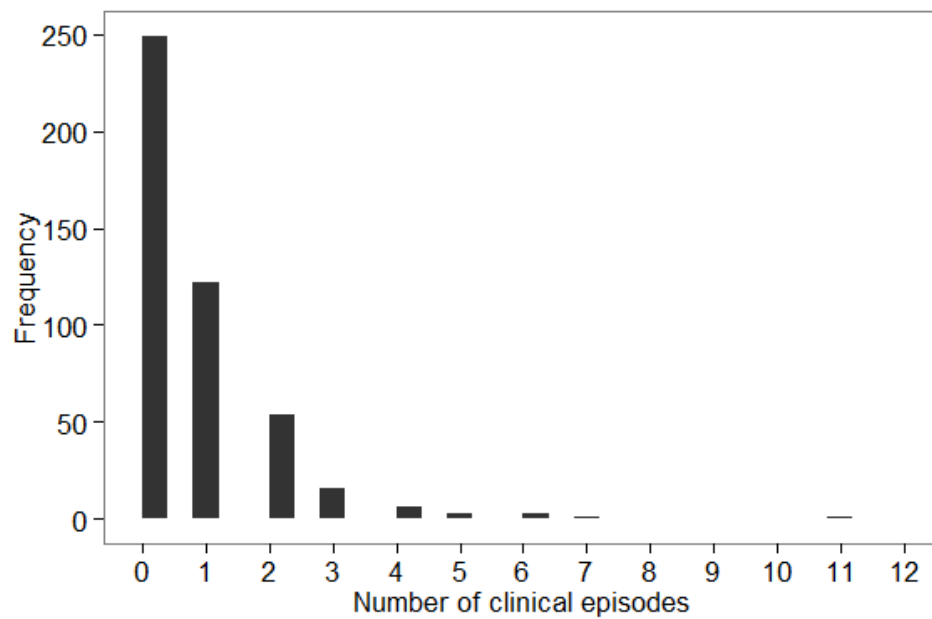
Figure D.3 Total number of clinical episodes a calf experienced during its time in the IDEAL study.

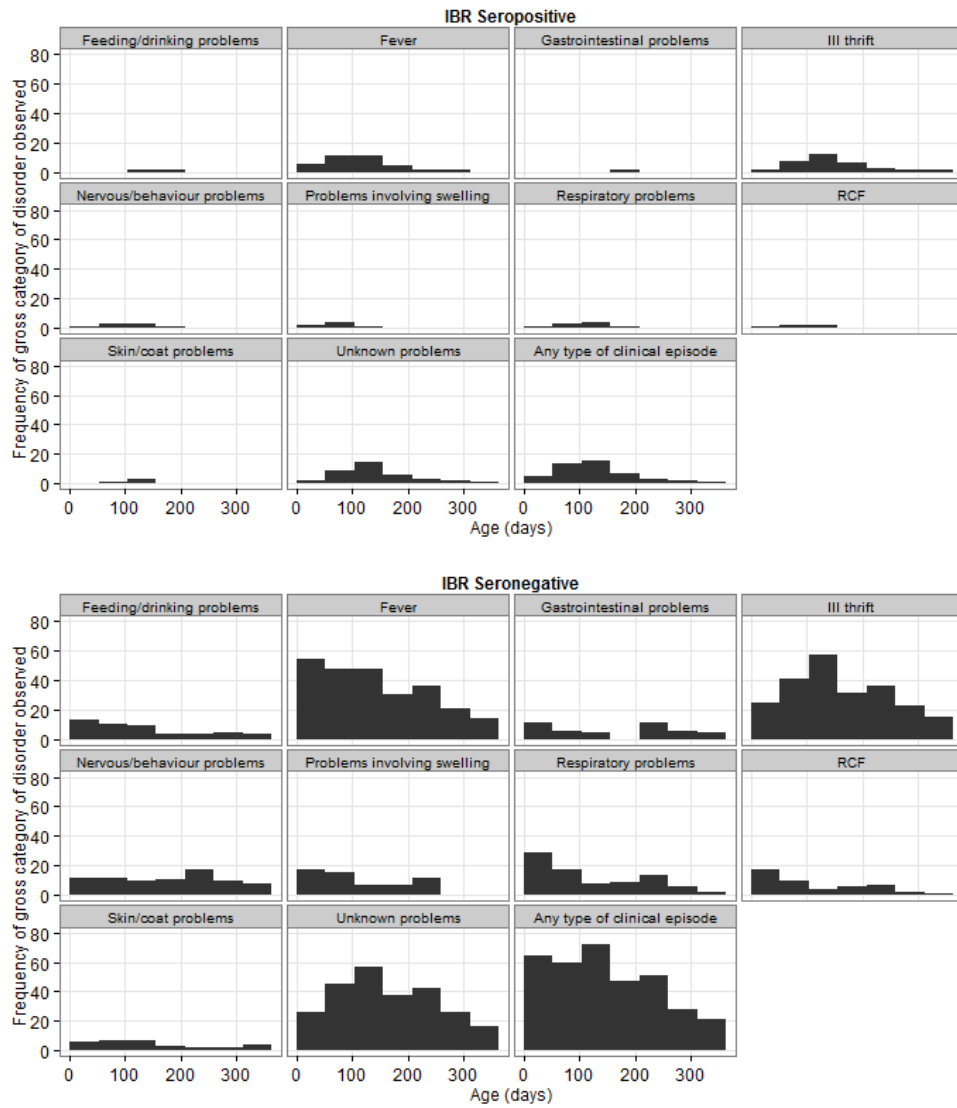**Figure D.4** Age at which calves experienced each type of gross category of clinical disorder during their enrolment in the IDEAL project given their IBR serostatus. RCF = Respiratory problems or cough and fever.
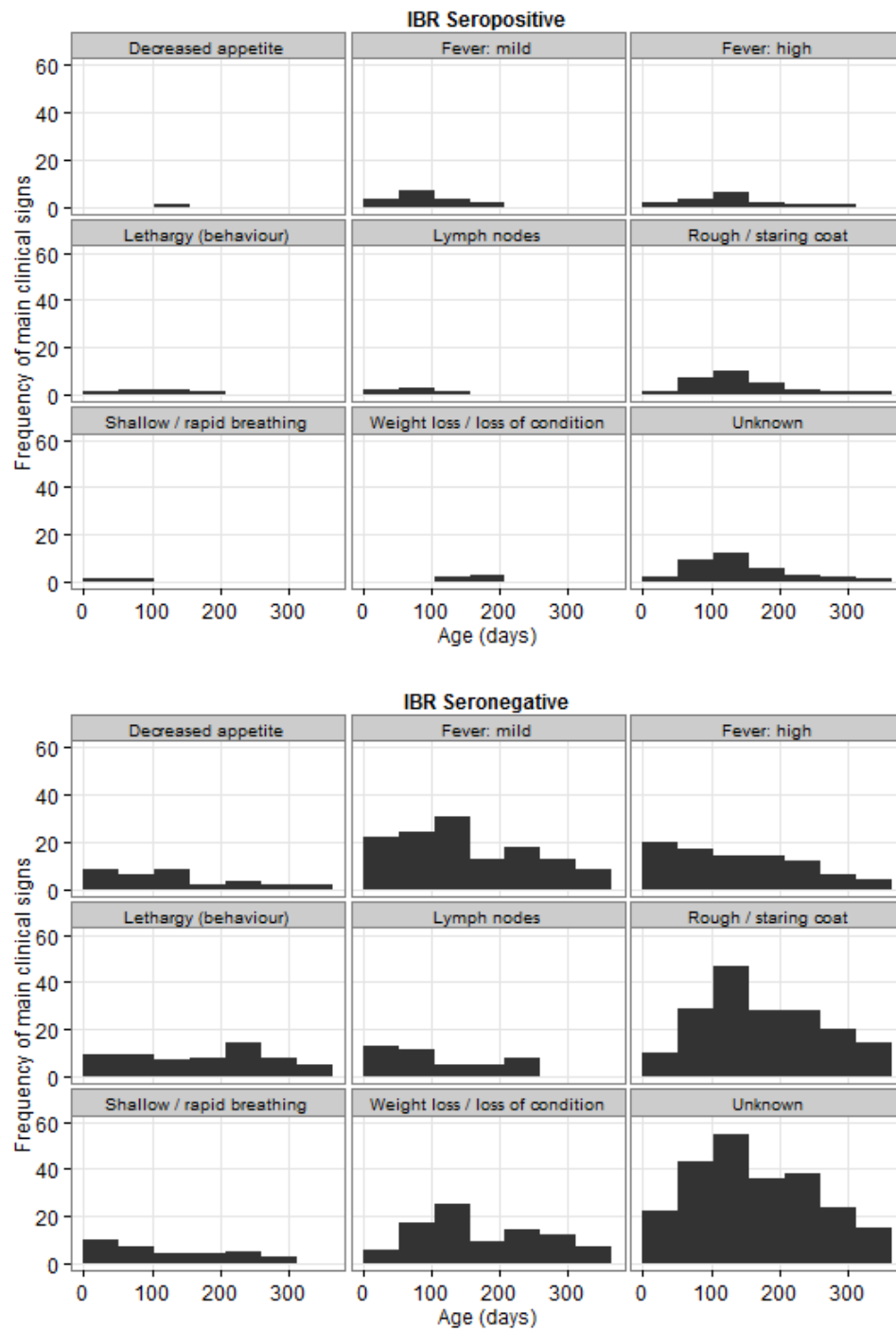
Figure D.5 Age at which calves experienced specific clinical signs during their enrolment in the IDEAL project during their enrolment in the IDEAL project given their IBR serostatus.

Appendix E

Variation and covariation in strongyle infection in East African shorthorn zebu calves

This appendix forms the supplementary materials referred to in the manuscript from Appendix B on "Variation and covariation in strongyle infection in East African shorthorn zebu calves".

E.1 Impact of different quality control parameters and marker density on heritability estimates

Different SNP quality control checks result in different heritability estimates. A balance is needed to be found between quality control checking and removing too much of the variation from the population. Table E.2 presents the heritability estimates for strongyle EPG in the univariate animal model as a result of varying the parameters in the quality control checks on the SNP data. All quality control cut-offs used in each section of Table E.2 produce very similar heritability estimates. However removal of the European taurine introgressed calves produce lower heritability estimates with higher standard errors than models which include the European taurine introgressed calves, this is discussed in more detail in the "Effect of including European taurine introgressed calves" section of Appendix E.

A SNP call rate cut-off = 0.9; individual call rate=0.9, identity by state (IBS) threshold of 0.9 and a Hardy-Weinberg Equilibrium cut-off of $1.00e^{-8}$ was chosen to build the models within this paper. The minor allele frequency for SNPs was set to 0.005, to include all SNPs where the minor allele count was 5 or more.

E.1 Impact of different quality control parameters and marker density on heritability estimates

To assess the effect of the number of markers on the heritability, I randomly sampled sets of 25, 50 and 75% of all autosomal markers corresponding to 10530, 21060, 31589 autosomal SNPs and examined their effect upon the heritability estimates of strongyle EPG. I show that the estimate of heritability of strongyle EPG increased with increasing marker density, and it was only in the model involving 100% of the markers which passed quality control that estimates were significantly different from zero (Table E.1).

Table E.1 Effect of changing the number of markers on the heritability of strongyle EPG. Estimates are followed by their standard errors in brackets.

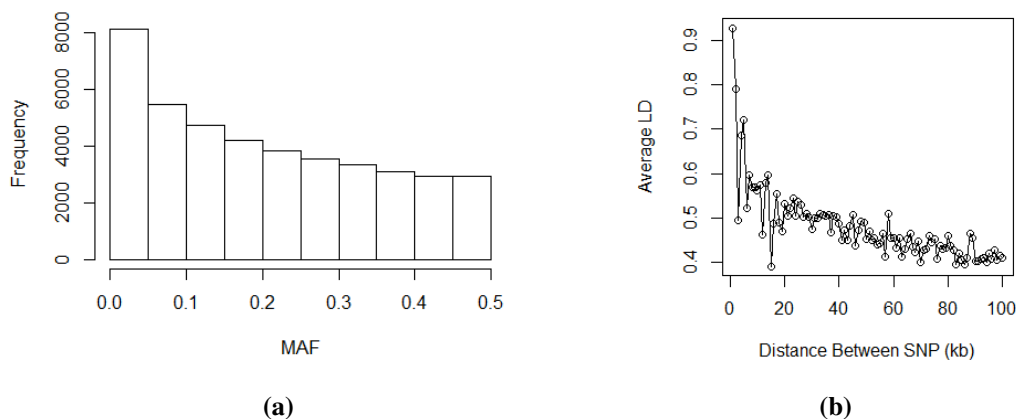| Percentage of Autosomal Markers Sampled	Number of Markers Sampled	Heritability (%)
25	10530	9.27 (9.83)
50	21060	17.58 (10.81)
75	31589	19.43 (11.64)
100	42119	23.92 (11.83)

(a)　　　　　　　　　　　　(b)

Figure E.1 a) The distribution of minor allele frequencies (MAF) at SNP markers and b) The average linkage disequilibrium (LD) in 1kb bins between pairs of SNPs plotted against the physical distance between SNPs in the pair.

Table E.2 Effect of changing the quality control cut-offs on the heritability of strongyle EPG using a univariate animal model with a negative binomial errors. The estimate is followed by its standard error in brackets. The impact of quality control on a study population with European taurine introgressed calves either a) included or b) excluded is also presented. See section of Appendix E called "Effect of including European taurine introgressed calves" for an explanation of European taurine introgressed calves. $N_{individuals}$ = Number of individuals left after quality control; $N_{markers}$ = Number of markers left after quality control.

a) All calves

SNP call rate	Individual call rate	Minor allele frequency	P value of Hardy-Weinberg Equilibrium	Identity by state cut-off	$N_{individuals}$	$N_{markers}$	Heritability (%)
0.9	0.9	0.005	$1.00e^{-08}$	0.9	446	42119	23.92 (11.83)
0.9	0.9	0.001	$1.00e^{-08}$	0.9	446	45241	27.12 (11.89)
0.99	0.99	0.01	$1.00e^{-08}$	0.99	440	36416	22.31 (12.16)

b) Without European taurine introgressed calves

SNP call rate	Individual call rate	Minor allele frequency	P value of Hardy-Weinberg Equilibrium	Identity by state cut-off	$N_{individuals}$	$N_{markers}$	Heritability (%)
0.9	0.9	0.005	$1.00e^{-08}$	0.9	353	40381	13.25 (13.37)
0.9	0.9	0.001	$1.00e^{-08}$	0.9	353	43460	14.30 (13.99)
0.99	0.99	0.01	$1.00e^{-08}$	0.99	348	35678	9.18 (13.23)

E.2 Effect of including European taurine introgressed calves

Previous research by Mbole-Kariuki et al. (2014) has indicated that there is evidence for European taurine introgression being present in the study population. Mbole-Kariuki et al. (2014) used the Ward clustering method (Ward Jr, 1963) to identify clusters in the distribution of European taurine ancestry. The first of these categories included calves with $\geq$12.5% European taurine background (representing animals with "substantial" European introgression, $N = 29$); the second included calves with between 1.56% to 12.5% European taurine introgression (representing the "moderate" European taurine introgressed, $N = 94$) and the third category included calves with less than $\leq$1.56% European taurine background (representing the "non-European introgressed", $N = 425$), Mbole-Kariuki et al. (2014). Principal component analysis showed that the main structuring was between the taurine and zebu breeds and the least genetic differentiation was observed between the European taurine breeds (Mbole-Kariuki et al., 2014). Therefore it was important to assess the effect of including European taurine (ET) introgressed cattle on my heritability estimates.

In order to assess the effect of including ET introgressed cattle on the heritability estimates, the heritability of strongyle EPG with and without the introgressed calves included was calculated (Table E.2). Removal of the 'introgressed' calves from the study resulted in a lower heritability estimate and larger standard errors (with ET introgressed calves included $h^2 = 23.9\%$, SE $= 11.8\%$, N calves= 446; with ET introgressed calves excluded $h^2 = 13.3\%$, SE $= 13.4\%$, N calves $= 353$). The decrease in heritability is likely to be due to the European introgressed calves having a higher genetic variance whilst the larger standards errors are possibly due to a decrease in sample size. Consequently, I choose to include the ET introgressed calves in my study population to maximise the same size and amount of variation in the population.

The effect of including ET introgression can also be accounted for by including ET introgression as a continuous fixed effect in my models. The genomic relationship matrix and model structures for these analyses are the same as the one used in the main section of this paper, and so the only difference is the additional ET term. The impact of including ET as a fixed effect in all my analyses is presented in Table E.3.

Table E.3 shows that the heritability estimates for all traits investigated are very similar to those from the models without ET introgression fitted explicitly. ET has little effect on any of results and its effect estimate does not significantly differ from zero. For example the heritability of strongyle EPG without ET fitted as a fixed effect equals 23.9% (SE = 11.8, Table E.2) and with ET fitted as a fixed affect the heritability

of strongyle EPG equals 25.7% (SE = 11.9, Table E.3). Furthermore, since this paper is interested in investigating the causes of variance, it is important to include as much variability in the population as possible.

Table E.3 Impact of including European Taurine introgression as a continuous fixed effect in the univariate animal model with negative binomial errors distribution for strongyle EPG and the univariate animal models with Gaussian errors for the physiological traits. The rest of the model has exactly the same structure as that used in Table 1. The genomic relationship matrix includes all calves, both introgressed and not. WBC = White Blood Cell Count; RBC = Red Blood Cell Count; TSP = Total Serum Protein; EO = Transformed Absolute Eosinophil Count ($\log_{10}(EO + 1)$); Weight = Transformed Body Weight ($\log_{10}$(Weight)); V_{SL} = sublocation variance; V_A = additive genetic variance; V_{PE} = permanent environment variance; V_{RES} = residual variance; h^2 = heritability; r^2 = repeatability.

Trait	V_{SL}	V_A	V_{PE}	V_{RES}
Strongyle EPG	0.06 ± 0.03	0.49 ± 0.23	0.05 ± 0.22	1.3 ± 0.03
WBC	0.21 ± 0.14	3.07 ± 1.25	0.93 ± 1.2	7.17 ± 0.16
RBC	0.12 ± 0.06	0.62 ± 0.41	0.55 ± 0.4	1.95 ± 0.04
TSP	0.002 ± 0.003	0.07 ± 0.04	0.03 ± 0.04	0.38 ± 0.01
EO	0.0001 ± 0.0001	0.0000 ± 0.0000	0.002 ± 0.000	0.016 ± 0.000
Weight	0.0004 ± 0.0002	0.0008 ± 0.0021	0.0055 ± 0.0021	0.0029 ± 0.0001

Trait	h^2 (%)	r^2 (%)	ET effect estimate	ET effect estimate P value
Strongyle EPG	25.74 ± 11.85	31.52 ± 2.16	0.35 ± 1.93	0.86
WBC	26.96 ± 10.83	36.98 ± 1.97	-7.31 ± 4.85	0.14
RBC	19.15 ± 12.56	39.72 ± 2.11	-1.91 ± 2.4	0.44
TSP	15.33 ± 8.25	22.09 ± 1.64	-0.52 ± 0.79	0.52
EO	0.00 ± 0.00	11.17 ± 1.42	-0.16 ± 0.09	0.06
Weight	8.57 ± 21.77	69.72 ± 1.67	0.26 ± 0.13	0.10

E.3 Comparison of model assuming Gaussian or negative binomial errors

Previous work of this kind, for example Beraldi et al. (2007); Bishop et al. (1996); Coltman et al. (2001) and Stear et al. (1990) all used log transformations in their estimations of the heritability of faecal egg counts in different species. In Table E.4, I present results of univariate models of strongyle EPG with both Gaussian and negative binomial errors. The Gaussian model uses a $\log_{10}$ (Strongyle EPG + 50) transformation. Both methods produce similar estimates of heritability of strongyle EPG (Table E.4 though, interestingly, the SEs are much larger with the GLMM).

Table E.4 Results of the univariate animal models for the heritability and variance components of Strongyle EPG across all ages using Generalised Linear Mixed Models (GLMM) fitted with a negative binomial distribution and a Linear Mixed Model (LMM) with a Gaussian distribution in which the strongyle EPG has been log-transformed.

Component	GLMM		LMM	
	Estimate	SE	Estimate	SE
Sublocation Variance	0.061	0.031	0.006	0.003
Permanent Environment Variance	0.082	0.218	8.34×10^{-8}	1.98×10^{-9}
Additive Genetic Variance	0.452	0.226	0.049	0.005
Residual Variance	1.296	0.031	0.168	0.004
Heritability (%)	23.92	11.83	22.09	1.79

E.4   Age related trends in the other traits of interest
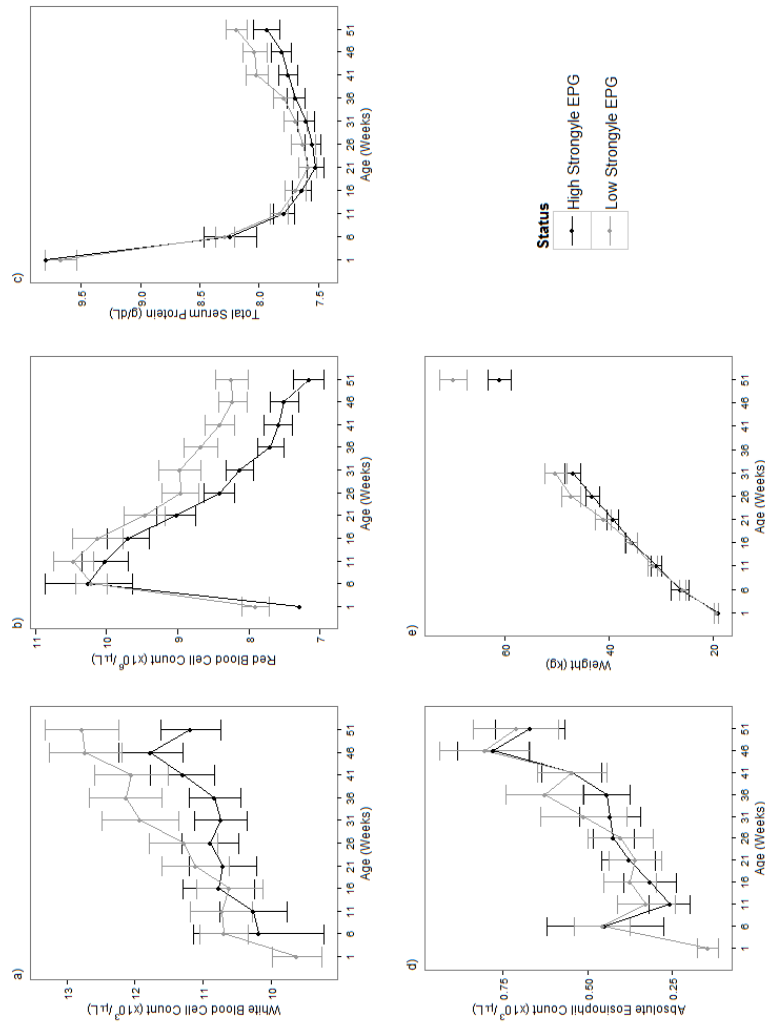
Figure E.2 : The age-related changes in the other traits for calves with a high (black line) and low (grey line) strongyle EPG. A high strongyle EPG is defined as being above the overall median strongyle EPG and a low EPG is below the median strongyle EPG. The median is the overall median taken across all visits, and is equal to 200 EPG. Error bars represent the 95% confidence intervals. Traits shown are (a) White Blood Cell Count; (b) Red Blood Cell Count; (c) Total Serum Protein; (d) Absolute Eosinophil Count; (e) Body Weight.

267

Appendix F

Genome-wide association study of East Coast Fever death and packed cell volume at the time of seroconversion to *Theileria parva*

F.1 FAmily-based Score Test for Association (FASTA) GWAS

Table F.1 SNPs associated with East Coast Fever death with a P value less than 10^{-4} from the FAmily-based Score Test for Association (FASTA) GWAS. SNPs are ordered in terms of their significance. The Q value is the P value corrected for a false discovery rate (FDR) of 0.005 using the *qvaluebh95* function in GenABEL. The effect size is given by the effect estimate of that allele/genotype in the allelic/genotypic test.

SNP Name	Chromosome	Position	P value	Q value
ARS-BFGL-BAC-18388	1	53489802	1.96×10^{-7}	9.96×10^{-5}
BTA-25932-no-rs	12	54218542	1.99×10^{-7}	9.96×10^{-5}
ARS-BFGL-NGS-78606	24	48899795	4.12×10^{-7}	1.37×10^{-4}
BTA-79451-no-rs	7	60414842	9.19×10^{-7}	2.30×10^{-4}
ARS-BFGL-NGS-62008	5	43253141	4.05×10^{-6}	8.09×10^{-4}
BFGL-NGS-112172	23	5145823	1.05×10^{-5}	1.76×10^{-3}
ARS-BFGL-NGS-48933	23	6517437	2.03×10^{-5}	2.70×10^{-3}
ARS-BFGL-NGS-36574	13	40031719	2.16×10^{-5}	2.70×10^{-3}
ARS-BFGL-NGS-100189	1	53296440	2.54×10^{-5}	2.77×10^{-3}
BTB-00893217	24	61931908	2.77×10^{-5}	2.77×10^{-3}
BTB-01968129	5	43315469	3.04×10^{-5}	2.77×10^{-3}
ARS-BFGL-NGS-56660	13	28095457	4.40×10^{-5}	3.14×10^{-3}
ARS-BFGL-NGS-79738	13	28149879	4.40×10^{-5}	3.14×10^{-3}
ARS-BFGL-NGS-67591	13	28183389	4.40×10^{-5}	3.14×10^{-3}
ARS-BFGL-NGS-40530	16	51154678	4.84×10^{-5}	3.23×10^{-3}
ARS-BFGL-NGS-24963	22	52056494	5.65×10^{-5}	3.53×10^{-3}
BTB-01477690	12	14358955	6.60×10^{-5}	3.88×10^{-3}
Hapmap46704-BTA-48392	2	85427167	7.32×10^{-5}	4.05×10^{-3}
ARS-BFGL-NGS-31451	13	37164944	7.70×10^{-5}	4.05×10^{-3}
ARS-BFGL-NGS-84142	18	18449746	8.99×10^{-5}	4.49×10^{-3}
ARS-BFGL-NGS-22854	21	11680036	9.45×10^{-5}	4.50×10^{-3}

Table F.2 Top 20 SNPs from the FASTA packed cell volume at time of seroconversion *T. parva* GWAS. SNPs are ordered in terms of their significance. The Q value is the P value corrected for a false discovery rate (FDR) of 0.005 using the *qvaluebh95* function in GenABEL. The effect size is given by the effect estimate of that allele/genotype in the allelic/genotypic test.

SNP Name	Chromosome	Position	P value	Q value
UA-IFASA-4025	13	33366445	4.14×10^{-5}	4.12×10^{-4}
BTB-01529431	20	50044555	8.16×10^{-5}	4.12×10^{-4}
ARS-BFGL-NGS-57388	10	4742149	9.84×10^{-5}	4.12×10^{-4}
ARS-BFGL-BAC-19924	15	48905274	1.77×10^{-4}	4.12×10^{-4}
BTB-01347410	26	1577113	1.87×10^{-4}	4.12×10^{-4}
ARS-BFGL-NGS-14970	5	55865567	1.99×10^{-4}	4.12×10^{-4}
BTB-01619529	26	29658789	2.09×10^{-4}	4.12×10^{-4}
BTB-00388616	9	33442358	2.21×10^{-4}	4.12×10^{-4}
BTA-94786-no-rs	26	13969960	2.31×10^{-4}	4.12×10^{-4}
BTB-00636189	16	34820781	2.47×10^{-4}	4.12×10^{-4}
Hapmap38486-BTA-30602	1	41379300	2.60×10^{-4}	4.12×10^{-4}
ARS-BFGL-NGS-56584	1	145598398	3.21×10^{-4}	4.12×10^{-4}
ARS-BFGL-NGS-42106	7	97772240	3.32×10^{-4}	4.12×10^{-4}
Hapmap26409-BTA-143127	5	55803789	3.46×10^{-4}	4.12×10^{-4}
BTB-00867539	23	45971361	3.57×10^{-4}	4.12×10^{-4}
BTB-01930671	5	76956329	3.59×10^{-4}	4.12×10^{-4}
BTB-01223707	7	79012351	3.84×10^{-4}	4.12×10^{-4}
ARS-BFGL-NGS-87466	10	7830003	3.90×10^{-4}	4.12×10^{-4}
Hapmap60788-rs29017234	X	82408727	3.92×10^{-4}	4.12×10^{-4}
BTA-87269-no-rs	9	57949852	4.63×10^{-4}	4.63×10^{-4}

F.2 Effect of European taurine introgression on East Coast Fever death GWAS

Table F.3 Top 10 SNPs from the East Coast Fever death *egscore* GWAS (after accounting for outbreeding or inbreeding by including either a) outbreeding or b) inbreeding or c) both in the model as a fixed effect). SNPs are ordered in terms of their significance. The Q value is the P value corrected for a false discovery rate (FDR) of 0.005 using the *qvaluebh95* function in GenABEL.

Original ECF Death GWAS			a) Outbreeding as a fixed effect		
SNP	P value	Q value	SNP	P value	Q value
Chr.1.53489802	2.93×10^{-7}	5.86×10^{-6}	Chr.1.53489802	3.21×10^{-7}	3.53×10^{-6}
Chr.24.48899795	8.34×10^{-7}	8.34×10^{-6}	Chr.24.48899795	9.7×10^{-7}	5.34×10^{-6}
Chr.5.43253141	3.74×10^{-6}	2.49×10^{-5}	Chr.5.43253141	4.17×10^{-6}	1.53×10^{-5}
Chr.7.60414842	5.56×10^{-6}	2.78×10^{-5}	Chr.7.60414842	5.97×10^{-6}	1.64×10^{-5}
Chr.12.54218542	1.03×10^{-5}	4.13×10^{-5}	Chr.12.54218542	1.20×10^{-5}	2.64×10^{-5}
Chr.23.5145823	1.76×10^{-5}	5.88×10^{-5}	Chr.23.5145823	1.95×10^{-5}	3.46×10^{-5}
Chr.5.43315469	2.29×10^{-5}	6.54×10^{-5}	Chr.1.53296440	2.36×10^{-5}	3.46×10^{-5}
Chr.1.53296440	2.68×10^{-5}	6.71×10^{-5}	Chr.5.43315469	2.51×10^{-5}	3.46×10^{-5}
Chr.13.40031719	3.06×10^{-5}	6.80×10^{-5}	Chr.13.40031719	3.34×10^{-5}	4.09×10^{-5}
Chr.22.52056494	3.62×10^{-5}	7.23×10^{-5}	Chr.22.52056494	4.19×10^{-5}	4.60×10^{-5}

b) Inbreeding as a fixed effect			c) Outbreeding and inbreeding as a fixed effect		
SNP	P value	Q value	SNP	P value	Q value
Chr.1.53489802	3.12×10^{-7}	3.74×10^{-6}	Chr.1.53489802	3.20×10^{-7}	3.84×10^{-6}
Chr.24.48899795	9.79×10^{-7}	5.87×10^{-6}	Chr.24.48899795	9.69×10^{-7}	5.81×10^{-6}
Chr.5.43253141	4.00×10^{-6}	1.60×10^{-5}	Chr.5.43253141	4.12×10^{-6}	1.65×10^{-5}
Chr.7.60414842	6.04×10^{-6}	1.81×10^{-5}	Chr.7.60414842	5.98×10^{-6}	1.79×10^{-5}
Chr.12.54218542	1.17×10^{-5}	2.81×10^{-5}	Chr.12.54218542	1.19×10^{-5}	2.87×10^{-5}
Chr.23.5145823	2.13×10^{-5}	3.75×10^{-5}	Chr.23.5145823	2.00×10^{-5}	3.75×10^{-5}
Chr.1.53296440	2.32×10^{-5}	3.75×10^{-5}	Chr.1.53296440	2.35×10^{-5}	3.75×10^{-5}
Chr.5.43315469	2.50×10^{-5}	3.75×10^{-5}	Chr.5.43315469	2.50×10^{-5}	3.75×10^{-5}
Chr.13.40031719	3.31×10^{-5}	4.41×10^{-5}	Chr.13.40031719	3.32×10^{-5}	4.43×10^{-5}
Chr.22.52056494	4.24×10^{-5}	5.09×10^{-5}	Chr.22.52056494	4.20×10^{-5}	5.04×10^{-5}

Table F.4 Top 10 SNPs associated with East Coast Fever death from the *egscore* GWAS (with substantially ET introgressed calves excluded from the IBS matrix and cases and controls). SNPs are ordered in terms of their significance. The Q value is the P value corrected for a false discovery rate (FDR) of 0.005 using the *qvaluebh95* function in GenABEL. The effect size is given by the effect estimate of that allele/genotype in the allelic/genotypic test.

SNP Name	Chromosome	Position	P value	Q value	Effect of the B allele	SE of the effect of the B allele	Effect of the AB genotype relative to AA	Effect of the BB genotype relative to AA
ARS-BFGL-BAC-18388	1	53489802	1.37×10^{-7}	2.58×10^{-6}	0.07	0.01	0.21	8.76×10^{-20}
ARS-BFGL-NGS-78606	24	48899795	2.80×10^{-7}	2.58×10^{-6}	0.11	0.02	0.31	8.76×10^{-20}
BTA-79451-no-rs	7	60414842	3.87×10^{-7}	2.58×10^{-6}	0.18	0.04	0.55	-2.80×10^{-4}
ARS-BFGL-NGS-36574	13	40031719	5.08×10^{-6}	2.54×10^{-5}	0.06	0.01	0.19	8.76×10^{-20}
ARS-BFGL-NGS-100189	1	53296440	7.90×10^{-6}	3.16×10^{-5}	0.03	0.01	0.11	8.76×10^{-20}
ARS-BFGL-NGS-62008	5	43253141	1.22×10^{-5}	4.06×10^{-5}	0.06	0.01	0.2	-1.19×10^{-3}
BTA-25932-no-rs	12	54218542	1.64×10^{-5}	4.53×10^{-5}	0.2	0.05	0.6	8.76×10^{-20}
BTB-01477690	12	14358955	1.81×10^{-5}	4.53×10^{-5}	0.09	0.02	0.27	8.76×10^{-20}
BTB-01398226	10	99731658	2.61×10^{-5}	5.80×10^{-5}	0.05	0.01	0.17	8.76×10^{-20}
Hapmap46704-BTA-48392	2	85427167	4.06×10^{-5}	8.11×10^{-5}	0.13	0.03	0.37	8.76×10^{-20}

F.3 Effect of European taurine introgression on packed cell volume at the time of seroconversion to *Theileria parva* GWAS

Table F.5 SNPs associated with packed cell volume at the time of seroconversion to *T. parva* with a P value less than 10-4 from the *egscore* GWAS (after accounting for outbreeding or inbreeding by including either a) outbreeding or b) inbreeding or c) both in the model as a fixed effect). SNPs are ordered in terms of their significance. The Q value is the P value corrected for a false discovery rate (FDR) of 0.005 using the *qvaluebh95* function in GenABEL.

Original PCVTP GWAS			a) Outbreeding as a fixed effect		
SNP	P value	Q value	SNP	P value	Q value
Chr.13.33366445	2.57×10^{-5}	1.76×10^{-4}	Chr.13.33366445	2.80×10^{-5}	1.49×10^{-4}
Chr.20.50044555	3.52×10^{-5}	1.76×10^{-4}	Chr.20.50044555	2.97×10^{-5}	1.49×10^{-4}
Chr.26.29658789	1.58×10^{-4}	2.67×10^{-4}	Chr.10.4742149	8.02×10^{-5}	2.41×10^{-4}
Chr.5.76956329	1.68×10^{-4}	2.67×10^{-4}	Chr.26.1577113	1.46×10^{-4}	2.41×10^{-4}
Chr.16.34820781	1.81×10^{-4}	2.67×10^{-4}	Chr.5.76956329	1.66×10^{-4}	2.41×10^{-4}
Chr.26.1577113	1.90×10^{-4}	2.67×10^{-4}	Chr.1.41379300	1.88×10^{-4}	2.41×10^{-4}
Chr.10.4742149	1.96×10^{-4}	2.67×10^{-4}	Chr.16.34820781	1.89×10^{-4}	2.41×10^{-4}
Chr.10.7830003	2.47×10^{-4}	2.67×10^{-4}	Chr.26.29658789	1.93×10^{-4}	2.41×10^{-4}
Chr.1.41379300	2.60×10^{-4}	2.67×10^{-4}	Chr.10.7830003	2.76×10^{-4}	3.07×10^{-4}
Chr.7.97772240	2.67×10^{-4}	2.67×10^{-4}	Chr.10.4766547	3.37×10^{-4}	3.37×10^{-4}

b) Inbreeding as a fixed effect			c) Outbreeding and inbreeding as a fixed effect		
SNP	P value	Q value	SNP	P value	Q value
Chr.20.50044555	2.83×10^{-5}	1.43×10^{-4}	Chr.20.50044555	2.83×10^{-5}	1.43×10^{-4}
Chr.13.33366445	2.85×10^{-5}	1.43×10^{-4}	Chr.13.33366445	2.85×10^{-5}	1.43×10^{-4}
Chr.10.4742149	8.98×10^{-5}	2.39×10^{-4}	Chr.10.4742149	8.98×10^{-5}	2.39×10^{-4}
Chr.26.1577113	1.50×10^{-4}	2.39×10^{-4}	Chr.26.1577113	1.50×10^{-4}	2.39×10^{-4}
Chr.1.41379300	1.70×10^{-4}	2.39×10^{-4}	Chr.1.41379300	1.70×10^{-4}	2.39×10^{-4}
Chr.26.29658789	1.73×10^{-4}	2.39×10^{-4}	Chr.26.29658789	1.73×10^{-4}	2.39×10^{-4}
Chr.5.76956329	1.83×10^{-4}	2.39×10^{-4}	Chr.5.76956329	1.83×10^{-4}	2.39×10^{-4}
Chr.16.34820781	1.91×10^{-4}	2.39×10^{-4}	Chr.16.34820781	1.91×10^{-4}	2.39×10^{-4}
Chr.10.7830003	3.03×10^{-4}	3.33×10^{-4}	Chr.10.7830003	3.03×10^{-4}	3.33×10^{-4}
Chr.15.48905274	3.33×10^{-4}	3.33×10^{-4}	Chr.15.48905274	3.33×10^{-4}	3.33×10^{-4}

Table F.6 Top 10 SNPs associated with packed cell volume at the time of seroconversion to *T. parva* from the 50K *egscore* GWAS (with substantially ET introgressed calves excluded from the IBS matrix and PCV$_{TP}$). SNPs are ordered in terms of their significance. The Q value is the P value corrected for a false discovery rate (FDR) of 0.005 using the *qvaluebh95* function in GenABEL. The effect size is given by the effect estimate of that allele/genotype in the allelic/genotypic test.

SNP Name	Chromosome	Position	P value	Q value	Effect of the B allele	SE of the effect of the B allele	Effect of the AB genotype relative to AA	Effect of the BB genotype relative to AA
BTB-01529431	20	50044555	1.76×10^{-5}	1.30×10^{-4}	0.70	0.16	2.24	-1.54×10^{-16}
UA-IFASA-4025	13	33366445	2.61×10^{-5}	1.30×10^{-4}	0.26	0.06	0.26	5.22×10^{-1}
Hapmap30258-BTA-143119	5	56661587	1.00×10^{-4}	2.79×10^{-4}	1.23	0.31	3.78	-1.54×10^{-16}
BTB-00388616	9	33442358	1.31×10^{-4}	2.79×10^{-4}	-0.22	0.06	-0.22	-4.37×10^{-1}
ARS-BFGL-NGS-57388	10	4742149	1.92×10^{-4}	2.79×10^{-4}	-0.36	0.09	-0.36	-7.26×10^{-1}
BTB-01619529	26	29658789	2.27×10^{-4}	2.79×10^{-4}	0.34	0.09	0.34	6.83×10^{-1}
BTB-00636189	16	34820781	2.27×10^{-4}	2.79×10^{-4}	0.29	0.08	0.29	5.83×10^{-1}
ARS-BFGL-NGS-14970	5	55865567	2.31×10^{-4}	2.79×10^{-4}	0.50	0.13	1.7	-1.54×10^{-16}
ARS-BFGL-NGS-33139	20	62593419	2.71×10^{-4}	2.79×10^{-4}	-0.44	0.12	-1.47	-1.54×10^{-16}
ARS-BFGL-BAC-19924	15	48905274	2.79×10^{-4}	2.79×10^{-4}	-0.34	0.09	-0.34	-6.72×10^{-1}

F.4   Minor allele frequencies from SNPs surrounding the candidate SNPs

East Coast Fever Death

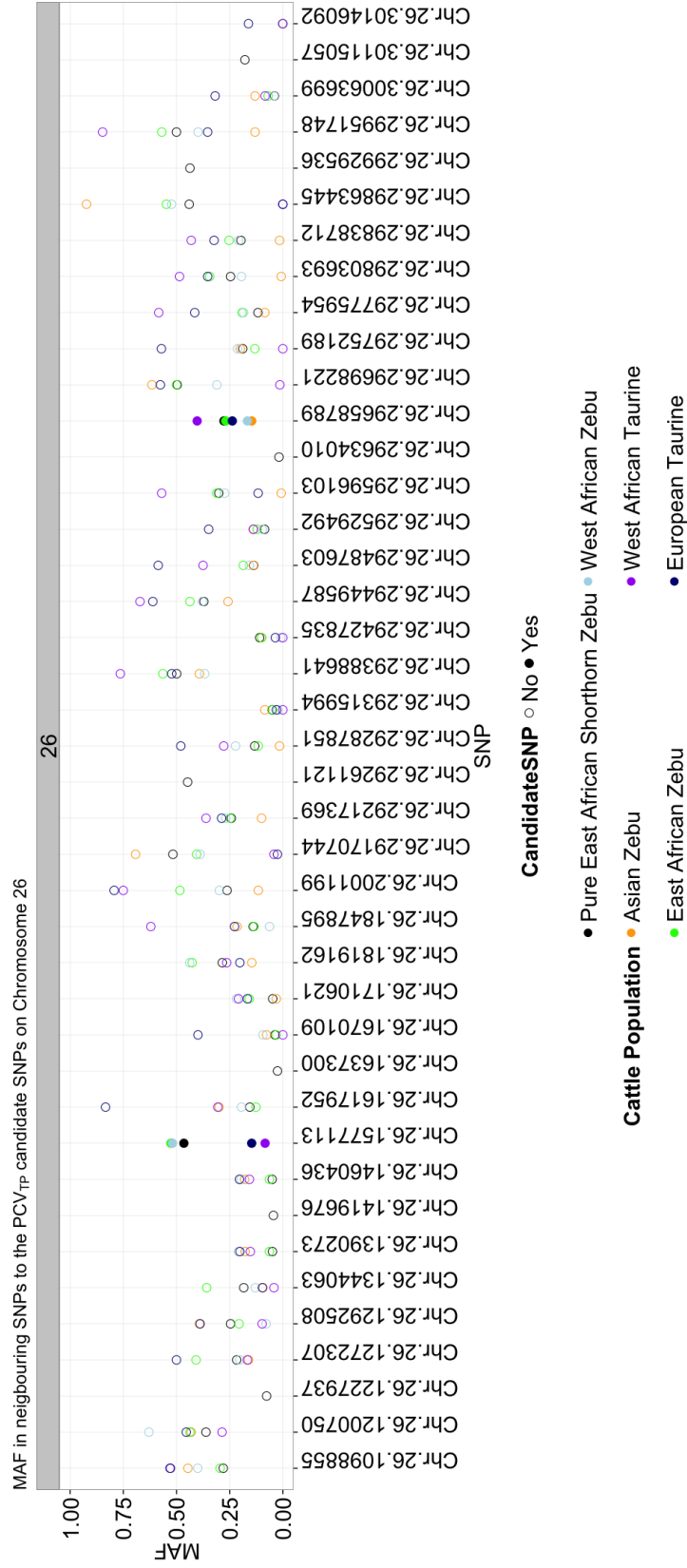MAF in neigbouring SNPs to the PCV$_{TP}$ candidate SNPs on Chromosome 26

Figure F.1 Continued on next page

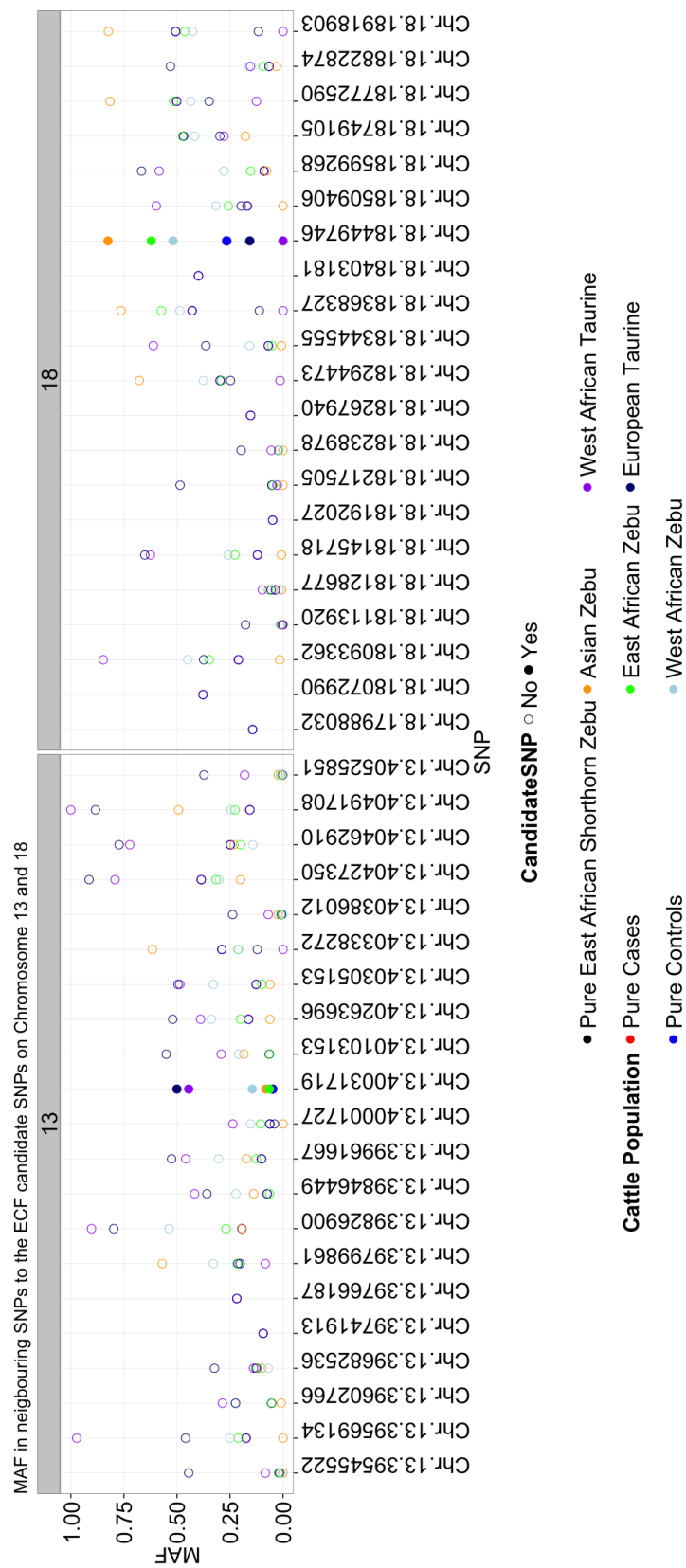**F.4 Minor allele frequencies from SNPs surrounding the candidate SNPs**

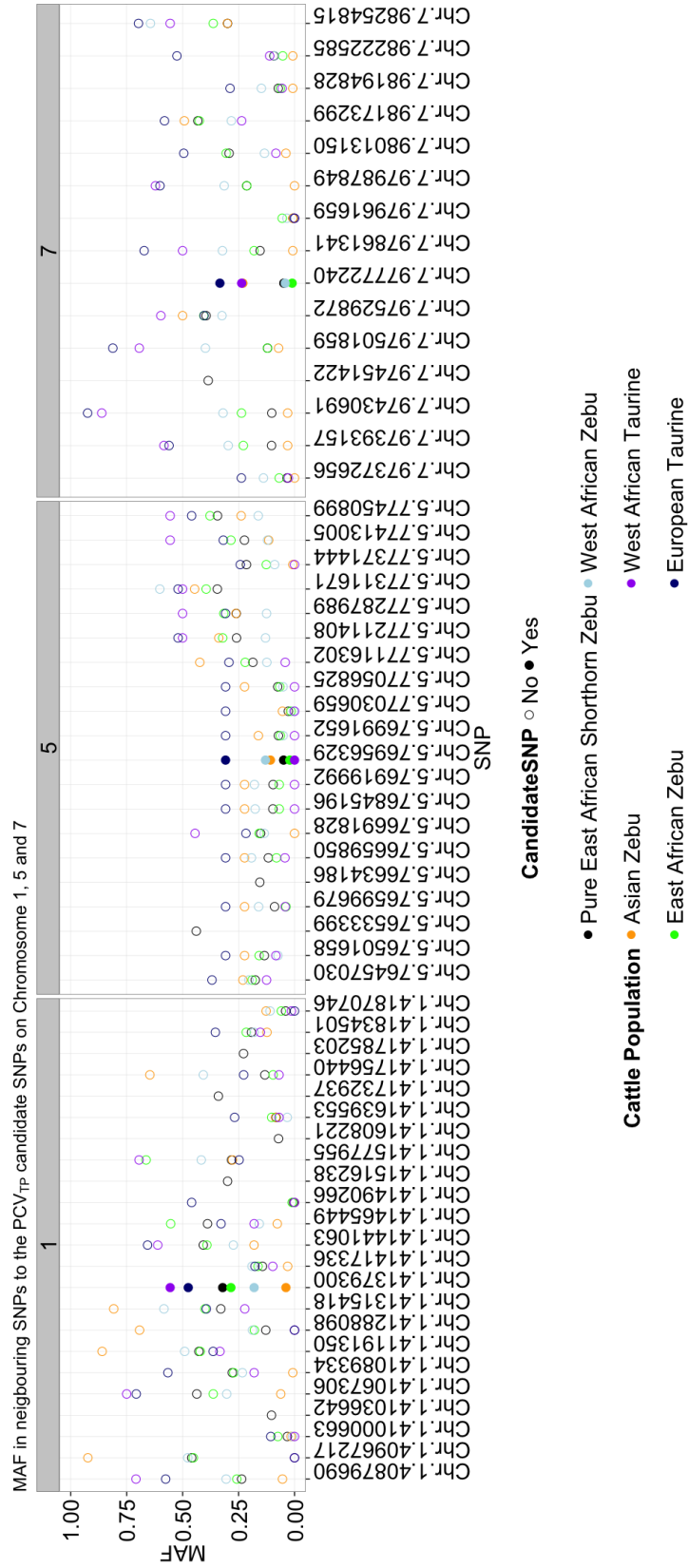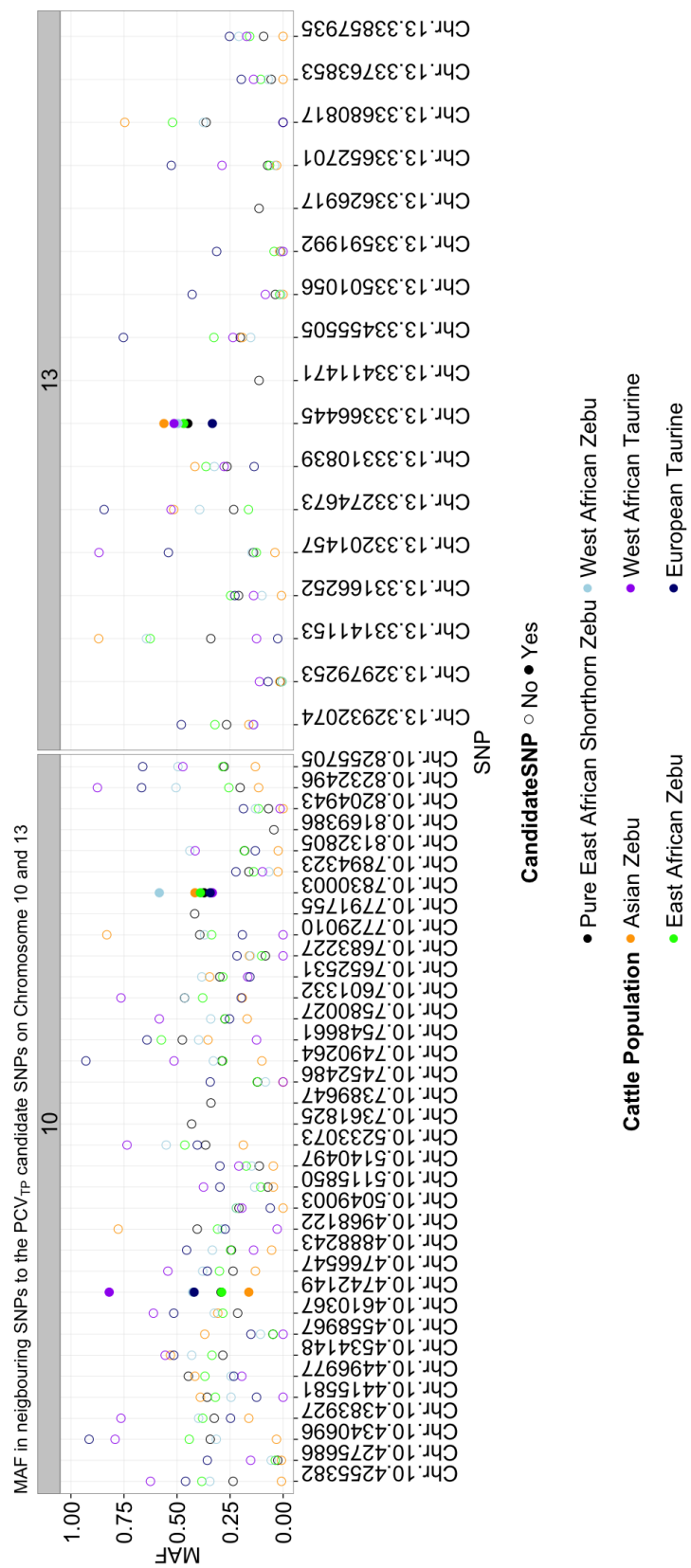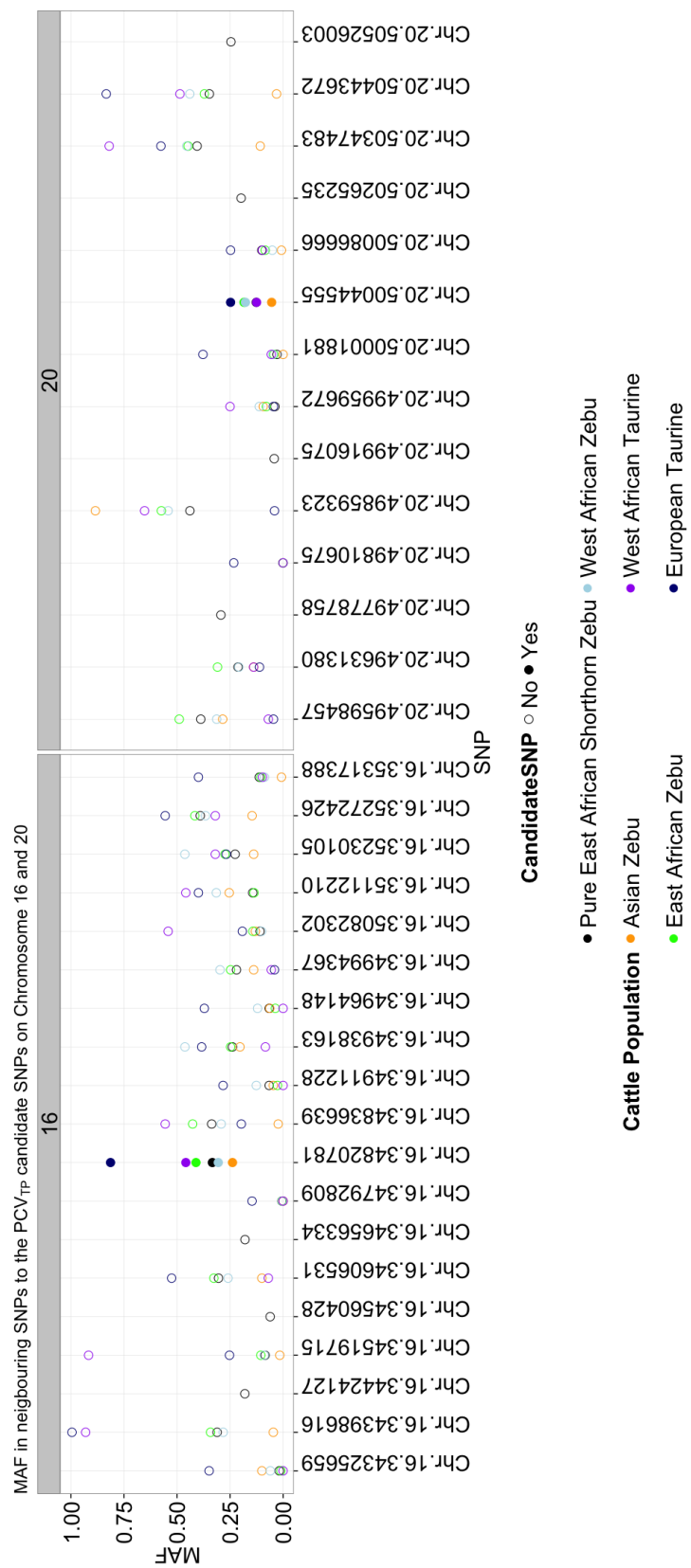Figure F.1 Continued on next page

Figure F.1 Continued on next page
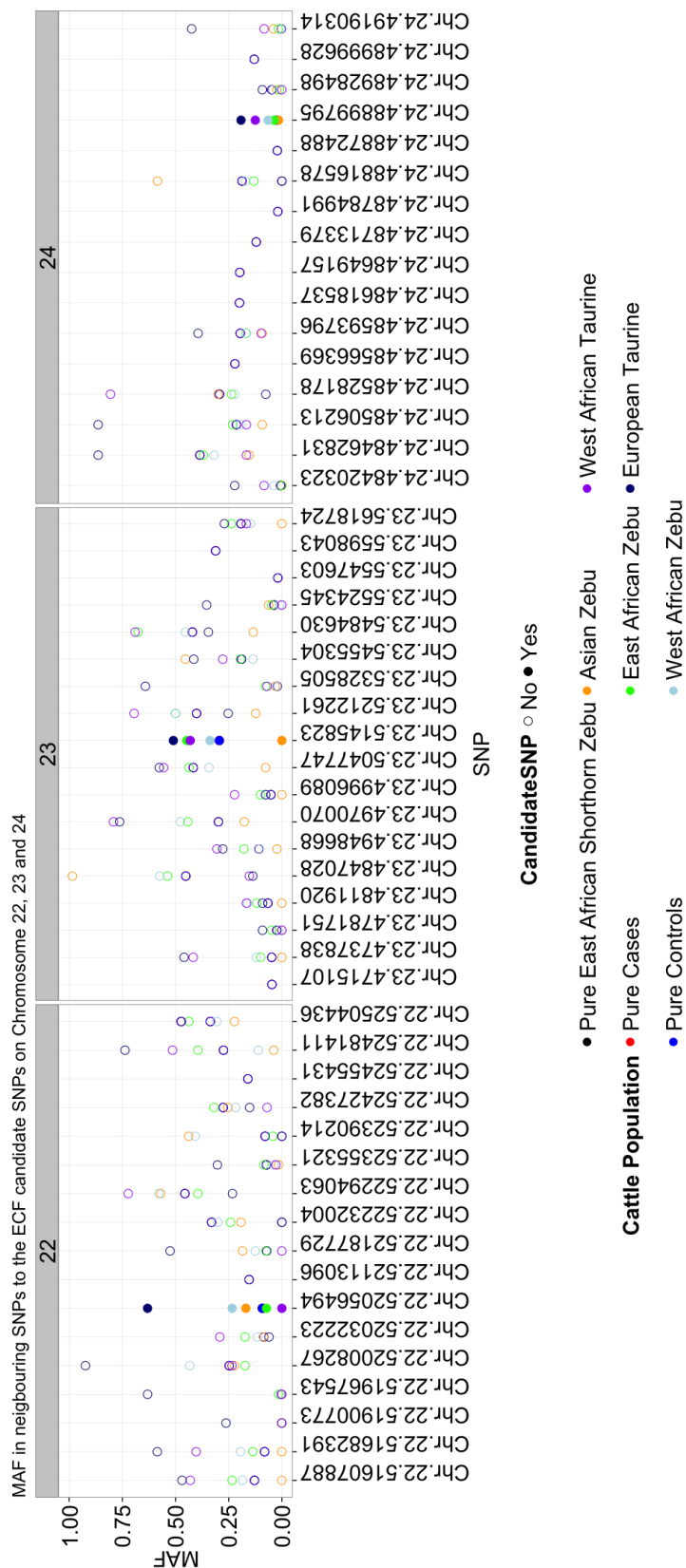
Figure F.1 Minor allele frequency of SNPs surrounding the East Coast Fever death candidate SNPs in each cattle population. Solid dots indicates the candidate SNP.

Packed cell volume at the time of seroconversion to *Theileria parva*

MAF in neigbouring SNPs to the PCV$_{TP}$ candidate SNPs on Chromosome 1, 5 and 7

Figure F.2 Continued on next page

Figure F.2 Continued on next page

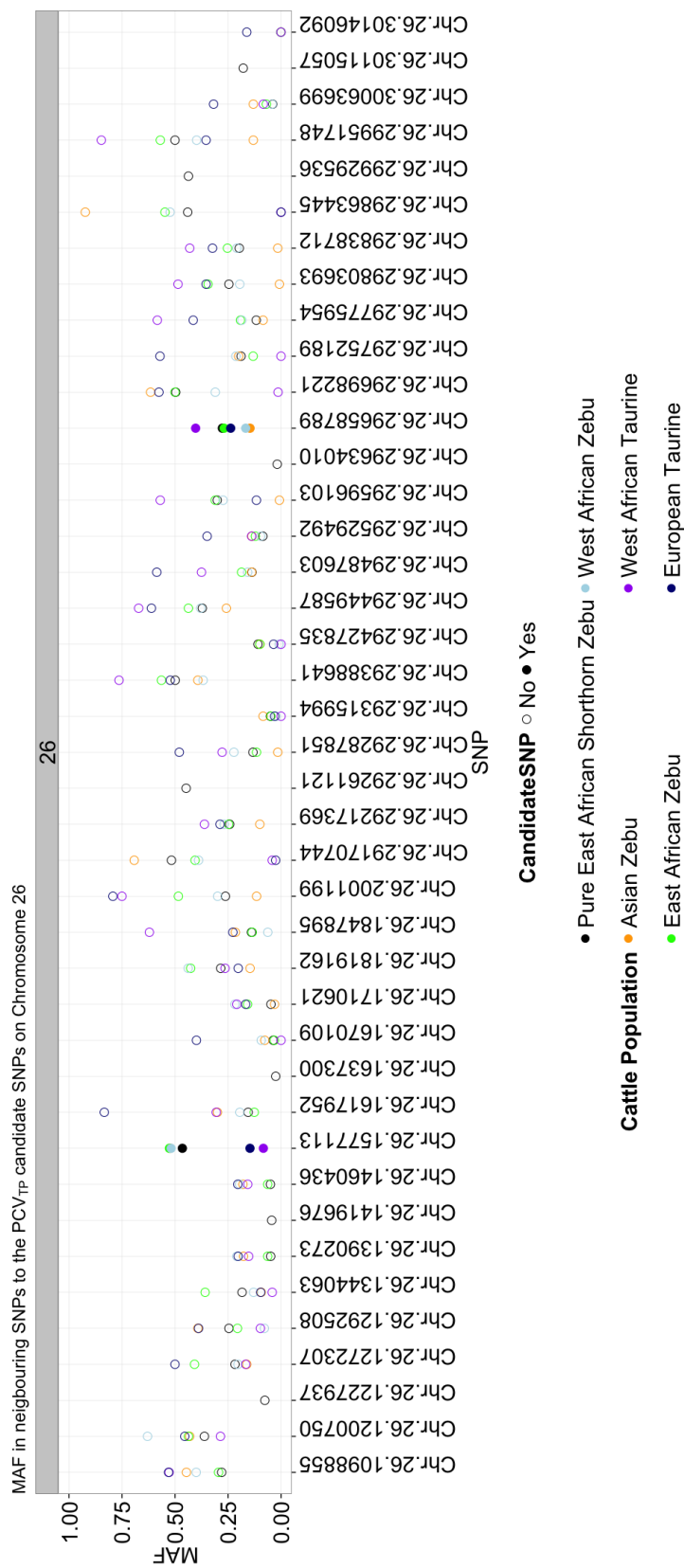Figure F.2 Continued on next page

Figure F.2 Minor allele frequency of SNPs surrounding the packed cell volume at the time of seroconversion to *Theileria parva* candidate SNPs in each cattle population. Solid dots indicates the candidate SNP.